



# Quaestio facti

Revista Internacional sobre Razonamiento Probatorio  
International Journal on Evidential Legal Reasoning

# Qf

N. 3 | 2022

ISSN: 2604-6202

Presentación, Diego Dei Vecchi, Jordi Ferrer Beltrán

## ENSAYOS

*Derechos fundamentales, discrecionalidad judicial y proceso penal: cómo la reglamentación de los derechos puede afectar los objetivos del proceso*, Pablo Larsen

*Filosofía de la mente y prueba de los estados mentales: Una defensa de los criterios de «sentido común»*, Daniel González Lagier

*Justificación epistémica, evidencialismo robusto y prueba jurídica*, Edgar Aguilera

*Evidence, Standards and Burdens of Proof in the European Asylum System: a New Field for Evidence Scholars?*, Elena Marchese

*«Déjame atravesar el viento con documentos»: propuestas para replantear la prueba documental del siglo XXI*, Vitor De Paula Ramos

*La distinción entre esclarecimiento probatorio y esclarecimiento preparatorio en el proceso civil*, Thomas Vogt Geisse

*Juez de amparo. Un garante de la corrección epistémica de los hechos*, Verónica Lorena Osornio Plata

*Sobre la obligación de aplicar los Protocolos de Estambul y Minnesota en la jurisprudencia de la Corte Interamericana de Derechos Humanos*, María Elena Lumiento

## CONJETURAS Y REFUTACIONES

*Naturalized Epistemology and the Law of Evidence: A Reply to Pardo, Spellman, Muffato, and Enoch*, Ronald J. Allen

## INFORME DEL PCAST: CIENCIA FORENSE EN LOS TRIBUNALES PENALES

*Presentación de la traducción al castellano del informe del PCAST sobre la ciencia forense en los tribunales penales*, Carmen Vázquez

*Informe al Presidente. Ciencia forense en los tribunales penales: asegurando la validez científica de los métodos forenses basados en comparación de características*

## CIENCIA PARA EL PROCESO

*(Un-)interpretability in Expert Evidence: An Inquiry into the Frontiers of Evidential Assessment*, Alex Biedermann, Kyriakos N. Kotsoglou

## IURIS PRUDENTIA

*La prueba de los motivos discriminatorios según la Corte Suprema argentina: ¿cómo mejorar su regulación?*, Alejo Joaquín Giles, Olivia Minatta



Qf

# Quaestio *facti*

Revista Internacional sobre Razonamiento Probatorio  
International Journal on Evidential Legal Reasoning

N. 3 | 2022

[www.quaestiofacti.com](http://www.quaestiofacti.com)

Quedan rigurosamente prohibidas, sin la autorización escrita de los titulares del «Copyright», bajo las sanciones establecidas en las leyes, la reproducción total o parcial de esta obra por cualquier medio o procedimiento, comprendidos la reprografía y el tratamiento informático, y la distribución de ejemplares de ella mediante alquiler o préstamo públicos.

© CÁTEDRA DE CULTURA JURÍDICA

© MARCIAL PONS  
EDICIONES JURÍDICAS Y SOCIALES, S. A.

San Sotero, 6 - 28037 MADRID

☎ (91) 304 33 03

*[www.marcialpons.es](http://www.marcialpons.es)*

ISSN: 2604-6202

Depósito legal: M. 3.518-2020

Fotocomposición: Ene Estudio

Impresión: ELECÉ, Industria Gráfica, S. L.

Polígono El Nogal - Río Tiétar, 24 - 28110 Algete (Madrid)

MADRID, 2022

## DIRECTORES

Diego Dei Vecchi  
Universitat de Girona, España

Jordi Ferrer Beltrán  
Universitat de Girona, España

## COMITÉ DE REDACCIÓN

Daniela Accatino  
Universidad Austral de Chile, Chile

Daniel González Lagier  
Universidad de Alicante, España

Edgar Aguilera  
Universidad Autónoma del Estado de México.  
Universitat de Girona, España

Carmen Vázquez  
Universitat de Girona, España

## COMITÉ EDITORIAL

Christian Dahlman  
Lund University, Suecia

Antonio Manzanero  
Universidad Complutense de Madrid, España

Mauricio Duce  
Universidad Diego Portales, Chile

Paul Roberts  
University of Nottingham, Reino Unido

Mercedes Fernández  
Universidad de Alicante, España

Sarah Summers  
University of Zurich, Suiza

Raymundo Gama  
Instituto Tecnológico Autónomo de México

Giovanni Tuzet  
Università Bocconi, Italia

Ho Hock Lai  
National University of Singapore, Singapur

Jonatan Valenzuela  
Universidad de Chile, Chile

## CONSEJO ASESOR

- Ronald Allen  
Northwestern University, EEUU
- Amalia Amaya  
Universidad Nacional Autónoma de México, México
- Perfecto Andrés Ibáñez  
Magistrado emérito de la Sala Segunda del Tribunal  
Supremo Español
- José María Asencio  
Universidad de Alicante, España
- Lorena Bachmaier  
Universidad Complutense de Madrid, España
- Zhang Baosheng  
University of Political Science and Law (CUPL),  
China
- Juan Carlos Bayón  
Universidad Autónoma de Madrid, España
- Lorenzo Bujosa  
Universidad de Salamanca, España
- Rodrigo Coloma  
Universidad Alberto Hurtado, Chile
- Margarita Diges  
Universidad Autónoma de Madrid, España
- Gary Edmond  
University of New South Wales, Australia
- Luigi Ferrajoli  
Università degli Studi di Roma, Italia
- Paolo Ferrua  
Università degli Studi di Torino, Italia
- Marina Gascón  
Universidad de Castilla-La Mancha, España
- Joaquín González  
Universidad Autónoma de Madrid, España
- Susan Haack  
University of Miami, EEUU
- Juan Igartua  
Universidad del País Vasco, España
- John Jackson  
University of Nottingham, Reino Unido
- Larry Laudan  
University of Texas, EEUU
- Giuliana Mazzoni  
University of Hull, Reino Unido
- Dale Nance  
Case Western Reserve University, EEUU
- Jordi Nieva-Fenoll  
Universitat de Barcelona, España
- Eduardo Oteiza  
Universidad Nacional de La Plata, Argentina
- Andrés Páez  
Universidad de los Andes, Colombia
- Jairo Parra Quijano  
Universidad Libre, Colombia
- Catherine Piché  
Université de Montréal, Canadá
- Joan Picó i Junoy  
Universitat Pompeu Fabra, España
- Geraldo Prado  
Universidade Federal do Rio de Janeiro, Brasil
- Giovanni Priori  
Pontificia Universidad Católica del Perú, Perú
- Vitor Lia de Paula Ramos  
Uniritter, Brasil
- Frederick Schauer  
University of Virginia, EEUU
- Paulo de Sousa Mendes  
Universidade de Lisboa, Portugal
- Michele Taruffo  
Università degli studi di Pavia, Italia
- William Twining  
University College of London, Reino Unido
- Giulio Ubertis  
Università Cattolica del Sacro Cuore, Italia
- Alan Uzelac  
University of Zagreb, Croacia
- Adrian Zuckerman  
University of Oxford, Reino Unido

## ÍNDICE

DEI VECCHI, DIEGO; FERRER BELTRÁN, JORDI: <i>Presentación</i> .....	7
---	---

### ENSAYOS

LARSEN, PABLO: <i>Derechos fundamentales, discrecionalidad judicial y proceso penal: cómo la reglamentación de los derechos puede afectar los objetivos del proceso</i> .....	13
GONZÁLEZ LAGIER, DANIEL: <i>Filosofía de la mente y prueba de los estados mentales: Una defensa de los criterios de «sentido común»</i> .....	49
AGUILERA, EDGAR: <i>Justificación epistémica, evidencialismo robusto y prueba jurídica</i> .....	81
MARCHESE, ELENA: <i>Evidence, Standards and Burdens of Proof in the European Asylum System: a New Field for Evidence Scholars?</i> .....	103
DE PAULA RAMOS, VITOR: <i>«Déjame atravesar el viento con documentos»: propuestas para replantear la prueba documental del siglo XXI</i> .....	131
VOGT GEISSE, THOMAS: <i>La distinción entre esclarecimiento probatorio y esclarecimiento preparatorio en el proceso civil</i> .....	159
OSORNIO PLATA, VERÓNICA LORENA: <i>Juez de amparo. Un garante de la corrección epistémica de los hechos</i> .....	175
LUMIENTO, MARÍA ELENA: <i>Sobre la obligación de aplicar los Protocolos de Estambul y Minnesota en la jurisprudencia de la Corte Interamericana de Derechos Humanos</i> .....	217

## CONJETURAS Y REFUTACIONES

- ALLEN, RONALD J.: *Naturalized Epistemology and the Law of Evidence: A Reply to Pardo, Spellman, Muffato, and Enoch*..... 253

## INFORME DEL PCAST: CIENCIA FORENSE EN LOS TRIBUNALES PENALES

- VÁZQUEZ, CARMEN: *Presentación de la traducción al castellano del informe del PCAST sobre la ciencia forense en los tribunales penales*..... 275
- Informe al Presidente. Ciencia forense en los tribunales penales: asegurando la validez científica de los métodos forenses basados en comparación de características*..... 285

## CIENCIA PARA EL PROCESO

- BIEDERMANN, ALEX; KOTSOGLU, KYRIAKOS N.: *(Un-)interpretability in Expert Evidence: An Inquiry into the Frontiers of Evidential Assessment*..... 483

## IURIS PRUDENTIA

- GILES, ALEJO JOAQUÍN; MINATTA, OLIVIA: *La prueba de los motivos discriminatorios según la Corte Suprema argentina: ¿cómo mejorar su regulación?*... 521

## PRESENTACIÓN

Con gran satisfacción presentamos este tercer paso hacia la consolidación de *Quaestio facti. Revista Internacional sobre Racionamiento Probatorio*. La revista nació hace tres años ante la necesidad de contar, en el marco de la cultura jurídica hispanohablante, con un espacio de discusión especializado, riguroso, en materia de razonamiento probatorio.

El objetivo inicial, invariable hasta hoy, ha sido el de sortear fronteras territoriales, por ejemplo, entre los ámbitos del *civil law* y del *common law*, pero también entre diferentes disciplinas. La revista se propuso desde el inicio derribar barreras entre la epistemología, el derecho procesal, el derecho internacional de los derechos humanos y diferentes ciencias relevantes para la práctica forense.

Esta iniciativa se apoya en poderosas razones que indican que derrotar esas barreras redundará siempre en beneficio creciente de la práctica jurídica decisoria y motivacional. Por cierto, afortunadamente no se trataba entonces, ni se trata ahora, de una iniciativa inusitada y solitaria. En septiembre de 2016, por ejemplo, durante la administración de Barack Obama, el *President's Council Advisors on Science and Technology* (PCAST) emitió el informe titulado: «Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods». Ese informe constituyó un hito fundamental, y de un carácter muy peculiar, dentro de la gama de estudios guiados por la iniciativa que también impulsa a *Quaestio facti*. Se trata de un trabajo que claramente desafía las barreras entre disciplinas pero que, a pesar de su importancia, había quedado hasta ahora circunscrito a un ámbito cultural muy concreto. De allí que hayamos decidido publicar en este tercer número la traducción al castellano de dicho informe, que estimamos aporta información extraordinariamente valiosa para académicos y juristas prácticos. La traducción fue llevada a cabo por José Juan Lucena Molina y está aquí antecedida por una esclarecedora introducción de Carmen Vázquez Rojas.

La extensión del informe nos motivó a suspender, solo por esta ocasión, la sección *Conjeturas y refutaciones*. En el presente número dicha sección se reduce a la contrarréplica a sus críticos de Ronald Allen, autor del texto principal de *Conjeturas y refutaciones* 2021.

Para concluir, aprovechamos la ocasión para agradecer a quienes nos brindan su apoyo para poder editar la revista cada año. En relación con este tercer número en particular, agradecemos especialmente a los autores y autoras por confiar en la revista para publicar sus trabajos y a todas las personas que aceptaron realizar la crucial y ardua labor de dictaminar textos anónimamente. Por último, por su invaluable y generosa ayuda a efectos de corregir pruebas de imprenta, agradecemos a Jorge Baquerizo, Renzo Cavani, Marianela Delgado Nieves, Carlo Vittorio Giabardo, Jesús Juan Gutiérrez Estrada, Sidney Marcos Escobar, Margarita Rosa Martínez, Jaime Oportus Maino, Sebastián Rebolledo, Henry Reyes Garcés y Marco Segatti.

Diego Dei Vecchi y Jordi Ferrer Beltrán

# ENSAYOS



## DERECHOS FUNDAMENTALES, DISCRECIONALIDAD JUDICIAL Y PROCESO PENAL: CÓMO LA REGLAMENTACIÓN DE LOS DERECHOS PUEDE AFECTAR LOS OBJETIVOS DEL PROCESO\*

Pablo Larsen

Universidad Torcuato Di Tella/Universidad de Buenos Aires  
pablolarsen93@gmail.com

**RESUMEN:** Este trabajo analiza las formas en que pueden reglamentarse los derechos fundamentales y el modo en que esto puede incidir en el cumplimiento de los objetivos del proceso penal. Para esto, primero se muestra que el análisis del tema debe ser sensible a los desacuerdos que pueden generarse al precisar las exigencias de los derechos y discutir sobre sus posibles reglamentaciones. Luego, mediante la tipología conocida como «reglas» (normas más precisas) y «estándares» (normas más vagas), se exponen dos modelos que permiten identificar las principales alternativas disponibles según reduzcan o amplíen, respectivamente, la discrecionalidad judicial para tomar decisiones en el proceso. Asimismo, se argumenta que un modelo puede ser atractivo para imponer límites más fuertes al poder estatal, mientras que el otro puede encontrarse en mejores condiciones para aumentar la eficiencia del proceso en la búsqueda de la verdad.

**PALABRAS CLAVE:** derechos fundamentales; discrecionalidad judicial; reglas; estándares; objetivos del proceso penal.

---

\* Una versión de este trabajo fue presentada como tesis para obtener el grado de Magister en Derecho Penal por la Universidad Torcuato Di Tella (Argentina), bajo la dirección del Prof. Dr. Máximo Langer, y ante un jurado integrado por los Prof. Dres. Alejandra Verde, Hernán Bouvier y Alejandro Chehtman. Agradezco a los cuatro por sus agudas críticas y comentarios. Asimismo, agradezco a María Belén Iorizzo, Alan Limardo, Pablo Rovatti, Leandro Días, Diana Velela, Ignacio Racca, Matías Díaz y Agustín Varela por las críticas y sugerencias realizadas luego de la lectura de algunas versiones previas del texto. El agradecimiento se extiende a las/los árbitras/os que evaluaron el artículo por sus valiosas recomendaciones.

FUNDAMENTAL RIGHTS, JUDICIAL DISCRETION  
AND CRIMINAL PROCESS: HOW THE REGULATION OF RIGHTS AFFECTS  
THE ACHIEVEMENT OF THE PROCEDURE GOALS

**ABSTRACT:** This paper explores how fundamental rights can be regulated, and analyzes how this activity can have an impact in the achievement of the goals of the criminal procedure. In order to do so, the paper first shows that the analysis of those rights must pay close attention to the disagreements that arise in discussions about their possible regulations. After that, it shows that the distinction between «rules» and «standards» can be used to identify two models that represent the main alternatives available, and shows that those models differ in how much discretion they grant to the courts. Also, the paper argues that one model can be related with the idea of imposing strong limits to the state's power, while the other can be related with the improvement of the fact-finding accuracy of the procedure.

**KEYWORDS:** fundamental rights; judicial discretion; rules; standards; goals of the criminal process.

**SUMARIO:** 1. Introducción.—2. Tomarse en serio los desacuerdos sobre el contenido de los derechos fundamentales en el proceso penal.—3. La discrecionalidad judicial como eje de los desacuerdos.—4. Una terminología para los desacuerdos: reglas, estándares y dos modelos para reglamentar derechos fundamentales. 4.1. La estructura de las normas jurídicas que reglamentan derechos fundamentales de la persona imputada: la dimensión de la «precisión». 4.2. Reglas y estándares en la reglamentación de los derechos fundamentales de la persona imputada.—5. La precisión de las normas jurídicas como herramienta: entre la imposición de límites fuertes al poder estatal y la eficiencia en la búsqueda de la verdad. 5.1. Reglas e imposición de límites fuertes al poder estatal. 5.2. Estándares y eficiencia epistémica del proceso. 5.3. Balance—6. Conclusiones.

**RECOMMENDED CITATION:** LARSEN, PABLO, 2021: «Derechos fundamentales, discrecionalidad judicial y proceso penal: cómo la reglamentación de los derechos puede afectar los objetivos del proceso», in *Quaestio facti*, 3: 13-47. Madrid: Marcial Pons Ediciones Jurídicas y Sociales. DOI: [http://dx.doi.org/10.33115/udg\\_bib/qf.i3.22598](http://dx.doi.org/10.33115/udg_bib/qf.i3.22598)

## 1. INTRODUCCIÓN

Este trabajo se ocupa de analizar las formas que pueden asumir las normas que reglamentan los derechos fundamentales de la persona imputada (aquellas que precisan de forma más concreta su contenido y exigencias) y cómo esto puede incidir en el cumplimiento de los fines del proceso penal. Puntualmente, propone focalizar ese análisis en el grado de discrecionalidad con el que cuentan los tribunales al aplicar esas normas según el grado de precisión con el que hayan sido redactadas: normas más precisas (reglas) o más vagas (estándares).

De ese modo, el trabajo se concentra en la clase de discusiones que se generan al debatir, por ejemplo, cómo los tribunales deberían decidir si la admisibilidad de la declaración previa de un testigo ausente en el juicio es compatible con el derecho

fundamental a interrogar a los testigos de cargo. Al analizar este asunto es usual encontrar dos posiciones contrapuestas. Por un lado, existen defensas de una reglamentación que establezca de forma precisa que los tribunales deben limitarse a verificar si la defensa contó con una oportunidad previa de interrogar al testigo, y que sólo en ese caso estos testimonios pueden ser admisibles. Por el otro, también existen argumentaciones a favor de una reglamentación que otorgue más libertad a los tribunales y les permita admitir esas declaraciones luego de analizar, en cada caso concreto, algunos factores como las razones por las cuales el testigo está ausente o los indicios de confiabilidad que pueda presentar su testimonio pese a que la defensa no haya tenido posibilidad de controlarlo, entre otros<sup>1</sup>.

El artículo busca mostrar que ese tipo de discusiones, vinculadas con la libertad que puede otorgarse a los tribunales para tomar decisiones durante el proceso penal, se repite en muchas más oportunidades de lo que se ha notado en la literatura tradicional. Asimismo, intenta ofrecer una herramienta que pueda dar cuenta de la complejidad que existe detrás de esos debates y que permita alcanzar dos objetivos concretos: identificar cuáles son las principales técnicas con las que pueden reglamentarse los derechos fundamentales de la persona imputada, y detectar las razones que podrían respaldar la implementación de alguna de esas alternativas. El primero de esos objetivos permitirá advertir que algunos asuntos importantes del proceso penal pueden ser abordados de formas variadas, que pueden ser distintas a las que se conozcan en un determinado ámbito, y ayudará a comprender dónde radican las diferencias entre esas alternativas. El segundo, por su parte, permitirá mostrar que la detección de esas alternativas y de las razones que pueden estar detrás de su elección puede ser útil para los debates que se generan en esta materia.

Esto último, a su vez, permitirá realizar un aporte concreto a la discusión sobre el rol que desempeñan los derechos fundamentales de la persona imputada en el cumplimiento de los fines del proceso penal. Tal como se verá más adelante, si bien los derechos en alguna medida constituyen límites a la búsqueda de la verdad, esos límites no son necesariamente claros, existen diferentes formas de establecerlos, y algunas alternativas pueden ser más favorables a la eficiencia epistémica del proceso que otras. El resultado que esto arroja puede ser útil para quienes consideren que los procesos penales actuales deben preocuparse más por la búsqueda de la verdad: si se superponen los dos objetivos de este trabajo, será posible advertir que algunos derechos pueden ser reglamentados válidamente de formas distintas a las que se conozcan en un ámbito determinado, y en algunas ocasiones de un modo tal que permita mejorar la eficiencia epistémica del proceso<sup>2</sup>. A la inversa, el desarrollo también puede ser útil

---

<sup>1</sup> El ejemplo más conocido de la primera alternativa es la regla adoptada por la Suprema Corte de Justicia de Estados Unidos en el precedente “Crawford v. Washington” (541 U.S. 36). Un caso paradigmático de la segunda es la regulación contenida en la Criminal Justice Act del Reino Unido (114, 1.d y 2). Se volverá sobre estos debates y estas regulaciones en las secciones 3 y 4.

<sup>2</sup> Esto puede ser atractivo para lo que en alguna ocasión ha sido denominado el «proyecto normativo» de la epistemología jurídica, es decir, aquel «consistente en proponer cambios en las reglas existentes

para quienes consideren que los límites al poder estatal que actualmente existen en algún contexto determinado deben ser más intensos de lo que hoy en día son.

La argumentación procederá de la siguiente manera. Primero (2) se exponen las razones por las cuales pueden generarse discusiones en esta materia, y se explica por qué se encuentra justificado un análisis más detenido sobre las posibles «reglamentaciones» de los derechos fundamentales de la persona imputada. Con ese concepto, en el marco de este trabajo, se hará referencia a la actividad consistente en precisar el contenido y las exigencias de los derechos para resolver casos concretos, algo que, como se verá más adelante, es realizado por las legislaturas o parlamentos y también por ciertos tribunales constitucionales o de derechos humanos. Aquí se intenta demostrar, en pocas palabras, que el análisis de varios temas no puede limitarse a afirmar que «el proceso penal debe respetar los derechos fundamentales de la persona imputada», porque esto es poco más que el inicio de los debates que pueden generarse. Para eso se muestra que los interrogantes que se generan al reglamentar muchos derechos fundamentales necesitan una respuesta que usualmente, y de forma deliberada, los textos constitucionales que los establecen omiten brindar.

Luego (3) se ofrece un punto de partida útil para desarrollar un análisis sensible a las complejidades del tema. Concretamente, se intenta mostrar que existen buenas razones para que el estudio de la reglamentación de los derechos fundamentales de la persona imputada esté centrado en el grado de libertad que se puede otorgar a los tribunales para tomar decisiones durante el proceso, es decir, en la «discrecionalidad judicial».

Posteriormente (4) se presentan dos modelos de reglamentación diferenciados por el grado de precisión que se utiliza al redactar las normas y, de ese modo, por la discrecionalidad que se otorga a los tribunales al reducir o ampliar los factores que pueden tener en cuenta al resolver casos. Aquí, con apoyo en la distinción conocida en la literatura anglosajona como «reglas» (normas más precisas) y «estándares» (normas más vagas), se busca mostrar que esos modelos permiten detectar las principales formas con las que pueden reglamentarse muchos derechos fundamentales de la persona imputada, tal como se encuentran en varias reglamentaciones de muchos países.

Finalmente (5) se esbozan algunas pautas que pueden ser tenidas en cuenta al decidir cuál de esos modelos implementar. Para esto se intenta mostrar que los modelos propuestos pueden ser herramientas útiles al servicio de los principales valores que suelen estar en juego en el proceso penal: la tradicional tensión entre la necesidad de imponer límites al poder estatal y la eficiencia del proceso en la búsqueda de la verdad. Concretamente, se argumentará que el modelo de reglas puede ser atractivo para quienes consideren necesario cumplir de forma más eficiente el primer objetivo, mientras que el modelo de estándares, en cambio, puede ser interesante para quienes

---

al efecto de modificar o eliminar aquellas que constituyan impedimentos graves para la búsqueda de la verdad» (LAUDAN, 2013: 23).

se inclinen preponderantemente por el segundo. Con esto se busca demostrar, en síntesis, que la identificación de los principales beneficios y costos de estos modelos puede orientar de forma clara la decisión de quienes reglamentan los derechos fundamentales en el proceso penal.

El desarrollo del trabajo, por razones de espacio, no pretende analizar en detalle la forma en que pueden reglamentarse cada uno de los derechos fundamentales que la mayoría de las constituciones estatales, cartas de derechos o tratados internacionales establecen para la persona imputada. Sea que se trate de los derechos fundamentales que asumen la modalidad de «inmunidades» —aquellos que establecen lo que el Estado *no puede hacer* a la persona imputada en el proceso penal—<sup>3</sup> o aquellos que son denominados «derechos de intervención» —aquellos que establecen lo que la persona imputada *debe poder hacer* en el proceso penal—<sup>4</sup>, su reglamentación representa muchos asuntos importantes del proceso penal. Precisamente por este motivo es que la literatura de diferentes tradiciones jurídicas considera al derecho procesal penal como el «sismógrafo» de la Constitución (ROXIN, 2000: 10), o a sus normas como un “subgrupo” de las constitucionales (STUNTZ, 1997: 6). Un análisis detallado de cada uno de los derechos fundamentales de la persona imputada, por ende, requeriría un trabajo mucho más extenso. Por esa razón, el objetivo será proporcionar una herramienta con un grado suficiente de abstracción y generalidad que pueda ser aplicada a muchas de las discusiones que pueden generarse al analizar esas reglamentaciones.

Para esto, sin embargo, se expondrán algunas reglamentaciones concretas que permiten ilustrar los rasgos de los modelos que se presentan. De ese modo, los modelos serán expuestos mediante un análisis de la forma en que pueden reglamentarse los derechos de la persona imputada a «ser juzgada en un plazo razonable» o «sin dilaciones indebidas», a «interrogar a los testigos de cargo», a «no ser juzgada más de una vez por un mismo hecho o delito», a «no ser obligada a declarar contra sí misma», a que «se presuma su inocencia mientras no se establezca legalmente su culpabilidad» y a un «debido proceso». Con esto se mostrará que los modelos pueden ser atractivos para analizar asuntos importantes como la duración de los procesos, el problema de los testigos de cargo que por alguna razón no pueden estar presentes en el juicio, la posibilidad de enjuiciar nuevamente a una persona cuando el primer proceso tuvo defectos fundamentales, la voluntariedad de las confesiones o declaraciones auto-

---

<sup>3</sup> Dentro de esta categoría se encuentran el derecho fundamental a «no ser condenada sin un juicio previo», a un «debido proceso», a «ser juzgada en un plazo razonable o sin dilaciones indebidas», a «ser juzgada por un tribunal imparcial», a que «se presuma su inocencia mientras no se establezca legalmente su culpabilidad», a «no ser obligada a declarar contra sí misma» o a «no ser juzgada más de una vez por un mismo hecho». Sobre el concepto de «inmunidades» NINO (2007: 29-31).

<sup>4</sup> Aquí se ubican los derechos fundamentales de la persona imputada a «ser oída», a «interrogar a los testigos de cargo», a «recurrir la sentencia condenatoria» o el derecho a contar con una defensa técnica, usualmente considerado como un derecho destinado a garantizar el ejercicio de los demás derechos de intervención o participación. La denominación «derechos de intervención» es utilizada, entre otros, por STEIN (2006: 31-33).

incriminatorias, las exigencias probatorias que se consideran suficientes para tener por probado un hecho y condenar a una persona, y la exclusión de la prueba ilícita. El análisis de esos temas, pese a no cubrir la totalidad de las reglamentaciones de derechos fundamentales que se efectúan al diseñar un proceso penal, puede ser útil para detectar o imaginar con mayor facilidad cómo los modelos que aquí se exponen pueden ser trasladados a otros asuntos que, por razones de espacio, no podrán ser estudiados en este trabajo.

## 2. TOMARSE EN SERIO LOS DESACUERDOS SOBRE EL CONTENIDO DE LOS DERECHOS FUNDAMENTALES EN EL PROCESO PENAL

Es posible afirmar que en la actualidad existe un consenso generalizado en torno a la necesidad de garantizar determinados derechos fundamentales a la persona imputada en un proceso penal. Esto puede verse reflejado, por ejemplo, en los catálogos que los principales tratados internacionales de derechos humanos incluyen sobre el tema<sup>5</sup>, o en la manera en que las constituciones o cartas de derechos de muchos países se encargan de establecerlos dentro de disposiciones dedicadas exclusivamente a la persona imputada<sup>6</sup>. De ese modo, la mayoría de los sistemas jurídicos comparten un mínimo común denominador según el cual ella tiene derecho a «ser juzgada en un plazo razonable» o «sin dilaciones indebidas», a «interrogar a los testigos de cargo», a «no ser juzgada más de una vez por un mismo hecho o delito», a «no ser obligada a declarar contra sí misma», a que «se presuma su inocencia mientras no se establezca legalmente su culpabilidad», y a un «debido proceso», entre otros.

Sin embargo, ese consenso puede volverse más borroso cuando se trata de precisar el contenido de esos derechos y determinar cuáles son sus exigencias concretas. A diferencia de lo que podría suceder en un debate reducido a determinar si la persona imputada es titular de algún derecho en abstracto, las normas jurídicas que se encargan de establecer los derechos fundamentales —típicamente constituciones, cartas de derechos o tratados internacionales— no suelen brindar respuestas a este tipo de interrogantes de carácter concreto. Por el contrario, en varias ocasiones se destaca que es poco usual contar con respuestas claras al analizar qué implican específicamente muchos derechos. Según una idea muy difundida, esta área no se encuentra domi-

---

<sup>5</sup> En esta tendencia se encuentran la Convención Americana sobre Derechos Humanos (8), el Convenio Europeo de Derechos Humanos (6), la Carta Africana sobre los Derechos Humanos y de los Pueblos (7) y el Pacto Internacional de Derechos Civiles y Políticos (14). En lo que sigue estos textos serán citados mediante siglas.

<sup>6</sup> De esa manera, por ejemplo, Argentina (18), Estados Unidos (enmiendas 5.<sup>a</sup>, 6.<sup>a</sup> y 8.<sup>a</sup>), Canadá (11), Chile (19), Colombia (28 a 34), España (24.2), Alemania (103, 104), Italia (111), Nueva Zelanda (24) y Sudáfrica (35).

nada por los consensos generalizados sino, en cambio, por los grandes desacuerdos (WALDRON, 2005: 67-68; SUNSTEIN, 2007: 1-2). Tal como explica WALDRON:

Es innegable que tanto las disposiciones de la carta de derechos tienen injerencia sobre cómo debe resolverse cada una de esas cuestiones, como que sus disposiciones no determinan por sí mismas una solución a la cuestión de un modo que supere la disputa razonable [...] El desacuerdo no impide la promulgación de una carta de derechos. Pero los desacuerdos continuarán sin resolverse, dejándonos en una situación en la que —cuando surja un debate sobre la posible violación de derechos— si bien las disposiciones de la carta de derechos serán sin duda pertinentes para resolver la cuestión, subsistirá un debate entre personas razonables sobre en qué sentido lo serán y en qué medida prohíben (o limitan la aplicación de) las disposiciones legislativas que estén en tela de juicio (2006: 1368-1369).

Al analizar las causas de ese fenómeno, frecuentemente se explica que se trata de una consecuencia derivada de los términos particularmente indeterminados que suelen utilizarse al redactar los textos constitucionales que establecen derechos fundamentales. De acuerdo con este paradigma, «las disposiciones que establecen derechos constitucionales tienden a resolver menos de lo que hacen algunas de las disposiciones más determinadas de una constitución; estas, de hecho, necesitan más especificaciones antes de poder ser aplicadas» (WEBBER, 2009: 13-14; de modo similar SUNSTEIN, 2018: 35, WALDRON, 2006: 1369, nota 60, SCHAUER, 1982: 801, y ALEXANDER, 2003: 374). También suele explicarse que ese estilo característico de los textos constitucionales no es fruto de casualidades. En el caso de las constituciones estatales, por ejemplo, se afirma que acudir a términos particularmente indeterminados al establecer derechos fundamentales permite que las personas coincidan en grandes abstracciones aun cuando no estén de acuerdo sobre lo que esos derechos implican en situaciones concretas, y se advierte que esto es lo que, en efecto, permite contar con constituciones (SUNSTEIN, 2018: 35-36). Esa explicación parece ser aún más clara en el caso de tratados internacionales de derechos humanos aplicables a Estados de tradiciones jurídicas diversas, pues «para ser universales, los derechos humanos deben ser básicos» y para esto «un nivel de abstracción es necesario, una búsqueda de un estándar mínimo que pueda ser alcanzable a nivel universal» (HO, 2012: 262).

Estas circunstancias pueden considerarse como las predominantes en los textos que establecen derechos fundamentales. Constituyen, además, buenas razones para entender que los debates vinculados con el contenido de esos derechos y la forma en que deberían resolverse casos que los involucren son difíciles de zanjar de forma definitiva. La explicación, en términos sencillos, puede ser resumida de la siguiente manera: las normas jurídicas que establecen muchos derechos fundamentales son «diamantes en bruto» que usualmente no proveen todo lo que se necesita para resolver casos relacionados con esos derechos. Como explica GUASTINI, usualmente las disposiciones constitucionales «a causa de su indeterminación, no son susceptibles de aplicación sin previa concretización» (2012: 219). Por esa razón, para cumplir con esta tarea es necesario que los derechos sean precisados de forma más concreta —si se desea continuar la metáfora, un diamante en bruto no puede ser utilizado en esas

condiciones, pues necesita ser «tallado y pulido»—, y en esto último se pueden tomar varios caminos alternativos sin que el texto constitucional establezca de antemano cuál de estos es el único correcto —podemos «pulir y tallar un diamante en bruto» de diferentes maneras y elaborar productos muy distintos con él—.

En cierta medida, se trata de distinguir entre las normas jurídicas que *establecen* un derecho fundamental y aquellas que lo *reglamentan* y le brindan un contenido más concreto. De acuerdo con esa distinción, los textos constitucionales suelen incorporar el primer tipo de normas jurídicas y lo hacen mediante la utilización de términos particularmente indeterminados. Las segundas, en las que estará puesta la atención en este trabajo, en cambio, pese a ser imprescindibles para tomar decisiones en muchos casos relacionados con estos derechos, suelen quedar reservadas a otros entes infraconstitucionales. De lo primero se ocupan las asambleas constituyentes o los Estados al asumir compromisos internacionales en materia de derechos humanos. De lo segundo, en cambio, se encargan las legislaturas o parlamentos y también ciertos tribunales constitucionales o de derechos humanos<sup>7</sup>.

Esto último obliga a realizar algunas precisiones adicionales. Si la actividad a la que aquí se hace referencia al hablar de «reglamentación» consiste en determinar cuál es el contenido concreto de los derechos fundamentales, es posible sostener que esa tarea requiere de interpretación (GUASTINI, 2011: 211-220). Y si se tiene en cuenta esta circunstancia, también se puede advertir que esa actividad no la realizan sólo los parlamentos —quienes suelen sancionar los códigos procesales, que contienen gran parte de las reglamentaciones de los derechos de la persona imputada— sino, además, cierto tipo de tribunales constitucionales o de derechos humanos, tal como se verá más adelante con varios ejemplos. El carácter compartido de esta actividad se advierte con claridad cuando algunas de las interpretaciones sobre las exigencias de los derechos entran en tensión. Considérese, por ejemplo, la situación que se genera cuando un tribunal constitucional concluye que una declaración autoincriminatoria puede ser admisible solo si la persona imputada fue informada previamente de sus derechos, y una legislatura, en cambio, sanciona una ley donde establece que esas declaraciones pueden ser admitidas si se concluye, de acuerdo a las particularidades del caso, que fueron efectuadas de forma voluntaria. Ambos entes, en ese caso, realizan una misma actividad: interpretan las exigencias derivadas de un enunciado normativo que establece un derecho fundamental —concretamente, aquel que esta-

---

<sup>7</sup> Una posible excepción a esto es el caso de aquellas reglamentaciones que no precisan el contenido de los derechos, sino que establecen restricciones a aquello que fue previamente delimitado. En esos supuestos, de acuerdo con algunas posiciones, se generan una serie de exigencias adicionales acerca de quienes pueden dictar esas reglamentaciones y las características concretas que estas deben reunir (concretamente, las que establecen que una restricción puede ser adoptada únicamente por un parlamento y que ese tipo de leyes debe satisfacer las exigencias derivadas del mandato de determinación o *lex certa*). Al respecto, en detalle, PÉREZ BARBERÁ (2015). No obstante, con determinados ajustes sensibles a esas exigencias, el desarrollo que aquí se propone también puede ser útil para analizar las reglamentaciones que constituyen restricciones.

blece el derecho a no ser «obligado» a declarar contra uno mismo— y de ese modo lo reglamentan<sup>8</sup>. Si alguna de esas interpretaciones se impone es algo que depende de cuestiones contingentes (por ejemplo, el grado de deferencia que ambos entes se tengan en cada sistema) y no de una cuestión conceptual que establezca que sólo uno de ellos puede interpretar las exigencias de los derechos fundamentales.

Concentrar la atención sólo en los parlamentos, en consecuencia, implicaría realizar una descripción inadecuada del fenómeno que interesa analizar en este trabajo. Esto, sin embargo, es conveniente por razones descriptivas. Si existen argumentos normativos para sostener que tanto legislaturas como los tribunales constitucionales deberían compartir la tarea de reglamentar los derechos fundamentales de la persona imputada es algo que no es necesario determinar aquí<sup>9</sup>.

Los textos de los tratados internacionales que establecen los principales derechos fundamentales de la persona imputada en el proceso penal, por supuesto, no son una excepción al fenómeno descrito. Por ejemplo, esos textos garantizan a la persona imputada un derecho a ser juzgada dentro de un «plazo razonable» o «sin dilaciones indebidas»<sup>10</sup>, pero guardan silencio acerca de cómo debe determinarse si la duración de un proceso fue razonable o si las demoras que se hayan verificado se encontraban justificadas. Del mismo modo, la persona imputada tiene un derecho a «no ser obligada a declarar contra sí misma»<sup>11</sup>, pero los textos de los tratados no indican la manera en que debe determinarse si una declaración autoincriminatoria fue realizada de forma voluntaria. La persona imputada también tiene un derecho a que «se presuma

<sup>8</sup> El ejemplo está tomado del debate generado en EE.UU luego de que el Congreso sancionó una ley donde propuso abandonar el test adoptado por la Suprema Corte de Justicia en el caso *Miranda v. Arizona* (384 U.S. 436), cuya constitucionalidad fue analizada en *Dickerson v. United States* (530 U.S. 428). Para un análisis de esta discusión véase STRAUSS (2001) y DORF y FRIEDMAN (2000), donde se desarrolla la idea de «interpretación constitucional compartida».

<sup>9</sup> De forma general, suele argumentarse que los parlamentos pueden encontrarse en mejor posición para crear normas de carácter general pues los tribunales usualmente adoptan sus decisiones en el marco de casos individuales que presentan muchas particularidades, y esto puede incidir negativamente en la creación de normas jurídicas de carácter general (SCHAUER, 2006; SUNSTEIN, 2018: 118, 200). También se argumenta que la reglamentación de los derechos fundamentales involucra una gran cantidad de factores que proyectan muchas consecuencias para el sistema, y por razones democráticas los tribunales quizá no posean las mejores credenciales para tomar ese tipo de decisiones (sobre esto, en general, WEBBER, 2009: 205-207). Con relación específica al derecho procesal penal, en cambio, muchas veces se argumenta a favor de la actividad de los tribunales en esta materia. Los argumentos que se ofrecen usualmente sostienen que «las legislaturas subestiman los derechos de la persona imputada» porque la gran mayoría de las personas «adopta la perspectiva de una potencial víctima antes que la de una imputada» (DRIPPS, 993: 1079) y que, por esa razón, «las reglas que limiten el poder público vendrán de los tribunales o no vendrán de ningún lado» (DRIPPS, 2001: 46). También se describe, a modo de primera aproximación, que la relación entre los poderes políticos y las cortes supremas es una en la que los primeros ignoran los intereses de las personas acusadas y las segundas cumplen la función de asegurarlos (STUNTZ, 2006: 781).

<sup>10</sup> CADH (8.1), CEDH (6.1), CADHP (7.1.d) y PIDCP (14.3.c).

<sup>11</sup> CADH (8.2.g) y PIDCP (14.3.g). El CEDH y la CADHP no contemplan expresamente este derecho.

su inocencia mientras no se establezca legalmente su culpabilidad»<sup>12</sup>, pero esa fórmula no permite saber cuáles son las exigencias probatorias que permitirían tener por probado un hecho y habilitar el dictado de una condena. Incluso en algunas ocasiones los textos de los tratados no se encargan siquiera de establecer una determinada cuestión, tal como sucede con la ausencia de cláusulas sobre la manera en que debe lidiarse en general con la prueba obtenida de forma ilícita<sup>13</sup>. Muchas constituciones o cartas de derechos nacionales han seguido un camino similar en todos esos asuntos<sup>14</sup>.

Esa falta de especificaciones puede persistir incluso en el caso de aquellos derechos fundamentales que parecen estar establecidos de forma más concreta, tal como sucede con el derecho de la persona imputada a «interrogar a los testigos de cargo»<sup>15</sup> y a «no ser juzgada más de una vez por un mismo hecho o delito»<sup>16</sup>. La posibilidad de generar discusiones vinculadas con cuál es el contenido concreto de esos derechos y, aun cuando se lo considere como precisamente definido, la posibilidad de permitir su reglamentación mediante restricciones debido a la tendencia a no considerarlos derechos «absolutos», genera que los debates sobre la forma en que deberían tomarse decisiones en casos que los involucren sean difíciles de zanjar incluso en esos supuestos.<sup>17</sup>

<sup>12</sup> CADH (8.2) y con redacciones similares CEDH (6.2), PIDCP (14.2) y CADHP (7.1.b).

<sup>13</sup> El CEDH, el PIDCP y la CADHP no prevén ninguna reglamentación expresa sobre la exclusión de la prueba ilícita, mientras que la CADH (8.3) establece ese remedio únicamente para los casos de confesiones obtenidas mediante coacciones. Por su parte, la Convención contra la Tortura y Otros Tratos o Penas Crueles, Inhumanos o Degradantes establece una regla de exclusión específica solo para declaraciones realizadas como resultado de torturas (15), y la misma reglamentación puede encontrarse en la Convención Interamericana para Prevenir y Sancionar la Tortura (10).

<sup>14</sup> Si bien no es posible realizar una revisión exhaustiva en este trabajo, pueden citarse como ejemplos la forma en que muchos Estados reglamentan el derecho de la persona imputada ser juzgada en un «plazo razonable» o «sin dilaciones indebidas» (Estados Unidos, 6.<sup>a</sup> enmienda; Canadá, 11.b; Italia, 111; Sudáfrica, 35.3.d; Colombia, 29; España, 24.2; Nueva Zelanda, 25.b), el derecho a no ser «obligada» a declarar contra sí misma (Estados Unidos, 5.<sup>a</sup> enmienda; Argentina, 18; Chile, 19.7.f; Colombia, 33; Canadá, 11.c; España, 17.3; Sudáfrica, 35.1.c; Nueva Zelanda, 25.d), la presunción de inocencia (Canadá, 11.e; España, 24.2; Italia, 27; Colombia, 29; Sudáfrica, 35.3.h) y la ausencia de una cláusula general vinculada con la exclusión de la prueba ilícita (Argentina, Chile, Estados Unidos, España, Italia y Alemania).

<sup>15</sup> CEDH (6.3.d) y PIDCP (14.3.e). El texto de la CADH solo se refiere a «interrogar a los testigos presentes en el tribunal» (8.2.f).

<sup>16</sup> CADH (8.4), CEDH (Protocolo Adicional n.º 7, 4) y PIDCP (14.7).

<sup>17</sup> Me refiero a esta posibilidad solo para describir una alternativa ampliamente difundida. Si ésta se encuentra justificada y si los derechos mencionados podrían ser considerados «absolutos» —es decir, que no admiten restricciones jurídicas de ningún tipo— es una pregunta que este trabajo deja abierta. Sin embargo, la tendencia a considerar que incluso esos derechos cuyo campo de aplicación parece más sencillo de delimitar pueden ser restringidos es una posición bastante difundida a nivel internacional. Por ejemplo, con relación al derecho de la persona imputada a «no ser juzgada más de una vez por un mismo hecho o delito», el CEDH expresamente contempla una cláusula de limitación (Protocolo Adicional n.º 7, 4.2), y la Corte IDH lo ha considerado como no absoluto (*Almonacid Arellano v. Chile*, 26/8/2006, §154). Algo similar sucede con el derecho a «interrogar o hacer interrogar los testigos de cargo», sobre el cual se ha advertido que «está sujeto a numerosas limitaciones y excepciones» y que «a pesar de la retórica, en ninguna de sus variadas formas [...] es absoluto» (DENNIS, 2010: 256). Algunos

El escenario descrito hasta este punto es particularmente relevante, y el gran abanico de posibilidades que ofrece al reglamentar los derechos fundamentales de la persona imputada justifica la existencia de un análisis más detenido sobre el tema. Sin embargo, se trata de un fenómeno que no ha recibido mucha atención en las exposiciones tradicionales sobre derecho procesal penal, donde es común que se ofrezca un análisis sobre los alcances de los derechos fundamentales y de cómo éstos han sido reglamentados en un contexto determinado pero, sin embargo, no suele advertirse que un derecho fundamental puede ser compatible con varias reglamentaciones, ni se analizan las metodologías que pueden utilizarse en esta materia. La literatura específica sobre derechos fundamentales de la persona imputada, por su parte, si bien se ha ocupado de analizar la existencia de diferentes formas de reglamentación, usualmente lo ha hecho en el marco de algún derecho concreto y de algún tema específico del proceso penal, sin ofrecer un análisis de los conceptos generales que pueden ser comunes a la mayoría de esa clase de discusiones.

Frente a esto, el objetivo del trabajo, tal como se adelantó, es tratar de llenar ese vacío y ofrecer una herramienta útil para canalizar los desacuerdos pueden generarse al discutir sobre la reglamentación de los derechos fundamentales de la persona imputada. Se mostrará que el fenómeno de los desacuerdos no es ajeno a ese tema, se identificarán cuáles son las principales técnicas con las que pueden reglamentarse esos derechos, y se analizará qué razones podrían respaldar la implementación de alguna de esas alternativas, con énfasis en la forma en que esto puede incidir en el cumplimiento de los fines del proceso penal.

---

ejemplos de esto son las tendencias de los tribunales internacionales a considerar admisible una restricción de ese derecho si se encuentra «suficientemente contrarrestada por medidas de contrapeso» y existen «medidas que permitan una evaluación justa y apropiada de la confiabilidad» de una declaración que la persona imputada no tuvo la posibilidad de confrontar (TEDH, *Gran Sala, Al-Khawaja y Tahery v. Reino Unido*, 15/12/2011, §147, con referencia a ambas cuestiones; Corte IDH, *Norín Catrimán v. Chile*, 29/5/2014, §246, con referencia solo a la primera). De forma más general, la argumentación a favor de considerar que la mayoría de los derechos fundamentales son susceptibles de restricciones usualmente se basa en las cláusulas que muchos textos constitucionales incorporan para permitirlo. Suele considerarse que esto vale tanto para los sistemas jurídicos que incluyen expresamente una cláusula de restricción aplicable a todos los derechos (por ejemplo, Canadá, 1; Alemania, 19; Sudáfrica, 36; y en alguna medida Argentina, 28) como en aquellos Estados cuyas constituciones no lo prevén explícitamente (BARAK: 2012: 133; WEBBER, 2009: 11-12). La existencia de esas cláusulas ha generado que el análisis de «proporcionalidad» sea una herramienta que ocupa un lugar central al analizar si la restricción de un derecho fundamental se encuentra justificada (MÖLLER, 2012: 178) y en algunas ocasiones se ha propuesto que las restricciones a los derechos de la persona imputada sean analizadas precisamente con esa herramienta (DENNIS, 2011: 346). Sobre todo esto, con énfasis en la literatura alemana y en la posibilidad de restringir derechos fundamentales más allá de la existencia de disposiciones que expresamente lo permitan, PÉREZ BARBERÁ (2014: 407-411).

### 3. LA DISCRECIONALIDAD JUDICIAL COMO EJE DE LOS DESACUERDOS

Un análisis del tema sensible a las complejidades descritas en la sección anterior requiere un punto de partida útil que permita comenzar a desarrollar las argumentaciones. Por esto, en este apartado se argumenta que existen buenas razones para que el análisis esté centrado en el grado de discrecionalidad que las reglamentaciones de los derechos fundamentales otorgan a los tribunales para tomar decisiones. Esto no implica negar la relevancia que posee el grado de discrecionalidad que se otorga a otros agentes en el proceso penal, como la policía o la fiscalía. Sin embargo, existen buenas razones para poner aquí el foco de la atención en los tribunales<sup>18</sup>.

La «discrecionalidad judicial», en el sentido al que interesa hacer referencia en este trabajo, puede ser definida como el grado de libertad —nulo, escaso, moderado o amplio— con el que cuentan los tribunales para decidir algún asunto<sup>19</sup>. De acuerdo con SCHAUER (2009: 190), el concepto de discrecionalidad, entendido de esa manera, lleva a preguntas como «¿bajo qué circunstancias se asignará a un agente estatal la libertad para aplicar su propio criterio y tomar sus propias decisiones, y bajo cuáles esa libertad será restringida o incluso eliminada casi en su totalidad?». Un ejemplo sencillo puede permitir explicar la idea con claridad. Piénsese en la situación en la que se encuentra quien debe resolver si una persona condujo su vehículo a una velocidad antirreglamentaria mediante la aplicación de una norma que fija con precisión cuál es la velocidad máxima en una determinada zona —por ejemplo, estableciéndola en cien kilómetros por hora—. En ese contexto, quien deba resolver el caso cuenta

---

<sup>18</sup> La relevancia de otros agentes diferentes a los tribunales ha sido advertida en la literatura. En algunas ocasiones se resalta, por ejemplo, respecto de la fiscalía, que «existen buenas razones para que la discrecionalidad de los fiscales constituya un mayor motivo de preocupación que la discrecionalidad judicial, ya que los jueces deben rendir cuentas públicamente en mayor medida que los fiscales» (BRAITHWAITE y PETTIT, 2015: 38). Algo similar sucede cuando se cuestiona la atención en los desarrollos constitucionales centrados en la etapa de juicio en desmedro de los acuerdos o condenas negociadas entre fiscalías y defensas, y se destaca que «en los juicios la Constitución es omnipresente. En los acuerdos, es casi invisible» (STUNTZ, 2006: 791). Sin embargo, la decisión de poner la atención exclusivamente en los tribunales obedece, en primer lugar, a que los tribunales son quienes, de hecho, tienen asignada la toma de decisiones en la mayoría de las materias que constituyen reglamentaciones de derechos fundamentales, tal como se verá en la sección 4 al analizar varios ejemplos de esta clase de normas. En segundo término, este recorte también obedece a que suele argumentarse que eso efectivamente debe ser así, principalmente por la tendencia a atribuir a los tribunales, en tanto autoridades neutrales, la función de resolver disputas (TRECHSEL, 2005: 46). Más allá de estas precisiones, la argumentación que se desarrolla en este trabajo puede ser trasladable a aquellos casos donde se asigna a las fiscalías, y no a los tribunales, la toma de decisiones en algunas materias que constituyen reglamentaciones de derechos fundamentales.

<sup>19</sup> Un análisis basado en la forma en que los sistemas jurídicos regulan la discrecionalidad judicial presupone compartir la idea según la cual los términos de las normas jurídicas puede efectivamente reducir la libertad de los tribunales. Esto puede ser considerado problemático para ciertas posiciones que niegan que eso sea posible. Sin embargo, al tratarse de un problema cuyo análisis excede el marco de este trabajo, ese debate no podrá ser abordado. Al respecto, entre muchos otros, SCHAUER (1988).

con escasa libertad al momento de tomar su decisión: debe limitarse sencillamente a verificar la velocidad en la que conducía la persona y contrastarla con la indicada en la norma. Sin embargo, considérese la situación en la que se encuentra quien debe resolver ese mismo asunto, pero mediante el uso de una norma que le indica que evalúe si la velocidad en la que conducía la persona era «razonable y prudente». Aquí, en cambio, quien deba tomar la decisión cuenta con un margen de discrecionalidad considerablemente elevado para analizar distintas circunstancias y arribar a soluciones diversas. Se trata, como puede advertirse, de uno de los usos del concepto que DWORKIN (1978:32) denominaba «débiles» al señalar que es posible hablar de discrecionalidad cuando las normas, por alguna razón, «no se pueden aplicar mecánicamente, sino que exigen valoración (*judgement*)»<sup>20</sup>.

Un interés particular en ese concepto se encuentra justificado porque varias de las discusiones específicas sobre la reglamentación de los derechos fundamentales de la persona imputada suelen, en cierta medida, girar alrededor de un gran desacuerdo sobre el grado de libertad que debería otorgarse a los tribunales para tomar decisiones. Esto es así, por ejemplo, en los debates acerca de cómo debería determinarse si la duración de un proceso penal es razonable<sup>21</sup>, y en las discusiones sobre qué podrían tener en cuenta los tribunales al analizar la admisibilidad de la declaración de un testigo ausente al momento del juicio<sup>22</sup> y de la prueba

---

<sup>20</sup> En derecho procesal penal siguen este uso del concepto, de forma explícita o implícita, ROBERTS y ZUCKERMAN (2010: 29), DUFF *et al.* (2007: 106) y REDMAYNE (2012: 305). El uso de esa definición, por ende, implica dejar de lado el otro uso «débil» del concepto, como aquel que lo asimila con «la autoridad final para tomar una decisión que no puede ser revisada ni anulada por otro funcionario» o aquel denominado «fuerte», es decir, cuando se lo utiliza para afirmar que un funcionario «en lo que respecta a algún problema, simplemente no está vinculado por estándares impuestos por la autoridad en cuestión» (DWORKIN, 1978:32). Sobre todo esto también CARACCILO (2009: 253-254).

<sup>21</sup> Cfr., entre otros, TRECHSEL (2005: 141), para quien «es esencial que cada caso sea analizado de acuerdo a sus particularidades», y PASTOR (2009: 318), quien considera, en cambio, que los criterios de esa posición «son completamente insatisfactorios para el derecho procesal penal, pues lejos de colaborar con él en su fin principal de impedir la arbitrariedad, la estimulan, al remitir a la voluntad de los jueces la determinación del plazo razonable y sus consecuencias, a través de pautas tan abiertas y difusas».

<sup>22</sup> Cfr., entre otros, REDMAYNE (2012: 305-306), para quien «la confrontación debe ser vista como un derecho positivo a examinar los testigos de cargo disponibles, y no como un derecho negativo que obligue a excluir la evidencia aportada por testigos ausentes», y considera que debería permitirse que los tribunales evalúen las razones por las cuales la persona cuya declaración se pretende excluir no se encuentra presente en el juicio; SPOTTSWOOD (2018: 598-599), para quien corresponde «examinar los costos y beneficios que generaría la decisión de excluir una declaración» y adoptar un enfoque que permita excluirlas «en los casos en los que es particularmente probable que no sean confiables»; ROVATTI (2020: 34), quien afirma que una reglamentación rígida de este derecho es «insostenible»; NANZER (2008: 367), quien considera que si la defensa no contó con la oportunidad de controlar una declaración la única consecuencia posible es la «absoluta, inexorable e irrestricta invalorableidad del testimonio»; Suprema Corte de Justicia de EE.UU., *Crauford v. Washington* (541 U.S. 36), opinión del juez SCALIA («los estándares vagos son manipulables» y «no ofrecen protecciones significativas»); FRIEDMAN (2004: 4-5), quien califica a los enfoques que proponen otorgar una mayor libertad a los tribunales para considerar indicios que demuestren la confiabilidad de la declaración que la defensa no tuvo posibilidad de confrontar como «complejos», «manipulables» e «impredecibles».

ilícita<sup>23</sup>. La lista podría aumentarse mediante el análisis de otras discusiones en las que ese contraste se repite en términos similares. A modo de ejemplo pueden mencionarse algunos desarrollos relacionados con el derecho probatorio —por ejemplo, al analizar el grado de libertad que deberían tener los tribunales para tener por probado un hecho—<sup>24</sup>, los debates acerca de si el *ne bis in idem* permite perseguir nuevamente a una persona cuando el primer proceso tuvo defectos fundamentales<sup>25</sup>, y las discusiones generadas al analizar cómo reglamentar la manera en que los tribunales deben evaluar si una declaración autoincriminatoria es contraria al derecho fundamental de la persona imputada a «no ser obligada a declarar contra sí misma»<sup>26</sup>.

Lo anterior constituye una buena razón para considerar que un análisis centrado en la discrecionalidad judicial puede permitir hacer explícito lo que muchas veces está implícito en los debates sobre la reglamentación de los derechos fundamentales de la persona imputada. Se trata, por esa razón, de un concepto que cuenta con una gran capacidad de rendimiento para expresar los desacuerdos que existen esas discusiones, y puede ofrecer una base firme para el análisis general del tema.

#### 4. UNA TERMINOLOGÍA PARA LOS DESACUERDOS: REGLAS, ESTÁNDARES Y DOS MODELOS PARA REGLAMENTAR DE DERECHOS FUNDAMENTALES

Para continuar con el desarrollo propuesto, el objetivo de esta sección será mostrar cómo los desacuerdos a los que se hizo referencia pueden ser trasladados a una

---

<sup>23</sup> Cfr., entre otros, PIZZI (1999: 43), quien argumenta que no es necesario abordar el problema como una cuestión de «todo o nada», y que existen buenas razones para adoptar una aproximación «más balanceada» que «permita excluir la evidencia cuando ha existido un abuso serio por parte de la policía» y «la niegue cuando la violación es proporcionalmente inferior al delito»; STUART (2010: 313-332), quien sostiene que no existe «sabiduría» en la exclusión «automática» de la prueba obtenida de forma ilícita, y afirma que la aproximación flexible a este problema es más «sensata» y «cuidadosamente balanceada»; MORGENSTERN (2014: 1551), quien considera que «la exclusión probatoria indiscriminada y automática es una solución equivocada» y propone la adopción de un esquema «basado en una ponderación»; PASTOR (2012: 326-327), quien critica a la otra posición con base en que «siempre será posible un margen de juego para un decisionismo judicial que justifique, a pesar de todo, la valoración de la prueba prohibida o ilícitamente obtenida»; THAMAN (2011: 697), quien considera que «un sistema jurídico que toma ventaja de principios de ponderación vagos para utilizar la evidencia obtenida directa o indirectamente en función de actos inconstitucionales de sus órganos de investigación sólo con dificultad incorpora el espíritu de un Estado de derecho»; ASHWORTH (1977:723), quien muestra una gran preocupación por la posibilidad de que la decisión sobre la admisibilidad de esas evidencias quede «librada al capricho de cada tribunal».

<sup>24</sup> Al respecto SCHAUER (2008), STEIN (1997), FERRER BELTRÁN (2013) y DUFRAIMONT (2013). Para un análisis de estos debates con una metodología similar a la propuesta en este trabajo véase LARSEN (2020a).

<sup>25</sup> Sobre este tema RUDSTEIN (2008), DENNIS (2014:231), MORGENSTERN (2015:196) y GUZMÁN (2008: 209-210).

<sup>26</sup> Al respecto KAMISAR (2012), SCHAUER (2013) y MARCUS (2006).

metodología vinculada con la creación de las normas jurídicas. Esto será realizado mediante la presentación de dos modelos diferenciados por la forma en que sus normas jurídicas regulan la discrecionalidad de los tribunales para tomar decisiones. Al analizar sus rasgos característicos, se mostrará que se trata de modelos que permiten identificar con claridad las principales tendencias que existen al reglamentar los derechos fundamentales de la persona imputada.

#### 4.1. La estructura de las normas jurídicas que reglamentan derechos fundamentales de la persona imputada: la dimensión de la «precisión»

Al igual que una enorme cantidad de normas jurídicas, aquellas que reglamentan los derechos fundamentales de la persona imputada constituyen «reglas» en sentido general. Si bien se trata de un concepto sobre el cual se realizarán algunas precisiones terminológicas más adelante, ese término, al menos en sentido amplio, suele ser utilizado en la teoría del derecho para hacer referencia a las reglas «prescriptivas». De acuerdo con algunas definiciones tradicionales, se trata de reglas que poseen «un contenido semántico normativo» y son utilizadas «para guiar, controlar o modificar la conducta de agentes con capacidad para tomar decisiones» (SCHAUER, 2004: 58), tratándose los agentes, en lo que aquí interesa, de los tribunales<sup>27</sup>.

Como toda regla prescriptiva, las normas jurídicas que reglamentan derechos fundamentales de la persona imputada poseen una estructura lógica particular. En términos sencillos, cualquier regla prescriptiva, más allá de cómo esté formulada, puede ser presentada como un enunciado condicional de la forma “si *p*, entonces *q*” ( $p \rightarrow q$ ), donde el antecedente indica las condiciones bajo las cuales la regla se aplica, mientras que el consecuente establece lo que habrá de ocurrir cuando esas condiciones se verifiquen (TWINING y MIERS, 2010: 90, MACCORMICK, 2018: 79, SCHAUER, 2004: 81-82). Las reglas prescriptivas que reglamentan derechos fundamentales también pueden ser fácilmente descompuestas de esa manera. Muchas de esas normas jurídicas, tal como se verá más adelante en detalle, están compuestas por un antecedente donde se describe qué es lo que debe verificarse para que un tribunal tome una decisión y un consecuente que establece cuál es la consecuencia aplicable —típicamente una decisión para la cual la existencia del derecho fundamental del que se trate podría ser un obstáculo—. Un ejemplo de esto es una norma jurídica que reglamente el derecho fundamental a «no ser obligado a declarar contra uno mismo» y prescriba que «si no se notificó a la persona imputada cuáles son sus derechos de forma previa a su declaración autoincriminante (antecedente), *entonces* esa evidencia no es admisible (consecuente)».

---

<sup>27</sup> Por esta característica —estar dirigidas a órganos estatales como los tribunales— las normas jurídicas que reglamentan el proceso penal suelen ser consideradas «reglas de decisión», por oposición a las «reglas de conducta», que tienen como destinatarios al público en general (DRIPPS, 2011: 409).

La exposición de los modelos a los que se hizo referencia estará enfocada en el antecedente de las reglas prescriptivas, es decir, en la parte que describe las condiciones en las cuales se aplican las normas jurídicas, usualmente denominada «hecho operativo» (MACCORMICK, 2018: 79)<sup>28</sup>. De manera más específica, por tratarse de una circunstancia directamente relacionada con la discrecionalidad judicial, aquello que se tendrá en cuenta es una dimensión concreta de esos hechos operativos: la *precisión* de los términos utilizados al redactar esa parte de las normas jurídicas, usualmente considerada como un «dispositivo muy efectivo para el manejo de la discrecionalidad» (SCHAUER, 2009: 190).

Un recurso muy útil para esto, como se verá enseguida, es clasificar a las normas jurídicas según el grado de precisión con el que cuentan. De acuerdo con una tipología ampliamente difundida en la literatura anglosajona, la regulación de una actividad puede ser realizada mediante dos grandes alternativas conocidas como «reglas» —o normas jurídicas más *específicas*— y «estándares» —o normas jurídicas más *vagas*—, consideradas como dos extremos de un continuo en el cual las primeras representan el máximo en materia de precisión y los segundos, en cambio, el mínimo<sup>29</sup>. Si bien se volverá enseguida sobre las características de ambas modalidades, un ejemplo permite ilustrar con claridad sus diferencias. Una norma según la cual «quien circule a más de 120 km/h será sancionado con una pena de diez días de prisión» contiene un hecho operativo redactado bajo la modalidad de una regla. En cambio, una norma que establece que recibirá la misma sanción «quien circule *de manera irrazonable o imprudente*» contiene un hecho operativo que asume la forma de un estándar.

El uso de la terminología «reglas» y «estándares», pese a resultar a primera vista lo suficientemente claro, requiere algunas precisiones adicionales para despejar ciertos equívocos, en particular aquellos generados cuando se acude a una terminología similar para relacionar algunos de esos conceptos con dimensiones diferentes a la precisión de las normas jurídicas, como el *peso* que éstas deberían tener en el razonamiento de los tribunales. Un ejemplo conocido de este uso de la terminología es la posición que contrapone a las «reglas» con los «principios» y afirma que las reglas son «precisas y absolutas» mientras que los principios son «vagos y derrotables» (DWORKIN, 1978: 22). Sin embargo, existen buenas razones para dejar de lado este

---

<sup>28</sup> Esta decisión obedece a que es en los hechos operativos o antecedentes de las normas (y no en los consecuentes) donde las reglamentaciones de derechos fundamentales presentan las diferencias que interesa analizar en este trabajo. Como se verá enseguida, el consecuente de estas reglamentaciones usualmente es el mismo más allá del grado de precisión que se utilice al redactar el antecedente, y se encuentra definido con precisión. Por mencionar algunos ejemplos, si bien muchas reglamentaciones presentan notables diferencias al establecer qué pueden tener en cuenta los tribunales antes de decidir si un determinado elemento de prueba es admisible (por ejemplo, una declaración autoincriminatoria, una prueba obtenida de forma ilícita, o un testimonio de cargo no controlado por la defensa), el consecuente, en todos esos casos, es el mismo: o bien la evidencia es admisible o inadmisible.

<sup>29</sup> Al respecto, SCHAUER (2003: 309) explica que «aunque por supuesto que hay casos intermedios, la existencia de casos en la frontera entre reglas y estándares no invalida la distinción entre ellos más de lo que la existencia del atardecer destruye la distinción entre el día y la noche».

tipo de clasificaciones. Como advierte SCHAUER (2006: 889, nota 32), atribuir un determinado *peso* a un nivel de *precisión* oscurece el análisis en lugar de aclararlo, pues ambas dimensiones operan de forma separada. En términos de ese autor, esto promueve «una considerable cantidad de confusión» al «asumir que las dimensiones de precisión y peso operan en tándem, cuando de hecho parecen ser considerablemente independientes una de otra», pues «existen reglas precisas pero derrotables, y también principios vagos pero muy estrictos o incluso absolutos, como sucede con el vago, pero no derrotable, imperativo categórico de Kant»<sup>30</sup>. De todos modos, al tratarse de una estipulación terminológica, también puede hablarse de «reglas precisas» y «reglas vagas» para referir a la idea que se quiere reflejar con la distinción entre «reglas» y «estándares» (así, por ejemplo, TWINING y MIERS, 2010: 84). Lo que debe quedar en claro es que, cualquiera sea la terminología que se utilice, la argumentación estará referida únicamente a la *precisión* de los términos utilizados en la redacción de las normas jurídicas, y no al *peso* que estas pueden tener en la toma de decisiones de los tribunales.

#### 4.2. Reglas y estándares en la reglamentación de los derechos fundamentales de la persona imputada

Una alternativa disponible al regular la manera en que los tribunales deben tomar sus decisiones consiste en utilizar «reglas». Esta modalidad de normas jurídicas se caracteriza por utilizar términos considerablemente precisos o, dicho de otra manera, por emplear palabras «detalladas, específicas, concretas y determinadas» (SCHAUER, 2009: 189; de modo similar ALEXANDER y SHERWIN, 2001: 29, BRAITHWAITE, 2002: 47, PARCHOMOVSKY y STEIN, 2015: 166). Cuando se procede de esa manera, suele afirmarse que se trata de reglamentaciones que adoptan una perspectiva predominantemente *ex ante*, donde los creadores de las normas se adelantan a la aparición de los casos concretos y establecen con precisión, de forma previa, cuál es la forma en que deberán ser resueltos (KAPLOW, 1992: 559-560, SUNSTEIN, 2018: 22; SCHAUER, 2003: 305, WALDRON, 2011: 5). En esos contextos, la función de quienes aplican las normas suele quedar limitada a constatar los aspectos concretamente descritos en la regla y aplicar la solución allí prevista. Por esa razón, la característica que define a este modelo es una *reducción* considerable de la discrecionalidad.

En muchas ocasiones los derechos fundamentales de la persona imputada son reglamentados mediante esa clase de normas. Cuando esto sucede, quienes reglamentan los derechos, mediante la adopción de reglas estrictas, se encargan de describir de forma considerablemente precisa los hechos operativos que los tribunales deben tener en cuenta al tomar sus decisiones y, de ese modo, reducen la cantidad de circunstancias que pueden ser consideradas relevantes para resolver casos. Un ejemplo

---

<sup>30</sup> Del mismo modo TWINING y MIERS (2010: 83), con el ejemplo de la prohibición *absoluta* de algo considerablemente *indeterminado* como la «tortura».

de esa alternativa es una forma bastante habitual de reglamentar las ocasiones en que la declaración de un testigo ausente al momento del juicio puede ser utilizada como evidencia en contra de la persona imputada —una reglamentación relacionada con el derecho fundamental a «interrogar a los testigos de cargo»—. Al reglamentar esta cuestión, muchos sistemas procesales suelen adoptar una regla cuyo hecho operativo establece que lo único que los tribunales pueden tener en cuenta para tomar esa decisión es si la defensa «contó con la posibilidad de interrogar al testigo» pues, de lo contrario, esa evidencia será inadmisibile. Considérese, por ejemplo, el siguiente ejemplo paradigmático de este tipo de reglamentación:

Quando se trate de prueba testimonial, [el derecho a la confrontación] requiere la no disponibilidad del testigo y una oportunidad previa de interrogatorio<sup>31</sup>.

La forma en que una reglamentación como la mencionada puede reducir la discrecionalidad judicial es fácil de advertir. Al especificar de manera considerablemente clara qué es aquello que puede ser valorado para tomar una decisión, estas reglamentaciones privan a los tribunales de incluir en sus razonamientos otras circunstancias que podrían ser tenidas en cuenta para arribar a una solución distinta —en este caso, la admisión del testimonio—. Al prescribir que lo único que interesa para la decisión del caso es si la defensa contó con «una oportunidad previa de interrogatorio», esas reglamentaciones tornan irrelevante, por ejemplo, el análisis acerca de las razones por la cual el testigo no puede estar presente en el juicio, o los indicios de confiabilidad que pueda presentar su declaración anterior.

Sin embargo, las reglas no son la única alternativa disponible en esta materia. Como se señaló más arriba, otro de los caminos que se pueden seguir al regular un asunto consiste en acudir a «estándares», es decir, a una modalidad de normas caracterizada por usar términos «amplios, vagos, generales e imprecisos» (SCHAUER, 2009: 189; de modo similar ALEXANDER y SHERWIN, 2001: 29, BRAITHWAITE, 2002: 47, PARCHOMOVSKY y STEIN, 2015: 166). A diferencia de lo que sucede con las reglas, este tipo de normas se inclina por no establecer de antemano ni de manera precisa cuál es la solución aplicable al caso concreto *ex ante*, sino que, en cambio, adopta una perspectiva predominantemente *ex post* que deja la mayor parte de la tarea para ser realizada «caso a caso» por los aplicadores (KAPLOW, 1992: 559-560, SUNSTEIN, 2018: 28, SCHAUER, 2003: 305, WALDRON, 2011: 5). Cuando esto es así, el papel de quienes aplican las normas, a diferencia de lo que sucede con las reglas, adquiere una mayor relevancia en la solución de los casos concretos. La nota característica de este modelo, en consecuencia, es una *ampliación* notable de la discrecionalidad.

Algunos sistemas suelen incluir reglamentaciones de los derechos fundamentales de la persona imputada que adoptan la modalidad de estándares. En esas oportu-

<sup>31</sup> Suprema Corte de Justicia de EE.UU., *Crawford v. Washington* (541 U.S. 36). Pueden encontrarse reglamentaciones similares en las Reglas de Procedimiento y Prueba de la CPI (68.2, texto previo a la modificación introducida mediante ICC-ASP/12/Res.7), en el CPP Federal argentina (ley n. 27.063, 289.c), y en la jurisprudencia del Tribunal Constitucional de España (sentencia n. 134/2010).

nidades, los hechos operativos de las normas son descriptos de una forma considerablemente vaga que amplía las circunstancias que los tribunales pueden tener en cuenta al tomar sus decisiones, y que suelen ser excluidas cuando la misma cuestión es reglamentada mediante reglas. Esto puede advertirse con claridad si se contrasta el ejemplo analizado más arriba —la decisión sobre la admisibilidad de la declaración de un testigo ausente en el juicio— con una reglamentación del mismo asunto que, en lugar de acudir a reglas, procede, en cambio, a utilizar estándares. En ese escenario, las reglamentaciones pueden describir su hecho operativo de una forma tal que no se limite a establecer que los tribunales sólo pueden tener en cuenta si la defensa contó con la posibilidad de interrogar al testigo, sino que les permita valorar circunstancias descriptas de forma particularmente vaga como los «indicios de fiabilidad» con los que cuente la declaración, las «razones por las cuales el testigo no puede estar presente en el juicio», o el «contexto en que efectuó su declaración anterior». Un ejemplo notable de una reglamentación basada en estándares como los mencionados es el siguiente:

La declaración de un testigo ausente es admisible solo si cuenta con indicios de confiabilidad<sup>32</sup>.

No resulta difícil advertir que en este modelo un tribunal cuenta con un grado de discrecionalidad más elevado que aquel que le asignaría una reglamentación mediante reglas. La forma considerablemente vaga en que se describen las circunstancias que los tribunales pueden incluir en sus razonamientos les permite tener en cuenta mucho de aquello que resulta excluido en la reglamentación mediante reglas y, en consecuencia, les otorga la posibilidad de arribar a soluciones diversas en cada caso.

Todo lo desarrollado hasta aquí podría tratarse de una disquisición terminológica poco útil si la distinción propuesta fuese apta solo para describir formas alternativas de reglamentar unos pocos derechos fundamentales de la persona imputada, o si se tratase de una metodología que permite captar la forma de reglamentar solo algunos asuntos marginales del proceso penal. Sin embargo, un análisis de varias reglamentaciones de países de distintas tradiciones jurídicas permite advertir que el contraste se repite en un número nada despreciable de ocasiones.

La contraposición entre reglas y estándares se advierte, por ejemplo, en las formas posibles de reglamentar la forma en que los tribunales deben tomar decisiones en casos relacionados con la duración del proceso (regla que establezca que si el proceso no puede continuar si superó una determinada cantidad de tiempo vs. estándares que permitan arribar a esa solución si la duración del proceso es razonable, teniendo en cuenta la complejidad del caso, el comportamiento del imputado y la conducta de las autoridades judiciales)<sup>33</sup>. Lo mismo sucede con la reglamentación de la admisibi-

<sup>32</sup> Suprema Corte de Justicia de EE.UU en *Ohio v. Roberts* (448 U.S. 56). En sentidos similares la Criminal Justice Act del Reino Unido (114, 1.d y 2), las Reglas de Procedimiento y Prueba de la CPI (68.2, texto según la modificación introducida mediante ICC-ASP/12/Res.7), y la Evidence Act de Australia (65).

<sup>33</sup> Sobre la primera alternativa cf. CPP Colombia (175). Sobre la segunda cf. TEDH, *Philis v. Grecia*, 27/6/1997, §35; Corte IDH, *Genie Lacayo vs. Nicaragua*, 29/1/1997, §77; Suprema Corte de

lidad de la prueba ilícita (regla que la considere inadmisibles si fue obtenida mediante violación a las formas procesales o derechos fundamentales vs. estándares que establecen esa solución solo para los casos donde la admisión de esa evidencia afecte la integridad del proceso, teniendo en cuenta la magnitud de la afectación de derechos, la gravedad del delito y la conducta de las autoridades)<sup>34</sup>; y con la toma de la decisión final sobre los hechos probados (reglas que impiden tener por probado un hecho si solo se cuenta con un determinado tipo de prueba vs. estándares que permiten tener por probado un hecho si el tribunal alcanzó certeza más allá de toda duda razonable, o si lo considera acreditado luego de apreciar las pruebas según su conciencia, o de valorarlas mediante las reglas de la sana crítica)<sup>35</sup>. Otros ejemplos son las formas en que puede reglamentarse la manera en que los tribunales deben decidir si es posible permitir una nueva persecución penal (regla según la cual esto no es posible si existió un proceso penal previo sobre el mismo hecho vs. estándar que lo permite si el primer proceso fue incompatible con la intención de someter a la persona imputada a la acción de la justicia)<sup>36</sup>; y el modo en que deben evaluar si una declaración autoincriminatoria fue voluntaria (regla que obligue a excluir esa declaración si fue efectuada sin respetar determinadas formalidades vs. estándar según el cual esa declaración es admisible si, de acuerdo a las circunstancias particulares del caso, fue voluntaria)<sup>37</sup>.

Esos ejemplos permiten apreciar que las principales alternativas al momento de reglamentar derechos fundamentales de la persona imputada pueden reducirse a la contraposición de dos modelos caracterizados por el grado de discrecionalidad que en cada uno se asigna a los tribunales. Se trata de modelos contrapuestos, precisamente, en cuanto a la precisión de los términos empleados al redactar los hechos operativos de las normas jurídicas: reglas o estándares. La identificación de esas características y su clasificación mediante modelos permite, entonces, contar con una terminología útil para canalizar los desacuerdos que pueden generarse al discutir sobre el tema.

---

Justicia de EE.UU, *Barker v. Wingo* (407 U.S. 514).

<sup>34</sup> Sobre la primera alternativa cf. CPP Colombia (23 y 360); CPP Federal argentina (129); Ley Orgánica del Poder Judicial España (11.1). Sobre la segunda cfr. Estatuto CPI (60.7); Police and Criminal Evidence Act del Reino Unido (78.1); Evidence Act de Nueva Zelanda (30); Tribunal Federal Supremo Alemania, BGHSt 38 (214, 219 s.).

<sup>35</sup> Sobre la primera alternativa cfr. ley n. 27.304 de Argentina, 15 (imputado arrepentido); Corte IDH, *Ruano Torres v. El Salvador*, 5/10/2015, §133 (coimputado); Código Penal Modelo de EE.UU, 213.6, 5 (testimonio no corroborado de la presunta víctima). Sobre la segunda cfr. CPP Nación Argentina, 398 (sana crítica); Ley de Enjuiciamiento Criminal de España, 471 (según su conciencia); Estatuto CPI, 66.3 (más allá de toda duda razonable).

<sup>36</sup> Sobre la primera alternativa cfr. CPP Chile (1); CPP Federal argentina (5). Sobre la segunda cfr. Estatuto CPI (20.3); CPP Colombia (192); Criminal Procedure and Investigations Act del Reino Unido (54).

<sup>37</sup> Sobre la primera alternativa cfr. Suprema Corte de Justicia de EE.UU, *Miranda v. Arizona* (384 U.S. 436) (si no se notificaron a la persona imputada de sus derechos de forma previa); CPP Federal argentina (74). Sobre la segunda cfr. legislación federal estadounidense (18 U.S. Code § 3501); Police and Criminal Evidence Act del Reino Unido (76.2); Suprema Corte de Justicia de EE.UU, *Brown v. Mississippi* (297 U.S. 278).

Al tratarse de modelos, por supuesto, siempre existe la posibilidad de que éstos no representen la totalidad de las reglamentaciones imaginables o posibles. Por un lado, nada impide, por ejemplo, que algunos sistemas adopten reglamentaciones mixtas que incorporen aspectos de ambos modelos<sup>38</sup>. Por el otro, en la medida en que el análisis aquí propuesto pone el foco de la atención en la *forma* y no en el *contenido* de las normas, es posible concebir reglamentaciones basadas en reglas o estándares con un contenido completamente distinto a los de las normas que aquí se expusieron como muestras de cada modelo. Por ejemplo, una reglamentación del derecho a un juzgamiento en un plazo razonable o sin dilaciones indebidas puede establecer que los tribunales deben resolver los casos mediante la evaluación de factores como «la situación económica de la persona imputada» o «el contexto social del país» y, pese a esto, debe ser considerada igualmente una reglamentación basada en estándares. Sin embargo, si se presupone que quienes reglamentan los derechos son agentes razonables, esas reglamentaciones, sea cual sea la forma que asuman, no deberían tener un contenido arbitrario sino, en cambio, relacionado con las razones subyacentes a los derechos que se reglamentan. Si se parte de esa idea, los modelos que aquí se exponen permiten cubrir gran parte de las alternativas disponibles, al menos tal como se encuentran en las principales tendencias en la materia.

## 5. LA PRECISIÓN DE LAS NORMAS JURÍDICAS COMO HERRAMIENTA: ENTRE LA IMPOSICIÓN DE LÍMITES FUERTES AL PODER ESTATAL Y LA EFICIENCIA EPISTÉMICA DEL PROCESO

En la sección anterior se ofrecieron dos modelos útiles para identificar cuáles son las principales formas disponibles al momento de reglamentar derechos fundamentales de la persona imputada. En este apartado, para completar el desarrollo propuesto en el trabajo, se analizarán las razones que pueden ofrecerse para optar por alguno de esos modelos.

---

<sup>38</sup> Precisamente por este motivo es que en alguna ocasión se ha advertido que la dicotomía «reglas» y «estándares», pese a su amplia difusión, no agota la totalidad de formas de regulación posibles y que existen otras como los «catálogos» (PARCHOMOVSKY y STEIN, 2015). Algunos ejemplos de reglamentaciones de derechos «mixtas» son aquellas que reglamentan el derecho a ser juzgado en un plazo razonable mediante una regla que fija con precisión los plazos máximos de duración del proceso y, a su vez, incorporan un estándar que permite continuar el proceso si, pese a haber transcurrido los plazos indicados en la regla, existe una «buena causa» para no finalizarlo (Código Penal del Estado de California, Estados Unidos, sección 1382). Otro ejemplo es la forma en que la Corte IDH ha interpretado las exigencias derivadas de la presunción de inocencia, pues ese tribunal, si bien ha concluido que ese derecho fundamental obliga a adoptar el estándar de «más allá de toda duda razonable», también ha adoptado una regla que impide fundar una condena exclusivamente en la declaración de un coimputado (Corte IDH, *Ruano Torres v. El Salvador*, 5/10/2015, §126 y ).

Para esto, a diferencia de algunos enfoques que presentan a una de las alternativas como íntimamente relacionada con el Estado de Derecho o *Rule of Law*<sup>39</sup>, se analizará la posibilidad de adoptar un enfoque «instrumental» al determinar cuál es el grado de precisión con el que deben redactarse las normas. Desde esta perspectiva, la precisión con la que se reglamenta una determinada cuestión —y, en función de esto, la forma en que se amplía o reduce la discrecionalidad de quienes toman decisiones— no es algo inocuo sino que, por el contrario, puede ser un instrumento útil para alcanzar distintos objetivos. Tal como advierte SCHAUER (1987: 37), «una vez que reconocemos que tenemos control sobre la especificidad de nuestras instituciones, podemos comenzar a utilizar la especificidad como una herramienta». Si se sigue ese enfoque y se tienen en cuenta los rasgos característicos de los modelos desarrollados en este trabajo, el interrogante que debería formularse al reglamentar los derechos fundamentales de la persona imputada podría ser el siguiente: ¿qué objetivos podemos tener en mente, cuando se trata del proceso penal, al proponer la adopción de reglamentaciones de derechos más precisas que reducen la discrecionalidad judicial (reglas) o, en cambio, más vagas que la amplían (estándares)?

Para responder ese interrogante puede ser útil acudir a dos valores muy conocidos en la literatura sobre derecho procesal penal. De acuerdo con una idea muy difundida, la principal tensión en el proceso penal, a grandes rasgos, oscila entre la búsqueda de la verdad y la imposición de límites al poder estatal en esta materia. Así, por ejemplo, se destaca que el proceso penal «debe respetar la dignidad de los ciudadanos protegiéndolos de la opresión del Estado» (DUFF *et. al.*, 2007: 95) y que «la estructura del proceso, las reglas con las que se lo organiza, reflejan este objetivo tanto como lo hacen con el objetivo de buscar la verdad. La búsqueda de este objetivo [...] se encuentra restringida por la exigencia de que los ciudadanos sean protegidos de la opresión y que se respete su dignidad» (DUFF *et. al.*, 2007: 95)<sup>40</sup>.

En lo que sigue, con apoyo en esa descripción tradicional sobre los objetivos del proceso penal, se argumentará que quienes quieran acercarse más a la imposición de límites al poder estatal cuentan con buenas razones para acudir al modelo de reglas, mientras que quienes quieran aproximarse más al objetivo de la búsqueda de

---

<sup>39</sup> Al tratarse de una decisión que implica ampliar o reducir la discrecionalidad judicial, la alternativa que busca reducirla —en este caso, el modelo basado en reglas— suele ser vinculada con alguno de los valores tradicionalmente asociados al Estado de Derecho o *Rule of Law* (SCALIA, 1989). De acuerdo con esta idea, la existencia de normas jurídicas poco precisas, que amplían la discrecionalidad judicial, debería ser entendida inevitablemente como un problema. Sin embargo, por un lado, la postura contraria también justifica la ampliación de la discrecionalidad de quienes aplican las normas con los valores que emergen del Estado de Derecho o *Rule of Law* (WALDRON, 2016: 328) y, por el otro, existe una gran dificultad para determinar precisamente qué debe entenderse comprendido por ese concepto (WALDRON, 2008). Ambas son, pues, buenas razones para, al menos en este trabajo, no optar por ese camino. Sobre esta discusión también CARACCILO (2009).

<sup>40</sup> Pueden encontrarse descripciones similares tanto en autores de tradición continental-europea (ROXIN, 2000: 4-5, VOLK, 2016: 45, MAIER, 2016:81, FERRAJOLI, 2014:604) como anglosajona (PACKER, 1964, DUFF *et. al.*, 2007: 61-62, 94-95, ASHWORTH y REDMAYNE, 2010: 23).

la verdad, en cambio, cuentan con buenas razones para acudir al modelo de estándares. Los argumentos a favor y en contra de la adopción de reglas y estándares, por supuesto, exceden a esa dicotomía<sup>41</sup>. Sin embargo, al tratarse de valores íntimamente vinculados con algunas preocupaciones relevantes para el diseño del proceso penal, esto justifica que la atención esté puesta sobre ellos.

Si bien se volverá enseguida con más detalle sobre estos objetivos y la forma en que los modelos pueden ser herramientas útiles para alcanzarlos, es necesario precisar dos cuestiones antes de continuar. En primer lugar, es importante tener en cuenta que esa tensión entre valores existe y que cumplir con ambos de forma óptima, en simultáneo, puede ser algo muy difícil o imposible de lograr. Esto obedece a que, como se verá más adelante, en algunos casos la imposición de límites muy fuertes al poder estatal puede disminuir la eficiencia epistémica del proceso, mientras que, para aumentar esto último, en determinadas ocasiones puede ser necesario disminuir la intensidad de esos límites. En segundo lugar, precisamente por lo que se acaba de afirmar, es necesario advertir que ambos valores representan dos extremos a los que los modelos pueden acercarse en mayor o menor medida, de forma gradual. Por esa razón sería un error afirmar, de forma general, que un modelo cumple con un objetivo en su totalidad y renuncia al otro, y viceversa. Esto es, no sería correcto concluir que un modelo que adopta reglas, por reducir la discrecionalidad judicial e imponer límites más fuertes al poder estatal, renuncia por completo a la búsqueda de la verdad, ni tampoco sería acertado sostener que un modelo basado en estándares, por ampliar la discrecionalidad de los tribunales, renuncia a la imposición de límites al poder estatal. De lo que se trata, con relación a esto último, es de la *intensidad* de esos límites, y no de si estos existen o no<sup>42</sup>. Sí puede ser correcto, en cambio, tal como se intentará mostrar, que cada uno de estos modelos muestra una inclinación preferencial por el cumplimiento de uno de esos objetivos. Y esta posibilidad de aproximarse más al cumplimiento de un valor o el otro sin que la adopción de una de esas alternativas pueda ser considerada *prima facie* inválida es lo que, según se sostuvo en la introducción, puede mostrar la utilidad de estos modelos para las discusiones sobre el tema.

---

<sup>41</sup> Para un análisis general véase SCHAUER (2009: 188-202) y SUNSTEIN (2018: 21-31, 127-133).

<sup>42</sup> Sí sería problemático, en cambio, adoptar un modelo que otorgue a los agentes estatales una «discrecionalidad sin trabas» para la búsqueda de la verdad, y les confiera la «capacidad de usar el poder estatal de la forma en que lo elijan, mediante las consideraciones que quieran tener en cuenta, valoradas del modo en que las quieran valorar» (SUNSTEIN, 2018: 21). Pero los estándares, como ya se vio, efectivamente ponen límites a la discrecionalidad, pese a que sean más débiles que otros. Por esa razón, no pueden ser encasillados en esta categoría.

### 5.1. Reglas e imposición de límites fuertes al poder estatal

La necesidad de imponer límites al poder del Estado suele ser una preocupación compartida a nivel general. Un ejemplo de esto son las ideas asociadas con el constitucionalismo, al cual, al menos en su versión tradicional, suele atribuírsele la misión de «controlar, limitar y restringir el poder del Estado» (WALDRON, 2018: 33). Cuando se trata del poder penal, esta preocupación suele acentuarse de forma notable. La idea que subyace a esta actitud es fácil de comprender: si se considera necesario imponer límites al poder estatal en general, esos límites deberían ser aún más fuertes cuando se trata de la manifestación más intensa de ese poder, el cual puede ser utilizado para oprimir a la persona imputada, usualmente considerada como la parte más débil en un proceso penal (TRECHSEL, 2005: 96). De acuerdo con esa forma de aproximarse al asunto:

La combinación de estigma y pérdida de la libertad que representa el resultado final del proceso penal es vista como la privación más grande que el gobierno puede imponer a un individuo [...] El poder siempre está sujeto a abusos, algunas veces sutiles, y otras, como en el proceso penal, explícitos y peligrosos. Precisamente por su potencia para sujetar al individuo al poder coercitivo del Estado, el proceso penal debe [...] ser sujeto a controles y salvaguardas que prevengan que opere con la máxima eficiencia (PACKER, 1964: 16).

Quienes compartan estas preocupaciones por razones de principio, tal como suele enfatizarse muchas veces en la literatura<sup>43</sup>, pueden contar con buenas razones para mostrarse escépticos acerca de la posibilidad de otorgar mucha discrecionalidad a los tribunales cuando se trata de materias relacionadas con derechos fundamentales de la persona imputada. Lo mismo puede suceder con quienes puedan asumir esa actitud con respaldo en circunstancias contingentes y propias de algún contexto particular, tal como sucede, por ejemplo, en países que se encuentran en una transición de gobiernos dictatoriales a democráticos<sup>44</sup>.

<sup>43</sup> Cfr. DUFF *et al.* (2007: 94), quienes destacan la «necesidad de proteger a los ciudadanos contra el poder opresivo del Estado»; GRECO (2007: 19), quien sostiene que «para el pensamiento liberal parece casi constitutiva una desconfianza sana frente al Estado», al cual lo define como «el Leviatán, o sea, una amenaza constante para los derechos de sus sometidos»; ANDRÉS IBÁÑEZ (2017: 102), para quien «un rasgo central del Estado de derecho es que ‘desconfía de sí mismo’, hasta el punto de que, en su arquitectura y en su funcionamiento ideales, podría verse sin dificultad [...] la plasmación de un articulado sistema de desconfianzas»; y FERRAJOLI (2014: 210), quien afirma que el poder penal estatal representa «la manifestación más violenta, más seriamente lesiva de los intereses fundamentales del ciudadano y en mayor medida susceptible de degenerar en arbitrariedad».

<sup>44</sup> SCHAUER (2009: 192-193), cita como ejemplo la adopción de reglas en determinadas áreas del derecho de Sudáfrica y su posible explicación en la desconfianza en la discrecionalidad en el ámbito del sistema penal por los abusos cometidos por policías, fiscales y jueces durante la época del Apartheid. En un sentido similar, THAMAN (2011: 733) destaca que «sólo unas pocas jurisdicciones, notablemente aquellas que han emergido recientemente de regímenes totalitarios o autoritarios [...] han anulado todo tipo de ponderación cuando se trata del uso de cualquier evidencia obtenida de forma ilegal».

Las reglas pueden ser una herramienta muy atractiva para trasladar esas preocupaciones al diseño del proceso penal e imponer límites fuertes al uso del poder estatal. Si se trata de optar entre una alternativa que reduce el poder de los tribunales (reglas) y una que lo amplía (estándares), y se considera que es necesario imponer límites muy fuertes al poder estatal en esta materia, inclinarse por reglas parece una decisión sumamente intuitiva. En muchas ocasiones, de hecho, suele presentarse a esta idea como un argumento central a favor del uso de reglas. Así, por ejemplo, se afirma que «el escepticismo hacia la discrecionalidad llevará naturalmente a seleccionar reglas en lugar de estándares» (SCHAUER, 2009: 193), que las reglas pueden ser preferibles para «minimizar los abusos de poder» (KAPLOW, 1992: 609), que un estándar «al otorgar una amplia discreción a los encargados de aplicarlo, abre el camino a los abusos» (POSNER, 1990: 44), o que las reglas pueden reducir la posibilidad de que «consideraciones ilegítimas influyeran en las decisiones» (SUNSTEIN, 2018: 130).

¿Cómo permiten las reglas materializar este objetivo? Al establecer de manera precisa aquello que puede ser tenido en cuenta al momento de tomar una decisión, las reglas excluyen otro tipo de circunstancias que los tribunales, en ausencia de reglas, podrían incluir en su razonamiento para arribar a una solución distinta. Las reglas, tal como se ha destacado, permiten alcanzar ese tipo de objetivos porque «excluyen de la consideración algunas propiedades del hecho particular que un procedimiento de decisión particularista reconocería» (SCHAUER, 2004: 138) y se diferencian de éste precisamente porque «algunas características del hecho, que de otra manera serían relevantes en la decisión, son actual o potencialmente ignoradas por la generalización sub y sobreincluyente que constituye el predicado fáctico de cualquier regla» (SCHAUER, 2004: 138).

Al analizar las características del modelo basado en reglas se expusieron varias manifestaciones de ese fenómeno. Por ejemplo, una reglamentación del derecho a interrogar a los testigos de cargo que excluya cualquier declaración que el imputado no haya podido controlar impide que los tribunales, para admitir esas declaraciones, tengan en cuenta circunstancias como los «indicios de fiabilidad» que puedan presentar. Del mismo modo, una reglamentación del derecho a no ser juzgado más de una vez por un mismo hecho que impida iniciar un nuevo proceso por el solo hecho de que la persona ya haya sido sometida a uno con anterioridad impide que los tribunales, para permitir el inicio de una nueva persecución, valoren circunstancias considerablemente indeterminadas como si en el proceso previo hubo «un interés real en someter al acusado a los intereses de la justicia». Asimismo, una reglamentación del problema de la exclusión de la prueba ilícita mediante una regla estricta que obligue a excluir esa prueba cuando haya sido obtenida mediante una violación a las formas procesales o los derechos fundamentales de la persona imputada priva a los tribunales de considerar, para admitir esa evidencia, factores como «la gravedad del hecho» y similares.

Por esas razones, el modelo de reglas puede ser una alternativa ideal para quienes deseen acercarse más al cumplimiento del objetivo de imponer límites muy fuertes

al poder estatal. Si bien tanto las reglas como los estándares imponen límites, los estándares, al tratarse de límites poco precisos que operan caso a caso, pueden ser insatisfactorios para quienes consideren que, cuando se trata del proceso penal, esos límites deben ser particularmente fuertes. Y es aquí donde las reglas pueden cumplir un papel central para diseñar el proceso penal de un modo tal que se acerque más a ese objetivo.

Al sostener que las reglas pueden ser una herramienta atractiva para imponer límites fuertes al poder estatal y proteger a la persona imputada, por supuesto, no se pretende afirmar que existe una relación de carácter conceptual entre ambas cuestiones. Por el contrario, esa relación es contingente, pues las reglas, al tratarse de un instrumento, en algunas ocasiones pueden estar al servicio del objetivo directamente opuesto. Esto puede suceder, por ejemplo, si quienes reglamentan los derechos desconfían de los tribunales por el temor a que favorezcan mucho a la persona imputada, y deciden adoptar soluciones drásticas para evitar que esto suceda. Sin embargo, estas alternativas, que llevarían a concebir reglas con un contenido muy distinto a las que fueron expuestas al analizar este modelo, no se encuentra entre las principales tendencias en esta materia y, por ende, no es necesario ingresar aquí en esta discusión<sup>45</sup>.

---

<sup>45</sup> Ejemplos de estas reglas pueden ser reglamentaciones que prohíban categóricamente la exclusión de la prueba ilícita en todos los casos, permitan la admisión automática de declaraciones testimoniales que la persona imputada no haya tenido posibilidad de confrontar, o incluso obliguen a dictar una condena si se cuenta con un determinado tipo de evidencia. La posibilidad de concebir reglamentaciones con ese contenido permite advertir que la reducción de la discrecionalidad judicial mediante el uso de reglas no solo es útil para proteger fuertemente a la persona imputada, sino que puede tratarse de un «arma de doble filo». Si bien esta investigación no puede ocuparse en detalle de esa problemática, es posible afirmar que existen buenas razones para considerar que este tipo de reglamentaciones cuentan con problemas para justificar su validez jurídica. En primer lugar, existe la posibilidad de considerar algunas de estas posibles reglas como reglamentaciones que no satisfacen un posible «contenido mínimo» del derecho fundamental que reglamentan. Esto podría afirmarse respecto de la existencia de reglas que, por ejemplo, reglamenten el derecho fundamental a un «debido proceso» de modo tal que se impida excluir, en cualquier circunstancia, la prueba ilícita. En ese caso, puede considerarse que una reglamentación que cierre los ojos ante cualquier tipo de violación a las formas procesales al analizar la admisibilidad de esa prueba no cumple con el contenido mínimo del derecho a un debido proceso. Similares críticas podrían formularse a la existencia de reglas que, al reglamentar el derecho fundamental a «interrogar a los testigos de cargo», permitan incluir automáticamente cualquier declaración de testigos ausentes en el juicio y que la persona imputada no haya tenido posibilidad de confrontar. En este supuesto, permitir la admisión de esas declaraciones sin la existencia de un análisis sobre los factores que pueden ser relevantes de acuerdo a la razón de ser de ese derecho —por ejemplo, si ese testimonio cuenta con indicios de fiabilidad— podría afectar la validez de la reglamentación de la que se trate (un argumento similar es el que parece desprenderse de la forma en que el TEDH actualmente considera que una reglamentación del derecho a interrogar a los testigos de cargo es válida siempre que existan «medidas que permitan una evaluación justa y apropiada de la confiabilidad» de una declaración que la persona imputada no tuvo la posibilidad de confrontar (*Gran Sala, Al-Khawaja y Tahery v. Reino Unido*, 15/12/2011, §147). En segundo término, algunas de esas posibles reglas pueden ser criticadas desde una perspectiva diferente. Me refiero a aquellas reglas que imponen tener por probadas determinadas proposiciones fácticas. En este campo podrían encontrarse las reglamentaciones que, a diferencia de las reglas analizadas en este trabajo, no *impiden* a los tribunales tener por probado un hecho sino que, en cambio, lo *obligan* a hacer-

## 5.2. Estándares y eficiencia epistémica del proceso

La necesidad de imponer límites al poder estatal, por supuesto, no es la única preocupación que pueden tener en mente quienes reglamentan los derechos fundamentales. Por el contrario, en algunas ocasiones la preocupación por diseñar un proceso penal que permita arribar a un mayor número de decisiones que se correspondan con la verdad de los hechos —y que permita, en particular, aumentar las condenas a personas culpables— puede llevar a adoptar reglamentaciones de derechos que flexibilicen esos límites. Para entender por qué razones esto es así es necesario, primero, advertir cómo el establecimiento de límites muy fuertes mediante reglas puede afectar negativamente la búsqueda de la verdad.

El problema tradicional de las reglas reside en los costos que generan al reglamentar una determinada cuestión. Ese costo se deriva de la forma en que esa clase de normas jurídicas actúan para reducir la discrecionalidad de quienes las aplican. Tal como se vio anteriormente, las reglas acuden a términos más precisos, se adelantan a la solución de los casos concretos, e impiden que los tribunales consideren muchas circunstancias que en ausencia de las reglas podrían ser relevantes. Sin embargo, esto solo puede lograrse a cambio de generar un fenómeno conocido como el carácter «sobreinclusivo» de ese tipo de normas jurídicas. Esa expresión suele ser utilizada para destacar situaciones en las que una regla es aplicable a un caso pese a que no exista ninguna conexión con la que puede ser considerada su razón subyacente, y donde las circunstancias cuya consideración fue excluida al formularla parecen volverse pertinentes. Esto se debe a que:

El predicado fáctico de una regla es una generalización probabilística respecto de alguna justificación (habitual aunque no necesariamente implícita) [y] es sobreincluyente. Comprende estado de cosas que, en casos particulares, podrían no producir la consecuencia que representa la justificación de la regla, aunque el estado de cosas, en tanto tipo, esté relacionado de manera probabilística con la plausibilidad o incidencia de la justificación. El predicado fáctico de la regla posee una relación probabilística con el objeto de la regla pero esa relación deja abierta la posibilidad de que en casos particulares la conexión entre la justificación y la consecuencia esté ausente (SCHAUER, 2004: 90-91).

Si bien la discusión sobre las razones subyacentes a los derechos fundamentales de la persona imputada es un área muy amplia, un ejemplo puede ayudar a comprender el punto. Considérese el caso de la reglamentación del derecho fundamental a interrogar a los testigos de cargo mediante una regla que impide admitir las declaraciones

---

lo. Esto podría suceder, por ejemplo, si un creador de las normas jurídicas adopta reglas que obliguen a tener por probado un hecho cuando se cuente con determinado tipo de evidencias —reglas de prueba legal positiva—, u obliguen a considerar voluntaria una declaración autoincriminatoria efectuada en determinadas situaciones. En este tipo de escenarios, donde las reglas obligan a tener por probada una circunstancia para perjudicar a la persona imputada, las críticas a la validez de esas reglamentaciones pueden encontrarse en sintonía con las que en muchas ocasiones se han formulado a las presunciones irrefutables de esta clase (DEI VECCHI, 2017: 189).

de testigos ausentes en el juicio y que la defensa no tuvo posibilidad de interrogar. Si se considera que la razón subyacente a ese derecho es la necesidad de contar con evidencias confiables, existen casos en los cuales una regla como la mencionada puede llevar a excluir declaraciones que, pese a no haber sido confrontadas por la persona imputada, pueden, de todos modos, ser particularmente confiables (ROVATTI, 2020: 40-42).

Esto puede repetirse en muchas ocasiones con las reglamentaciones expuestas como ejemplos del modelo basado en reglas. Si la exclusión de la prueba ilícita obedece a la necesidad de preservar determinados valores éticos en el sistema de justicia penal, existen casos, como la imposibilidad de condenar a una persona en un caso muy grave porque la prueba decisiva fue obtenida mediante una infracción comparativamente muy leve, en los que esos valores pueden verse afectados, precisamente, por la exclusión de la evidencia. Por otro lado, si el derecho de la persona imputada a no ser juzgada más de una vez por un mismo hecho se justifica por la necesidad de protegerla del riesgo de una condena y de los sufrimientos que genera el sometimiento a un proceso penal, pueden existir casos donde el proceso previo no generó esos riesgos y sufrimientos porque fue manipulado para no dirigirse a la búsqueda de la verdad. Del mismo modo, si la prohibición de condenar cuando solo se cuenta con un determinado tipo de evidencia se justifica por la necesidad de no exponer a la persona imputada al riesgo de una condena errónea, pueden existir casos en los cuales esa prueba cuente con un valor considerablemente elevado como para superar un umbral probatorio exigente.

Esas consecuencias de la reglamentación mediante reglas pueden no ser consideradas problemáticas si lo único que se desea es establecer límites muy fuertes al poder estatal que protejan a la persona acusada de posibles abusos de poder. Sin embargo, si se considera que el diseño del proceso penal no debe quedar limitado exclusivamente a lograr ese objetivo, la discusión del tema adquiere otra complejidad que difícilmente pueda ser dejada de lado. Esto es particularmente cierto si se tiene en cuenta que el carácter sobreinclusivo de las reglas suele impactar negativamente en la eficiencia del proceso penal para arribar a decisiones que se correspondan con la verdad de los hechos, algo considerado como el fin institucional de los procesos judiciales en general (FERRER BELTRÁN, 2007: 30), y especialmente importante para el fin del proceso penal (DUFF *et al.*, 2007: 61-62, ROBERTS y ZUCKERMAN, 2010: 19-20), en particular por la relación que suele establecerse entre justicia retributiva y verdad (PÉREZ BARBERÁ, 2020). Muchas de las reglamentaciones basadas en reglas, de acuerdo con algunas posiciones, podrían ser criticadas por poner «serios obstáculos a la averiguación de la verdad» (LAUDAN, 2013: 297).

En varias ocasiones ese resultado obedece a que las reglamentaciones mediante reglas, al ser sobreinclusivas y generar resultados como los señalados más arriba, tiene un efecto «contraepistémico». Esto puede generarse por varias razones. En algunas oportunidades la aplicación de ciertas reglas genera la exclusión de evidencias relevantes y que pueden incluso ser particularmente confiables. Un ejemplo de ese

fenómeno es la ya mencionada reglamentación mediante reglas del derecho a «interrogar a los testigos de cargo» y la exclusión de testimonios que, pese a que la persona imputada no haya tenido posibilidad de confrontar, cuenten con buenos indicios de fiabilidad (ROVATTI, 2020: 40-42). Del mismo modo, la reglamentación rígida del problema de la prueba ilícita puede llevar a excluir elementos de prueba que en algunos casos son, además de relevantes, particularmente confiables (LAUDAN, 2013: 259-260), y la exclusión automática de las declaraciones autoincriminatorias por haber sido adquiridas sin respetar determinadas formalidades puede eliminar del conjunto de elementos de prueba algunas declaraciones que, pese a esto, sí sean confiables (LAUDAN, 2013: 246).

En otros casos algunas reglamentaciones pueden ser consideradas contraepistémicas por impedir que los tribunales tengan por probado un hecho pese a que el valor probatorio de la evidencia disponible pueda ser muy elevado, tal como sucede, por ejemplo, con las reglas que impiden dictar una condena si solo se cuenta con un determinado testimonio. Finalmente, otras veces ese carácter contraepistémico es generado por los obstáculos que las reglamentaciones mediante reglas ponen al inicio de un proceso penal. Por ejemplo, la protección contra el doble juzgamiento mediante una regla estricta puede impedir el inicio de un nuevo proceso penal pese a que el procedimiento anterior haya tenido defectos fundamentales que permitan concluir que no fue conducido con la finalidad de descubrir la verdad.

Quienes adviertan este tipo de problemas derivados del uso de reglas pueden tener buenas razones para reglamentar los derechos fundamentales de la persona imputada de una forma distinta. Si se considera, por ejemplo, que no existen motivos que justifiquen pagar el precio derivado del carácter sobreinclusivo de las reglas, las reglamentaciones de los derechos pueden ser realizadas mediante estándares y, de ese modo, otorgar a los tribunales un mayor grado de discrecionalidad para tomar decisiones considerando aquellas circunstancias relevantes que las reglas excluyen. Esto no significa que los tribunales, solo por esta circunstancia, vayan a tomar decisiones que permitan aumentar la eficiencia epistémica del proceso. De lo que se trata es de permitirles tener en cuenta varias circunstancias para que eso suceda, a diferencia de lo que, al menos explícitamente, sucede con las reglas como las analizadas en este trabajo. Posiblemente entren en juego aquí algunas consideraciones vinculadas con la gestión de la confianza (SHAPIRO, 2014: 409-411) que se tenga en los tribunales dentro de cada sistema y con la forma en que se considere que éstos harán uso de la mayor discrecionalidad que se les otorgue.

En pocas palabras, la existencia de un modelo basado en estándares permite advertir que muchos problemas contraepistémicos son «evitables», y pueden ser solucionados sin que esto implique adoptar reglamentaciones prohibidas por la existencia

de los derechos fundamentales de la persona imputada. Esto, por supuesto, podría servir para tratar de evitar *muchos*, aunque no *todos*, esos problemas<sup>46</sup>.

Las reglamentaciones mediante estándares, dicho de otro modo, pueden ser atractivas para impedir que los derechos fundamentales sean considerados, en todos los casos, como un obstáculo insuperable para quienes quieran aumentar la eficiencia epistémica del proceso. La existencia de esta alternativa permite incorporar esta preocupación sin renunciar al lenguaje de los derechos, a diferencia de lo que en alguna oportunidad —a mi juicio, erróneamente— se ha denunciado<sup>47</sup>. Podría tratarse, en suma, de un «garantismo medido» (AGUILERA, 2014), pero garantismo al fin.

La selección de estándares aumenta el poder de los tribunales y, junto con esto, incrementa las posibilidades de que se generen aquellas situaciones que se pretenden evitar con la imposición de límites fuertes al poder estatal. Por ende, una posición que defienda la adopción de estándares no puede evitar la consideración de esa circunstancia, ya sea para afirmar que no existen razones para mostrar las preocupaciones que exhiben los partidarios de las reglas, o para demostrar que esas razones, pese a existir en alguna medida, no justifican tolerar las consecuencias derivadas del carácter sobreinclusivo de ese tipo de normas.

### 5.3. Balance

En esta sección se analizaron algunos argumentos que pueden ser utilizados al discutir si la reglamentación de un derecho fundamental de la persona imputada debe ser realizada mediante un modelo basado en reglas o, en cambio, a través de uno basado en estándares. Para esto se tomaron los principales valores que suelen estar en tensión en el proceso penal y se los relacionó con los modelos que se expusieron anteriormente. Concretamente, se argumentó que si aquello que predomina es una preocupación por establecer límites fuertes al poder penal estatal y se desea proteger de forma robusta a la persona imputada, existen buenas razones para adoptar reglas como las que se mencionaron al exponer los rasgos de este modelo. Por otro lado, se sostuvo que si se tienen en cuenta los problemas que puede generar el modelo basado en reglas de cara a la eficiencia epistémica del proceso penal, se cuenta con un argumento de peso para acudir a reglamentaciones mediante estándares.

---

<sup>46</sup> La eliminación de todos los defectos epistémicos que tiene el proceso penal implicaría la renuncia a muchos derechos fundamentales de la persona imputada que son un obstáculo para la búsqueda de la verdad y, por ende, se trata de reglamentaciones que no serían jurídicamente admisibles en la mayoría de los sistemas jurídicos. Un ejemplo extremo de esto es la utilización de la tortura como medio de investigación, ya que, pese a existir prohibiciones jurídicas categóricas sobre su uso, puede, de todos modos, facilitar la búsqueda de la verdad. Sobre el problema de concebir la prohibición de la tortura únicamente por la escasa confiabilidad que tendría la información obtenida mediante su uso, GRECO (2010: 8-9).

<sup>47</sup> Así, por ejemplo, GUZMÁN (2018). Para una crítica cfr. LARSEN (2020b).

Existen, por supuesto, otras razones relevantes que deben ser tenidas en cuenta al adoptar reglas o estándares, pese a que no se vinculen estrictamente con las particularidades del proceso penal. Por ejemplo, es posible sostener que en determinadas áreas no es siquiera mínimamente viable optar por alternativas distintas a los estándares<sup>48</sup>. Sin embargo, poner el foco de la atención en aquellas preocupaciones centrales para el proceso penal permitió advertir cómo algunos de los argumentos a favor y en contra de estas alternativas se acentúan cuando se los relaciona con las especificidades del proceso penal.

De forma resumida, el desarrollo realizado buscó llamar la atención sobre la necesidad prestar una especial atención a los objetivos que se tengan en mente al concretar esa tarea y a los costos y beneficios que genera cada una de las alternativas disponibles. No se trató, en cambio, de ofrecer una respuesta concreta a la pregunta de qué grado de discrecionalidad debería otorgarse a los tribunales para tomar decisiones. Esto es, no se pretendió, por ejemplo, presentar una defensa de un modelo que limite la discrecionalidad judicial de forma fuerte, o brindar una argumentación en favor de uno que la amplíe de manera notable. Por el contrario, el objetivo fue intentar arrojar algo de luz sobre la forma en la cual puede estructurarse la discusión del problema y tratar de ofrecer algunas pautas para que los debates puedan llevarse adelante de forma productiva.

## 6. CONCLUSIONES

A lo largo de este trabajo se trató de ofrecer un desarrollo basado, esencialmente, en tres puntos. Dicho de forma breve, se buscó llamar la atención sobre un fenómeno usualmente poco atendido, se propuso una forma de analizarlo, y se ofreció una forma de ordenar las discusiones que pueden generarse en torno a él. En primer lugar (secciones 1 y 2) se intentó demostrar que el análisis de los derechos fundamentales de la persona imputada debe ser uno sensible a las complejidades que el tema posee

---

<sup>48</sup> Existen asuntos que, por su complejidad, más allá de cuál sea el objetivo que tenga en mente el creador de las normas, difícilmente puedan ser regulados de una forma que no otorgue a los tribunales un grado de discrecionalidad considerable para resolver casos concretos. Esto se debe, para explicarlo de forma gráfica, a que «somos hombres y no dioses» (HART, 2012: 160-162, en sentido similar SCHAUER, 2009: 194-196). Una explicación general de las razones por las que esto es así puede verse en ENDICOTT (2003: 31-57). La dificultad para diseñar reglas más precisas en determinados campos puede explicar por qué razones incluso los sistemas jurídicos que adoptan algunas reglamentaciones basadas en reglas no lo hacen en la totalidad del asunto que se busca reglamentar sino, en cambio, en una porción considerablemente menor. Un buen ejemplo de esto son las reglamentaciones vinculadas con la valoración de la prueba: la dificultad de establecer de forma precisa y de antemano cómo deben tomarse todas las decisiones en ese ámbito puede explicar que solo existan reglamentaciones basadas en reglas para determinados casos que pueden considerarse problemáticos. Esto también puede explicar por qué aquellos sistemas jurídicos que tratan de evitar los problemas generados por el carácter sobreinclusivo de las reglas opten por estándares y no, en cambio, por reglas precisas con un contenido diverso que permitan contemplar aquello que excluyen las reglamentaciones basadas en reglas.

y, fundamentalmente, a los desacuerdos que pueden generarse al discutir cómo deberían reglamentarse esos derechos. Luego, en la parte central (secciones 3 y 4), se trató de mostrar que una forma de realizar un análisis sensible a esas complejidades es uno que adopta como eje a la discrecionalidad judicial. Para esto se sostuvo que es necesario prestar una especial atención al grado de precisión con el cual se redactan las normas jurídicas, y se mostró que la tipología «reglas» vs. «estándares» permite identificar adecuadamente las principales formas en que pueden reglamentarse esos derechos. Finalmente (sección 5), se intentó ofrecer una serie de pautas para estructurar la discusión sobre cuál de esas formas adoptar. Allí se buscó demostrar que la reducción o ampliación de la discrecionalidad judicial puede ser una herramienta útil al servicio de los principales valores en tensión en el proceso penal, como la necesidad de imponer límites fuertes al poder estatal y la búsqueda de la verdad.

Tal como se adelantó en la introducción, estos desarrollos fueron puestos a prueba mediante un análisis de la forma en que pueden reglamentarse varios asuntos vinculados con derechos fundamentales de la persona imputada. Pese a su importancia, por supuesto, se trata de reglamentaciones que no abarcan la totalidad de los debates similares que pueden generarse. Sin embargo, la pretensión de generalidad de los modelos aquí propuestos y de los argumentos ofrecidos al analizar cuál de éstos adoptar permite trasladar ese desarrollo a otros asuntos que, por razones de espacio, no pudieron ser analizados en este trabajo.

## BIBLIOGRAFÍA

- AGUILERA, Edgar, 2014: «¿Garantismo extremo o mesurado? La legitimidad de la función jurisdiccional penal: construyendo el debate Ferrajoli-Laudan», en *Isonomía*, n. 40, pp. 61 y ss.
- ALEXANDER, Larry, 2003: «Constitutional Rules, Constitutional Standards, and Constitutional Settlement», *Constitutional Commentary*, n. 20, pp. 369 ss.
- ALEXANDER, Larry y SHERWIN, Emily, 2001: *The Rule of Rules. Morality, Rules, and the Dilemmas of Law*, Durham-Londres, Duke University Press.
- ANDRÉS IBÁÑEZ, Perfecto, 2017: «Sobre argumentación probatoria y su expresión en la sentencia (penal)», *Jueces para la Democracia*, n. 88, pp. 102 ss.
- ASHWORTH, Andrew, 1977: «Excluding evidence as protecting rights», *Criminal Law Review*, n. 3, pp. 723 ss.
- ASHWORTH, Andrew y REDMAYNE, Mike, 2010: *The Criminal Process*, 4.<sup>a</sup> ed., Oxford, Oxford University Press.
- BARAK, Aharon, 2012: *Proportionality. Constitutional Rights and Their Limitations*, Cambridge, Cambridge University Press.
- BRAITHWAITE, John, 2002: «Rules and Principles: a Theory of Legal Certainty», *Australian Journal of Legal Philosophy*, n. 27, pp. 47 ss.
- BRAITHWAITE, John y PETTIT, Philip, 2015: *No sólo su merecido. Por una justicia penal que vaya más allá del castigo* [trad. de Elena Odriozola], Siglo Veintiuno, Buenos Aires.
- CARACCILO, Ricardo, 2009: «Discreción, respuesta correcta y función judicial», en EL MISMO, *El derecho desde la filosofía*, Centro de Estudios Políticos y Constitucionales, Madrid, pp. 251 y ss.
- DEI VECCHI, Diego, 2017: *La decisión de encarcelar preventivamente y otros peligros procesales. Una aproximación a la prisión preventiva desde la óptica de la justificación de decisiones judiciales*, Buenos Aires, Ad-Hoc.

- DENNIS, Ian, 2014: «Rethinking double jeopardy: justice and finality in criminal process», en VAN KEMPEN, Piet Hein (ed.), *Criminal law and human rights*, Ashgate, pp. 219 ss.
- 2011: «The Human Rights Act and the Law of Evidence: Ten years on», *Sydney Law Review*, n. 33, pp. 333 ss.
- 2010: «The right to confront witnesses: meanings, myths and human rights», *Criminal Law Review*, n. 4, pp. 255 ss.
- DORF, Michael C. y FRIEDMAN, Barry, 2000: «Shared Constitutional Interpretation», *The Supreme Court Review*, n. 2000, pp. 61 ss.
- DRIPPS, Donald, 2011: «The substance-procedure relationship in criminal law», en DUFF, R. A./GREEN, Stuart (eds.), *Philosophical Foundations of Criminal Law*, Oxford, Oxford University Press, pp. 409 ss.
- 2001: «Constitutional Theory for Criminal Procedure», *William and Mary Law Review*, n. 43, pp. 409 ss.
- 1993: «Criminal Procedure, Footnote Four, and the Theory of Public Choice; Or, Why Don't Legislatures Give a Damn About the Rights of the Accused?», *Syracuse Law Review*, n. 44, pp. 1079 ss.
- DUFF, Antony, FARMER, Lindsay y MARSHALL, Sandra/TADROS, Victor, 2007: *The Trial on Trial*, t. III, Oxford-Portland, Hart Publishing.
- DUFRAIMONT, Lisa, 2013: «Realizing the Potential of the Principled Approach to Evidence», *Queen's Law Journal*, n. 39, pp. 11 ss.
- DWORKIN, Ronald, 1978: *Taking rights seriously*, Cambridge MA, Harvard University Press.
- ENDICOTT, Timothy, 2003: *Vagueness in Law*, Oxford, Oxford University Press.
- FERRAJOLI, Luigi, 2014 [1989]: *Derecho y Razón. Teoría del garantismo penal*, 10.ª ed. [trad. de Perfecto Andrés Ibáñez y otros], Madrid, Trotta.
- FERRER BELTRÁN, Jordi, 2013: «La prueba es libertad, pero no tanto: una teoría de la prueba cuasi-benthamiana», en VÁZQUEZ, Carmen (ed.), *Estándares de prueba y prueba científica*, Madrid, Marcial Pons, pp. 21 ss.
- 2007: *La valoración racional de la prueba*, Madrid, Marcial Pons.
- FRIEDMAN, Richard D., 2004: «Face to face: rediscovering the right to confront prosecution witnesses», *The International Journal of Evidence and Proof*, n. 8, pp. 1 ss.
- GUASTINI, Riccardo, 2012: «*Juristenrecht*. Inventando derechos, obligaciones y poderes» [trad. de Diego Dei Vecchi], en FERRER BELTRÁN, Jordi, MORESO, José Juan y PAPAYANNIS, Diego M. (eds.), *Neutralidad y teoría del derecho*, Madrid, Marcial Pons, pp. 207 ss.
- GUZMÁN, Nicolás, 2008: «El neopunitivismo en la jurisprudencia de la Corte Suprema de Justicia de la Nación y de la Corte Interamericana de Derechos Humanos: un pronóstico incierto para el *ne bis in idem* y la cosa juzgada», en PASTOR, Daniel (dir.), *Neopunitivismo y neoinquisición*, Buenos Aires, Ad-Hoc, pp. 203 ss.
- 2018: «La teoría del estándar objetivo de prueba y el garantismo penal: dos perspectivas antagónicas sobre el proceso penal», en *Jueces para la Democracia*, n. 92, pp. 85 y ss.
- GRECO, Luís, 2010: «Conveniencia y respeto: sobre lo hipotético y lo categórico en la fundamentación del Derecho penal» [trad. de Nuria Pastor Muñoz], *InDret Penal*, n. 4.
- 2007: «Las reglas detrás de la excepción» [trad. de Eduardo Javier Riggi], *InDret Penal*, n. 4.
- HART, H. L. A., 2012 [1961]: *El concepto de Derecho* [trad. de Genaro Carrió], Buenos Aires, Abeledo-Perrot.
- HO, Hock Lai, 2012: «The presumption of innocence as a human right», en ROBERTS, Paul y HUNTER, Jill (eds.), *Criminal evidence and human rights. Reimagining Common Law Procedural Traditions*, Oxford-Portland, Hart Publishing, pp. 259 ss.
- KAMISAR, Yale, 2012: «The Rise, Decline and Fall (?) of Miranda», *Washington Law Review*, n. 87, pp. 965 ss.
- KAPLOW, Louis, 1992: «Rules versus standards: an economic analysis», *Duke Law Journal*, n. 42, pp. 557 ss.
- LARSEN, Pablo, 2020a: «Reglas, estándares y dos modelos de derecho probatorio para el proceso penal», *InDret Penal*, n. 1, pp. 300 ss.

- 2020b: «El uso problemático de los derechos fundamentales en algunas discusiones sobre derecho probatorio», en ROVATTI, Pablo y LIMARDO, Alan (dirs.), *Pensar la Prueba*, n. 1, Del Sur, Buenos Aires, 2020b, pp. 91 ss.
- LAUDAN, Larry, 2013: *Verdad, error y proceso penal. Un ensayo sobre epistemología jurídica* [trad. de Carmen Vázquez y Edgar Aguilera], Marcial Pons, Madrid.
- MACCORMICK, Neil, 2018 [1978]: *Razonamiento jurídico y teoría del derecho* [trad. de José Ángel Gascón Salvador], Lima, Palestra.
- MAIER, Julio B. J., 2016 [1996]: *Derecho procesal penal*, t. I, 2.ª ed., Buenos Aires, Ad-Hoc.
- MARCUS, Paul, 2006: «It's not just about 'Miranda': determining the voluntariness of confessions in criminal prosecutions», *Valparaiso University Law Review*, n. 40, pp. 601 ss.
- MÖLLER, Kai, 2012: *The global model of constitutional rights*, Oxford, Oxford University Press.
- MORGENSTERN, Federico, 2015: *Cosa juzgada fraudulenta. Un ensayo sobre la cosa juzgada irrita*, Buenos Aires, BdeF.
- 2014: «Argumentos a favor de la relativización de la regla de exclusión de prueba ilegal», *Revista de Derecho Penal y Procesal Penal*, Abeledo-Perrot, n. 8, pp. 1551 ss.
- NANZER, Alberto, 2008: «La prueba derivada bajo el prisma del derecho a la confrontación», en PASTOR, Daniel (dir.), *Neopunitivismo y Neoinquisición*, Buenos Aires, Ad-Hoc, pp. 339 ss.
- NINO, Carlos Santiago, 2007 [1989]: *Ética y derechos humanos. Un ensayo de fundamentación*, 2.ª ed., Buenos Aires, Astrea.
- PACKER, Herbert L., 1964: «Two models of the criminal process», *University of Pennsylvania Law Review*, n. 113, pp. 1 ss.
- PARCHOMOVSKY, Gideon y STEIN, Alex, 2015: «Catalogs», *Columbia Law Review*, n. 115, pp. 165 ss.
- PASTOR, Daniel, 2012: «El principio de la descalificación procesal del Estado en el derecho procesal penal», en *Tendencias. Hacia una aplicación más imparcial del derecho penal*, Buenos Aires, Hammurabi, pp. 317 ss.
- 2009 [2002]: *El plazo razonable en el proceso del estado de derecho*, Buenos Aires, Ad-Hoc.
- PÉREZ BARBERÁ, Gabriel, 2015: «Reserva de ley, principio de legalidad y proceso penal», en *Letra Derecho Penal*, n. 1, pp. 42-86.
- 2014: «El principio general de reserva de ley (o principio general de legalidad) en la doctrina alemana», en RIVERA (H), Julio César/ELIAS, José Sebastián/GROSMAN, Lucas Sebastián/LEGARRE, Gonzalo (dirs.), *Tratado de los derechos constitucionales*, t. I, Abeledo-Perrot, Buenos Aires, pp. 330-424.
- 2020: «Prueba legítima y verdad en el proceso penal I: la independencia metafísica de la verdad», en *Isonomía*, n. 52, pp. 5 y ss.
- PIZZI, William T., 1999: *Trials without truth. Why our system of criminal trials has become an expensive failure and what we need to do to rebuild it*, New York, New York University Press.
- POSNER, Richard, 1990: *The problems of jurisprudence*, Cambridge MA, Harvard University Press.
- REDMAYNE, Mike, 2012: «Confronting confrontation», en ROBERTS, Paul y HUNTER, Jill (eds.), *Criminal evidence and human rights. Reimagining Common Law Procedural Traditions*, Oxford-Portland, Hart Publishing, pp. 283 ss.
- ROBERTS, Paul/ZUCKERMAN, Adrian, 2010: *Criminal Evidence*, 2.ª ed., Oxford, Oxford University Press.
- ROVATTI, Pablo, 2020: «Testigos no disponibles y confrontación: fundamentos epistémicos y no epistémicos», *Quaestio Facti*, n. 1, Madrid, Marcial Pons, pp. 31 ss.
- ROXIN, Claus, 2000 [1998]: *Derecho procesal penal* [trad. de Gabriela Córdoba y Daniel Pastor], Buenos Aires, Del Puerto.
- RUDSTEIN, David, 2008: «Retrying the acquitted in England, part II: the exception to the rule against double jeopardy for tainted acquittals», *San Diego International Law Journal*, n. 9, pp. 217 ss.
- SCALIA, Antonin, 1989: «The rule of law as a law of rules», *University of Chicago Law Review*, n. 56, pp. 1175 ss.
- SCHAUER, Frederick, 2013: «The Miranda warning», *Washington Law Review*, n. 88, pp. 155 ss.
- 2009: *Thinking like a lawyer. A new introduction to legal reasoning*, Cambridge MA, Harvard University Press.

- 2006: «Do cases make bad law?», *University of Chicago Law Review*, n. 73, pp. 883 ss.
- 2004 [1991]: *Las reglas en juego. Un examen filosófico de la toma de decisiones basada en reglas en el derecho y en la vida cotidiana* [trad. de Claudia Orunesu y Jorge Rodríguez], Madrid, Marcial Pons.
- 2003: «The convergence of rules and standards», *New Zealand Review*, pp. 303 ss.
- 1988: «Formalism», *Yale Law Journal*, n. 97, pp. 509 ss.
- 1987: «Authority and indeterminacy», *Nomos*, n. 29, pp. 28 ss.
- 1982: «An essay on constitutional language», *UCLA Law Review*, n. 29, pp. 797 ss.
- SHAPIRO, Scott J., 2014: *Legalidad*, [trad. de Diego M. Papayannis y Lorena Ramírez Ludueña], Madrid, Marcial Pons.
- SPOTSWOOD, Mark, 2018: «Truth, Lies and the Confrontation Clause», *University of Colorado Law Review*, n. 89, pp. 566 ss.
- STEIN, Alex, 1997: «Against Free Proof», *Israel Law Review*, n. 31, pp. 573 ss.
- 2006: *Foundations of Evidence Law*, Oxford, Oxford University Press.
- STRAUSS, David, A., 2001: «Miranda, the Constitution, and Congress», *Michigan Law Review*, n. 99, pp. 958 ss.
- STUART, Don, 2010: «Welcome flexibility and better criteria from the Supreme Court of Canada for exclusion of evidence obtained in violation of Canadian Charter of Rights and Freedoms», *Southwestern Journal of International Law*, n. 16, pp. 313 ss.
- STUNTZ, William J., 1997: «The Uneasy Relationship Between Criminal Procedure and Criminal Justice», *Yale Law Journal*, n. 107, pp. 1 ss.
- 2006: «The Political Constitution of Criminal Justice», *Harvard Law Review*, n. 119, pp. 780 ss.
- SUNSTEIN, Cass, 2018: *Legal reasoning and political conflict*, 2.<sup>a</sup> ed., Oxford, Oxford University Press.
- 2007: «Incompletely theorized agreements in constitutional law», *Social Research*, n. 74, pp. 1 ss.
- THAMAN, Stephen, 2011: «Constitutional rights in the balance: modern exclusionary rules and the toleration of police lawlessness in the search of truth», *University of Toronto Law Journal*, n. 61, pp. 691 ss.
- TRECHSEL, Stefan, 2005: *Human Rights in Criminal Proceedings*, Oxford, Oxford University Press.
- TWINING, William y MIERS, David, 2010: *How to do things with rules*, 5.<sup>a</sup> ed., Cambridge, Cambridge University Press.
- WALDRON, Jeremy, 2018: «Constitucionalismo: una postura escéptica», en EL MISMO, *Contra el gobierno de los jueces*, Siglo XXI, Buenos Aires, pp. 25 y ss.
- 2016: «Clarity, Thoughtfulness, and the Rule of Law», en KEIL, Geert y POSCHER, Ralf (eds.), *Vagueness and Law. Philosophical and Legal Perspectives*, Oxford, Oxford University Press, pp. 317 ss.
- 2011: «Thoughtfulness and the rule of law», *British Academy Review*, n. 18, pp. 1 ss.
- 2008: «The concept and the Rule of Law», *Georgia Law Review*, n. 43, pp. 1 ss.
- 2006: «The core of the case against judicial review», *Yale Law Journal*, n. 115, pp. 1346 ss.
- 2005: *Derecho y desacuerdos* [trad. de José Luis Martí y Águeda Quiroga], Madrid, Marcial Pons.
- WEBBER, Grégoire, 2009: *The Negotiable Constitution. On the Limitation of Rights*, Cambridge, Cambridge University Press.



## FILOSOFÍA DE LA MENTE Y PRUEBA DE LOS ESTADOS MENTALES: UNA DEFENSA DE LOS CRITERIOS DE «SENTIDO COMÚN»

Daniel González Lagier

Catedrático de Filosofía del Derecho de la Universidad de Alicante  
daniel.gonzalez.lagier@gmail.com

**RESUMEN:** La prueba de los hechos psicológicos o estados mentales plantea dificultades de carácter ontológico (¿existen realmente hechos mentales, distintos de los hechos físicos?), epistemológico (¿cómo puedo conocer los estados mentales de los demás?) e ideológico (¿son jurídicamente relevantes?). Una parte de la doctrina y de la jurisprudencia ha dicho que, en realidad, los estados mentales no son —o no deben ser tratados como— “hechos” propiamente, por lo que no se prueban (esto es, no se demuestra que hayan ocurrido o hayan existido), sino que para el Derecho son entidades ficticias que se imputan o adscriben a partir de ciertos criterios normativos. Este escepticismo contrasta con el hecho de que todos atribuimos con frecuente éxito estados mentales a los demás y conseguimos adaptar y coordinar nuestra conducta a la de los otros gracias a esas atribuciones. Sabemos atribuir estados mentales, pero no sabemos cómo lo hacemos. En este trabajo trato de analizar qué luz puede arrojar la filosofía de la mente (en particular, las teorías acerca del llamado “problema de las Otras Mentes”) a la prueba de los hechos psicológicos en el Derecho.

**PALABRAS CLAVE:** Prueba; estados mentales; filosofía de la mente; otras mentes; atribución psicológica.

**ABSTRACT:** The proof of psychological or mental facts poses challenges of an ontological nature (do mental facts really exist as distinct —and not reducible to— physical facts?) epistemological (how can we know others’ mental states?) and ideological (are those facts legally relevant?). Part of legal doctrine and case law holds that states of mind are actually not —or ought not to be approached as— proper “facts”, therefore they are not to be proved (their occurrence or existence cannot be

demonstrated). From a legal perspective, they are fictional entities to be ascribed or imputed using certain normative criteria. This scepticism is openly in contrast with the fact that we quite successfully and frequently attribute states of mind to others, as well as we all manage to adapt and coordinate our behavior following those attributions. We know how to attribute states of mind, however we ignore how we do it. In this paper, I try to explore how Philosophy of Mind (theories on the so-called “problem of other minds”) can contribute to shed some light on the proof of psychological facts in the realm of Law.

**KEYWORDS:** Proof; evidence; mental states; Philosophy of Mind; Other Minds; psychological attribution.

**SUMARIO:** 1. UNA APROXIMACIÓN A LA FILOSOFÍA DE LA MENTE: 1.1. ¿Qué son los «estados mentales»? 1.2. ¿Existen los estados mentales?. 1.3. ¿Existen otras mentes, además de la mía?. 1.4. ¿Cómo conozco mis propios estados mentales?. 1.5. ¿Cómo conozco los estados mentales de los demás?.— 2. LA PRUEBA DE LOS ESTADOS MENTALES EN EL DERECHO: 2.1. El argumento ontológico contra la prueba de los estados mentales. 2.2. El argumento epistemológico contra la prueba de los estados mentales. 2.3. El argumento ideológico contra la prueba de los estados mentales. 2.4. Entonces... ¿prueba o imputación?. 2.5. La prueba de los estados mentales con contenido representacional. 2.6. La prueba de los estados mentales de carácter fenomenológico. 2.7. La prueba de los estados mentales mixtos (las emociones). 3. CONCLUSIONES.— BIBLIOGRAFÍA.

**RECOMMENDED CITATION:** DANIEL GONZÁLEZ LAGIER, 2022: «Filosofía de la mente y prueba de los estados mentales: Una defensa de los criterios de «sentido común»», in *Quaestio facti*, 3: 49-79. Madrid: Marcial Pons Ediciones Jurídicas y Sociales. DOI: [http://dx.doi.org/10.33115/udg\\_bib/qf.i3.22731](http://dx.doi.org/10.33115/udg_bib/qf.i3.22731)

La prueba de los hechos psicológicos o estados mentales es un problema complejo de la teoría de la prueba jurídica, un problema que plantea profundas dificultades y perplejidades de carácter ontológico (¿existen realmente hechos mentales, distintos de los hechos físicos?), epistemológico (¿cómo puedo conocer los estados mentales de los demás?) e ideológico (¿son jurídicamente relevantes?). Hasta tal punto parecen irresolubles estas dificultades que una parte de la doctrina y de la jurisprudencia ha dicho que, en realidad, los estados mentales no son —o no deben ser tratados como— “hechos” propiamente, por lo que no se prueban (esto es, no se demuestra que hayan ocurrido o hayan existido), sino que para el Derecho son entidades ficticias que se imputan o adscriben a partir de ciertos criterios normativos<sup>1</sup>. Este

<sup>1</sup> Lo cual tiene importantes consecuencias prácticas. Por ejemplo, en algunas sentencias del Tribunal Supremo y el Tribunal Constitucional español se ha argumentado que, al no tratarse propiamente de hechos, no les alcanza la presunción de inocencia: “Los juicios de valor sobre intenciones no pueden someterse a las exigencias de la presunción de inocencia por ser circunstancias inaprehensibles por los sentidos, razón por la cual solo los hechos en sí, sobre los que se base la inducción, pueden ser objeto de prueba...” (STS de 17 de febrero de 1995); “Tales juicios de valor sobre intenciones o juicios de inferencia sobre pensamientos, querer o deseos escondidos en el intelecto humano, no son hechos en sentido estricto porque no constituyen datos aprehensibles, por ello no pueden ser objeto de prueba propiamente”.

escepticismo contrasta con el hecho de que todos atribuimos con frecuente éxito estados mentales a los demás y conseguimos adaptar y coordinar nuestra conducta a la de los otros gracias a esas atribuciones. Adscribir estados mentales a los demás es la estrategia que ponemos en práctica para explicar y predecir su comportamiento, lo que a su vez es necesario para coordinarnos con ellos y participar de la vida social. Que en buena medida logremos comunicarnos, entendernos, colaborar, emprender tareas comunes, construir el mundo de reglas, convenciones e instituciones en el que vivimos requiere poder conocer los estados mentales de los otros. Por supuesto, abundan los desencuentros y malentendidos; pero eso no debe ocultar la importancia de los aciertos. Esta es una circunstancia que me parece que no debe perderse de vista cuando discutimos sobre la prueba de este tipo de hechos, porque, frente a las posturas escépticas, nos recuerda que el sentido común parece, en muchos casos, una estrategia adecuada para adscribir hechos mentales.

De hecho, determinar qué intención tiene un agente o cuál es su estado de ánimo es algo tan natural en nuestra vida cotidiana que puede considerarse una habilidad en la que somos expertos desde niños, aunque la ejerzamos de una manera intuitiva. Sabemos atribuir estados mentales, pero no sabemos cómo lo hacemos. Explicar cómo lo hacemos es un problema propio de disciplinas como la filosofía y las ciencias cognitivas (psicología, neurociencia, inteligencia artificial, etc.), cuyas respuestas quizá puedan contribuir a resolver algunos de los problemas que se les plantean a los juristas. En este trabajo voy a tratar de analizar qué luz puede aportar la filosofía de la mente (en particular, las discusiones acerca del llamado “problema de las otras mentes”) a la prueba de los hechos psicológicos en el Derecho.

## 1. UNA APROXIMACIÓN A LA FILOSOFÍA DE LA MENTE

En la filosofía de la mente podemos distinguir varios tipos de cuestiones:

a) La cuestión conceptual: ¿Qué son los estados mentales?

b) La cuestión ontológica:

(b.1) ¿Existen los estados mentales?

(b.2) ¿Existen otras mentes, además de la mía?

c) La cuestión epistemológica:

(c.1) ¿Cómo conozco mis propios estados mentales?

(c.2) ¿Cómo puedo conocer los estados mentales de los demás?

Las preguntas b.2 y c.2 constituyen el aspecto ontológico y epistemológico, respectivamente, del llamado problema de las “otras mentes”. Es la vertiente epistemo-

---

te dicha y quedan fuera de la garantía constitucional” (STC 195/1993). Sobre la cuestión de si la prueba del dolo es genuina prueba o imputación puede verse RAGUÉS, 1999 y 2002 y GONZÁLEZ LAGIER, 2004.

lógica de este problema lo que nos interesa aquí, pero digamos algo también sobre el resto de cuestiones, dado que todas están entrelazadas.

### 1.1. ¿Qué son los «estados mentales»?

Comencemos con la cuestión conceptual. Son estados mentales —usando la elegante enumeración de Richard Wollheim— “las percepciones, tales como oír un coro de maitines o ver una constelación de estrellas; las sensaciones, como el dolor, los picores, las punzadas de hambre o sed; los sueños y los ensueños; los momentos de desesperación, aburrimiento o lujuria; los momentos de inspiración; las evocaciones; las imágenes fantaseadas y las melodías oídas en la cabeza; y los pensamientos, tanto los pensamientos que pensamos, como los que se cuelan en nuestra cabeza sin invitación” (WOLLHEIM, 2006: 24). Con una clasificación menos lírica, pero algo más sistemática, podríamos distinguir entre (a) voliciones (como los deseos y las intenciones), (b) cogniciones (creencias, conocimientos), (c) percepciones y sensaciones, (d) estados afectivos (emociones, estados de ánimo) y (e) actos mentales (planear, deliberar, decidir, etc.).

Pese a su diversidad, parece haber una serie de propiedades que permiten diferenciar este tipo de hechos de los hechos físicos (MOYA, 2006: Cap. I; SEARLE, 1990).

*a) Consciencia:* Normalmente, somos conscientes de los estados mentales que tenemos. Podemos entender la *consciencia* como una especie de conocimiento “de lo que sucede dentro de nosotros” (GARCÍA CARPINTERO, 1995: 46) no fundado en la evidencia empírica acerca de rasgos del mundo externo: Si me pellizco el brazo, experimento dolor y por ello sé que el brazo me duele, y lo sé con independencia de que me vea a mí mismo sacudirlo o gritar; si deseo un vaso de cerveza, sé que lo deseo, sin necesidad de esperar a ver si me levantaré e iré a la nevera; y si estoy preocupado, sé —porque lo siento— que lo estoy.

*b) Qualia:* Ahora bien, hay cierta diferencia cualitativa en cómo experimento mi dolor, mi deseo o mi preocupación. Decimos entonces que estos estados mentales se corresponden con diferentes sensaciones internas o *qualia*. Los *qualia* son la manera peculiar como cada estado mental emerge a mi consciencia, el componente fenomenológico de los estados mentales. La consciencia (y los *qualia*) establecen una importante asimetría entre la manera como conocemos nuestros estados mentales y la manera como conocemos los de los demás; este último conocimiento sí requiere una inferencia a partir de la observación de la conducta del otro. Un ejemplo que ayuda a entender la noción de *qualia* puede ser el siguiente: supongamos que María tiene una extraña enfermedad que no le permite ver los colores; todo lo ve en blanco y negro, como en las antiguas televisiones<sup>2</sup>. María se ha interesado mucho por el fenómeno de los colores y ha estudiado e investigado todo lo que hay que saber sobre

<sup>2</sup> Adapto un ejemplo de Frank Jackson, comentado por MOYA, 1996 .

ellos, hasta convertirse en una eminencia en el tema. Desde el punto de vista físico y neurofisiológico sabe todo lo que hay que saber de los colores. Pero aún así podemos decir que le falta algo: le falta la experiencia de haber visto el rojo, el verde, el azul... Si de repente se curara de su enfermedad, aprendería algo más sobre los colores (si realmente lo sabe todo sobre los colores desde un punto de vista físico, incluyendo las reacciones neurofisiológicas correspondientes, pero le falta algo, ese algo no puede ser físico; por ello este ejemplo se ha usado también para mostrar que las propiedades mentales no son reducibles a propiedades físicas).

*c) Contenido mental:* Otra característica de muchos estados mentales consiste en poseer un *contenido representacional*, esto es, en tener la capacidad de versar o ser sobre objetos y estados de cosas del mundo (que, además, no tienen por qué estar presentes o existir en realidad, sino que pueden pertenecer al pasado o al futuro o incluso ser ficticios) distintos de ellos mismos (SEARLE, 1992: 18). La creencia de que mañana lloverá se refiere a un contenido (“mañana lloverá”). Dado un estado mental con contenido representacional, podemos distinguir entre su contenido y su modo psicológico: “Creo que mañana lloverá” y “Deseo que llueva mañana” son dos estados mentales que se diferencian por su modo psicológico, pero no por su contenido (a los estados mentales con contenido mental podemos llamarlos también actitudes proposicionales).

*d) Subjetividad:* Otra característica propia de los estados mentales es su *modo subjetivo de existencia*. Mis dolores, temores, odios, deseos y creencias son exclusivamente míos. En opinión de John Searle se trata no solo de una subjetividad epistemológica (la consciencia), sino también ontológica: “Consideremos, por ejemplo, el enunciado ‘tengo dolor de espalda’. Tal enunciado es completamente objetivo en el sentido de que lo convierte en verdadero la existencia de un hecho real y no depende de las actitudes de los observadores. Sin embargo, el fenómeno mismo, el dolor real mismo, tiene un modo subjetivo de existencia” (SEARLE, 1992:21)

*e) Causalidad mental:* A todas estas propiedades hay que añadir el hecho de que los estados mentales no son causalmente inertes, sino que interactúan entre sí y con estados físicos; esto es lo que los filósofos de la mente llaman *causalidad mental*. Los estados mentales parecen causar hechos físicos: nuestras emociones, creencias y deseos causan nuestro comportamiento. Mi miedo a que un perro me muerda causa que salga corriendo y mi deseo de saciar la sed, junto con las creencias adecuadas, causa que vaya a la nevera a por una cerveza. Pero para que se hable de causalidad mental el efecto debe producirse precisamente en virtud del contenido mental. Una manera de explicar que los estados mentales causan hechos físicos consiste en afirmar que tienen un componente o sustrato físico (neuronal) y atribuir su poder causal a ese sustrato. Por ejemplo, el miedo a que el perro me muerda causa mi conducta porque se corresponde con ciertos cambios neurofisiológicos que disparan señales eléctricas que estimulan los músculos. Sin embargo, la relación entre esos estados neuronales y la conducta no sería un caso de causalidad mental, porque en esta explicación el contenido de la emoción (el miedo al perro, la creencia de que ese perro pone en riesgo

mi integridad física) no juega ningún papel. Supongamos una soprano que hace estallar un vaso cantando el aria *Il dolce suono* de Donizetti. El aria posee un significado (“*Il dolce suono mi colpi di sua voce! Ah, quella voce m’e qui nel cor discesa!...*”), pero no es el significado, sino las vibraciones provocadas por la voz de la soprano, las que rompen la copa (con total independencia de su significado), por lo que este no sería un caso de causalidad mental. El desafío que plantea la causalidad mental es explicar que nuestras razones para actuar tienen influencia o producen una diferencia en la probabilidad de que actuemos de una u otra manera que varía según su contenido.

Tomemos como ejemplo los estados mentales que llamamos intenciones: La consciencia de la intención consiste en la experiencia, no basada en evidencia empírica externa, de nuestra pretensión de hacer algo: si se trata de la intención de hacer algo en un futuro próximo o lejano, la certeza de que –a menos de que algo me lo impida o cambie de opinión– efectivamente lo haré; si se trata de la intención en el momento de actuar, el conocimiento del tipo de acción que estoy realizando (HAMP-SHIRE y HART, 1958: 1 y ss.). El contenido representacional de la intención consiste en las acciones u objetivos que procuro realizar o alcanzar. La subjetividad ontológica consiste en el carácter privado y particular de las intenciones (hay cierto sentido en que otro sujeto y yo no podemos compartir la misma intención, porque mi intención es solo mía: aunque tengamos intenciones de hacer lo mismo). Por último, la intención tiene alguna relación (en virtud de su contenido) con mi conducta física, ejerce algún control sobre ella.

Es importante observar que estas propiedades no son condiciones necesarias y suficientes, ni individual ni conjuntamente, que sirvan para completar una definición de la noción de “hecho psicológico”. En realidad, no todos los estados mentales tienen estas propiedades. Por ejemplo, quizá tenga sentido hablar de estados mentales no conscientes. También es dudoso que todos los estados mentales conscientes posean un contenido representacional, o que todos posean un aspecto fenomenológico o *qualia*. Es discutible si las sensaciones, como los dolores, la picazón, etc., tienen en algún sentido un contenido mental (hay quien dice que un dolor en el brazo tiene como contenido la parte del brazo que te duele, pero resulta ciertamente forzado), o acerca de si las actitudes proposicionales (como las intenciones, creencias, etc.) tienen *qualia* o carecen por completo de él. Una definición sugerente de estado mental es la que establece una condición necesaria disyuntiva:  $x$  no puede ser considerado un estado mental si (a) no tiene un contenido representacional o (b) no posee carácter fenomenológico (*qualia*). En todo caso, estas dos propiedades de lo mental son sumamente importantes y se puede trazar una distinción entre aquellos estados mentales cuya principal característica es que poseen un contenido representacional articulado o conceptual y aquellos cuya característica más llamativa es poseer un componente fenomenológico. Esto permitiría clasificar los deseos, las intenciones y las creencias en el primer grupo; las sensaciones en el segundo, y las emociones (que, muy a grandes rasgos, se pueden ver como asociaciones entre creencias y sensaciones) se clasificarían como fenómenos mixtos.

## 1.2. ¿Existen los estados mentales?

Las propiedades de lo mental son difíciles de encajar en nuestra imagen científica y materialista del mundo. ¿Cómo puede un cerebro, algo completamente material, tener consciencia? ¿Cómo puede versar sobre algo, esto es, tener un contenido semántico? ¿Cómo puede ser real algo esencialmente subjetivo? ¿Y cómo puede lo mental interactuar con lo físico? Cuando nos planteamos estas cuestiones, surge el problema ontológico de lo mental: Si los estados mentales están correctamente caracterizados por las anteriores propiedades, ¿existen realmente? Las distintas concepciones de la mente pueden interpretarse como intentos de responder a esta pregunta<sup>3</sup>.

Las concepciones dualistas consideran que existen dos tipos de sustancias o entidades: las entidades materiales o físicas y las entidades mentales. Para estas teorías, los estados mentales existen, con las propiedades señaladas anteriormente, y tienen un tipo de existencia distinto e independiente de las entidades físicas. Tampoco plantea un problema su existencia para el monismo idealista (que, sin embargo, niega la existencia de los hechos físicos). El monismo materialista, por el contrario, al sostener que solo existen entidades físicas, se enfrenta al problema de dar cuenta de las propiedades de lo mental. Por ello, dentro del materialismo podemos distinguir tres tipos de actitudes ante los estados mentales: las posiciones eliminacionistas, las posiciones reductivistas y las no reductivistas.

El *materialismo eliminacionista* cree que no hay estados mentales, tal como lo considera la psicología y el sentido común. Hay estados cerebrales, actividad neuronal, pero presuponer que la neurociencia permitirá establecer una correlación entre estados cerebrales y mentales es dar erróneamente carta de existencia a estos últimos: la psicología está equivocada, trata de dar explicaciones de la conducta que la investigación científica mostrará que no logran decir nada verdadero sobre la realidad. Como la teoría del flogisto (una supuesta sustancia, en realidad inexistente, que se creía que se liberaba en los procesos de combustión u oxidación), las teorías y los conceptos mentales están condenados a desaparecer, simplemente porque hacen referencia a cosas que no existen. A medida que avance la neurociencia, podremos sustituir el lenguaje de los estados mentales por el lenguaje de los estados cerebrales, de igual manera que el descubrimiento de los virus y las bacterias permitió rechazar las explicaciones de las enfermedades en términos de demonios, maldiciones y espíritus. Este sería, simplemente, un ámbito más donde la extensión de la ciencia permitiría la expulsión de la superstición.

El *materialismo reduccionista* no impugna el sentido de los conceptos mentales (mantiene la psicología como una disciplina epistemológicamente adecuada o conveniente, al menos en cierto grado), pero cree que las teorías mentales ganarán en precisión y poder explicativo si son traducidas o reducidas a un lenguaje en última

---

<sup>3</sup> Una presentación de estas concepciones y sus problemas puede verse en MOYA, 2006.

instancia no mental. Una teoría es reducida a otra, de acuerdo con Paul Churchland, cuando “una teoría nueva y eficaz logra abarcar un conjunto de proposiciones y principios que reflejan perfectamente (o casi perfectamente) las proposiciones y principios de una teoría o marco conceptual anteriores” (CHURCHLAND, 1999: 43). Si el ejemplo preferido de los eliminacionistas es el flogisto, el de los reduccionistas es la explicación del calor en términos de promedio de energía cinética de las moléculas (a diferencia de lo que ocurría con el flogisto, las teorías o leyes que usan el concepto de calor, como la que afirma que el calor dilata los metales, son útiles, pero su reducción a la nueva teoría proporciona una descripción más precisa del fenómeno). El conductismo (reducción a conducta), el materialismo de la identidad de tipos (reducción a actividad cerebral) y el funcionalismo (reducción a funciones y roles causales) son concepciones reduccionistas.

El *materialismo no reduccionista* sostiene que la traducción del lenguaje mental al lenguaje físico es imposible sin pérdidas significativas: aunque haya una sola sustancia, lo físico, podemos hablar de estados o propiedades cerebrales y de estados o propiedades mentales. El materialismo no reduccionista puede verse como un monismo respecto de las sustancias y, simultáneamente, un dualismo respecto de las propiedades. Su desafío, por tanto, es explicar cómo de lo físico pueden surgir las propiedades de lo mental. Una estrategia interesante (dejo de lado el monismo anómalo de Davidson) es la del emergentismo. El *emergentismo* es una postura materialista que identifica los estados mentales con propiedades emergentes o sistémicas, es decir, propiedades que surgen a partir de cierto grado de complejidad de un sistema. Son propiedades del conjunto del sistema, pero no de cada una de sus partes. Para explicar qué es una propiedad emergente John Searle —cuya concepción, que él mismo llama *naturalismo biológico*, es una teoría emergentista— recurre al siguiente ejemplo: “Supongamos que tenemos un sistema S, compuesto de los elementos *a, b, c...* Por ejemplo, S podría ser una piedra y los elementos podrían ser moléculas. En general, habrá rasgos de S que no son, o no necesariamente, rasgos de *a, b, c...* Por ejemplo, S podría pesar 20 kilogramos, sin que las moléculas individualmente pesen 20 kilogramos. Denominemos a estos rasgos ‘rasgos del sistema’. Algunos rasgos del sistema pueden ser deducidos, o determinados, o calculados a partir de los rasgos *a, b, c...*, simplemente por la forma en que se componen y ordenan (y, a veces, por sus relaciones con el entorno). Ejemplos de ellos serían la forma, el peso y la velocidad. Pero algunos otros rasgos del sistema no pueden ser determinados solo a partir de los elementos que los componen y de las relaciones con el entorno: han de ser explicados a partir de las relaciones entre los elementos. Llamémosles ‘rasgos del sistema causalmente emergentes’. La solidez, la liquidez y la transparencia son ejemplos de rasgos del sistema causalmente emergentes” (SEARLE, 1996: 121). De manera que, para Searle, una propiedad emergente es una propiedad que surge a partir de los elementos que componen el sistema, de las relaciones del sistema con el entorno y —esto parece lo decisivo— de las relaciones de esos elementos entre sí. Los estados mentales se identifican con propiedades emergentes de los cerebros. Los cerebros son sistemas físicos

que han ido adquiriendo complejidad a lo largo de su historia evolutiva. El naturalismo biológico es descrito por Searle a partir de cuatro tesis (SEARLE, 2006: 147 y ss.):

1) Los estados mentales, con su ontología subjetiva de primera persona, son fenómenos reales del mundo real.

2) Los estados mentales son causados en su totalidad por procesos neurobiológicos de nivel inferior localizados en el cerebro.

3) Los estados mentales son rasgos del sistema cerebral en su conjunto y existen, por tanto, en un nivel superior al de las neuronas y sus sinapsis (“Por sí misma —dice Searle—, una neurona no es consciente, pero las partes del sistema cerebral compuestas por ella sí lo son”).

4) Los estados mentales, en la medida en que son reales, tienen eficacia causal (“Mi sed consciente, por ejemplo, me lleva a tomar agua”).

El materialismo emergentista no es reductivista y, por tanto, no elimina la perspectiva de primera persona, a diferencia del conductismo o la tesis de la identidad de instancias. De acuerdo con Searle, aunque los estados mentales pueden ser reducidos causalmente a —en el sentido de explicados por— los componentes del cerebro y sus relaciones, esta explicación no es suficiente, porque lo característico de lo mental es que es un fenómeno en primera persona, requiere un lenguaje subjetivo que no puede ser descrito con el lenguaje objetivo de la neurociencia.

El siguiente esquema trata de dar cuenta de todas estas concepciones de la mente:

1. Dualismo (dos sustancias: materia y mente)
2. Monismo (una sustancia).
  - 2.1. Monismo idealista (todo es espíritu/mente).
  - 2.2. Monismo materialista (todo es materia).
    - 2.2.1. Materialismo eliminacionista (no hay estados mentales, es una descripción falsa).
    - 2.2.2. Materialismo reductivista (los estados mentales son estados de un tipo distinto).
      - 2.2.2.1. Conductismo (reducción a conducta).
      - 2.2.2.2. Tesis de la identidad de tipos (reducción a eventos neurofisiológicos)
      - 2.2.2.3. Funcionalismo (reducción a roles causales).
    - 2.2.3. Materialismo no reductivista.
      - 2.2.3.1. Tesis de la identidad de instancias (los estados mentales son estados físicos, pero no hay correlación entre tipos).
      - 2.2.3.2. Emergentismo (los estados mentales son propiedades emergentes. Monismo de sustancias, dualismo de propiedades).

En lo que sigue se dará por sentado que es posible conciliar, a la manera del emergentismo, la existencia de estados mentales (tal como han sido caracterizados) con una concepción básicamente materialista del mundo (de acuerdo con la cual el sustrato físico es una condición de posibilidad de la existencia de los estados mentales, que son generados por él).

### 1.3. ¿Existen otras mentes, además de la mía?

El problema de las “otras mentes” comienza a plantearse cuando asumimos que tiene sentido hablar de la existencia de estados mentales y nos preguntamos si es posible que, además de mi mente, existan otras mentes, esto es, si el resto de sujetos poseen también una mente. Como hemos visto, parece que debe aceptarse una asimetría entre la forma como conocemos nuestros propios estados mentales (conocimiento sin observación ni inferencia) y la forma como conocemos los estados mentales de los demás (conocimiento indirecto, que requiere una inferencia a partir de su comportamiento) (DAVIDSON, 2003: 25 y ss.; MOYA, 1996: 157). Además, en cierto sentido, el conocimiento que tenemos de nosotros mismos, aunque puede discutirse si es infalible, está claramente dotado de una autoridad de la que carece el conocimiento de los estados mentales de terceros. No puedo dudar de que yo pienso, siento, percibo..., pero no puedo tener esa misma convicción respecto de los demás. ¿Cómo puedo saber si existen “otras mentes”, además de la mía?

La respuesta clásica a esta pregunta, defendida por autores como Agustín de Hipona o John Stuart Mill, es que lo sabemos (lo suponemos) por analogía (CASTRO, 2017: 48). Del mismo modo que mi mente y mi conducta están relacionadas, de manera que sé que son mis estados mentales los que producen mi comportamiento, puedo asumir que la conducta que observo en los demás, que es básicamente análoga a la mía, está causada igualmente por sus estados mentales. John Stuart Mill lo expuso de la siguiente manera:

Soy consciente en mí mismo de una serie de hechos conectados por una secuencia uniforme, el principio de la cual es la modificación de mi cuerpo, el punto medio son las sensaciones, y el final, la conducta exterior. En el caso de los otros seres humanos, tengo la evidencia de mis sentidos para el primer y último eslabón de esa serie, pero no para el eslabón intermedio. Sin embargo, descubro que la secuencia entre el primero y el último es tan regular y constante en esos casos como en el mío. En mi propio caso, sé que el primer eslabón produce el último a través del eslabón intermedio y que no lo podría producir sin él. Por tanto, la experiencia me obliga a concluir que debe existir un eslabón intermedio; que debe ser en otros, o bien el mismo que en mí, o bien otro diferente; [...] al suponer que el eslabón es de la misma naturaleza [...] actúo de acuerdo con las reglas legítimas de la investigación experimental (MILL, 1867; tomado de DANCY, 1993: 87).

El argumento proyecta a los demás una relación entre estados mentales y comportamiento a partir de mi conocimiento fiable de que mis propios estados mentales están detrás de mi propio comportamiento; sin embargo, se trata de un argumento

demasiado débil para satisfacer al escéptico, que puede señalar que, en realidad, se trata más bien de una inducción a partir de un solo caso<sup>4</sup>. Se podría aducir en respuesta a esta objeción que no se trata en realidad de un único caso: lo que proyecto son *múltiples* experiencias en las que observo esta conexión entre mis estados mentales y mis acciones; no obstante, todos los casos están referidos a mí, lo que limita la fuerza de la conclusión.

La anterior crítica puede sortearse completando el argumento con una inferencia a la mejor explicación. Habría varias formas de hacerlo. Podríamos decir, por ejemplo, que la mejor explicación de que la conducta de los demás sea semejante a la mía, frente a alternativas del tipo de que sean autómatas o marionetas de algún dios, es que posean mentes como la mía; o podría decir que la mejor explicación del éxito de la estrategia de atribuir estados mentales como los míos a los demás para explicar y predecir su conducta es que efectivamente tienen estados mentales como los míos. En todo caso, la existencia de otras mentes, aunque es un importante desafío filosófico, suele aceptarse como un presupuesto de nuestra comprensión del mundo: no parece existir un solipsista convencido que lleve su convicción a su conducta cotidiana.

#### 1.4. ¿Cómo conozco mis propios estados mentales?

Pasemos al problema epistemológico de los estados mentales. Como hemos visto, se pueden plantear dos preguntas acerca del conocimiento de los estados mentales: Cómo conocemos nuestros estados mentales y cómo conocemos los estados mentales de los demás.

La primera cuestión suele responderse, como hemos adelantado, recurriendo a un conocimiento directo de nuestros estados mentales, un conocimiento que no requiere la observación de nuestro comportamiento por parte de nosotros mismos, ni ninguna inferencia a partir del mismo; y además un conocimiento privilegiado, en el sentido de dotado de una especial autoridad o una “presunción de veracidad” (MOYA, 1996). A veces se dice que este modo de conocimiento consiste en “volver los ojos hacia dentro”, “hacia uno mismo”, en un ejercicio de introspección. Pero esta forma de describir las cosas hace pensar que este autoconocimiento es el resultado de un esfuerzo intencional de examinarse a sí mismo; esto puede ser así en ocasiones, pero la consciencia de nuestros propios estados mentales suele ser más “directa”, sin requerir un esfuerzo consciente. Incluso puede sostenerse, como sugiere Carlos Moya, que en el momento en que me pregunto por mis estados mentales y hago el esfuerzo de sacarlos a la luz por medio de una inferencia, estos se convierten en algo distinto de lo que eran (MOYA, 1996: 182). Podemos hacer auto—atribuciones de estados mentales respecto del pasado, del presente, del futuro o respecto de situaciones

---

<sup>4</sup> Sobre los problemas planteados por este argumento puede verse DANCY, 1993: 87 y ss.

hipotéticas, pero cuando se postula un conocimiento sin inferencia este parece darse solo a propósito de los estados mentales presentes, los que ocurren en ese momento (las autoatribuciones hacia el pasado son recuerdos o reconstrucciones, y el resto de autoatribuciones parece resultado de inferencias; esto no excluye, sin embargo, que el conocimiento de nuestros estados mentales pasados o las predicciones hacia el futuro de los mismos carezcan necesariamente de una autoridad privilegiada, en la medida en que ya tenemos experiencias acerca de cómo normalmente pensamos y sentimos) (SKIDELSKY, 2011: 35).

Una forma sugerente (pero no exenta de problemas) de explicar este tipo de conocimiento consistiría en verlo como una relación entre estados mentales de diverso orden (SKIDELSKY, 2011: 36). Así, por ejemplo, si yo sé que creo que la tierra es redonda tengo dos estados mentales: la creencia de que la Tierra es redonda, que sería un estado mental de primer nivel, y el estado mental de segundo nivel consistente en ser consciente de que creo que la tierra es redonda (el conocimiento de que creo que la tierra es redonda). La consciencia, entonces, es un estado mental de nivel superior, un “pensamiento” sobre los estados de primer nivel. Pero ¿cómo se relacionan ambos niveles?

De acuerdo con el modelo de la consciencia como “percepción” entre ambos niveles hay una relación de carácter epistemológico y empírico, que consistiría en alguna relación causal entre los estados de nivel inferior y los de nivel superior. La consciencia sería una especie de “percepción” de estados mentales de primer nivel (MOYA, 1996: 159 y ss.). La presunción de veracidad de las autoatribuciones de estados mentales descansaría entonces en esta “percepción” especial que solo yo poseo respecto de mis estados mentales, en el hecho de que somos espectadores privilegiados de nuestra mente. Pero también se ha sugerido que la prioridad epistemológica de la primera persona sobre sus propios estados mentales tiene una dimensión constitutiva o conceptual (es parte de lo que significa tener un estado mental que el sujeto que lo tiene tenga una autoridad especial sobre el mismo) o normativa (las autoatribuciones implican aceptar el compromiso de actuar conforme al estado mental autoatribuido), o simplemente descansaría en que somos expertos en nosotros mismos (ya sabemos qué estados mentales hemos tenido en circunstancias semejantes en el pasado y cómo hemos actuado, lo que da más peso e inmediatez a las inferencias que hacemos sobre nuestros propios pensamientos presentes) (SKIDELSKY, 2011: 35 y ss.).

Considerar que nuestro conocimiento de los estados mentales es una especie de “percepción” (cuyo objeto no sería un hecho externo sino otro estado mental o ciertas propiedades del mismo) plantea varios problemas: por ejemplo, parece iniciar un regreso al infinito (¿cómo somos conscientes de que somos conscientes de tener una determinada creencia?) y, respecto de algunos estados mentales, parece redundante (puede decirse que no hay distinción entre tener un dolor y ser consciente de que tengo un dolor: son exactamente el mismo estado mental). Además, esta concepción del autoconocimiento parece casar bien con el tipo de estados mentales en los que predomina su aspecto fenomenológico, como los dolores, la sensación de hambre, las

punzadas, etc., pero no con las actitudes proposicionales o estados representacionales. Puede pensarse que el “*qualia*” asociado a los primeros nos “informa” del estado mental, pero no parece haber ninguna sensación asociada a los estados mentales representacionales. Creer que la tierra es redonda puede generar, en ciertas situaciones, estados mentales (emociones, por ejemplo) que sí tienen asociada alguna sensación, pero la creencia misma no la tiene. Luego el conocimiento de los estados mentales proposicionales no parece encajar bien con la idea de percepción causada por el estado mental subyacente, salvo que se establezca una correlación necesaria entre estos estados mentales y estados fenomenológicos. Por ejemplo, si se concluyera que creencias, deseos y emociones suelen estar asociadas, es decir, se presentan generalmente de manera simultánea, podría decirse que estas últimas cumplen un papel de “informantes” de nuestros estados mentales no fenomenológicos.

Sea como sea, vale la pena dejar establecido que si se rechaza el conductismo como explicación de la mente (y creo que hay buenas razones para hacerlo. Véase GONZÁLEZ LAGIER, 2003: 651), las autoatribuciones de conocimiento de nuestros estados mentales presentes no están basadas en la observación de nuestro comportamiento, a diferencia de lo que parece ocurrir con la atribución de estados mentales a terceros. Vale la pena también recordar que la autoridad de las primera persona respecto de sus propios estados mentales no implica necesariamente infalibilidad: respecto de los estados psicológicos presentes, solo tendría sentido predicar infalibilidad si son conscientes; cuando nos atribuimos estados mentales futuros o hipotéticos hacemos una inferencia que podría estar equivocada; y cuando nos atribuimos estados mentales pasados, la intermediación de los recuerdos y la posibilidad de reconstrucción de tales estados mentales (por ejemplo, interpretando nuestro comportamiento bajo su mejor luz) también pueden distorsionar la corrección de la atribución.

Sin embargo, a pesar de la asimetría, la capacidad de conocer nuestros propios estados mentales parece cumplir, como veremos, un papel importante en el conocimiento de estados mentales de terceros. Esto nos lleva al problema epistemológico de las otras mentes.

### 1.5. ¿Cómo conozco los estados mentales de los demás?

Aceptemos que existen los estados mentales, con las propiedades que hemos mencionado anteriormente, sean o no reducibles unas a otras; y aceptemos que, como yo, los demás agentes también tienen una mente: si es así, ¿cómo conocemos el contenido de sus mentes? ¿cómo conocemos los estados mentales de los demás? Esta es una cuestión que recientemente se ha revitalizado gracias a ciertos descubrimientos neurocientíficos, como veremos enseguida. Es, además, la cuestión que más nos interesa para la prueba de los estados mentales.

La filosofía y las ciencias cognitivas se han planteado esa pregunta desde la perspectiva de quien asume que los seres humanos hacemos atribuciones psicológicas

frecuentemente, con la finalidad de explicar y predecir la conducta de los demás, coordinarnos con ellos, entenderlos, cooperar y, en definitiva, construir una vida social. Desde esta perspectiva, lo que les interesa no es si podemos hacer atribuciones psicológicas exitosas, sino cómo lo hacemos. Las respuestas que se han dado se pueden clasificar en función de que se adopte una perspectiva de *tercera persona* o de *primera persona*. Para las primeras, el intérprete adopta una perspectiva externa respecto del sujeto interpretado, la perspectiva de un observador; para las segundas, el intérprete trata de ponerse en el lugar del interpretado (una exposición y discusión en detalle de las distintas teorías puede verse en GOLDMAN, 1989 y 2006; SKIDELSKY, 2001; BRUNSTEINS, 2008 y 2011; QUINTANILLA, 2019).

Las teorías que adoptan la perspectiva de *tercera persona* suelen recurrir a las siguientes estrategias:

a) La primera, que suele denominarse “teoría de la teoría”, hace una analogía entre la manera como conocemos el mundo físico y la manera como conocemos los estados mentales de los demás, en el sentido de que en ambos casos hacemos inferencias basadas en observaciones que subsumimos en teorías o generalizaciones empíricas. De la misma manera que inferimos que este trozo de metal se ha calentado del hecho de que observamos que se ha dilatado y de una ley de acuerdo con la cual “el calor dilata los metales”, o predecimos que el metal se va a dilatar a partir de la observación de que ha aumentado su temperatura y de la misma ley, explicamos que x echó a correr porque vio un perro, tiene miedo a los perros y aquellos que tienen miedo a algo normalmente lo rehúyen. La idea es que todos tenemos (al menos, a partir de cierta edad) una “teoría de la mente” (que a veces se denomina “psicología popular”). Esta teoría puede ser aprendida o innata (así, algunos autores postulan que, al igual que poseemos un módulo para el lenguaje, nuestro cerebro posee también un módulo o sistema encargado de la atribución psicológica), y se puede entender en un sentido más o menos articulado. Su contenido sería un conjunto de conceptos y “leyes” o regularidades nomológicas que correlacionarían estados mentales internos y conducta externa. A partir de estas “leyes” y de la observación de la conducta de los demás, podemos inferir sus estados psicológicos. Estas generalizaciones, que serían “de sentido común”, dirían cosas como “Las personas con dolor tienden a aliviar ese dolor”, “Las personas que desean determinado fin F y creen que M es un medio adecuado para dar lugar a F, intentarán hacer M” o, en general, “Cuando alguien está en una tal y cual combinación de estados mentales y recibe estímulos sensoriales del tipo tal y cual, tiende con tal y cual probabilidad a pasar a tales y cuales estados mentales y producir tales y cuales respuestas” (SKIDELSKY, 2001: 39).

La correlación entre estados mentales y conductas que expresan estas “leyes” no puede ser más que una aproximación probabilística, porque es posible que un mismo estado mental dé lugar a acciones distintas o que una misma acción pueda explicarse por estados mentales distintos. Por ello, la teoría de la mente puede completarse con una concepción de la racionalidad. Esto nos lleva a la segunda estrategia.

b) La segunda es la estrategia de la “racionalización”. Varios autores han sugerido que la atribución psicológica se basa en la idea de que los seres humanos son racionales y actúan siempre bajo parámetros explicables a la luz de la racionalidad.

Existen muchos modelos de interpretación de la conducta (y del significado) basados en la idea de que hay que tratar a los agentes como sujetos “racionalizadores”. Uno de los más conocidos y discutido es el modelo de Donald Davidson. De acuerdo con esta idea, la atribución de estados mentales es una interpretación o comprensión de la conducta de los otros basada en el “principio de caridad” (Davidson ofreció varias formulaciones de este principio, pero se considera clásica la que puede encontrarse en DAVIDSON, 1995: 281; véase también GOLDMAN, 1989: 163 y ss.). Este principio estaría compuesto por un conjunto de asunciones (relacionadas con la racionalidad del agente) que serán condición de posibilidad de la interpretación de la conducta. De acuerdo con Pablo Quintanilla (QUINTANILLA, 2004: 86), el principio de caridad requiere que quien atribuye los estados mentales al otro asuma los siguientes postulados:

- (i) El comportamiento del agente es consistente.
- (ii) El agente mantiene los mismos supuestos lógicos que el intérprete.
- (iii) Las creencias del agente son verdaderas.
- (iv) El agente se comporta siguiendo su mejor juicio.

Los postulados (i) y (iv) nos permitirían, por ejemplo, inferir las intenciones de un agente a partir de sus creencias y de su conducta, o las creencias a partir de su intención y su comportamiento; el postulado (iii) nos permitiría atribuirle las mismas creencias que tengo yo (porque son las que yo considero verdaderas); etc.

Es importante darse cuenta de que, desde una perspectiva racionalizadora, un rasgo relevante de los estados mentales es la coherencia que mantienen entre sí: intenciones, deseos, creencias y acción mantienen relaciones de consistencia, de manera que la posesión de uno de estos estados mentales lleva a, o presupone, la de otro. A veces se ha llamado a esta característica “el holismo de lo mental”. La atribución consiste en buena medida en encontrar un ajuste coherente (racional) entre los estados mentales implicados y la acción.

El principio de caridad, en definitiva, nos pide que aceptemos que el agente que tratamos de entender se comporta racionalmente. Su justificación puede buscarse en un argumento transcendental, como hacen muchos autores (es un presupuesto que hay que aceptar si queremos atribuir estados mentales a los demás y explicar su conducta) o puede tomarse como una generalización empírica (de hecho, los agentes suelen comportarse de forma racional). Más adelante volveremos sobre el papel de la racionalidad en la atribución de los estados mentales y en la justificación de esta atribución.

Frente a las teorías que adoptan una perspectiva de tercera persona, otros autores han propuesto una perspectiva de *primera persona*. Si desde la anterior perspectiva

las nociones centrales en la explicación de nuestro modo de atribuir estados mentales eran la noción de teoría o la de racionalidad, aquí son las de empatía e imaginación. El punto de partida de la perspectiva de primera persona es el autoconocimiento que las personas tienen de sus estados mentales y su capacidad de ponerse en lugar del otro. La atribución psicológica consistiría en un proceso de imaginarse en la situación del otro para atribuirle estados mentales. Esta atribución podría hacerse de dos formas: imaginando que yo, con mis características, estoy en la posición del otro y atribuyéndole por tanto al otro los estados mentales que yo creo que tendría en esa situación o tratando de ponerme completamente en el lugar del otro (es decir, asumiendo las características del otro) y atribuyéndole los estados mentales que creo que tendría si fuera él (SKILDESKY, 2011). Es decir: (a) imaginando que yo estoy en su lugar o (b) imaginando que yo soy él. En todo caso, el proceso constaría de tres fases: (a) Imaginarse en el lugar del otro, o siendo el otro; (b) determinación de los estados mentales que tendríamos en esa situación y (c) atribución de esos estados mentales al otro.

En todo caso, lo que esta teoría sostiene es que la atribución psicológica se realiza de una manera espontánea, sin inferencia o reflexión; de una manera semejante, por tanto, al proceso de autoconocimiento, pero simulando que soy el otro o que estoy en su lugar (por ello esta teoría recibe el nombre, no muy apropiado en español, de “teoría de la simulación”), lo que requiere capacidades empáticas del intérprete. Se postula que este conocimiento es más intuitivo que racional, más “sentido” que “inferido”. Esta concepción de la atribución de estados mentales se ha visto fortalecida con el descubrimiento de las neuronas espejo: neuronas que se activan tanto cuando se realiza un movimiento o se tiene una experiencia como cuando se observa que otro realiza ese movimiento o tiene esa experiencia (RIZZOLATI y CRAIGHERO, 2004, IACOBONI, 2013). Estas neuronas no se limitan a la percepción de las acciones, sino que se postula que activan las mismas intenciones y emociones que motivan la acción observada, aunque existe también un mecanismo que inhibe la realización de la acción (es decir, algún mecanismo que informa de que se trata de una situación ficticia). Además, se postula que las neuronas espejo no solo se activan cuando “vemos” a la otra persona, sino también en situaciones ficticias (por ejemplo, leyendo una novela).

Aunque la perspectiva de tercera y de primera persona se han presentado como excluyentes, es evidente que resulta fácil conciliarlas. Por ejemplo, se puede sostener que son recursos distintos que los sujetos podemos poner en marcha sucesivamente: los procesos directos y automáticos que postula la teoría de la simulación podrían proporcionar la primera información sobre los estados psicológicos de los demás, que luego se verían respaldados o corregidos por las inferencias basadas en generalizaciones empíricas y/o perspectivas racionalizadoras. Esto casaría bien con las concepciones de los neurocientíficos y psicólogos que postulan dos sistemas generales de conocimiento y toma de decisiones, uno más rápido e intuitivo y otro lento y reflexivo (KAHNEMAN, 2012; HAIDT, 2012). Otra manera de compatibilizar ambas

teorías consistiría en recordar la distinción entre estados mentales con contenido representacional y estados mentales en los que predomina el aspecto fenomenológico. Las teorías basadas en los principios de racionalidad o de caridad parecen más plausibles para el primer tipo de estados mentales que para los del segundo tipo, dado que la racionalidad parece requerir estados proposicionales. Una creencia o una intención pueden ser racionales o irracionales, pero un dolor o una punzada no. La teoría de la simulación y la teoría de la teoría sí serían útiles para la adscripción de estados fenomenológicos. La tendencia, en general, ha sido la de construir teorías mixtas que integran elementos de una y otra perspectiva.

Tampoco son teorías exhaustivas sobre los modos de atribución de estados mentales, esto es, cabe hablar de otros modos de conocimiento de la mente de los demás: por ejemplo, se puede plantear una “percepción directa” de sus pensamientos a partir de gestos y expresiones, que serían “significantes naturales” de los estados mentales (KALPOKAS, 2019). Y también se han propuesto teorías desde la perspectiva de la *segunda persona*, que ponen el acento en la relevancia de la interacción y comunicación entre los sujetos a la hora de interpretar mutuamente sus estados psicológicos.

Por último, las tres teorías acerca de cómo atribuimos estados mentales a los demás parecen presuponer la capacidad de conocer los propios estados mentales: las tres estrategias parecen presuponer una similitud entre los demás y nosotros, es decir, asumen el argumento de la analogía que hemos mencionado acerca de la existencia de las otras mentes. La construcción de las teorías basadas en regularidades empíricas o “leyes” que usamos para explicar la conducta de los demás, según la “teoría de la teoría”, probablemente combinan la observación de la conducta de los demás con la reflexión sobre nuestra propia conducta; la estrategia de la racionalización, como hemos visto, presupone que consideramos que los deseos de los demás son razonables y sus creencias verdaderas, esto es, que son aproximadamente los mismos deseos y las mismas creencias que tenemos nosotros; y la teoría de la simulación proyecta a terceros los estados mentales que, visto desde una perspectiva de primera persona, tendríamos contrafácticamente en determinadas circunstancias.

\*\*\*

Hasta aquí un rápido repaso a algunas cuestiones centrales de la filosofía de la mente. ¿Puede ser útil algo de todo esto para la prueba de los estados mentales en el Derecho?

## 2. LA PRUEBA DE LOS ESTADOS MENTALES EN EL DERECHO

### 2.1. El argumento ontológico contra la prueba de los estados mentales

Como hemos visto al principio de este trabajo, es una opinión bastante extendida entre los juristas que las dificultades para probar los estados mentales son de tal calibre que se debe renunciar a esa pretensión y sustituirla por una imputación a partir de criterios normativos. La prueba de la verdad no sería realmente un caso de prueba genuina, en el sentido de que no consistiría en una actividad cognoscitiva, orientada a establecer juicios que son verdaderos o falsos, sino una actividad dirigida a calificar la acción del agente de acuerdo con ciertos criterios que pueden satisfacer intereses muy variados (como el carácter ejemplificativo de la pena, la resolución de conflictos, etc.), pero no orientados al descubrimiento de una realidad psicológica. Se suele sostener esta posición con tres argumentos: El primero, de carácter ontológico, toma las peculiaridades de lo mental para negar su compatibilidad con una perspectiva materialista del mundo y sostener su inexistencia; el segundo, de carácter epistemológico, gira en torno a la imposibilidad de conocer los estados mentales de los demás, especialmente dadas las limitaciones de la prueba judicial; el tercero, de carácter ideológico, sostendría que los estados mentales no son realmente relevantes para el Derecho (aunque pudieran ser conocidos) (GONZÁLEZ LAGIER, 2006: 178 y ss.).

La filosofía de la mente esbozada en el anterior apartado proporciona objeciones contra los dos primeros argumentos. Así, frente al argumento eliminacionista, cabe aducir que negar la existencia de estados mentales es sostener algo muy implausible, lejos del sentido común, pues se estaría negando todas las experiencias que tenemos en primera persona: se estaría negando que tenemos consciencia de nuestros pensamientos, de nuestras emociones, de nuestros deseos; no se podría explicar, por ejemplo, la sensación que sentimos cuando nos pellizcamos un brazo. No podríamos decir tampoco que a María, la científica del anterior ejemplo con visión en blanco y negro que sabe todo sobre los colores desde un punto de vista físico, todavía le falta experimentar o aprender algo sobre ellos.

También muestra la filosofía de la mente que la alternativa entre eliminacionismo o reduccionismo, por un lado, y dualismo cartesiano, por otro (que es uno de los argumentos clásicos para negar la existencia de los estados mentales) es un falso dilema: como hemos visto a propósito de las distintas concepciones sobre la mente, cabe la posibilidad teórica de un materialismo no eliminacionista ni reductivista, una tercera vía que podría presentar los estados mentales como hechos que tienen propiedades peculiares no reductibles que supervienen a, o emergen de, hechos físicos.

Por último, puede decirse que el eliminacionismo deja sin explicar el aparente éxito de la estrategia de atribuir estados mentales a los demás para interpretar su

conducta y coordinarnos con ellos. ¿Por qué la estrategia de la racionalización o de la simulación parecen funcionar bien para explicar o predecir la conducta de los demás, si no existen tales estados mentales? Es cierto que hay ejemplos de teorías falsas que logran explicaciones y predicciones afortunadas; estas deben ser sustituidas por teorías más ajustadas y precisas, cuando se dispone de ellas. Pero no es posible tener teorías eliminacionistas o reductivistas que logren explicar la subjetividad y las experiencias en primera persona, precisamente porque adoptan perspectivas objetivizantes, de tercera persona.

## 2.2. El argumento epistemológico contra la prueba de los estados mentales

Esta última razón (el éxito de las atribuciones psicológicas como medio de interpretación y predicción de la conducta) sirve también para refutar el argumento epistemológico contra el carácter genuino de la prueba de los estados mentales. Simplemente, no parece cierto que no podamos conocer los estados mentales de los demás, aunque no tengamos un acceso directo a ellos. Más sólida es la versión del argumento que no niega que, en general, podemos hacer atribuciones satisfactorias de estados mentales, pero afirma que estas mismas atribuciones son demasiado débiles para superar criterios de suficiencia probatoria o estándares de prueba jurídicos. Paredes Castañón, por ejemplo, ha puesto de manifiesto las dificultades metodológicas que plantea la prueba del dolo en el proceso penal, que le parecen insuperables hoy en día (PAREDES CASTAÑÓN, 2001). Por ejemplo, a su juicio sería necesario dejar de lado la psicología popular, basada en generalizaciones “de sentido común” (el entrecomillado es suyo) y sustituirla por una psicología de carácter científico. Ello es así, entre otras razones, porque las atribuciones a partir de estas generalizaciones empíricas no son “demostraciones” de los hechos psicológicos y los indicios que suelen ser usados por los tribunales dejan abierta la posibilidad de que el dolo atribuido realmente no se haya dado (o viceversa: que el dolo no atribuido sí se hubiera producido en realidad). Sin embargo, no faltándole razón en sus propuestas constructivas, hay que advertir, por un lado, que la prueba nunca es tan concluyente como para poder considerarse una “demostración absoluta” que excluya totalmente el resto de posibilidades alternativas (GONZÁLEZ LAGIER, 2020) y, por otro lado, que si la prueba de los estados mentales nunca puede superar un estándar de prueba aceptable, entonces se complica también la prueba de cualquier acto humano, porque la prueba de la intención es necesaria (a) para determinar que se trata realmente de una acción, y no de un acto reflejo o involuntario (en el sentido de no dirigido por los estados mentales proposicionales del sujeto) y (b) para identificar qué acción realizó el sujeto (GONZÁLEZ LAGIER, 2013). Nada de esto sería posible si las generalizaciones empíricas “de sentido común” o alguna otra estrategia que sea factible en el proceso judicial no fueran mínimamente fiables.

### 2.3. El argumento ideológico contra la prueba de los estados mentales

De acuerdo con lo que he llamado el argumento ideológico, la prueba de los estados mentales *reales* no es necesaria o siquiera conveniente en el marco del proceso. Hay varias maneras de sostener esto: se puede negar que la verdad, en general, sea una finalidad relevante de la prueba, sosteniendo alguna concepción no cognoscitivistica de la misma. Esta idea ha sido discutida ya en muchas ocasiones, y no insistiré aquí sobre la importancia de una teoría fundamentalmente cognoscitivistica de la prueba judicial (TARUFFO, 2011; GASCÓN, 2004; FERRER, 2021). Otra forma de sostener el argumento ideológico es negar la relevancia, en particular, de la verdad sobre los estados mentales. Esto suele hacerse vinculando la finalidad de la sanción penal a un objetivo distinto del reproche o la retribución, como la prevención general. Si la finalidad de la pena —como sugiere Jakobs— es asegurar a la sociedad que la norma violada sigue vigente, o restaurar su vigencia frente al hecho que la lesiona (Jakobs 2001), no puede permitirse el caso de que la sociedad considere que una norma ha sido violada intencionalmente y no se siga la pena pertinente, con independencia de que hubiera intención real o no de violar la norma, de donde se infiere que lo importante no es la intención real del agente (o cualquier otro estado subjetivo), sino cómo va a interpretar la sociedad su conducta (“por lo que se refiere al dolo y a la imprudencia, estos, como hechos puramente psíquicos, carecen de interés a los efectos de la atribución”, escribe en JAKOBS, 1989: 639). En palabras de Laurenzo Copello (refiriéndose a Puppe y a Jakobs): “Lo importante, por consiguiente, de cara a determinar si se produce una contradicción con la norma jurídico penal, no son las actitudes internas del sujeto, sino lo que exprese la conducta desde la perspectiva de la comunicación entre seres inteligentes” (LAURENZO COPELLO, 1999: 225).

Sin embargo, esta manera de pensar está legitimando que se atribuyan consecuencias penales al sujeto al margen de que efectivamente se den en la realidad algunos requisitos (los elementos subjetivos del delito) que exigen las normas penales. Podemos preguntarnos por qué pensamos que tiene sentido atribuir responsabilidad a los seres humanos, y no a los animales. Parece plausible sostener que la dignidad especial que le atribuimos al ser humano, que constituye el fundamento de sus derechos y deberes, descansa —al menos, en buena parte— en la creencia de que este posee ciertas características mentales que lo diferencian de otros seres que, o bien no las tienen, o las tienen en un grado significativamente distinto. Por ejemplo, se ha sugerido por varios filósofos que un presupuesto de la responsabilidad por nuestras acciones es que seamos conscientes de lo que estamos haciendo intencionalmente, que conozcamos nuestras propias creencias, deseos e intenciones (MOYA, 1996). Esta misma dignidad exige tener en cuenta los estados mentales reales de los sujetos a la hora de adscribirles responsabilidad. Por el contrario, la objetivización de los elementos subjetivos del delito (y, por tanto, de su prueba) va en la dirección de una responsabilidad objetiva difícil de justificar en el ámbito penal.

#### 2.4. Entonces... ¿prueba o imputación?

Como hemos visto, una consecuencia de sostener la inexistencia de la mente, la imposibilidad de conocerla o su irrelevancia es la imputación de los estados mentales a partir de criterios normativos, construidos legal o jurisprudencialmente. Se habla por ello de que la prueba de los estados mentales consiste en adscribirlos a partir de ciertas reglas de imputación o presunciones (entendidas como reglas que establecen la obligación de aceptar como probado determinado hecho psicológico si se dan ciertas circunstancias). Pero también podría pensarse que el recurso a estas reglas no es consecuencia de que los estados mentales sean ficciones, sino la causa de que lo sean. Esto es, que el hecho de que la jurisprudencia haya establecido reglas para la valoración de la prueba de estos hechos, facilitando así su determinación, es ya un argumento suficiente para negar el carácter epistemológico y orientado hacia la verdad de estas inferencias. No es que los hechos psicológicos sean hechos normativos y por eso deban probarse de esa manera, es que es el hecho de que se determinen por medio de normas lo que los convierte (para el Derecho) en “hechos normativos” (o calificaciones jurídicas).

No obstante, en mi opinión, que un hecho se pruebe por medio de presunciones (reglas de valoración normativamente guiadas, indicios-tipo o como queramos llamarlas) no nos autoriza a decir sin más que al Derecho no le interesa si en realidad ocurrió o no ni lo convierte en un hecho normativo o institucional. Debemos tener en cuenta que hay tipos distintos de presunciones: Algunas presunciones (que desde la filosofía podrían considerarse las genuinas presunciones) pueden estar respaldadas por la observación de una asociación regular entre hechos (en cuyo caso son similares a máximas de experiencia, pero revestidas de autoridad normativa); otras, vienen justificadas como medios de proteger algún valor o principio que se considera relevante (por ejemplo, el de seguridad, el de protección de los intereses de la parte más débil, etc.). Las primeras, por tanto, tienen un fundamento epistemológico; las segundas, un fundamento normativo. Unas y otras, a su vez, pueden ser derrotables (*iuris tantum*) o no (*iuris et de iure*). Si una presunción no está respaldada por una regularidad empírica y no admite prueba en contrario, podríamos decir que efectivamente está prescindiendo de la realidad y podemos verla como una regla constitutiva que crea una ficción normativa. Pero si se limita a unificar criterios en torno a una regla de la experiencia confirmada o plausible y es derrotable (*iuris tantum*), entonces todavía mantiene un grado importante de “sensibilidad hacia la verdad” o hacia la dimensión empírica de estos hechos.

Las presunciones que la jurisprudencia ha ido introduciendo para la prueba de los hechos psicológicos (a) no cubren todos los supuestos (hay muchos casos en los que es necesario probar el propósito de un agente que no están resueltos por estas reglas y, por tanto, el juez debe valorar los indicios libremente), (b) por lo general, son derrotables y (c) en mi opinión, se basan en máximas de experiencia, en razonamientos

que en última instancia tienen una justificación empírica. Esto último explica que todas estas reglas nos parezcan razonables —no arbitrarias— desde el punto de vista de su eficacia para descubrir (por supuesto, dentro de un margen de error) los estados mentales que realmente tenía el sujeto en los casos a los que se aplican (cosa que no tendría por qué suceder si el objetivo fuera una imputación dirigida a satisfacer otros fines). Así ocurre, por ejemplo, con los criterios establecidos para la distinción entre el *animus necandi* y el *animus laedendi* (“entre otros: a) los antecedentes de hecho y las relaciones entre el autor y la víctima; b) la clase de arma utilizada; c) la zona o zonas del cuerpo a que se dirige la agresión; d) el número de golpes inferidos; e) las palabras que acompañaron al ataque; f) las condiciones de lugar, tiempo y circunstancias conexas o concomitantes con la acción; g) la causa o motivación de la misma; h) la entidad y gravedad de las heridas causadas” STS, 2.<sup>a</sup>, de 31-1-2011); o con los criterios para dar por probado el requisito, en el delito de receptación, del conocimiento del origen delictivo del bien comprado (la irregularidad de las circunstancias de la compra o modo de adquisición; la mediación de un precio vil o ínfimo, desproporcionado con el valor real de los objetos adquiridos; la clandestinidad de la adquisición; la inverosimilitud de las explicaciones aportadas para justificar la tenencia de los bienes sustraídos; la personalidad del adquirente acusado y de los vendedores o transmitentes de los bienes; la adquisición fuera de los cauces ordinarios del comercio. STS, 2.<sup>a</sup>, de 29-1-2000); o, por poner un último ejemplo, con el criterio (entre muchos otros) de la posesión de una cantidad de droga notablemente superior a la que un consumidor emplea en 5 días para la atribución de la intención de traficar con sustancias tóxicas (entre otras, STS de 5-6-1997). En todos estos casos, las atribuciones se basan en máximas de experiencia y generalizaciones empíricas, y no en criterios sin fundamento epistemológico.

Todo esto no quiere decir, no obstante, que no haya nada normativo en la prueba de los estados mentales. El Derecho usa o presupone necesariamente conceptos (también conceptos relativos a hechos psíquicos: el concepto de intención, de conocimiento, de emoción, etc.) que pueden tener un componente normativo importante. Estos conceptos tienen una función clasificatoria: tratan de “poner orden” en una realidad (en este caso, psicológica) que, antes de ellos, no está completamente estructurada. Al construir estos conceptos, los juristas deben tener en cuenta dos tipos de criterios de corrección: (1) deben tener en cuenta que la realidad objetiva tiene en sí misma algún grado de organización o estructura que los conceptos deben tratar de reflejar y (2) deben diseñarse de manera que satisfagan en la mayor medida posible su finalidad. Al fin y al cabo, los conceptos son en buena medida herramientas, y su finalidad, en el caso del Derecho, es permitir una aplicación adecuada de las normas y la adscripción correcta de responsabilidad de acuerdo con las mismas. Esto hace que los conceptos de acción, intención, causalidad, etc. que usan los juristas tengan un fuerte componente normativo en el momento en que se construyen. Ahora bien, una vez fijados estos conceptos, la comprobación de si ha ocurrido en la realidad un hecho subsumible en ese concepto no tiene por qué ser una ficción, sino que ya requiere “una mirada” a la realidad empírica (GONZÁLEZ LAGIER, 2007).

## 2.5. La prueba de los estados mentales con contenido representacional

¿Qué argumentos y qué criterios pueden usarse en el Derecho para la prueba de los hechos psíquicos? Como hemos visto, la filosofía de la mente (con su descripción de la estrategia de la racionalización, de las teorías basadas en “reglas de psicología popular” y la estrategia de la empatía/imaginación) también puede ofrecer algunas orientaciones generales útiles para responder a esta pregunta. Sin embargo, hay que tener en cuenta una diferencia importante entre los contextos que estudia la filosofía de la mente y el que nos interesa como juristas: Las teorías de atribución psicológica que hemos visto tratan de describir cómo atribuimos estados mentales, pero no están preocupadas por el problema de la justificación de esas atribuciones, esto es, por dar razones aceptables intersubjetivamente que muestren que esas atribuciones son correctas. La prueba judicial, por el contrario, requiere la motivación o justificación de la atribución de los hechos psicológicos, lo que, como sabemos, remite a un tipo especial de razones. Esto descarta el valor probatorio directo de las estrategias de atribución en primera persona (la teoría de la simulación), por un lado porque se conciben como estrategias no inferenciales y, por otro lado, porque el hecho de que el juez piense que tendría uno u otro estado mental en la situación en la que se encuentra el sujeto al que atribuye ese mismo estado mental no es, por sí sola, una razón justificativa (salvo que se “objetive” en alguna regla de experiencia compartida y generalizada). No obstante, que carezca de fuerza justificativa no le resta importancia en el “contexto de descubrimiento”, esto es, en el proceso de formación de la hipótesis del juez.

Teniendo esto en cuenta, puede avanzarse un poco más en la prueba de los estados mentales si se parte de la distinción entre las actitudes proposicionales y los estados mentales de carácter fenomenológico. Comencemos por los primeros. En este caso es plenamente aplicable la estrategia racionalizadora.

Hay una estrecha conexión entre lo que llamamos una acción intencional y lo que podemos llamar racionalidad estratégica o racionalidad de medio a fin. Actuar intencionalmente es actuar con un propósito, persiguiendo un fin. Pero solo podemos decir seriamente que perseguimos un fin si sabemos cómo obtenerlo, o si confiamos en llegar a saberlo; esto es, si sabemos qué hemos de hacer para lograr ese fin. Esto establece una fuerte conexión conceptual entre (1) las creencias acerca del mundo y de las relaciones de medio a fin, (2) las intenciones y deseos y (3) las acciones. Esta relación se puede expresar con lo que se ha llamado el Principio de Racionalidad Mínima (PRM) (GONZÁLEZ LAGIER, 2003 y 2004). El PRM puede enunciarse de la siguiente manera: «Si un agente actúa intencionalmente, tiende a realizar la acción que en ese momento cree más adecuada para lograr el fin que persigue».

El PRM es un principio emparentado con el principio de racionalidad o al de caridad (puede verse como una consecuencia del mismo), que hemos visto antes

como presupuesto de lo que hemos llamado la estrategia racionalizadora de atribución psicológica, pero en lugar de referirse a una racionalidad objetiva, pone el acento en lo que el sujeto cree, a la luz de su evaluación del contexto en un momento determinado, que es lo más racional o adecuado. Se trata, por tanto, de una racionalidad mínima o imperfecta, que no tiene por qué coincidir con lo que el sujeto consideraría racional en una situación “menos tensa” (es decir, sin urgencia por decidir) o con lo que sería “objetivamente racional”. El PRM tiene, en mi opinión, un fundamento, por un lado, empírico y, por otro, conceptual. Tiene un fundamento empírico porque este principio, o alguno semejante, es el que parece estar detrás de nuestras atribuciones de actitudes proposicionales que, como ya hemos visto, parecen ser en buena medida exitosas. Tiene también un fundamento conceptual en la medida en que hemos construido la noción de intención de manera que si un sujeto no elige en un momento determinado el medio que le parece más adecuado para lograr un fin concreto, no diríamos que tiene la intención de conseguir ese fin (no es infrecuente que una correlación empírica fundamente un concepto: por ejemplo, la constatación de que el calor dilata los metales fundamenta una definición de metal que incluye la propiedad de ser dilatado por el calor).

Si se acepta el PRM, entonces la prueba de la intención podría verse como una inferencia del siguiente tipo:

- (1) Los agentes realizan la acción que creen más adecuada para lograr el fin que persiguen (PRM).
- (2) S creía que la manera más adecuada de conseguir F consistía en hacer A en las circunstancias C
- (3) S hizo A en las circunstancias C

- 
- (4) S hizo A con la intención de conseguir F

Veamos un ejemplo: Alfredo y Herminio viven en huertas colindantes. Una tarde, Alfredo llama a grandes voces a Herminio, con el que anteriormente había mantenido múltiples y acaloradas discusiones. Al asomarse este a la puerta de su casa entablan una fuerte discusión sobre la propiedad de una vereda que separa ambas fincas. La discusión es interrumpida por el yerno de Herminio, quien asiéndole del brazo lo mete en la casa. A continuación, Alfredo entra en su propia casa, coge una escopeta IS, modelo PR, del calibre 12 y vuelve a salir, retando a Herminio a gritos; le sigue su hermana Arsenia, quien, agarrándole del brazo, trata de disuadirle. Cuando Herminio se asoma de nuevo a la puerta, Alfredo le dispara un cartucho de perdigones desde una distancia de unos 12 metros. En el momento en que Alfredo dispara, Arsenia tira de él hacia atrás para desviar la trayectoria del disparo. Los perdigones impactan en un muro de piedra que está detrás de Herminio (a un metro de altura), sin que este resulte lesionado. Se calcula que la trayectoria del disparo se ha desviado 70 cm. respecto de la que hubiera matado o herido a Herminio. Estos hechos son calificados por el tribunal que los juzga como un delito de homicidio en grado de tentativa, lo

que significa que se da por probado (a) que la acción de Alfredo es intencional y (b) que su intención es la de matar a Herminio y no meramente lesionarle (en cuyo caso se le podría haber condenado de un delito de lesiones en grado de tentativa).

¿Cómo sabemos que Alfredo disparó con la intención de matar a Herminio, y no meramente de lesionarle? La respuesta es que, de acuerdo con el Principio de Racionalidad Mínima, los sujetos procuran realizar la acción que, a su juicio, tiende a asegurarles el resultado que pretenden y la acción de Alfredo fue instrumentalmente adecuada para matar a Herminio (como sabemos por el calibre de la escopeta usada, el lugar del cuerpo de Herminio hacia donde apuntó, la distancia, etc.).

El razonamiento sería el siguiente:

- (1) Los agentes realizan la acción que creen más adecuada para lograr el fin que persiguen (PRM).
  - (2) Alfredo creía que la manera más adecuada de matar a Herminio consistía en disparar en las circunstancias C
  - (3) Alfredo disparó en las circunstancias C
- 
- (4) Alfredo disparó con el fin de matar a Herminio (esto es, tenía la intención de matar a Herminio).

Como puede apreciarse, una de las premisas que necesitamos para inferir la intención hace referencia a otro estado mental, las creencias del agente respecto de ciertas relaciones de medio a fin. Como ya vimos a propósito del principio de caridad, la estrategia racionalizadora requiere asumir que el agente al que atribuimos estados mentales tiene aproximadamente creencias razonables acerca de las relaciones de medio a fin. Esto puede justificarse a partir de reglas basadas en generalizaciones empíricas, como las siguientes, propuestas (entre otras) por Ramón Ragués (RAGUÉS, 1999: 521 y ss.):

a) La consideración de una persona como imputable lleva a atribuir a un sujeto todos aquellos conocimientos cuya ausencia solo se entiende posible en quien padece algún tipo de perturbación psíquica o sensorial, o en menores de edad.

b) El hecho de que una persona haya sido normalmente socializada hace que se le puedan atribuir todos aquellos conocimientos cuya ausencia solo se concibe en sujetos que no han tenido contacto con la civilización de la que se trate.

c) La circunstancia de que en un sujeto concurren determinadas características personales o de que ocupe una determinada posición social lleva a imputarle los conocimientos cuya ausencia haría impensable que tuviera esas características o que ocupara tal situación.

Por otra parte, la coherencia que la estrategia de la racionalización exige entre creencias, intenciones y acciones, tiene algunas consecuencias importantes para la prueba.

En primer lugar, también es posible inferir las creencias de un sujeto acerca de los medios para lograr un fin a partir del PRM si conocemos su intención y la acción realizada:

- (1) Los agentes realizan la acción que creen más adecuada para lograr el fin que persiguen (PRM).
- (2) S hizo A en las circunstancias C
- (3) S hizo A con la intención de conseguir F

- 
- (4) S creía que la manera más adecuada de conseguir F consistía en hacer A en las circunstancias C

Igualmente, es posible inferir (o predecir) que el sujeto realizó (realizará) cierta acción con el siguiente argumento:

- (1) Los agentes realizan la acción que creen más adecuada para lograr el fin que persiguen (PRM).
- (2) S creía que la manera más adecuada de conseguir F consistía en hacer A en las circunstancias C
- (3) S tenía la intención de conseguir F

- 
- (4) S hizo A en las circunstancias C

En definitiva, si el agente es racional, podemos asumir las siguientes reglas de inferencia entre sus acciones, sus intenciones y sus creencias:

- a) Si conocemos su intención y sus creencias (acerca de las relaciones medio-fin adecuadas para satisfacer su intención) podemos inferir (explicar, predecir) su acción.
- b) Si conocemos la acción realizada y sus creencias (acerca de las relaciones medio-fin) podemos inferir su intención<sup>5</sup>.
- c) Si conocemos la acción realizada y su intención, podemos inferir sus creencias instrumentales (de medio a fin).

Igualmente, podemos usar las siguientes reglas (también basadas en criterios de racionalidad) como criterios de atribución (*prima facie*):

- a) Si un agente tiene la intención de hacer A, y cree que B es incompatible con A, no tiene la intención de hacer B (Principio de no contradicción de las intenciones).

---

<sup>5</sup> Esto quizá arroja cierta luz sobre la discusión en la doctrina penal acerca de si el dolo es conocimiento o voluntad: no se trata en realidad de una alternativa, porque es conceptualmente imposible saber o ser consciente de que la acción x reúne los elementos del tipo penal, realizar esa acción y no tener la intención (voluntad) de dar lugar a una acción que reúne los elementos del tipo penal.

b) Si el agente cree que el resultado B se sigue necesariamente de la acción A, no puede tener la intención de A y no tener la intención de B (Principio de transmisión de la intención a las consecuencias necesarias o previsibles).

c) Si el agente tiene la intención de hacer B y cree que A es el único medio para dar lugar a B, tiene también la intención de hacer A (Principio de transmisión de la intención a los requisitos causales o convencionales).

## 2.6. La prueba de los estados mentales de carácter fenomenológico

En resumen, para la prueba de las actitudes proposicionales disponemos de una combinación de la estrategia racionalizadora y (especialmente para las creencias) de máximas de experiencia de sentido común (lo que los filósofos de la mente han llamado “teoría de la teoría” o “psicología popular”) que sirven como criterio de atribución psicológica. En la medida en que estas estrategias coinciden con las que usamos habitualmente para inferir los estados mentales de los demás, y en la medida en que tienen un éxito que podemos medir en nuestra capacidad de comunicarnos y entendernos (éxito cuya mejor explicación es que a menudo *acertamos* cuando atribuimos estados mentales, esto es, descubrimos una realidad), no se trata de criterios para una imputación en sentido estricto o fuerte (no sensible a la verdad). Dicho de otra manera, aunque se hable de “atribución” de estados mentales, los criterios tienen un fundamento empírico, están orientados a tratar de descubrir una realidad (psicológica) y son derrotables por una prueba en contrario de más peso. Exactamente como ocurre con la prueba de los hechos físicos.

Pero ¿qué ocurre con el resto de hechos psicológicos, con aquellos que carecen de un contenido proposicional, al menos en un sentido evidente? ¿Cómo se prueba el dolor, la somnolencia, la ansiedad, etc.? Es evidente que, en estos casos, no es aplicable la estrategia de la racionalización, dado que esta requiere estados mentales con contenido proposicional, y las sensaciones puras no lo tienen (otro caso, como veremos, es el de las emociones). Para este tipo de hechos, parece más adecuada la teoría de la empatía o simulación, de acuerdo con la cual el intérprete se imagina en el lugar del interpretado y le atribuye los estados mentales que cree que él tendría en esa situación o que él tendría si fuera el otro. No obstante, como ya he señalado, esta estrategia de atribución psicológica no permite obtener razones justificativas que sean útiles para motivar las decisiones sobre la prueba de este tipo de hechos, salvo que, de alguna manera, esas “intuiciones” del juez puedan respaldarse por medio de reglas de experiencia que correlacionen ciertos indicios con los estados mentales correspondientes. Estas reglas de experiencia pueden referirse a la relación entre ciertos hechos, que asumimos como causas, y los estados fenomenológicos, que asumimos como efectos (como, por ejemplo, cuando inferimos que alguien siente dolor porque hemos visto cómo se golpea un dedo con un martillo), o a la relación entre los estados

mentales, que tomaríamos ahora como causas, y la conducta subsiguiente, que sería el efecto (como cuando inferimos que a alguien le duele el dedo porque vemos que se lo agarra fuertemente y oímos sus quejidos). Dicho de otra manera, el juez puede apelar (como, de hecho, hace) a la idea de que cualquier persona razonable tendría tales sensaciones en tales circunstancias, o actuaría de tal manera si tuviera esos estados mentales. Las apelaciones a lo que sentiría o haría una “persona razonable” o “persona media” pueden verse como la objetivización de la estrategia de la empatía (de manera que la “teoría de la simulación” acaba reconduciéndose a la “teoría de la teoría”). Pero —de nuevo— esta no es una estrategia para una mera imputación ciega a lo que de hecho ocurriera realmente, si las generalizaciones usadas tienen un buen respaldo inductivo.

## 2.7. La prueba de los estados mentales mixtos (las emociones)

Las emociones son un estado mental del que puede decirse que tiene simultáneamente contenido representacional y también un importante aspecto fenomenológico. En cierto sentido, las emociones pueden verse como una combinación de creencias (o algún estado mental cognitivo) y sensaciones (GONZÁLEZ LAGIER, 2009: Capítulo II). No siempre se ha visto de esa manera, sino que por influencia de la tradición cartesiana se han identificado solo con el aspecto sensitivo: las emociones, pensaban autores como Descartes o William James, son percepciones causadas por determinados cambios fisiológicos que experimentamos en ciertas circunstancias (GONZÁLEZ LAGIER, 2009: Capítulo I). Para esta concepción, las emociones quedan fuera del ámbito de la racionalidad (más bien, la impiden). Pero hoy en día ha habido una fuerte rehabilitación de la noción aristotélica de “emoción”, que presenta a las emociones como un tipo de juicio sobre el mundo, seguido de ciertos cambios fisiológicos, una sensación (la percepción de los mismos) y una tendencia a la acción. Un contexto típico en el que surge una emoción sería el siguiente: Los sujetos tienen un trasfondo de deseos, objetivos, proyectos, planes,... Ciertos acontecimientos pueden enjuiciarse como positivos (facilitadores) o negativos (obstáculos) en relación con tales objetivos; si el sujeto juzga un acontecimiento como un obstaculizador de su acción, se desencadena cierta sensación (en este caso, negativa, esto es, no placentera) y una tendencia a actuar de manera que se pueda sortear o resolver el obstáculo; por el contrario, si el acontecimiento es juzgado como algo que facilita el objetivo, la sensación será positiva (placentera) y se desencadenará una tendencia a actuar de manera que se aproveche esa oportunidad (GONZÁLEZ LAGIER, 2009: Capítulo II). Como puede apreciarse, en las emociones se vinculan creencias (los juicios evaluativos), deseos (los objetivos), sensaciones y tendencias a la acción. Esta conexión con deseos, creencias y acciones hace que las emociones mantengan una cierta relación con la racionalidad. Las emociones son racionales si están basadas en creencias o juicios de valor correctos o justificados, y son irracionales si no lo están. Igualmente,

las emociones son racionales si generan tendencias a la acción instrumentalmente adecuadas para resolver el obstáculo o para aprovechar la oportunidad que representa el acontecimiento que las dispara (sobre la relación entre racionalidad y emociones véase GONZÁLEZ LAGIER, 2009: Capítulo IV).

Lo anterior sugiere que la estrategia de la racionalización es aplicable también para la atribución de emociones. A partir de los objetivos de un agente, de sus creencias evaluativas y de su acción, podemos inferir las emociones que probablemente tuvo en determinadas circunstancias. De nuevo, resulta fundamental aquí la coherencia entre la acción y los estados mentales, hasta el punto de que la prueba de las emociones, como la de las intenciones o la de las creencias, puede verse como la búsqueda de una explicación coherente de la conducta del agente en la que sus estados mentales y su comportamiento se ajustan uno a otro y se explican mutuamente.

Pero, como hemos visto, la emoción tiene un componente, la sensación, que tomado aisladamente no tiene esta misma conexión con la racionalidad. En ocasiones, es este componente el que necesita ser probado, por ejemplo, cuando nos preguntamos si la emoción bajo la cual ha actuado el agente ha sido suficientemente intensa como para considerar que ha “nublado” su juicio y, por tanto, podría constituir un caso de atenuante de la culpabilidad. Para estos supuestos, la prueba ha de basarse en máximas de experiencia acerca de qué intensidad suelen tener las emociones generadas en unas u otras circunstancias, en suposiciones acerca de cómo un sujeto razonable se sentiría en tales circunstancias.

### 3. CONCLUSIONES

En las páginas anteriores he intentado ofrecer un recorrido por algunos de los temas de la filosofía de la mente que más relación tienen con el problema de la prueba de los estados mentales, especialmente aquellos que son útiles para entender qué son y para replantear la discusión acerca de si son “hechos” que deban probarse o el resultado de juicios normativos de imputación. Esto es, si forman parte de la *quaestio facti* o de la *quaestio iuris* de la decisión judicial. Esta discusión no me parece trivial, pues de ella dependen cuestiones como si la determinación de los estados mentales está sujeta a garantías procesales como la presunción de inocencia, si es posible o no (o por qué vía) el recurso ante tribunales superiores por una determinación incorrecta de los mismos o cuáles son los criterios adecuados para su atribución (si los estados mentales son hechos, el criterio de corrección de tales atribuciones será la verdad o falsedad de las mismas, pero si no son hechos, dependerá de otros objetivos que no tendrán un carácter epistemológico). En mi opinión, como hemos visto, la filosofía de la mente aporta argumentos sólidos para aceptar una concepción cognoscitivistica de la prueba de los estados mentales: Frente a los argumentos ontológicos que la rechazan es posible poner de manifiesto (a) que negar la existencia de estados mentales deja sin explicar un importante aspecto de la realidad y (b) que entre el materialismo eliminacionista y el dualismo cartesiano hay alternativas plausibles. Los argumentos

epistemológicos que aducen la imposibilidad de conocer los estados mentales de los demás con suficientes garantías, por su parte, parecen olvidar que los seres humanos poseen la habilidad de adscribir estados mentales de una manera aparentemente exitosa. Por último (aunque esto ya queda fuera de la filosofía de la mente), frente al argumento de que los fenómenos psicológicos no son jurídicamente relevantes, cabe oponer los problemas de una objetivización de la responsabilidad penal para una aplicación justa del Derecho.

Es importante también advertir que, como hemos visto, pese a esta discusión, los criterios que, *de hecho*, usan los jueces para atribuir estados mentales —al igual que los que propone la doctrina o las presunciones que establece el legislador o la jurisprudencia— normalmente resultan razonables si los estados mentales se consideran “hechos” y su determinación se considera “prueba” en el sentido habitual; esto es, se trata de criterios epistemológicamente razonables si nuestro objetivo es que las atribuciones psicológicas sean verdaderas.

Los filósofos de la mente han propuesto varias teorías para explicar cómo funciona la habilidad de las personas de hacer atribuciones psicológicas: las principales son la “teoría de la teoría” (recurso a generalizaciones o leyes), la estrategia de la racionalización (asumir la racionalidad de los agentes, de manera que la coherencia entre sus estados mentales y entre estos y su conducta permite realizar inferencias en ambas direcciones) y la “teoría de la simulación” (que pone de manifiesto la importancia de la empatía como recurso para la atribución psicológica). En realidad, más que como teorías enfrentadas parece más útil verlas como estrategias complementarias, con mayor o menor adecuación según el tipo de estado mental que se pretende conocer. No creo que estas teorías aporten criterios nuevos a los ya usados en el Derecho, pero tener en cuenta estas estrategias puede servir para darles un mayor fundamento y, quizá también, para entenderlos mejor y hacerlos más precisos.

## BIBLIOGRAFÍA

- BRUNSTEINS, P. (2008), “Algunas reflexiones acerca de la simulación mental y la perspectiva de la primera persona”, en *Arete*, v. 20, n.1, Lima.
- (2011), “El rol de la empatía en la atribución mental”, en *Revista Argentina de Ciencias del Comportamiento*, vol. 3, núm.1, Córdoba.
- CASTRO, L.G. (2017) “¿Está presente el problema de las otras mentes en la filosofía de R. Descartes?”, en *Ingenium*, vol. 11, Madrid.
- CHURCHLAND, P. (1999), *Materia y conciencia*, Barcelona, Ed. Gedisa.
- DANCY, J. (1993), *Introducción a la epistemología contemporánea*, Madrid, Ed. Tecnos.
- DAVIDSON, D. (2003) *Subjetivo, intersubjetivo, objetivo*, Madrid, Ed. Cátedra.
- FERRER BELTRÁN, J. (2021), *Prueba sin convicción*, Madrid, Ed. Marcial Pons.
- GARCÍA CARPINTERO, M. (1995), “El funcionalismo”, en Fernando Broncano (ed.), *La mente humana*, Enciclopedia Iberoamericana de Filosofía, Madrid, ed. Trotta – Consejo Superior de Investigaciones Científicas.
- GASCÓN, M. (2004), *Los hechos en el Derecho. Bases argumentales de la prueba*, Madrid, Ed. Marcial Pons.

- GOLDMAN, A. (1989), "Interpretation Psychologyzed", en *Mind & Language*, vol. 4, núm. 3
- GONZÁLEZ LAGIER, D. (2003), "Buenas razones, malas intenciones (sobre la atribución de intenciones)", en *Doxa. Cuadernos de Filosofía del Derecho*, núm. 26, Alicante.
- (2004), "La prueba de la intención y el Principio de Racionalidad Mínima", en *Jueces para la democracia*, núm. 50, Madrid.
- (2006), "La prueba de la intención y la explicación de la acción", en *Isegoría*, núm. 35, Madrid.
- (2007), "Hechos y conceptos", en *Cuadernos Electrónicos de Filosofía del Derecho*, núm. 15.
- (2009), *Emociones, responsabilidad y Derecho*, Madrid, Ed. Marcial Pons.
- GONZÁLEZ LAGIER, D. (2013), *Las paradojas de la acción*, Madrid, Ed. Marcial Pons.
- (2020), "¿Es posible formular un estándar de prueba preciso y objetivo? Algunas dudas desde un enfoque argumentativo de la prueba", en *Revista Telemática de Filosofía del Derecho*, núm. 23.
- HAI DT, J. (2012), «El perro emocional y su cola racional: Un enfoque intuicionista social del juicio moral, juicio moral», en A. Cortina (Ed.), *Guía Comares de Neurofilosofía práctica*, Granada, Ed. Comares.
- HAMPSHIRE, Stuart; HART, H.L.A. (1958), "Decision, Intencion and Certainty", en *Mind*, vol. LXVII, núm. 265, Oxford.
- IACOBONI, M.: (2013), *Las neuronas espejo. Empatía, neuropolítica, autismo, imitación o de cómo entendemos a los otros*, Ed. Katz
- JAKOBS, G. (1989), "Sobre la función de la parte subjetiva del delito en Derecho penal", en *Anuario de derecho penal y ciencias penales*, Tomo 42, Fasc/Mes 2, Madrid.
- (2001) "¿Qué protege el Derecho penal: bienes jurídicos o la vigencia de la norma?", en *Cuadernos de Doctrina y Jurisprudencia Penal*, vol. 7, núm. 11.
- KAHNEMAN, D. (2012), *Pensar rápido, pensar despacio*, Barcelona, Ed. Mondadori.
- KALPOKAS, D. (2019), "Algunas reflexiones en torno al conocimiento de las otras mentes", en Carolina Scotto, Fernando G. Rodríguez y Irene Audisio (compiladores), *Los signos del cuerpo. Enfoques multimodales de la mente y el lenguaje*, Buenos Aires, Ed. Teseo.
- LAURENZO COPELLO, P. (1999), *Dolo y conocimiento*, Valencia, Ed. Tirant lo Blanch.
- MILL, J. S. (1867): *An Examination of Sir William Hamilton's Philosophy*, 3a ed., Londres, Longmans.
- MOYA, C. (1996), "El sujeto enunciado", en Manuel Cruz, *Tiempo de subjetividad*, Barcelona, Ed. Paidós.
- (2006), *Filosofía de la mente*, Universidad de Valencia.
- PAREDES CASTAÑÓN, J.M. (2001), "Problemas metodológicos en la prueba del dolo", en *Anuario de filosofía del derecho*, núm. 18.
- QUINTANILLA, P. (2004), "Comprender al otro es crear un espacio compartido. Caridad, empatía y triangulación", en *Ideas y valores*, núm. 125, Bogotá.
- (2019), *La comprensión del otro. Explicación, interpretación y racionalidad*, Lima, Pontificia Universidad Católica del Perú.
- RAGUÉS I VALLÉS, R. (1999), *El dolo y su prueba en el proceso penal*, Barcelona, J.M. Bosch Editor.
- (2002), "Consideraciones sobre la prueba del dolo", en *La Ley*, Año XXIII, núm. 5633, Madrid.
- RIZZOLATI, G; CRAIGHERO, L. (2004), "The Mirror Neuron-System", en *Annuary Review of Neuroscience*, 27.
- SEARLE, J. (1990), *Mentes, cerebros y ciencia*, Madrid, ed. Cátedra.
- (1992), *Intencionalidad*, Madrid, Ed. Tecnos.
- (1996), *El redescubrimiento de la mente*, Barcelona, Ed. Crítica.
- (2006), *La mente. Una breve introducción*, Barcelona, Ed. Norma.
- SKIDELSKY, L. (2011), "Autoconocimiento y Atribución de Estados Mentales en Teoría de la Mente", en *Revista Argentina de Ciencias del Comportamiento*, vol. 3, núm. 1, Córdoba.
- TARUFFO, M. (2011), *La prueba de los hechos*, Madrid, Ed. Trotta.
- WOLLHEIM, R. (2006), *Sobre las emociones*, Madrid, Ed. Antonio Machado Libros.



## JUSTIFICACIÓN EPISTÉMICA, EVIDENCIALISMO ROBUSTO Y PRUEBA JURÍDICA

Edgar Aguilera

Universidad de Girona

edgar.aguilera@udg.edu

<https://orcid.org/0000-0003-0119-0515>

**RESUMEN:** En el trabajo me propongo dar cuenta (y contribuir a un posible desarrollo) de un debate poco explorado en el ámbito de la teorización sobre la prueba jurídica de corte racionalista, mismo que se origina en las filas del «evidencialismo» en epistemología (y que ha cristalizado en la propuesta de un evidencialismo denominado «robusto»). Ello a los efectos de esbozar una visión de la prueba en el derecho, no estándar, pero también racionalista, que podría emerger de la consideración de dicho debate, la cual estaría integrada por un componente aretaico (que tiene que ver con el despliegue de ciertas virtudes intelectuales), una concepción argumentativa de la prueba en el derecho y un análisis peculiar de la fuerza y el sentido de los enunciados probatorios del tipo «está probado que “p”».

**PALABRAS CLAVE:** justificación epistémica; evidencialismo; prueba jurídica, tradición racionalista; estándares de prueba.

## EPISTEMIC JUSTIFICATION, ROBUST EVIDENTIALISM, AND LEGAL EVIDENCE AND PROOF

**ABSTRACT:** My purpose in this essay is twofold: 1) to introduce and elaborate on a debate originated within the province of evidentialism in epistemology (leading to «robust» versions of evidentialism) that has not received much attention from the rationalist approach to legal evidence and

proof, and 2) to sketch the contours of a conception of legal evidence and proof that emerges from that debate, which is to a certain extent alternative but nonetheless rationalist too. This conception is comprised of an aretaic component (which refers to certain epistemic virtues), an argumentative understanding of legal evidence and proof, and a particular analysis of the force and sense or meaning of the so-called evidentiary statements of the form «it is proven that “p”».

**KEYWORDS:** epistemic justification; evidentialism; legal evidence and proof; rationalist tradition; standards of proof.

**SUMARIO:** 1. INTRODUCCIÓN.— 2. PRESENTACIÓN DE UN DEBATE EN LAS FILAS DEL EVIDENCIALISMO: 2.1. Aportes al desarrollo del evidencialismo robusto: el deber de reflexionar. 2.2. Aportes al desarrollo del evidencialismo robusto II: la virtud de ejercer autoridad sobre nuestras creencias. 2.3. Aportes al desarrollo del evidencialismo robusto III: la idea de cautela epistémica. 2.4. Aportes al desarrollo del evidencialismo robusto IV: cautela epistémica y descarte de hipótesis alternativas. 2.5. Recapitulando la discusión.— 3. EVIDENCIALISMO ROBUSTO Y PRUEBA JURÍDICA.— REFERENCIAS.

**RECOMMENDED CITATION:** EDGAR AGUILERA, 2022: «Justificación epistémica, evidencialismo robusto y prueba jurídica», in *Quaestio facti*, 3: 81-102. Madrid: Marcial Pons Ediciones Jurídicas y Sociales. DOI: [http://dx.doi.org/10.33115/udg\\_bib/qf.i3.22733](http://dx.doi.org/10.33115/udg_bib/qf.i3.22733)

## 1. INTRODUCCIÓN

Entre la prueba jurídica y la epistemología puede surgir, entre otros, un tipo de vínculo semejante al que se establece entre cierto objetivo, meta o finalidad deseable y el o los medios idóneos para alcanzarlo, es decir, un vínculo de racionalidad instrumental. Esta clase de relación ha sido explícitamente sugerida por la conocida como «Tradición Racionalista de la Prueba Jurídica» (TRPJ)<sup>1</sup>.

Para dicha tradición, la relación de racionalidad instrumental en comento se obtiene, por un lado, como resultado de atribuirle a la institución jurídico-procesal de la prueba (o más específicamente, a la actividad probatoria en el derecho), la finalidad primordial de maximizar los episodios de coincidencia o de correspondencia entre el contenido de la premisa menor o fáctica de un silogismo judicial y la realidad (o de minimizar las ocasiones de falta o ausencia de esa coincidencia, que sería la otra cara de la misma moneda) y, por otro, de considerar que el medio idóneo para lograr la referida maximización de la verdad (o la minimización del error) en el contexto de la aplicación del derecho consiste en que principalmente los legisladores y los jueces, en sus respectivas labores, tanto de diseño y/o reforma del proceso del que se trate (en el caso de los primeros), como de valoración de las pruebas de las que disponen

<sup>1</sup> Para una excelente revisión de las tesis a las que adhieren tanto la vertiente anglo-sajona, como la vertiente latina de la TRPJ, véase ACCATINO, 2019.

para resolver un asunto concreto y también al motivar sus decisiones (en el caso de los segundos), se dejen guiar por, implementen o pongan en práctica los conceptos, nociones, teorías, reglas y principios provenientes justamente del terreno de la epistemología (o racionalidad empírica) general.

Pues bien, en sintonía con esta importante influencia que, de acuerdo con la TRPJ, la epistemología debería ejercer en el derecho y el razonamiento probatorios (así como en los desarrollos teóricos sobre estas cuestiones), me propongo a continuación dar cuenta (y contribuir a un posible desarrollo) de un debate, en mi opinión, poco explorado, al menos en los trabajos que se han vuelto referentes de la vertiente latina de la tradición racionalista. Se trata de un debate que se origina en las filas del denominado «evidencialismo» en epistemología (y que ha cristalizado en la propuesta de un evidencialismo más robusto), y a él haré referencia con el propósito de esbozar una visión de la prueba jurídica que podría emerger de su consideración, la cual estaría integrada por un componente aretaico (que tiene que ver con el despliegue de ciertas virtudes intelectuales), una concepción argumentativa de la prueba en el derecho y un análisis peculiar de la fuerza y el sentido de los enunciados probatorios del tipo «está probado que “p”».

## 2. PRESENTACIÓN DE UN DEBATE EN LAS FILAS DEL EVIDENCIALISMO

El evidencialismo es una postura que en epistemología puede adoptarse cuando se aborda el problema de determinar qué es lo que hace que un sujeto esté justificado al creer algo, por ejemplo (aunque no solo), algo acerca del mundo o de su entorno, es decir, cuando adopta creencias con contenido empírico. Como ya podrá intuirse solo por su nombre, la condición que de acuerdo con esta corriente debe satisfacerse para dotar a un sujeto de justificación cuando cree algo es la de contar con evidencia (o, como en los sistemas jurídicos de la tradición del *civil law* solemos decir, con pruebas) a favor o en apoyo del contenido de su creencia, es decir, a favor de una proposición empírica cualquiera, por ejemplo, “p”<sup>2</sup>.

Ahora bien, el debate al que quiero referirme (cuyo corolario es la propuesta de una versión más robusta de justificación epistémica de corte evidencialista), tiene su origen en ciertas dudas de que la condición que define al evidencialismo (consistente en que el sujeto en cuestión cuente con pruebas a favor de “p”) sea, además de necesaria, suficiente para generar justificación epistémica, tanto proposicional como doxástica<sup>3</sup>.

---

<sup>2</sup> Una sofisticada colección de ensayos que dan cuenta de los rumbos más actuales que ha tomado la discusión en el campo del evidencialismo se encuentra en DOUGHERTY, 2011.

<sup>3</sup> La distinción entre justificación proposicional y doxástica pretende capturar la intuición de que una cosa es contar, tener o disponer de evidencia/pruebas para creer que “p”, con lo cual “p” estaría proposicionalmente justificada para S en un momento dado m, y otra cosa es que esas pruebas hayan sido,

Por su parte, las dudas mencionadas emergen de la consideración de una serie de casos en los que nuestras intuiciones parecen indicarnos que los sujetos involucrados no están justificados en creer que “p” (o en todo caso, que si están justificados, lo están en un sentido mínimo, como se verá después), pese a haberse satisfecho la condición aludida. Los casos referidos han sido planteados por Jason BAEHR (BAEHR, 2011: 89-97), y son los siguientes:

*Casos de investigación/indagación deficiente  
(o de acopio deficiente de evidencial/pruebas)*

El caso de George: George es un tipo intelectualmente holgazán, apático e indiferente. Solo le interesan las cuestiones más inmediatas y prácticas: alimentarse, llegar a la oficina a tiempo, hacer su trabajo de forma mínimamente satisfactoria, pagar las cuentas, etc. Carece de la curiosidad intelectual mínima y en el día a día no está al tanto de ninguna noticia relevante. No es de sorprender que George crea muchas cosas que no debería y que no crea otras en las que sí que debería hacerlo. En la primera categoría tenemos su creencia de que la exposición al humo de cigarro no presenta ningún riesgo significativo para la salud. Dado su muy estrecho enfoque práctico, George no tiene idea de toda la investigación fiable y publicada que trata sobre los peligros de la exposición al humo de cigarro. Sin embargo, George cuenta con cierta evidencia positiva en apoyo de su creencia. Recuerda haber estado en contacto con alguna fuente fiable hace algunos años, que reportaba que las investigaciones que hasta ese momento se habían llevado a cabo no lograban establecer una correlación entre la exposición al humo y padecer de problemas de salud serios. Hasta donde George sabe, el estado de la investigación no ha cambiado y no tiene ninguna razón para suponer que sí. El resultado es que, desde su perspectiva, su evidencia apoya (o favorece) la creencia en cuestión.

El caso de Gerry: Tiene la misma creencia que George y más o menos las mismas razones para hacerlo (el recuerdo de haber leído algo al respecto). Sin embargo, Gerry tiene un poco más de interés por las noticias del día y además, disfruta de verificar las cosas por su cuenta. El problema (del que no está enterado) es que sus indagaciones suelen ser insuficientemente rigurosas y discriminativas, tienden a ser descuidadas, a dar por sentado muchas cosas y a llegar a conclusiones apresuradas. Al escuchar que en las noticias se habló de los peligros de la exposición al humo de cigarro, Gerry decide investigar más la cuestión. Con lo primero que se encuentra y termina conformándose es un reporte publicado por una organización con estrechos lazos financieros con la industria del tabaco, el cual, en lugar de tratar el asunto de forma equilibrada, exclusivamente intenta exponer las aparentes debilidades de la investiga-

---

en efecto, el fundamento de la creencia de S en “p”. En suma, la justificación doxástica requiere no solo que haya evidencia a favor de “p”, sino que el sujeto crea efectivamente que “p” y que lo haga *sobre la base de esa* evidencia (es decir, que crea que “p” *por* las razones correctas, es decir, por razones epistémicas).

ción respectiva. Para cualquier persona intelectualmente rigurosa y discriminativa, lo poco fiable de un reporte así sería evidente. Pero no para Gerry. El resultado es que su conjunto total de pruebas apoya su creencia.

### *Casos de tratamiento/valoración/procesamiento defectuoso de la evidencia/pruebas*

El caso de Daphne: También cree que la exposición al humo de cigarro no plantea ningún riesgo serio a la salud. Igualmente cuenta con alguna evidencia positiva respecto de su creencia. Sin embargo, no es intelectualmente perezosa y es razonablemente discriminativa. Al escuchar sobre las investigaciones realizadas al respecto que parecen apuntar a lo contrario de lo que cree, decide consultar fuentes por su cuenta, y casi toda la información que encuentra indica que exponerse al humo de cigarro es peligroso para la salud. Su problema es que es hipocondríaca y que fue criada por fumadores crónicos. Debido a la extrema ansiedad que experimenta con respecto a su estado de salud, no puede aceptar la información relevante. De hecho, rápida, pero genuinamente, la olvida o suprime de su mente (no recuerda ni siquiera que tuvo contacto con ella). El resultado es que, hasta donde sabe, su evidencia continúa apoyando su creencia.

El caso de Doris: Cree que la exposición al humo de cigarro no es dañina y cuenta con alguna evidencia positiva. Al escuchar sobre las investigaciones que apuntan a lo contrario, también realiza su propia indagación de forma razonablemente diligente. Ella termina encontrando información que amenaza con refutar su creencia y, como Daphne, tampoco la acepta, pero en su caso ello se debe a su fuerte apego a fumar. Es de las pocas cosas que le dan placer en la vida. Sin embargo, su constitución cognitiva es tal que no puede simplemente olvidar la información relevante. En lugar de ello, distorsiona (sin saberlo) los aspectos críticos de la evidencia que recabó. El resultado es que, desde su perspectiva, el caso a favor de que exponerse al humo de cigarro es peligroso es muy débil, por lo que su creencia permanece favorecida por el balance de las pruebas de las que dispone.

En su análisis de esta clase de casos, algunos de los autores más representativos del evidencialismo contemporáneo, como CONEE y FELDMAN (2004), sostienen que los sujetos de estos escenarios están, en efecto, justificados en creer lo que creen, por la sencilla razón de que ellos, como cualquier otra persona, han de trabajar, en términos de procesamiento cognitivo, con lo que en un momento dado tienen a la mano, es decir, con la evidencia de la que disponen (o que perciben/creen que disponen), *aquí y ahora*, sin importar si pudieron recabar más y/o mejor información (o si pudieron depurar o no los resultados de su valoración preliminar de esa información). Y la actitud doxástica con respecto a “p” que mejor se ajusta a esa evidencia (*fit*), o a la situación probatoria (*evidential situation*) en la que, desde su perspectiva, los sujetos de los ejemplos anteriores consideran que se encuentran, es la de creencia, no la actitud de suspensión de la creencia, ni la de creer que “no p” (*disbelieve* “p”).

De acuerdo con CONEE y FELDMAN, el que pensemos que quizá lo que mejor explica por qué, por ejemplo, George se encuentra en la situación probatoria en la que juzga encontrarse (es decir, solo con elementos de juicio favorables al contenido proposicional de su creencia) es que ha sido presa del sesgo de la «visión de túnel», o el haber empleado un método muy poco riguroso de indagación o acopio de evidencia para el caso de Gerry, puede ser relevante para evaluar su carácter intelectual o sus métodos doxásticos, pero no para negarles a sus creencias resultantes el estatus de estar justificadas.

Pues bien, siguiendo a BAEHR, aun si se concede que las creencias de estos sujetos están justificadas, hemos de conceder también que lo están solo en un sentido mínimo o débil. Y esto apunta a que puede haber una noción más interesante y robusta de justificación que, sin dejar de situarse en las coordenadas del evidencialismo, no obstante, decreta que sus creencias están, desde esta otra perspectiva, injustificadas.

Antes de explorar esa ruta, BAEHR destaca que, pese a lo poco exigente que nos parezca ese sentido mínimo de justificación epistémica, según el cual, un sujeto S está justificado en creer que “p” en un momento dado m, solo *si le parece* que la evidencial/pruebas de la(s) que dispone en m apoya(n) (*support*) “p”, tiene cierto valor epistémico: el equivalente al valor que tiene el menos malo de los cursos de acción que se abren para un sujeto cuando enfrenta un «dilema trágico». En el caso de los sujetos de los ejemplos, las dos malas opciones serían creer sin pruebas/evidencia y creer con base en la perspectiva probatoria en algún sentido contaminada o distorsionada que terminan adoptando. En principio, lo segundo es menos malo.

Pero el que los sujetos de nuestros ejemplos hayan creído de la manera en que lo hicieron sea mejor que creer algo sin pruebas, no es suficiente para neutralizar la intuición de que el estatus epistémico de sus creencias sigue mereciendo una *evaluación negativa*. Y eso es porque nos parece que los sujetos *pudieron haber hecho algo más* para evitar adoptar la perspectiva probatoria que terminaron adoptando. Y es justo eso, la etiología de la creencia (de que “p” y de que hay evidencia que, luego de su balance, favorece a, o indica que “p”), o el componente o aspecto diacrónico (de la justificación), como AXTELL (2011: 73-75) lo llama, lo que también le importaría a esa noción más robusta de justificación.

En este sentido, BAEHR (2011: 98) nos pide preguntarnos cuáles son las deficiencias en el proceder de los sujetos de los ejemplos que los conducen a la adopción de sus respectivas perspectivas probatorias. En su opinión, esas deficiencias consisten en la manifestación de ciertos *vicios intelectuales*, tales como la pereza, la inatención, su falta de discriminación en lo que respecta a las fuentes, su indiferencia a la verdad, ignorar y distorsionar la evidencia en contra, el autoengaño, etc.

Como ya se puede intuir, BAEHR piensa que el antídoto para contrarrestar (o que pudo haber contrarrestado) la influencia de los vicios referidos se halla en el cultivo

de las virtudes intelectuales como la rigurosidad en la indagación, mantener una mente abierta, la honestidad, la integridad, etc.<sup>4</sup>.

Así, el ejercicio de estas virtudes podría plantearse como una condición necesaria adicional del siguiente modo: S está justificado en creer que “p” en un momento dado m, solo si 1) le parece que la evidencia/pruebas de la(s) que dispone en m apoya(n) “p” y 2) S ejercita las virtudes relevantes de su carácter intelectual en la formación de esta creencia (BAEHR, 2011: 98).

El problema es que algunas creencias justificadas surgen del modo normal o natural en que funciona el equipamiento cognitivo de los sujetos. Tal es el caso, por ejemplo, de la creencia en que «la habitación se ha oscurecido» que un sujeto, sin pensarlo mucho, casi automáticamente, tendría si de pronto se apagara la luz que antes le iluminaba, creencia que le parecería (y estaría) justificada sobre la base de su experiencia visual inmediata (claro, si no tiene razones para dudar de la fiabilidad de su percepción visual). Esta clase de situaciones apunta a que no hay necesidad del ejercicio de las virtudes intelectuales en todos los casos de adquisición de creencias justificadas.

Pues bien, si se piensa en lo que distingue a los casos de nuestros ejemplos (los casos de George, Gerry, Daphne y Doris) de los casos de justificación automática (*brute justification*), puede verse, según BAEHR, que es en los primeros, y no en los últimos, en los que la *agencia* de los sujetos realiza una contribución notable/destacada/importante (*salient contribution*) a la perspectiva o situación probatoria de los sujetos. En gran medida es su agencia la que determina la amplitud de su banco de pruebas (o el peso de su acervo probatorio, o la riqueza de su conjunto de elementos de juicio), el cual es muy pobre como resultado de una indagación insipiente o descuidada en los casos de George y Gerry, o bien, la que determina cómo dichas pruebas son tratadas/valoradas/procesadas, por ejemplo, de un modo que las suprime o distorsiona en los casos de Daphne y Doris.

Tomando en cuenta lo anterior, BAEHR (2011: 99-101) propone el siguiente criterio asociado a una noción de justificación epistémica más robusta:

---

<sup>4</sup> BAEHR es un autor que se ubica en las coordenadas de la epistemología de la virtud de corte «responsabilista». Esto significa que, a diferencia de los autores que adhieren a la epistemología de la virtud de corte «fiabilista», para quienes las virtudes intelectuales son facultades (como la visión, la audición, la memoria, la razón, etc.) fiables o conducentes a la verdad en la mayoría de las ocasiones, BAEHR piensa que las virtudes intelectuales son más bien rasgos estables de la dimensión intelectual (complementaria de la dimensión moral) del carácter de una persona. Por otro lado, el propio BAEHR caracteriza su propuesta como una versión conservadora-débil (o conservadora-modesta) de responsabilismo. Conservadora porque a diferencia de los enfoques que llama autónomos, la suya es una propuesta en la que los problemas y preguntas propios de la agenda clásica de la epistemología tradicional (como la naturaleza del conocimiento, la refutación del escepticismo o la solución de los casos tipo Gettier) siguen teniendo una gran importancia. Y modesta porque BAEHR le asigna a la noción de virtud intelectual un papel a lo sumo complementario en la forma en la que se abordan esos problemas clásicos (BAEHR, 2011b: sección 1.2.2. del capítulo primero).

S está justificado en creer que “p” en un momento dado m, solo si le parece que la evidencia/pruebas de la(s) que dispone en m apoya(n) “p”, *siempre y cuando*, si su agencia realiza una contribución notable o sobresaliente a su perspectiva o situación probatoria, S ejerza (o despliegue) su agencia con respecto a esa contribución, de una manera consistente con un desempeño intelectual o cognitivo virtuoso.

BAEHR sostiene que su noción más robusta de justificación epistémica sigue siendo evidencialista, pues, en efecto, lo que genera justificación es la posesión de buenas razones epistémicas (es decir, la posesión de evidencia o de pruebas) a favor o en apoyo de la proposición respectiva; es solo que, en algunos casos (que pueden ser la mayoría en ciertos contextos institucionales), la perspectiva o situación probatoria que adopten los sujetos ha de surgir en el marco, como resultado o teniendo como telón de fondo, un *desempeño cognitivo virtuoso*.

Por su parte, otros autores se han unido al proyecto de ofrecer nociones más robustas de justificación epistémica de corte evidencialista. Tal es el caso de MATHESON (2021), quien, basándose en la intuición de que la calidad y amplitud de lo que llama los «cuerpos probatorios» (*bodies of evidence*) de que disponen los sujetos aunado al esfuerzo puesto en su acopio debe en ocasiones contar también en la evaluación epistémica de las creencias, propone el siguiente criterio:

La actitud doxástica D hacia “p” está robustamente justificada para S en un momento dado m, solo si 1) tener la actitud D hacia “p” se ajusta o cuadra con (*fit*) la evidencia de la que S dispone en m, y 2) S ha sido *responsable* en el acopio de la evidencia relevante (MATHESON, 2021: 154-155).

Cabe destacar, sin embargo, que la propuesta de BAEHR aspira a ser más ambiciosa, pues su criterio no se limita a exigir de los sujetos que sean responsables en el acopio de pruebas, sino que también comprende el hacer algo (*lo que haría un agente intelectualmente virtuoso*) para evitar adoptar una perspectiva probatoria distorsionada en la que la evidencia en contra o a favor de “p” nos parezca más débil o más contundente de lo que sería razonable pensar.

## 2.1. Aportes al desarrollo del evidencialismo robusto I: el deber de reflexionar

Ese algo que habría que hacer para evitar adoptar una perspectiva probatoria distorsionada o irrazonable puede volverse un poco más concreto si miramos lo que otros evidencialistas, como STAPLEFORD y MCCAIN (2021), han propuesto al situarse en el punto intermedio de lo que llaman evidencialismo reduccionista (como el de CONEE y FELDMAN) y evidencialismo expansionista (como el de BAEHR).

En la versión de evidencialismo de STAPLEFORD y MCCAIN, los sujetos cognoscentes tenemos el deber (epistémico) de reflexionar (*duty to reflect*) sobre la evidencia de la que disponemos (STAPLEFORD y MCCAIN, 2021: 116), y no tenemos el deber (epistémico al menos) de indagar o investigar más con el propósito de recabar más in-

formación o pruebas adicionales (*no duty to inquire*) (STAPLEFORD y McCAIN, 2021: 117-121).

Para explicar ese deber de reflexión, los autores parten de considerar las obligaciones o deberes de rol (que son las que adquirimos por el hecho de ocupar alguno). Uno de esos roles es el de ser agentes epistémicos. De hecho, no podemos evitar desempeñar este rol en el sentido de que nos formamos o adquirimos creencias y continuaremos haciéndolo ineludiblemente. Pues bien, hay ciertas cosas que, dado ese rol que desempeñamos, debemos hacer vinculadas con nuestras creencias, cosas que pueden ser llevadas a cabo de mejor o peor manera. El deber de *creer bien o apropiadamente* las englobaría. Para determinar qué es creer apropiadamente se debe partir de que la verdad es el objetivo constitutivo de las creencias (creer que “p” es pensar que es verdad que “p”). De modo que creer apropiadamente involucra *apuntar a, o poner la mira en* adquirir creencias verdaderas. La cuestión es que no disponemos de una ruta de acceso directo a la verdad. Si es que alguna vez llegamos a ese destino, lo hacemos a través de las pruebas de las que disponemos. Así, tenemos que, para cumplir con nuestro deber de creer apropiadamente, debemos creer de conformidad con nuestras pruebas, pues al proceder de este modo estamos haciendo lo razonable para adoptar creencias verdaderas<sup>5</sup>. Lo anterior es la base de nuestra obligación de creer solo las proposiciones que se ajustan a (o que son soportadas por) nuestra evidencia. Sin embargo, el perfecto ajuste entre nuestras pruebas y nuestras creencias es inalcanzable, podemos solo aproximarnos al estado de cosas en que creemos de conformidad con las pruebas de que disponemos. Y esa aproximación nos la da el reflexionar adecuadamente en nuestra evidencia y en lo que apoya, sin lo cual, cuando se trata sobre todo de creencias inferenciales, frecuentemente podemos terminar creyendo (o no creyendo) con base en perspectivas probatorias poco depuradas o irrazonables en distintos grados.

Con respecto a que no tenemos un deber epistémico de recabar más pruebas de las que en un momento dado disponemos, estos autores conceden que quizá ese deber existe, pero en todo caso sería un requerimiento moral o práctico. Creer bien, nos dicen, es solo un asunto de alineación entre nuestras actitudes doxásticas y las pruebas disponibles en el momento de que se trate (STAPLEFORD y McCAIN, 2021: 117-118).

Sin embargo, los autores referidos conceden que de la ejecución de nuestro deber de reflexionar en las pruebas de que disponemos y en lo que apoyan, puede surgir nueva evidencia (para nosotros mismos) en la forma de conexiones inferenciales (probatorias) de las que no nos habíamos percatado antes (STAPLEFORD y McCAIN, 2021: 119-121). En ese sentido, luego de reflexionar podemos percibir que nuestras pruebas apoyan “p” o que lo hacen en un grado mayor (lo cual puede ser contrario a

---

<sup>5</sup> Dicho de otro modo, cuando creemos que “p” sin considerar prueba alguna, pese a que “p” sea verdadera, habremos creído inapropiadamente, pues no hemos mostrado haber hecho lo razonable para adquirir una creencia verdadera.

lo que pudimos pensar antes de reflexionar), o bien, que las pruebas no apoyan “p” o que lo hacen en un grado menor (que puede también ser contrario a lo que pensábamos antes de reflexionar).

En suma, según STAPLEFORD y MCCAIN, no tenemos un deber, al menos epistémico, de recabar pruebas adicionales en el sentido de «salir al mundo» y recolectarlas, pero sí el deber epistémico de registrar o tener en cuenta las pruebas adicionales en términos de las «nuevas» conexiones inferenciales que podrían surgir del cumplimiento de nuestro deber de reflexionar en las pruebas con las que en un momento dado contamos, y en lo que apoyan. El deber de creer de conformidad con nuestra evidencia/pruebas termina siendo, por tanto, un deber de creer de conformidad con alguna de las versiones o perspectivas probatorias más depuradas que podemos adoptar como resultado de nuestra reflexión.

## 2.2. Aportes al desarrollo del evidencialismo robusto II: la virtud de ejercer autoridad sobre nuestras creencias

Ahora bien, cuando queremos avanzar en la especificación de la forma que podría adoptar esa reflexión (misma que, siguiendo el pensamiento de BAEHR, consideraré un componente fundamental del desempeño intelectual o cognitivo virtuoso por parte del sujeto), no solemos llegar muy lejos consultando la literatura sobre la denominada epistemología de la virtud (o *virtue epistemology*). Y es que lo que abunda, por ejemplo, en el campo de la epistemología de la virtud de corte responsabilista, son análisis conceptuales, sin duda muy valiosos y sofisticados, pero demasiado abstractos, de ciertos rasgos del carácter intelectual de las personas, en los cuales, en general, no hay un esfuerzo por dilucidar de qué manera dichos rasgos, como la apertura de mente, la humildad intelectual, etc., se manifestarían en términos del razonamiento probatorio que debería desplegarse en determinados episodios doxásticos<sup>6</sup>.

Una excepción a lo anterior puede hallarse en el trabajo de Federico PICINALI, quien hace referencia a lo que llama la virtud de *ejercer autoridad sobre nuestras creencias* (PICINALI, 2015: 143-148). De acuerdo con el autor en comento, dicha virtud consiste en adoptar una «postura deliberativa» (*deliberative stance*) con respecto a la pregunta «¿creo que “p”?». A diferencia de lo que ocurre cuando un agente epistémico adopta una «postura teórica» (*theoretical stance*) —misma que lo hace desentenderse de la precisión o acierto de sus creencias, pues su principal interés al adoptar esta postura reside meramente en detectar si experimenta o no alguna de esas creencias—, cuando adopta la postura deliberativa, de entrada, se plantea si ha

<sup>6</sup> Esto ocurre incluso en ciertos casos en los que se intenta vincular el análisis conceptual de algún rasgo intelectual a cuestiones probatorias. En este sentido, Sharon RYAN define a la apertura de mente del siguiente modo: S es abierto de mente siempre que 1) S quiera y valore tener una adecuada comprensión de la realidad, 2) S concienzudamente recabe buena evidencia/pruebas, y 3) S sea justo en su interpretación y uso de la evidencia/pruebas de la(s) que dispone (RYAN, 2021: p.182).

de creer o no en la proposición respectiva. En este sentido, el agente epistémico trata a la pregunta «¿creo que “p”?» como equivalente a «¿es verdad que “p”?». Así, está dispuesto a asumir la responsabilidad de formarse una actitud cognitiva acertada mediante la consideración de las razones que vuelven más probablemente verdadero su contenido, es decir, mediante la consideración de las pruebas relevantes, y no (o no solo) haciendo un ejercicio de introspección que lo conduzca a percatarse del estado mental-doxástico en que se encuentra (como supone la postura teórica).

Siguiendo a PICINALI (2015: 143-148), la autoridad que ejercemos sobre nuestras creencias al adoptar la postura deliberativa es una cuestión gradual. Esta idea puede plantearse a la manera de un continuo flanqueado por dos estados ideales: de un lado, por la adopción plena, por parte del agente, de la postura teórica y, de otro, por la adopción plena de la postura deliberativa. En principio, los casos que involucran el ejercicio de esta clase de autoridad se ubicarían en diversos puntos, todos ellos cercanos, pero algunos más que otros, a la adopción plena de la postura deliberativa. Por su parte, mientras más cerca se ubique el agente de adoptar plenamente la postura teórica (es decir, mientras menos autoridad ejerza sobre sus creencias), menos atención pondrá en las razones que justifican creer que algo es el caso, su interés en considerar seriamente y ponderar dichas razones irá decreciendo progresivamente, y la equivalencia entre las preguntas «¿creo que “p”?» y «¿es verdad que “p”?» irá difuminándose (hasta un punto en el que prácticamente se haya perdido).

Es momento ahora de preguntarnos por las operaciones intelectuales que podríamos en práctica (o en las que se traduciría) la virtud de ejercer autoridad sobre nuestras creencias, virtud que, como hemos visto, tiene que ver básicamente con la consideración seria de las razones epistémicas o pruebas a nuestra disposición. PICINALI (2015: 143-148) sostiene que dichas operaciones consisten en 1) articular la generalización empírica que permite inferir la ocurrencia de un hecho a partir de otro, 2) determinar la fuerza (*evidential strength*) con la que dicha generalización correlaciona las clases de hechos en cuestión (o el grado de probabilidad con el que una clase de hecho se sigue de la otra), y 3) determinar la fiabilidad (*reliability*) de la información concerniente a (o en la que se apoya la afirmación de la existencia de) dicha correlación.

Estas operaciones dotan de un contenido más específico a ese deber de reflexionar sobre las pruebas de las que disponemos, al que antes aludimos. Y tienen el potencial de contribuir a que depuremos o refinemos las perspectivas probatorias que podemos formarnos prematuramente, debido a que su ejecución al menos pone de relieve ciertas dimensiones o atributos con los que deberían contar nuestras pruebas y las inferencias que con su contenido realizamos, que por distintas razones pudieron pasarse por alto en una primera aproximación a nuestro caudal probatorio.

Dicho de otro modo, uno puede adoptar una postura teórica a la que se refiere PICINALI, no solo con respecto a la proposición de que se trate, sino también con respecto a las pruebas disponibles (en las que, en principio, dicha proposición se sustenta), conduciéndonos a su consideración irreflexiva, apresurada o prematura, como si las pruebas fueran objetos dados, incuestionables. Y la ejecución de las ope-

raciones en comento trazarían una ruta a seguir para adoptar la postura deliberativa, de nuevo, no solo con respecto a la proposición en cuestión, sino primordialmente con respecto a las pruebas con las que contamos.

Ahondando en este punto, la adopción de la postura teórica implicaría sucumbir a lo que podríamos denominar la tentación de cosificar las pruebas de las que disponemos, es decir, al impulso de tratarlas como ítems dados, respecto de los que no cabe más que percatarse de su presencia o ausencia y, en caso de estar presentes, considerar incuestionable la impresión inicial que nuestra mente se forma concerniente al apoyo que le brindan a la proposición de que se trate. Esta, curiosamente, parece corresponderse con la noción mínima de justificación epistémica que defienden los evidencialistas reduccionistas u ortodoxos como CONNE y FELDMAN, a la cual invocan en su análisis de los ejemplos presentados por BAEHR, en el que concluyen que los protagonistas de casos como esos están justificados en creer lo que creen. Por su parte, la adopción de la postura deliberativa implicaría no solo percatarse de la presencia de las pruebas respectivas, sino ser críticos con ellas y, por tanto, preguntarse por su fiabilidad, cuyo grado es una conclusión (y énfasis «conclusión» para hacer más evidente el carácter argumentativo de este proceder), que, de acuerdo con PICINALI, se alcanza determinando la fuerza de la correlación de la que da cuenta la generalización empírica empleada, así como la solidez o calidad del fundamento epistémico de aquella.

### 2.3. Aportes al desarrollo del evidencialismo robusto III: la idea de cautela epistémica

Ahora bien, dado su potencial de conducirnos a refinar o depurar nuestra perspectiva o situación probatoria (alejándonos con ello más y más de asumir una postura teórica con respecto a las pruebas), sostengo que las operaciones intelectuales que PICINALI considera involucradas en la adopción de una postura deliberativa pueden concebirse como elementos importantes de una suerte de repertorio o arsenal de medidas epistémicas cautelares que, junto (aunque no solo, como veremos a continuación) con el acopio responsable del conjunto de pruebas relevantes (responsabilidad que conlleva una tendencia a enriquecerlo lo más posible), estarían orientadas a disminuir el riesgo de que terminemos adquiriendo creencias falsas.

Dicho riesgo, por supuesto, nunca puede eliminarse del todo, pues ningún conjunto de pruebas con el que podamos contar es apto para fundar certezas empíricas racionales. Por su parte, el grado tolerable del mismo es una cuestión contextual, en el sentido de que la magnitud deseable de este riesgo es sensible principalmente a la importancia de las acciones que queremos emprender sobre la base de la o las creencias de que se trate, y sensible también a las consecuencias o costos que se seguirían si lleváramos a cabo esas acciones y la o las creencias de partida fuesen falsas. Así, mientras más importantes sean dichas acciones y más costosas las consecuencias del error,

más dispuestos deberíamos estar a emplear con mayor intensidad (y a reflejar en la argumentación respectiva que podemos hacer para nosotros mismos y/o para otros), las operaciones intelectuales y los actos (como el de recabar más pruebas) que conforman el repertorio referido de medidas epistémicas cautelares, so pena de arriesgarnos a ser receptores del reproche (epistémico-moral e incluso, jurídico) al que estaríamos expuestos si la cautela en comento fuese considerada insuficiente<sup>7</sup>.

#### 2.4. Aportes al desarrollo del evidencialismo robusto IV: cautela epistémica y descarte de hipótesis alternativas

Sin embargo, el repertorio de cautelas epistémicas que viene emergiendo de la discusión, prescinde de una muy importante medida para reducir el riesgo de adquirir creencias falsas, que se implementa de manera más explícita, por ejemplo, en muchas de las prácticas científicas y en las prácticas diagnósticas de los médicos: me refiero a la consideración y descarte o refutación de hipótesis alternativas. A efecto de incluir dicha medida recurriremos a la denominada justamente «teoría de las alternativas relevantes» (*Relevant Alternatives Theory*), que se ha solido emplear en algunos de

---

<sup>7</sup> Considerar las operaciones asociadas a la adopción de la postura deliberativa (o a la virtud de ejercer autoridad sobre nuestras creencias) como acciones (mentales en este caso) cautelares epistémicas tendientes a reducir el riesgo de adquirir alguna(s) creencia(s) falsa(s) en un momento determinado, y al sujeto que no las lleva a cabo (con la intensidad o grado apropiado) como merecedor de un reproche epistémico, moral y quizá, jurídico, conlleva, en mi opinión, el emparentamiento de la concepción robusta-aretaica de evidencialismo que estamos desarrollando aquí con base en las ideas de BAEHR, con las denominadas concepciones «deontológicas» de la justificación epistémica, las cuales plantean el fenómeno de la justificación epistémica en términos de derechos, deberes, obligaciones, responsabilidad, culpa, etc. Si se acepta la alegación frecuente de que no es posible hablar de deberes doxásticos porque debe implicar puede y no parece que podamos simplemente decidir tener (o no tener) una creencia, las concepciones deontológicas de la justificación epistémica no parecen llevarnos por la ruta adecuada. Sin embargo, autores contemporáneos como Rik PEELS (2017), muestran que no hemos entrado a un callejón sin salida. Su propuesta deontológica de justificación epistémica (PEELS, 2017: 237-250) parte de conceder la objeción anterior, es decir, de asumir que no tenemos ningún tipo de control (directo, indirecto o compatibilista) sobre nuestras creencias (PEELS, 2017: 52-88). Sin embargo, sí que tenemos control sobre ciertos factores a los que llama «*belief-influencing factors*», cuya manipulación por parte del sujeto cognoscente hace una diferencia respecto de lo que termina creyendo. Dichos factores no son otra cosa que un conjunto de acciones (como recolectar pruebas, cultivar las virtudes y desmontar o aminorar los vicios intelectuales, así como mejorar la fiabilidad de nuestros procesos de formación de creencias, como la percepción, la memoria y especialmente nuestro razonamiento) que, de no llevarse a cabo (con la intensidad contextualmente relevante), muy probablemente nos conducirían a formarnos creencias falsas. Ese destino al que muy probablemente nos conduce su falta de ejecución, aunado al control que sobre dichas acciones tenemos, son la base del deber epistémico (y moral, si se incluye en el análisis las acciones que se realizan sobre la base de las creencias en cuestión) de llevarlas a cabo, y del elogio o reproche que merecemos respectivamente por su cumplimiento o por su inobservancia. A dicho deber PEELS se refiere en plural mediante la expresión «obligaciones intelectuales» (PEELS, 2017: 89-131), y en su cumplimiento (o incumplimiento sin culpa) radica nuestro grado de justificación epistémica.

los análisis de las condiciones bajo las cuales un sujeto *sabe* que “p”<sup>8</sup>. Sin embargo, ello no obsta para que la incorporemos en el análisis de una noción más robusta de justificación epistémica evidencialista.

De acuerdo con esta teoría (proyectada, como hemos dicho, al ámbito de la justificación), un sujeto se encontraría justificado en creer que “p” en un grado cada vez mayor o, dicho de otro modo, se encontraría cada vez más a salvo de equivocarse, mientras más hipótesis contextualmente relevantes e incompatibles con “p” (incompatibles en el sentido que, si fueran verdaderas, “p” sería falsa), pero consistentes, como “p” lo es, con alguna(s) de las pruebas disponibles, queden, desde la perspectiva de S, descartadas a la luz de la consideración de subconjuntos más amplios de las pruebas de las que se dispone. Tal vez se entienda mejor esta propuesta con los siguientes ejemplos:<sup>9</sup>

Supongamos que Ana, bióloga de profesión, estando un día en el zoológico con su hija pequeña, llama su atención para comunicarle el contenido de la creencia que experimenta diciendo «ahí tenemos una cebra». Desde un nivel alto de abstracción podríamos decir que la evidencia con la que cuenta consiste en percibir un animal grande de cuatro patas y cola. Descrita así, la evidencia disponible es consistente no solo con la hipótesis de que el animal sea una cebra, sino también con las hipótesis de que se trate de un camello, un yak, un venado, etc., y dado que los zoológicos suelen también exhibir esos especímenes, ignorar esas posibilidades o alternativas sería inapropiado, pues si se ignoran (es decir, si no se determina si la evidencia es o no consistente con ellas), uno no estaría suficientemente a salvo del error. Pues bien, descendiendo en el nivel de abstracción, el contenido de su percepción visual podría especificarse más hasta decir que lo que Ana observa es un équido con rayas blancas y negras por todo su cuerpo. Esta evidencia parece suficiente para descartar que se trate de un camello, un yak o un venado, pero sigue siendo consistente con que delante se tenga, por ejemplo, un okapi. Debido a los conocimientos especializados que posee, Ana sabe (predice) que, si ese fuera el caso, las rayas de la creatura no continuarían a lo largo del torso, pero al echar un segundo vistazo, se percata de que sí lo hacen. Esta observación descarta la hipótesis del okapi, pues no se confirma la predicción anterior; la evidencia no es consistente con esa alternativa. En este punto, todavía sería posible conjeturar escenarios alternativos en los que, con base en la evidencia de la que dispone (la percepción visual de un équido cuyas rayas blancas y negras continúan también por el torso), Ana crea y diga «ahí tenemos una cebra» y, no obstante, ello fuese falso. Podría ser, por ejemplo, que el animal efectivamente fuera un okapi, pero con una coloración anormal (quizá por alguna anomalía o modificación genética), o bien que se tratara de una mula meticulosamente disfrazada como cebra,

<sup>8</sup> GARDINER recurre a esta teoría para identificar las operaciones cognitivas que subyacen a la determinación de la satisfacción de algún estándar de prueba jurídico y para explicar cómo podría emplearse a los efectos de especificar estándares de prueba progresivamente más severos o demandantes (GARDINER, 2019).

<sup>9</sup> Tomados, con leves modificaciones, de GARDINER, 2019.

o de un sofisticado robot con la misma apariencia y comportamiento que los de una cebra. Sin embargo, al menos en este contexto, sería absurdo exigir que la evidencia de Ana tenga que ser inconsistente con (o que tenga que descartar) dichos escenarios posibles. Cuando la evidencia alcanzó para descartar la hipótesis del okapi, ya fue razonable atribuirle un grado de justificación contextualmente adecuado a su creencia de que lo que tiene enfrente es una cebra, pues en ese momento (y repetimos, en este contexto) ya estaba lo suficientemente a salvo (o lo suficientemente protegida) del riesgo de equivocarse.

Consideremos ahora el caso del arquitecto Mario. En algún momento de su paseo por el mismo zoológico se detiene para decirle a su hija menor, tal como Ana lo hizo, «mira, una cebra». Como la mayoría de los adultos, Mario puede distinguir entre una cebra y un camello, un yak o un venado, razón por la cual, su experiencia visual es suficiente para descartar esas alternativas. No obstante, dado que el poder de discriminación de sus observaciones no se asemeja a la «mirada de experto» que Ana posee sobre estas cuestiones, por más que mire y vuelva a mirar, no podría diferenciar entre una cebra, un cebroide y un okapi; parece entonces razonable que todavía no podamos atribuirle a Mario un grado de justificación significativo al creer que es una cebra lo que él y su hija tienen enfrente, pues la sombra del error no se ha difuminado lo suficiente. En este punto Mario, como casi cualquier otra persona podría hacer en el contexto de la visita a un zoológico, realiza la predicción de sentido común de que, si no se tratara de una cebra, no habría un letrero en alguna parte de la puerta de la jaula en el que se lea «cebra». Al percatarse de la presencia de un letrero con esas características, su predicción de lo que cabría observar si no se tratase de una cebra, no se cumple, con lo cual, en este contexto, es razonable pensar que queda descartada o eliminada la hipótesis del okapi (aunque Mario simplemente albergara la duda genérica de que se tratara de una cebra y no hubiera contemplado la hipótesis más específica de que pudiera tratarse de un okapi). Podría ser, por supuesto, que, por accidente o deliberadamente, el letrero estuviera mal colocado donde está. Sin embargo, en este contexto, el que la evidencia con la que se cuenta (la percepción del animal con rayas blancas y negras y el letrero que dice «cebra») no descarte esos escenarios, es compatible con afirmar que Mario está justificado en un grado considerable al creer que hay una cebra en esa jaula. Por otro lado, el que la evidencia de la que disponen Ana y Mario sea consistente con (o no sea apta para descartar) hipótesis extravagantes como que el mundo surgió apenas hace cinco minutos, que puedan estar alucinando, o que en realidad Ana y Mario son solo cerebros conectados a una computadora que les suministra los estímulos eléctricos necesarios para hacerlos pensar en cada instante que tienen un cuerpo que interactúa con su entorno, también es compatible con (o no le resta mérito a) afirmar que los protagonistas de ambos ejemplos creen, con un grado contextualmente apropiado de justificación, que hay una cebra frente a ellos.

Como puede constatarse en estos ejemplos, la teoría de las alternativas relevantes pone el acento en la habilidad de los sujetos cognoscentes de determinar si ciertos escenarios alternativos al inicialmente contemplado han quedado o no descartados

a la luz de la consideración de evidencia adicional a aquella con la cual todos los escenarios (el inicial y los alternativos) son consistentes en un momento dado. Pero esta operación intelectual, a su vez presupone, por un lado, la habilidad de conjeturar diversos escenarios consistentes con, o explicativos de, los mismos datos o pruebas iniciales (es decir, la habilidad de razonar abductivamente) y, por otro, la habilidad de predecir qué otra información, pruebas o datos deberían también poderse recabar o, de algún modo, registrar, si fuesen verdaderos cada uno de esos escenarios, lo cual abre la posibilidad de contrastar dichas predicciones con el mundo a los efectos de establecer si las expectativas probatorias u observacionales respectivas se satisfacen o no. Si el sujeto en cuestión determina que no se satisfacen, en ello podrá fundar su consideración de que un escenario particular ha sido descartado o eliminado, tal y como lo hicieron los protagonistas de nuestros ejemplos, Ana y Mario, al descartar la hipótesis del okapi, y como suelen hacerlo, como antes dijimos, los médicos con los llamados diagnósticos diferenciales que no son consistentes con los resultados de los *tests* a los que son sometidos<sup>10</sup>. Sin embargo, esta consideración (de que un escenario ha quedado descartado) no puede concebirse como concluyente, y ello debido a que el incumplimiento (así como el cumplimiento) de las predicciones referidas constituye un fenómeno que a su vez puede ser explicado por múltiples hipótesis, lo cual pone de manifiesto el carácter recursivo de la operación de descartar hipótesis alternativas, así como el carácter ampliativo del razonamiento abductivo.

Una cuestión más a la que quiero referirme con respecto a la teoría de las alternativas relevantes es que la hipótesis de que cada una de las pruebas recabadas (o algunas de ellas) no sea fiable (en el grado deseado) podría considerarse a su vez como hipótesis relevante, es decir, como una más de las hipótesis que deberían ser descartadas a través de la determinación de que no fueron confirmadas las expectativas probatorias que serían consistentes con su carencia de (el grado deseado de) fiabilidad. Dicho de otro modo, las operaciones intelectuales a las que se refiere PICINALI al desarrollar la idea de la adopción de la postura deliberativa pueden replantearse como un episodio más del recursivo o reiterativo despliegue de la operación básica (o de la operación maestra) de descartar escenarios alternativos.

## 2.5. Recapitulando la discusión

Recapitulemos, para cerrar esta sección, algunos de los aspectos más destacables de la reconstrucción que hemos hecho de la discusión surgida a propósito de los casos propuestos por BAEHR:

Los casos referidos condujeron a BAEHR a plantear los cimientos de una noción más robusta de justificación epistémica de corte evidencialista, para la cual, cuando

---

<sup>10</sup> Para una explicación de cómo figuran la abducción y la eliminación de hipótesis alternativas en el razonamiento médico-clínico, véase LAWSON y DANIEL, 2011.

se trata de creencias con un fuerte componente inferencial —que contrastan con las creencias cuya justificación es comúnmente considerada automática, como el caso de muchas de nuestras creencias perceptuales, y que con frecuencia son producto (me refiero a las creencias inferenciales) de un proceso mayor o menormente riguroso de indagación (a lo que en la literatura de habla inglesa se le denomina «*inquiry*»)—, resulta sumamente relevante determinar si lo que los sujetos hicieron para adoptar su respectiva situación o perspectiva probatoria es consistente con un nivel o grado adecuado de comportamiento virtuoso, que incluye su desempeño intelectual o cognitivo. Dicho comportamiento tiene que ver, al menos, con la búsqueda o acopio diligente de pruebas y con el cumplimiento de un deber de reflexión que idealmente debería llevar al sujeto a contar con alguna modalidad más refinada de su perspectiva probatoria. En los términos de PICINALI, dicha reflexión (a la que llama la virtud de ejercer autoridad sobre nuestras creencias), se encuentra vinculada a la adopción de una postura deliberativa con respecto a una determinada proposición, misma que contrasta con la denominada postura teórica, bajo la cual el sujeto en cuestión hace énfasis meramente en detectar si experimenta o no una creencia cuyo contenido coincide con la proposición de que se trate, y/o desde la cual, una vez detectada la creencia y si es que hay pruebas a su favor, el sujeto trata dichas pruebas como ítems dados o estructuras acabadas, es decir, sin cuestionar la impresión inicial que se forma con respecto al apoyo que las pruebas le brindan a la proposición (caso en el cual, la postura teórica en comento se corresponde con la noción mínima de justificación evidencialista defendida por CONNE y FELDMAN). Echando mano de la teoría de las alternativas relevantes, y dado que es posible replantear en los términos de esta a las operaciones que, siguiendo a PICINALI, conducen a establecer el grado de fiabilidad de las pruebas disponibles, sostengo que la ejecución con la intensidad contextualmente apropiada de la operación intelectual de descartar escenarios alternativos en gran medida condensa o materializa lo que significa exhibir un desempeño cognitivo virtuoso con respecto a las pruebas y a las inferencias que con su contenido pueden hacerse, y que, junto a la diligente confección del conjunto de pruebas relevantes (diligencia que se traduce en una propensión a enriquecer dicho conjunto con más y/o mejores pruebas), constituye una cautela epistémica muy importante para reducir el riesgo de adquirir creencias falsas. El grado apropiado de tal cautela, como hemos dicho, depende del (o es sensible al) contexto de acción en el que el sujeto se halle, es decir, de su valoración moral de la gravedad o importancia de las acciones que quiere emprender y de los costos de llevarlas a cabo sobre la base de creencias falsas. Mientras más importantes sean esas acciones y más costosos los errores, mayor debería ser el grado de la cautela epistémica a emplearse. Por su parte, mientras mayor sea el grado de cautela epistémica considerado apropiado, con mayor intensidad habría de desplegarse la operación de descartar hipótesis o escenarios alternativos (ya sea al nivel macro de las hipótesis principales, y/o al nivel micro, que correspondería al nivel de la fiabilidad de las pruebas en que las hipótesis se sustentan).

Demos paso a continuación al esbozo de la visión de la prueba en el derecho que podría surgir a partir de estas consideraciones.

### 3. EVIDENCIALISMO ROBUSTO Y PRUEBA JURÍDICA

De acuerdo con el evidencialismo robusto delineado en la sección previa, para estar justificado (en cierto grado) el sujeto cognoscente debe encauzar su agencia de tal suerte que la contribución que esta realice a la configuración de su perspectiva o situación probatoria sea consistente con un desempeño cognitivo virtuoso. Dicha contribución debería manifestar el grado de cautela epistémica considerado (moralmente) apropiado; cautela que, por su parte, tiene que ver fundamentalmente con la diligente recolección de pruebas relevantes y con el descarte de las hipótesis alternativas (entre las cuales podría contemplarse incluso la hipótesis de que las pruebas disponibles o algunas de ellas, no fuesen fiables) que, en el contexto de que se trate, se considere necesario determinar que son inconsistentes con las pruebas recabadas a los efectos de que el sujeto se halle lo suficientemente a salvo de la adquisición de una creencia errónea.

Pues bien, si la implementación en el ámbito jurídico de esta versión de evidencialismo robusto se considera plausible, el entorno institucional en el que el juzgador se desenvuelve (el proceso judicial) debería estar en condiciones de potenciar el desempeño cognitivo virtuoso en el que esa versión de evidencialismo hace énfasis (de ahí el elemento aretaico de este esbozo de la prueba jurídica que se anunció en la introducción).

En este orden de ideas, dado que el acopio diligente de las pruebas relevantes (acopio que tiende a su enriquecimiento continuo) constituye un componente fundamental de dicho desempeño virtuoso, un posible punto de partida para la creación de las condiciones aludidas podría ser el reconocimiento serio de la necesidad de que los distintos tipos de proceso cuenten con mecanismos aptos para la adecuada gestión de lo que Dale NANCE llama la «optimización del peso keynesiano» del conjunto o acervo de pruebas que será tomado en cuenta en la decisión probatoria final (NANCE, 2016)<sup>11</sup>.

No es esta la ocasión de profundizar en la discusión de los mecanismos de los que NANCE da cuenta (como la imposición a las partes de la carga de producir prueba con respecto a diversos puntos de la controversia, ciertas reglas de exclusión probatoria, la figura del descubrimiento probatorio, etc.). No obstante, considero que vale la pena tomar nota de que entre ellos se encuentran algunas propuestas más innovadoras y controvertidas que otras, como la de permitirle al juzgador (juez o jurado) intervenir en la práctica de las pruebas, por ejemplo, haciendo también preguntas a los testigos. A ello podríamos agregar la posible implementación de modos de obtener informa-

---

<sup>11</sup> Un conjunto de pruebas tiene mayor peso mientras más información o evidencia relevante incluye, con independencia de a cuál de las hipótesis contempladas dicha evidencia le confiere corroboración (o a cuál de ellas contribuye a refutar).

ción testimonial más consistentes con la ciencia relevante (como la psicología del testimonio) y, en ese sentido, más fiables. Tales modos podrían hacernos replantear las prácticas del contrainterrogatorio —de manera de que fuesen menos hostiles o menos orientadas a exhibir y hasta ridiculizar a las personas—. De hecho, la consideración seria de los mecanismos de los que se podría echar mano para la optimización del peso keynesiano de las pruebas podría, de forma más general, llevarnos a matizar uno de los rasgos más prominentes de nuestras culturas jurídico-procesales; me refiero al carácter adversarial del proceso. Dicha matización parece ser, cuando menos, compatible con el proyecto de crear las condiciones aptas para que este entorno institucional potencie un comportamiento, de parte de quienes en él participan, consistente con un desempeño cognitivo virtuoso.

Ahora bien, la implementación de mecanismos aptos para la optimización del peso keynesiano del acervo probatorio del que dispondrá el juzgador constituye uno de los componentes de la cautela epistémica a la que hace referencia nuestra versión robusta de justificación evidencialista. Toca el turno entonces de abordar el segundo componente, o sea, la operación de descartar escenarios alternativos contextualmente relevantes.

Esto nos remite al ámbito de los estándares de prueba o umbrales de suficiencia probatoria que en distintas ramas del derecho procesal pretenden establecerse mediante reglas que contienen, entre otras, frases como «la preponderancia de la evidencia», «prueba clara y convincente», «más allá de toda duda razonable» o «convicción íntima». Referirnos a este componente de la cautela epistémica nos conduce al ámbito mencionado dado que se podría aprovechar la práctica de formular dichos estándares para sugerir que esto se haga justamente en términos del descarte de hipótesis o escenarios alternativos relevantes. Siguiendo a GARDINER (2019), la exigencia de que sean descartados solo los escenarios cercanos, o bien, los cercanos y los distantes, o bien, los cercanos, los distantes y los remotos, se correspondería con la intención de implementar acumulativamente estándares cada vez más exigentes y, por tanto, con la intención de que sean desplegados niveles de cautela epistémica progresivamente más altos. Por su parte, y como hemos dicho en múltiples ocasiones, el nivel apropiado de este componente de la cautela epistémica dependería del contexto de que se trate, es decir, primordialmente del momento procesal en cuestión, de la decisión a tomarse (por ejemplo, imponer una medida cautelar o condenar a alguien) y de los costos de que esa decisión se tome sobre la base de una determinación fáctica errónea.

Es importante destacar que la formulación de estándares de prueba en términos del descarte de hipótesis o escenarios alternativos no puede más que hacer referencia de manera aproximada u orientativa al nivel de cautela que se espera que sea incorporado en el razonamiento que antecede a declarar probada cierta proposición. Dicho de otro modo, los estándares de prueba así articulados meramente dirigen (aunque de forma muy importante) la atención del juzgador al rango de argumentaciones que

podrán usarse válidamente para tener por probada una determinada clase de hecho<sup>12</sup>. Y estas serán las que conduzcan razonablemente a la conclusión (o las que incluyan la determinación razonable) de que han quedado descartadas las hipótesis o escenarios alternativos cuyo descarte o refutación haya sido considerada consistente con el (o equivalente al) grado de cautela epistémica deseado<sup>13</sup>.

Lo anterior significa que los estándares de prueba, cuando son articulados siguiendo la sugerencia de FERRER (2021: 29-33), es decir, apelando a criterios relativos a la capacidad justificativa del acervo probatorio disponible respecto de las conclusiones probatorias alcanzadas (como sostengo que es el caso cuando se centran en el despliegue, con distintos grados de intensidad, de la operación de descarte de hipótesis o escenarios alternativos relevantes), no resultan ser puntos o parámetros fijos. Más bien, se trata de pautas en buena medida elusivas en el sentido de que la determinación de que dichos estándares han sido o no satisfechos no es concluyente. La determinación referida no es concluyente pues, con todo y las ventajas que acarrea la articulación de algún estándar de prueba en estos términos<sup>14</sup>, ello no es apto para eliminar el espacio en que podrían tener lugar desacuerdos razonables, es decir, desacuerdos en los que las posiciones enfrentadas se fundan en argumentos plausibles. Dichos desacuerdos pueden girar en torno a distintas cuestiones, como pueden ser, al menos, las relativas a la amplitud, alcance o cobertura de las categorías sugeridas de hipótesis o escenarios alternativos «cercanos», «más distantes» y «remotos», y/o las relativas a las condiciones en que han de considerarse descartados o refutados, es decir, en torno a las predicciones o expectativas probatorias que habrían de incumplirse.

Lo dicho hasta aquí sobre los estándares de prueba puede tener, en mi opinión, ciertas repercusiones adicionales en cuanto al análisis que puede hacerse acerca de la fuerza y el sentido de los denominados enunciados probatorios del tipo «está probado que “p”» (es decir, sobre lo que se hace y se dice cuando se emiten enunciados de este tipo). Para cerrar, procedo entonces a plantear someramente dichas repercusiones:

Con respecto a su fuerza, la perspectiva desarrollada en este trabajo conduce a sostener como más adecuado un análisis de corte expresivista de los enunciados probatorios<sup>15</sup>. En este orden de ideas, dichos enunciados:

---

<sup>12</sup> Lo cual me parece que va en la línea seguida por Diego DEI VECCHI en muchos de sus trabajos. Véase por ejemplo DEI VECCHI, 2014.

<sup>13</sup> Como se ha dicho previamente, las hipótesis o escenarios a descartar podrían ser los cercanos, los cercanos y los más distantes, o bien, los cercanos, los más distantes y los remotos.

<sup>14</sup> Ventajas que se aprecian mejor si contrastamos dichos estándares posibles con las frases extremadamente vagas anteriormente referidas que, en última instancia, parecen invitar al juzgador a adoptar la postura teórica de la que nos advierte PICINALI, que consiste meramente en detectar si se experimenta o no la creencia firme en la proposición de que se trate.

<sup>15</sup> Sigo de cerca en este punto la reconstrucción que TOH hace, y de la que da cuenta RAPETTI (2017: 44 y ss.), de los que HART llamó «enunciados jurídicos internos» del tipo «El derecho dispone que...», que difieren de los enunciados del tipo «En México reconocen como derecho que...», denominados «enunciados jurídicos externos». Dicha reconstrucción, como explica RAPETTI, es análoga a la que hacen de los enunciados internos del discurso moral algunos autores defensores del expresivismo en

1) Expresan la aceptación por parte del juzgador, del argumento que considera que razonablemente incorpora el grado de cautela epistémica moralmente apropiado para el contexto de que se trate, es decir, para la clase de proposiciones en cuestión (grado de cautela cuya elucidación y aplicación forma parte de la responsabilidad moral del juzgador, misma que no desaparece o que no puede suplantarse ni siquiera cuando se implementan reglas sobre estándares de prueba, pues dichas reglas solo lo orientan a este respecto)<sup>16</sup>;

2) Y, a su vez, los enunciados probatorios acarrear:

2.1) El compromiso del juzgador de hacer explícito dicho argumento si se le es requerido (compromiso que se vuelve más latente cuando el ordenamiento jurídico en cuestión contempla el deber de motivación de las decisiones judiciales);

2.2) Su exhortación al auditorio relevante (principalmente compuesto por las partes y por el órgano superior jerárquico que eventualmente podría conocer de la cuestión) a que se acepte de forma compartida dicho argumento, y;

2.3) Su disposición a ser racionalmente persuadido de lo contrario (o sea, de que su argumento no incorpora el grado de cautela apropiado).

Con respecto a su sentido, los enunciados probatorios significarían que, mediante un argumento construido sobre la base de la información probatoria disponible, que el juzgador considera que razonablemente incorpora el grado de cautela epistémica contextualmente apropiado, dicho juzgador ha establecido preliminarmente que “p”. El carácter preliminar (o prima facie) del establecimiento de que “p” está ligado a la disposición del juzgador a ser persuadido de que el argumento en cuestión no incorpora el grado apropiado de cautela epistémica, disposición dilucidada en el análisis de la fuerza de los enunciados probatorios.

## REFERENCIAS

ACCATINO, D., 2019: «Teoría de la prueba: ¿somos todos “racionalistas” ahora?», *Revus: Journal for Constitutional Theory & Philosophy of Law*, vol. 39, pp. 85-102.

---

el campo de la metaética contemporánea, quienes, como GIBBARD, ofrecen un análisis oblicuo de tales enunciados en el sentido de centrarse en lo que los hablantes hacen al formularlos, decantándose por la opción de sostener que lo que hacen es precisamente *expresar su aceptación de ciertas normas*.

<sup>16</sup> La aceptación en juego sería una actitud práctica, conativa, no cognitiva. En este sentido, lo que un enunciado probatorio expresa es tanto la aprobación por parte del juzgador del argumento en cuestión, como su recomendación o exhortación (como inmediatamente se verá) a que dicho argumento sea también aprobado por otro(s). Claro está que hay una creencia involucrada, es decir, un estado cognitivo; la creencia del juzgador de que su argumento razonablemente incorpora el grado de cautela apropiado. Se podría decir que esta creencia sería el fundamento o la razón de su aceptación, no obstante, en el análisis que se está ofreciendo, el foco es puesto en la función de los enunciados probatorios, primordialmente expresiva de esa actitud práctica.

- AXTELL, G., 2011: «From internalist evidentialism to virtue responsibilism», en Dougherty, T. (coord.), *Evidentialism and its discontents*, Oxford University Press.
- BAHR, J., 2011a: «Evidentialism, vice, and virtue», en Dougherty, T. (coord.), *Evidentialism and its discontents*, Oxford University Press.
- , 2011b: *The inquiring mind. On intellectual virtues and virtue epistemology*, Oxford University Press.
- CONNOR, E. y FELDMAN, R., 2004: *Evidentialism: Essay in epistemology*, Oxford University Press.
- DEI VECCHI, D., 2014: «Acerca de la fuerza de los enunciados probatorios», *DOXA, Cuadernos de Filosofía del Derecho*, núm. 37, pp. 237-261.
- DOUGHERTY, T. (coord.), 2011: *Evidentialism and its discontents*, Oxford University Press.
- FERRER, J., 2021: *Prueba sin convicción: Estándares de prueba y debido proceso*, Madrid, Marcial Pons.
- GARDINER, G., 2019: «The reasonable and the relevant: Legal standards of proof», *Philosophy and Public Affairs* 47, núm. 3, pp. 288-318.
- LAWSON, A. y DANIEL, E., 2011: «Inferences of clinical diagnostic reasoning and diagnostic error», *Journal of Biomedical Informatics*, núm. 44, pp. 402-412.
- MATHESON, J., 2021: «Robust justification», en McCain, K. y Stapleford, J. (coords.), *Epistemic duties, new arguments, new angles*, Routledge, pp. 146-160.
- NANCE, D., 2016: *The burdens of proof: Discriminatory power, weight of evidence, and tenacity of belief*, Cambridge University Press.
- PEELS, R., 2017: *Responsible belief: A theory in ethics and epistemology*, Oxford University Press.
- PICINALI, F., 2015: «The threshold lies in the method: instructing jurors about reasoning beyond reasonable doubt», *International Journal of Evidence and Proof*, vol. 19, núm. 3, pp. 139-153.
- RAPETTI, P., 2017: «Expresivismo metajurídico, enunciados internos y aceptación plural: Una exploración crítica», *Isonomía*, núm. 47, pp. 39-80.
- RYAN, S., 2021: «Wisdom, open-mindedness, and epistemic duty», en McCain, K. y Stapleford, J. (coords.), *Epistemic duties, new arguments, new angles*, Routledge, pp. 174-190.
- STAPLEFORD, S. y MCCAIN, K., 2021: «Bound by the evidence», en McCain, K. y Stapleford, J. (coords.), *Epistemic duties, new arguments, new angles*, Routledge, pp. 113-124.

## EVIDENCE, STANDARDS AND BURDENS OF PROOF IN THE EUROPEAN ASYLUM SYSTEM: A NEW FIELD FOR EVIDENCE SCHOLARS?

Elena Marchese

Research fellow at Bocconi University  
elena.marchese@unibocconi.it

**ABSTRACT:** This article presents the main evidential features implied in the European Union directives and case law regarding the recognition of refugee status. It focuses, first, on the challenges posed to judicial decision making by the absence of a unitarian European proof system, by the significant lack of evidence in the relevant judgments, and by the conflicting interests at stake. The text highlights, then, the «subjective» connotation of international protection evidence and inserts it into the debate on the standards of proof which characterizes the reflections of some legal philosophers.

**KEYWORDS:** evidence; asylum; standard of proof; CEAS; UNHCR.

**SUMMARY:** 1. INTRODUCTION.— 2. THE NOTION OF EVIDENCE.— 3. THE *THEMA PROBANDUM*: 3.1. Evaluation of evidence. 3.2. Burdens of proof. 3.3. The standards of proof. 3.4. On the objectivity. 3.5. Standards as tools of evidential policy.— 4. THE CREDIBILITY OF THE APPLICANT AND THE RELIABILITY OF THE APPLICATION: 4.1. Credibility indicators. 4.2. Credibility: special cases.— 5. CONCLUSIONS.— BIBLIOGRAPHY

**RECOMMENDED CITATION:** ELENA MARCHESE, 2022: «Evidence, Standards and Burdens of Proof in the European Asylum System: a New Field for Evidence Scholars?», in *Quaestio facti*, 3: 103-129. Madrid: Marcial Pons Ediciones Jurídicas y Sociales. DOI: [http://dx.doi.org/10.33115/udg\\_bib/qf.i3.22737](http://dx.doi.org/10.33115/udg_bib/qf.i3.22737)

## 1. INTRODUCTION

One of the contexts to which legal epistemologist and evidence experts seem to have still paid little attention is that concerning the study of the evidential activities involved in the procedure for the attribution of refugee *status*.

There are four orders of reasons that make the study of the probative issues in this area particularly stimulating and make this assessment remarkably complex: 1) the plurality of the regulatory levels involved (European legislation and national systems; administrative and judicial jurisdiction)<sup>1</sup>; 2) the peculiar condition of evidence scarcity on the basis of which judicial bodies have to decide in this sector, 3) the subjective characterization of evidence and probative instruments; 4) the humanitarian and political reasons that inform such practices.

Given the purposes of this work, I will be able to concentrate only on the third of these reasons, leaving the others to subsequent works. First, I will attempt to provide a synthetic picture of European secondary law and literature produced by the European Asylum Support Office (EASO) and the United Nations High Commissioner for Refugees (UNHCR) and, subsequently, I will focus on one of the debates that, in recent years, has gained great fortune among legal philosophers and which concerns the preferability, possibility and necessity to conceive the evidential activity in more objective terms in order to allow a better *ex post* control of decisions. The field seems relevant not only for the importance of the epistemological, political and legal challenges posed by the context of analysis to the European Union and to the evidence scholars, but also because it can offer new tools for the reflection of the legal philosophers engaged in the growing disagreement regarding which conception of evidence and proof is more suitable for the legal domain, and regarding the formulation of standards of proof.

## 2. THE NOTION OF EVIDENCE

There is no general or unambiguous definition of *evidence* in EU law. Article 4 of the 2011 Qualification Directive, entitled Assessment of Facts and Circumstances,

---

<sup>1</sup> In Italy, for example, there is a two-level system: the first is of an administrative nature and the second of a judicial nature. In the first level, the examination of the applications is carried out by the so-called Territorial Commissions. They are made up of a college of four members (the president of the Commission, a commissioner designated by UNHCR and two officials from the Ministry of the Interior one of whom is the person conducting the interview). The second level, on the other hand, has a judicial nature and is the remedy granted by our legal system against the decisions of the Territorial Commission. The appeal can be proposed before the specialized Section of the Court of Appeal. Depending on the case, the appeal must be lodged within thirty or fifteen days of the Commission's decision. The procedure is handled by the college in the council chamber.

however, deals with the «necessary elements» for the assessment of an international protection application<sup>2</sup> which, according to §2, consist of «declarations [...] and in all the documentation in the possession of the applicant». Thus, it would seem that the term *elements* includes already proven facts and<sup>3</sup> those facts useful for proving a given hypothesis.

A systematic reading of the Qualification Directive and European Court of Justice (henceforth CJEU) case law, however, suggests a broader notion of *evidence* capable of including more than the simple statements and documentation provided by the applicant. In *A, B and C v. Staatssecretaris* case, in fact, the CJEU (§54) speaks of «assessing statements and documentary or *other evidence*»,<sup>4</sup> that is, of a broader notion, consistent with the international sources solid orientation, which consider evidence «whatever asserts, confirms, supports, refutes or in any case refers to the relevant facts in question».<sup>5</sup> The jurisprudence of the CJEU has very rarely dealt with evidential issues, so these issues remain largely unexplored fields for the legal theorist.

### 3. THE *THEMA PROBANDUM*

The assessment concerning the determination of refugee *status* has its «key element» in establishing the «well-founded fear of being persecuted» for reasons relating<sup>6</sup> to «race», «religion», «nationality», «membership of a social group» or «political opinion».

In this sentence, present in art. 1(A) of the Refugee Convention, there are two fundamental elements: one objective (the substantiation) and one subjective (the «fear») and both must be proven through a joint assessment, functional to establish the general reliability (the fear), and both of the application and to grant, therefore, protection.<sup>7</sup> This definition replaces and expands the previous approach that defined refugees only on the basis of the fact that they do not enjoy the protection of their own country and therefore through a purely external and objective criterion<sup>8</sup> that was considered insufficient and inadequate for the particular condition of the applicants. This new definition, on the other hand, makes the subjective element a constitutive criterion of refugee *status* to be placed side by side, however, always with the objective one.<sup>9</sup>

<sup>2</sup> Directive 2011/95/EU: Art. 4, §1; see EASO 2018, 17.

<sup>3</sup> See EASO 2018, 18

<sup>4</sup> CJEU, judgment of 2 December 2014, joined cases C-148/13 to C-150/13, *A, B and C v. Staatssecretaris van Veiligheid en Justitie*, EU: C: 2014: 2406. Italics mine.

<sup>5</sup> UNHCR 2013, 28; see EASO 2018, 18

<sup>6</sup> UNHCR 2019, p. 19, §37.

<sup>7</sup> See UNHCR 1998, p. 3, §13.

<sup>8</sup> See UNHCR 1998, p. 4, §15.

<sup>9</sup> UNHCR 2019, p. 19, §37.

In practice, however, it seems that the justifying function of the objective dimension in this definition is inhibited or at least strongly hindered by international guidelines and practices which, as we shall see, favour a *discovery* and probative activity mostly in subjective terms, i.e. referring mainly to judge's mental states and to the applicant's ones.

In fact, UNHCR itself seems to have progressively changed its orientation in 2019, when it highlighted that:

Since fear is subjective, the definition [of refugee] involves [the recognition of] a subjective element in the subject requesting recognition as a refugee. The determination of refugee *status* will therefore mainly require an assessment of the applicant's statements rather than a judgment on the situation prevailing in his country of origin.<sup>10</sup>

In this passage, UNHCR seems to subtract probative value from the «objective» evidence in favour of the «subjective» ones in a sort of *favour* towards the applicant.

For example, «fear» is interpreted by UNHCR as a «state of mind» and a «subjective condition»<sup>11</sup> which is expressed through the applicant's «belief» or «prediction» that he is «subject to persecution». It seems interesting to note that this aspect is generally established on the basis of *what the person manifests as his state of mind at the time of departure* from his country and that, according to the UNHCR, such statements should be *considered in themselves as proof of the existence of fear* as long as there are no elements giving rise to serious doubts with respect to the credibility of the reporting subject.

However, this orientation seems to be in contrast with contemporary EU policies which, on the other hand, tend to increasingly restrict the mesh of reception, in contrast to the massive migratory phenomena of the last decades.<sup>12</sup>

### 3.1. Evaluation of evidence

The 1998 *Note on Burden and Standard of Proof* is the first UNHCR document addressing evidential issues related to the refugee *status* determination and aims to present a general overview of the nature and peculiarities of the application process of international protection.

The document highlights that the heart of the evaluation I am dealing with concerns the fear «foundress» and that, although it is in part an evaluation of an «intrinsically speculative» nature, it is not, by that very fact, «pure conjecture».<sup>13</sup> The text underlines how, because of this, in this area there will be no «rigorous juridical inferences» and the decision will be based mainly on the «probability» or «possibility»

<sup>10</sup> UNHCR 2019, p. 19, §37.

<sup>11</sup> UNHCR 2019, p. 19, §38.

<sup>12</sup> Commission, 23 September 2020, COM (2020) 611 final.

<sup>13</sup> UNHCR 1998, p. 5, §21; UNHCR 1998, p. 4, §18.

that the feared event will happen.<sup>14</sup> The arguments involved will occupy an intermediate place between rigorous inferences and mere conjectures, and will have to be supported by valid reasons.<sup>15</sup>

Thus, recognizing that the inferences that will characterize this kind of evaluation will be predominantly of a non-deductive type,<sup>16</sup> the *Note* also states that *the* discrimination with respect to their goodness or usability is the solidity of the justifications placed at their basis and not their certainty or conclusiveness. The document underlines, however, that these reflections will in any case be balanced with the fact that, unlike what happens in other types of proceedings (for example, in civil or criminal ones), the assessment involved in this area has eminently a humanitarian aim and that this element profoundly shapes the evidentiary activities employed in this area. In fact, they are mainly aimed at the protection of the subjects and not at the search for certainty or truth.<sup>17</sup>

According to the CJEU, the process of evaluating applications for international protection has «two distinct phases»: the first regards the «ascertainment of the factual circumstances that may constitute evidence in support of the application» while the second «the legal evaluation of such elements, which consists in deciding whether, in the light of the facts characterizing a particular case, the requirements [...] for the recognition of international protection are met».<sup>18</sup>

The first is of a purely epistemic nature and will concern the collection, examination of the genuineness and information value (and therefore of the relevance) of ascertained facts<sup>19</sup>, while the second has a normative-evaluative nature and consists in making a choice on the legal «legitimacy» of the taken evidence and on the overcoming of the required standard of proof for the assignment of refugee *status*. The two phases are distinguished for reasons of conceptual clarity but, obviously, they manifest themselves in a highly interrelated manner.

The jurisprudence of the CJEU also seems to point to this line, stressing that the procedure for assessing applications should never be reduced to a purely mechanistic activity. According to the Court, in fact, the determining authority will have to «modify its methods of evaluating declarations and other evidence [...] taking into account the specific characteristics of each category of asylum request, in compliance

<sup>14</sup> UNHCR 1998, p. 5, §21.

<sup>15</sup> UNHCR 1998, p. 5, §21.

<sup>16</sup> Specifically abductive and inductive: TUZET, 2006; TUZET, 2016; CANALE & TUZET, 2020.

<sup>17</sup> UNHCR 1998, p. 1, §2: «In examining refugee claims, the particular situation of asylum-seekers should be kept in mind and consideration given to the fact that the ultimate objective of refugee status determination is humanitarian. On this basis, the determination of refugee status does not purport to identify refugees as a matter of certainty, but as a matter of likelihood».

<sup>18</sup> CJEU, Judgment of 22 November 2012, case C-277/11, *MM v. Minister for Justice, Equality and Law Reform, Ireland*, Attorney General, §64 et seq.

<sup>19</sup> See EASO 2018, p.19 which also includes in the first phase an evaluation of the legitimacy of the evidence collected with respect to material facts.

with the rights guaranteed by Charter [of fundamental rights]»<sup>20</sup> required by the examination of individual circumstances required by art. 4 §3 of the Qualification Directive 2011/95/EU. This disposition provides for the assessment of:

- a) All relevant facts concerning the country of origin at the time of the decision on the application, including the laws and regulations of the country of origin and the relative methods of application;
- b) the relevant declarations and documentation submitted by the applicant who must also disclose whether he has already suffered or is likely to suffer persecution or serious harm;
- c) the individual situation and personal circumstances of the applicant, in particular background, sex and age, in order to assess whether, based on the applicant's personal circumstances, the acts to which he has been or could be exposed constitute persecution or serious harm;
- d) whether the activities carried out by the applicant after leaving the country of origin have aimed exclusively or mainly at creating the conditions necessary for the presentation of an application for international protection, in order to establish whether these activities expose the applicant to persecution or to serious damage in case of return to the country;
- e) the possibility that the applicant can reasonably be expected to seek the protection of another country of which he could declare himself a citizen.

### 3.2. Burdens of proof

EU secondary law does not prescribe a particular procedural form for ascertaining refugee *status* and it is difficult to rigidly frame the nature of this procedure in one of the classic procedural categories (administrative or judicial; inquisitorial or *adversary*).<sup>21</sup> The main characteristics will then be outlined, starting with the distribution of the probative burdens.

The duty to prove the facts on which the request is based rests primarily with the applicant but in a way that can be defined as «attenuated» compared to what happens in the Italian civil process where the *principio dispositivo* is in force (art. 2697 cc and 115 co.1 c.p.c.) according to which parties —and in particular those who assert a right— must indicate the necessary evidence to support the facts attached to the court. Here, indeed, the decision maker has powers of inquiry only in specific cases and is in force the principle of correspondence between requested and pronounced (pursuant to art. 112 c.p.c.) according to which the judge will have to «pronounce on the entire request and not beyond its limits», not being able to decide on aspects not proposed by the parties. In the criminal field, on the contrary, the burden falls entirely on the accusation in compliance with the principle of presumption of innocence (art. 27 co. 2 of the Constitution) which, in principle, would allow the accused to even remain inert.

<sup>20</sup> CJEU, Judgment of 2 December 2014, Cases joined by C - 148/13 to C - 150/13, *A, B and C v. Staatssecretaris van Veiligheid en Justitie*, §54. See also: EASO 2018, p. 19.

<sup>21</sup> See also: EASO 2018a, p. 29. UNHCR 1998, p. 1, §2.

Thus, in the context of a procedure for the recognition of international protection it is generally the applicant who has the burden of demonstrating his right to benefit from the protection, but given the difficulties (or impossibility) that usually affect the applicant's probative activity, this obligation can be shared with the ascertaining authority. Thus, generally, while the burden of proof rests with the applicant, the duty to establish the relevant facts is shared between the applicant and the person in charge of examining the application in a sort of «ascertainment cooperation» aimed at assessing the relevant facts.<sup>22</sup>

Therefore, the examiner therefore has the right to use any means at his disposal to produce, through an *ex officio* investigation, the evidence necessary to support the application. This task has special practical relevance given that, as previously mentioned, in these procedures it is very common that the *evidence* is particularly incomplete or completely absent.

Usually, in fact, the investigation activity of the proceeding authority has greater possibilities and resources to reach information that is more difficult to find by the applicant.

Those that, in general, may present the greatest difficulties in finding the applicant—but which at the same time are among the most relevant for the purposes of the assessment—concern the situation in his country of origin<sup>23</sup> and pertain to the «objective» sphere of evidence of fear.

The judge will therefore have both a duty to ascertain and evaluate the general plausibility of the request (aimed at assessing the foundness of the fear), and the duty to assess the general legitimacy of the request with respect to the regulations in force.

The assessment of the credibility of the applicant's statements is part of the assessment of the plausibility of the application, which has recently become a central point in the evidentiary debate on asylum issues. As I will show shortly with «credibility», in this context, we should not mean «sincerity» but «overall reliability» of the statements in support of an application.<sup>24</sup>

This division of the burden of proof can be justified, however, not only by virtue of the humanitarian purpose but also in recognition of the probative asymmetry that, indeed, exists between the applicant and the authorities in the ability to present

<sup>22</sup> UNHCR 1998, p. 2: In the process of evaluating applications for international protection, the judge is called to *cooperate* with the applicant «in the duty to ascertain and evaluate all the relevant facts». UNHCR 2013 p. 35, 87-88; 104 et seq.; EASO 2018, pp. 28 and ff.

<sup>23</sup> The aforementioned COI (Country Origin Information).

<sup>24</sup> EASO 2018 p. 81, UNHCR 2013, p. 49, UNHCR 2019, p. 44, MINNITI, 2020, §2.2. and ff. but also: ECtHR: *JK and others / Sweden*, question 59166/12, note 20; judgment of 10 September 2015, *RH v Sweden*, application n. 4601/14, paragraph 58; judgment of 20 July 2010, *N / Sweden*, application no. 23505/09, paragraph 53; judgment of 9 March 2010, *RC v Sweden*, application no. 41827/07, paragraph 50. See art. 4 of the 2011 Qualification Directive at §1.

evidence. Thus, this distribution rule is also understandable in the light of a criterion of «proximity to the source of evidence».<sup>25</sup>

In recent decades, this rule on the distribution of the burdens of proof among the parties is also significantly orienting Italian jurisprudence in those «areas where the protection of [...] protected rights [...] finds an obstacle in the profound inequality/ asymmetry, even —but not only— information, constituting a serious obstacle to the demonstration and therefore to the protection of the right».<sup>26</sup> The Cassazione Civile, Section VI, sent. n. 17923 of 12 September 2016, denying, in the present case, the possibility of applying the principle of proximity of evidence, however, specified that it cannot simply follow a disparity of «strength» of the parties in the process, but requires that the necessary «proof» is not symmetrically acquirable by one of the parties, given an impossibility or strong disparity of acquisition and access to it. That is, very frequently, what happens to applicants for international protection, too.

### 3.3. The standards of proof

Given that the activity of ascertaining and judging international protection applications often takes place on very limited epistemic elements, one of the crucial issues become the establishment of a probability threshold that is adequate for the attribution of refugee *status*, as not every risk of persecution or damage can give rise to international protection and because there are many and relevant opposing interests that weigh on this evaluation.<sup>27</sup>

Already in 1998, UNHCR promoted the use of standards of proof, emphasizing their importance especially in relation to those *civil law* systems where, historically, doctrine and jurisprudence had preferred to focus on burdens of proof and intimate judicial conviction.<sup>28</sup> Even today, however, the regulatory instruments of the CEAS do not provide any information on the degree of evidential sufficiency to be achieved in order to consider an application as generally plausible or reliable (and therefore acceptable).<sup>29</sup>

<sup>25</sup> MINNITI, 2020, §4.1.

<sup>26</sup> MINNITI, 2020, §4.1. See Cass. Civ., N. 24051 of 26 September 2019 and SS.UU. n. 13533 of 30/10/2001.

<sup>27</sup> There are conflicts concerning the relationship between States and the EU (those relating mainly to the quantity of hospitality to be granted) but also conflicts between national States (regarding the distribution of reception loads), then there are conflicts between subjective interests and humanitarian interests of the individual and economic interests of the States.

<sup>28</sup> UNHCR 1998, p. 2, §III. FERRER & TUZET, 2018, p. 458, also attribute to the discipline of burdens of proof the function, typical of the standards, of distributing the risk of error. On the standards of proof see also: LAUDAN, 2006.

<sup>29</sup> EASO 2018b, p. 87 and EASO 2018a, p. 82.

By the term *standard of proof*, however, UNHCR means the threshold «which must be reached by the applicant in *persuading* the judge of the truth of his allegations». <sup>30</sup> The dimension of «conviction» is then further specified by UNHCR and EASO <sup>31</sup> through an analysis that national Member States jurisprudence differentiated on the basis of the legal system type (*common law* or *civil law*).

In general, it is noted that, in both cases, the standard commonly used in terms of international protection is closer to a civil standard of proof (i.e. a «probability balance») rather than a criminal one, since, for the granting of international protection, a conviction «beyond a reasonable doubt» or a state of certainty of the decision maker is not necessary. <sup>32</sup>

Indeed, according to these sources, in fact, in the international protection field, the standard should only serve to place the decision maker in a position to affirm that, given the evidence provided, «it is probable that the applicant's request is *credible*». <sup>33</sup> Thus, the fear will be considered «well founded» if the applicant can «demonstrate, to a *reasonable extent*», that the stay in his own country has become *intolerable*. <sup>34</sup> As it is easy to notice, the terms *intolerable*, *fear*, *persuasion*, *belief* and *conviction* are all related to emotional or mental states which, from a purely epistemological point of view, clearly poses a series of relevant questions about the difficulty of their proof and justifiability in decision making.

By analysing more specifically the standards of proof used in the individual European *civil law* systems, EASO highlights the principle —developed by German jurisprudence— according to which in cases where a certain degree of probability is already present in substantive law (as it would be in the case of international protection with the terms *well-founded fear*, *effective risk of serious damage*), the deciding authority will have to operate taking into account this degree of probability as part of the same rules that bind its decision-making activity. <sup>35</sup> For this reason, both for the attribution of refugee *status* and for the granting of subsidiary protection, <sup>36</sup> the validity of the fear must be established by ascertaining the presence of a «substantial

<sup>30</sup> UNHCR 1998, p. 2, §III. Italics mine. Concordantly, EASO 2018a p. 83, states: «judge must be *convinced* that the applicant actually meets the criteria for the recognition of refugee *status or for being entitled to benefit from subsidiary protection*». Italics mine.

<sup>31</sup> UNHCR 2013; EASO 2018a.

<sup>32</sup> UNHCR 1998, p. 2, §III.

<sup>33</sup> UNHCR 1998, p. 2, §8.

<sup>34</sup> UNHCR 2019, p. 20. Italics mine.

<sup>35</sup> This principle is interesting as it seems to have the purpose of justifying the presence of imprecise formulations.

<sup>36</sup> EASO 2018a, p. 84. See Federal Administrative Court (Germany), judgment of 10 June 2011, BVerwG 10 C 25.10, BVerwG: 2011: 010611U10C25.10.0, §22, available in English on the website: [www.bverwg.de](http://www.bverwg.de).

probability»: a probability that is not merely theoretical, but not necessarily higher than 50%.<sup>37</sup>

As for the European *common law* systems, on the other hand, EASO analyzes the standard used by the Irish *High Court* and the British House of Lords. The former recognizes as adequate the balance of probabilities combined, in cases where it is necessary, with the benefit of the doubt,<sup>38</sup> while the latter considers it sufficient to ascertain a level of «reasonable possibility». <sup>39</sup> This level is described as «a real possibility», «well-founded reasons for believing» or a «reasonable possibility» of suffering persecution, a level which in any case may remain lower than that of the simple balance of probabilities.<sup>40</sup>

Similarly, the ECtHR in the *Saadi v. Italy*, argued that the applicant is not obliged to «[demonstrate] that submission to ill-treatment is “more likely than unlikely”»<sup>41</sup> while in *Y and Z*,<sup>42</sup> the CJEU clarified that when the competent authorities assess the foundness of the fear, they «are required to ascertain whether or not the circumstances established constitute such a threat that the person concerned may reasonably fear, in the light of his individual situation, that he will in fact be subject to acts of persecution».<sup>43</sup>

The CJEU then specified: in the *Abdulla* judgment that, in the field of asylum, the «assessment of the extent of the risk must, in all cases, be carried out with vigilance and care, since what are at issue are issues relating to the integrity of the person and

<sup>37</sup> EASO 2018, p. 84. This guidance is developed by the Federal Administrative Court (Germany), judgment of 7 February 2008, BVerwG 10 C 33.07, BVerwG: 2008: 070208B10C33.07.0, §37, available in English at: [www.bverwg.de](http://www.bverwg.de). See also Supreme Court (Spain), judgment of 9 December 2015, question n. 1699/2015, ECLI: ES: TS: 2015: 5211, p. 9.

<sup>38</sup> High Court (Ireland), Judgment of 17 January 2017, *ON v. Refugee Appeals Tribunal et al*, 2017, JEHC 13, §63: Cf. EASO 2018a, p. 84. The principle of the «benefit of the doubt» will be dealt with in §3.6.

<sup>39</sup> This way also UNHCR 2019, p. 20, §42.

<sup>40</sup> House of Lords (United Kingdom), judgment of 16 December 1987, *R v. Secretary of State for the Home Department*, Ex parte Sivakumaran, 1988, AC 958.EASO 2018, p. 84.

<sup>41</sup> ECtHR, judgment of 28 February 2008, Grand Chamber, *Saadi v. Italy*, appeal n. 37201/06, §140.

<sup>42</sup> CJEU, judgment of 5 September 2012, Grand Chamber, joined cases C-71/11 and C-99/11, *Bundesrepublik Deutschland v. Y and Z*, EU: C: 2012: 518 §76. In that judgment, the CJEU, at §80, takes up this formulation and states that: «the applicant’s fear of being persecuted is well founded when the competent authorities [...] consider it reasonable to believe that [...] he will carry out acts [...] which will expose him to a real risk of persecution». EASO 2018b, p. 87 and EASO 2018a, p. 82.

<sup>43</sup> CJEU, judgment of 5 September 2012, Grand Chamber, joined cases C-71/11 and C-99/11, *Bundesrepublik Deutschland v. Y and Z*, EU: C: 2012: 518. See also CJEU, judgment of 2 March 2010, Grand Chamber, joined cases C-175/08, C-176/08, C-178/08 and C-179/08, *Aydin Salahadin Abdulla and others v. Bundesrepublik Deutschland*, EU: C: 2010: 105, §89; CJEU, judgment of 7 November 2013, joined cases from C-199/12 to C-201/12, *X, Y, and Z v. Minister voor Immigratie en Asiel*, EU: C: 2013: 720, §72.

to individual liberties, issues which relate to the fundamental values of the Union». <sup>44</sup> The jurisprudence of the ECtHR refers to careful scrutiny and rigorous evaluation. <sup>45</sup>

It can be noted that all the standards just mentioned are united by at least two characteristics: 1) they present terms such as *reasonableness* or *probability* (words that seem to provide with objective criteria); 2), they also refer to the judge's or applicant (persuasion, belief, conviction, etc.).

These points allow us to make some reflections on the margins of a debate which, starting precisely from specific criticisms regarding the use and formulation of standards of proof, has also been strengthened on elements of a more general nature, which concern the correct shape of a proof conception. <sup>46</sup>

The just mentioned points 1) and 2) show profiles of theoretical interest as they highlight a very close link between the applicant's mental states (fear, intolerability), probative standards of proof and mental states of the decision maker (beliefs, persuasion, conviction). This aspect could give arguments to those who, in literature, believe that Jordi Ferrer's standards of proof aren't able to account for the reality of the judicial decision and legal evidentiary reasoning and are, therefore, inadequate just as his rational conception of legal evidence and proof. One of the problems would also be the lack of harmonization between the subjective and objective spheres of reasoning on evidence and judicial decision. <sup>47</sup>

Furthermore, according to Gama, the works of those who adhere to the rationalist approach to legal evidence and proof would discount the distortions of an excessively analytical methodology which, by atomistically splitting the phase of gathering, evaluating and deciding on evidence, would offer a misleading account of the procedural practices <sup>48</sup> together with giving a too essential (or wrong) reading of the historically developed concept of proof. <sup>49</sup>

<sup>44</sup> CJEU, judgment of 2 March 2010, Grand Chamber, joined cases C-175/08, C-176/08, C-178/08 and C-179/08, *Aydin Salahadin Abdulla and others v. Bundesrepublik Deutschland*, EU: C: 2010: 105, paragraph 90.

<sup>45</sup> ECtHR, judgment of 12 April 2005, *Shamayev and others v. Georgia and Russia*, question no. 36378/02, paragraph 448 and ECtHR, JK and others v. Sweden, cit., Footnote 20, paragraph 77. ECtHR, judgment of 17 July 2008, NA v. United Kingdom, application no. 25904/07, paragraph 111. See EASO 2018a, p. 71.

<sup>46</sup> The term obviously recalls the work of Jordi FERRER BELTRÁN (see for example: 2005, 2007, 2019, 2021) who, over the years, has developed in a systematic and original way many of the insights present in the works of FERRAJOLI, 1989; GASCÓN, 1999 and TARUFFO (among many: 1979, 1982, 2009) on proof and evidential systems. His work on the «rational» way of conceiving proof and evidence system has inspired the reflections of a large number of authors such as Perfecto Andrés IBÁÑEZ, GONZALEZ LAGIER, BAYON, TUZET, DEI VECCHI. It is beyond the scope of this paper to present a detailed reconstruction of how attention to evidential issues was born and established itself in recent years. I will cite among many the reconstructions made by RICAURTE, 2019 and ACCATINO, 2019.

<sup>47</sup> GAMA, 2021, pp. 13 and 17.

<sup>48</sup> See TWINING, 2020; GAMA 2020 16-17.

<sup>49</sup> GAMA, 2020, pp. 7 and ff.

Points 1) and 2), however, may also give rise to considerations in the opposite sense. The standards present in this area, although cloaked in terminology with a neutral and objective reference (reasonableness, probability), would however be resolved in a merely subjective decision criterion (belief, conviction, certainty) since there is no specific literature that exemplify or interpret these terms and they are not, in fact, subject to specific control.

Nevertheless, the presence of these terms suggests that, in such a complex area of assessment there is a certain urgent need to stem and justify the simple judicial conviction. From this point of view then (and in the current state of affairs), such words would seem to operate as simple warnings or generic references to the prudence of the judge.

A further consideration pertains to the weight that these formulations have on the effectiveness of the «objective» dimension of the proof required by the refugee definition. Do these standards really allow for compliance with this justification requirement? What kind of rational elements can an applicant appeal to against the denial of protection when, as is very frequent in the immediacy of his arrival and in situations of severe discomfort, he is unable to convey the right emotional state to exceed the standard?<sup>50</sup> Many are the reasons why providing the validity of the fear for the applicant can be very difficult, including evidential asymmetry, cultural, linguistic, emotional, psychological reasons.<sup>51</sup> These aspects also significantly affect the credibility assessment.

It is therefore reasonable to ask whether 1) the formulation of these standards is adequate (as Ferrer's critics would seem to suggest) or 2) there is a need to reformulate these decision rules. I will try to answer these questions in the following section.

In the meantime, however, it seems relevant to consider how the national jurisprudential results return a bad image in terms of transparency, access to protection and worsening of conflicts of interest in relations between European Member States and between them and the Union. Moreover, this situation seems to have been worsened by the variegated mosaic of national evidential systems which does not seem to guarantee equal access to the same level of rights for all applicants in the various European States.

For this reason, I believe that an in-depth study of the evidentiary issues in this area would be useful and fruitful both at the judicial level, but already in the ad-

---

<sup>50</sup> It is very frequent that applicants fail to communicate in a resilient manner in the first contact with authorities. In this crucial moment, in fact, applicants are requested to provide all the elements necessary for the judgment in a very short time and paper space. There are also a series of other dynamics which affect the applicant's ability to express his fear and which usually generate in the subjects a closed and hostile attitude towards the authorities: v. EASO 2018a, pp. 176 and following; BLACK, 1994.

<sup>51</sup> See UNHCR 2013, pp. 53-75; pp. EASO 2018a, pp. 33 and ff. but also: HERLIHY & TURNER, 2009.

ministrative one (Commissione Territoriale) where problems of transparency and justification arise with negative effects of even greater vehemence.

### 3.4. On the objectivity

As already said, Ferrer's version of the rationalist approach to legal evidence and proof is the one that in recent years has obtained the greatest consensus in Europe and Latin America. I believe that *mutatis mutandis*, most of his reflections are, by reason of their generality, widely applicable also to the context that belongs to us.

Ferrer's position, however, has recently been criticized in a more systematic way both with reference to some of his basic theoretical assumptions and in relation to the proposed construction of «precise and objective» standards, that is, less vague and free of references to psychological states.<sup>52</sup> I will concentrate especially on examining the first of these two criticisms (that of assumptions and objectivity) leaving those concerning the question of the «precision» of standards to another work.

As for the theoretical assumptions, the critique of the characterization of the rational conception proposed by Ferrer in opposition to the so-called «persuasive conception» of proof stands out due to its importance.<sup>53</sup> It has rightly been pointed out that persuasion and judicial conviction, in contemporary legal systems, are in fact always linked also to elements of an epistemic nature and that, therefore, it would be wrong to equate a system of evidentiary activities and instruments that aim to produce judicial conviction to a totally irrational system.<sup>54</sup> In my opinion, this criticism presents two problems.

The first is that these theses seem to presuppose the opposition to the «persuasive conception» as an identity trait of the rational conception of proof while I believe that a charitable reading of Ferrer's texts does not allow this conclusion or even the identification between «persuasive» and «irrational» *tout court* by the author. Ferrer's words cannot be interpreted correctly without referring them to the central purpose of his conception (which, instead, is a strongly identifying element), namely that of expanding the possibilities of an *ex post* control over decisions.

The connection between persuasive and irrational used by Ferrer is therefore evidently to be understood in a consequentialist sense, as having to do with the demo-

---

<sup>52</sup> DEI VECCHI, 2016; GAMA, 2020. See also FERRER, 2019, p. 257.

<sup>53</sup> DEI VECCHI, 2016; SUCAR & CERDIO, 2017; GAMA, 2020; relating to the conceptualization of intimate conviction: TUZET 2020 pp. 101-102. *Intimate conviction, conviction, certainty, doubt* and other such phrases have all been linked in a more or less direct way to the persuasive conception I will discuss shortly.

<sup>54</sup> TUZET, 2020, p. 100; DEI VECCHI, 2016, p. 278 et seq.; GAMA, 2020, p. 8, the author attributes this thesis to Ferrer, but it is doubtful that Ferrer would ever accept this.

cratic effects of the lack of possibility of accounting (even by the judges themselves) of subjective mental states.

Indeed, to date, in fact, the only means to jointly account for the subjective and objective sphere of decision and evaluation of the evidence, in more or less controllable terms, would be to provide the judge with joint support from neuroscientists, forensic psychologists and sociologists at the moment, say, of the drafting of the sentence. However, such supports would not seem so accessible and it is not certain that the effective yield of the use of such supports (in terms of efficiency and transparency) is greater than the system devised by Ferrer.

The second problem with such criticisms is that of conceiving Ferrer's position in a descriptive rather than a prescriptive sense. It seems to me that there is a clear difference in scientific intent here (for example, between Gama and Ferrer) which makes the two positions complementary and not necessarily in conflict. Ferrer's proposal should be accepted for what it is: a prescriptive proposal for the construction of a more controllable and more empowering evidential system for decision makers, not a realistic description of procedural practices.

If this is the case, then, Gama should make an effort to clarify and concretely specify which real contributions would provide to Ferrer's normative proposal a greater reference to the history of evidential thought<sup>55</sup> and a greater harmonization of the «descriptive» and «subjective» sphere.<sup>56</sup>

Moreover, if it is true that after years of studies on these issues, the reference to a persuasive conception *tout court* borrowed from Taruffo, may now turn out to be a bit naive, it is also true that this reference, for Ferrer, does not constitute the founding axiom of his theory. Instead, it seems to be only functional to the explanation of the effects that his conception aims to avoid, just as it happens with reference to an elementary paradigm or model whose distorting effects are to be avoided.<sup>57</sup> Simplicity is precisely the element that allows the quality and harmfulness of the effects to be highlighted with particular clarity in relation to some variables that have been considered relevant. Obviously, a margin of error must be considered also in this case.

Something similar normally occurs also in other scientific fields: think, for example, of economic models, constructs that are always associated with an economic doctrine, used to «simplify» reality and to highlight some aspects (e.g. cause and effect between two or more economic variables).

---

<sup>55</sup> GAMA, 2020, spec. pp. 6-7, where we read: «The theorists of proof have placed emphasis on the assumptions of a rational study of proof, but they have not paid sufficient attention to evidence studies in our tradition from an historiographical point of view».

<sup>56</sup> GAMA, 2020, pp. 9 and ff.

<sup>57</sup> Which, it should be remembered, are well known since procedural systems of this type really existed. Think for example of ordeals.

The fact, then, that in real legal systems the irrational effects always appear with different intensity and complexity, does not necessarily invalidate the usability or effectiveness of the model, given that the gap between reality and the margin of error are natural variables of every intellectual processes. For this reason, the criticisms on this point seem to me not so crucial for Ferrer's position.

All this does not mean that Ferrer's model cannot be enriched or improved, but that critics should make a more concrete positive effort in suggesting which benefits would be introduced through their suggestions, not forgetting the aims of the concept they are analysing.

The «objectivity» pursued by Ferrer and achieved through the choice of the acceptance lexicon<sup>58</sup> seems to me, therefore, a quality element of his conception, although it is certainly possible to continue to work to make this proposal more refined or more «inclusive» towards other perspectives and problems that systematically afflict decision makers such as, for example, systematic cognitive biases.

Having considered this, I believe that formulating standards that tend to be devoided of references to mental states is desirable not only to guarantee a more equal *treatment* to refugee but also, as a consequence, to gain equality weights of hospitality between states.

UNHCR gives an interesting example about Netherlands. Here legislation and jurisprudence until 2015, adopted two different standards for establishing credibility: normally the applicant's statements must merely be «plausible» but in cases of simple inconsistency, ambiguity or gaps in the narrative, there was a presumption of fraudulent conduct and an extremely higher standard had to be satisfied. The applicant's statements, indeed, had to be evaluated as «positively persuasive»<sup>59</sup> but this, in most cases, led to a denial of protection and to the practical impossibility for the applicants to oppose it, given the subjective nature of this standard.<sup>60</sup>

The case is interesting because the Netherlands choose to move from a subjective standard of proof to a *methodological criteria* of evaluation of proof, the so called «integral credibility assessment» closer to the indications of the international literature.<sup>61</sup> This method has a holistic nature because only through the application of a

<sup>58</sup> FERRER, 2021, pp.198-200, 2006 pp. 303-305; 2005 pp. 35 ff.

<sup>59</sup> Known as the standard of «positive conviction». IND Working Instruction 2010/14, §4 and Art. 31 (2) (a) to (f), Aliens Act of 2000; see UNHCR 2013, pp. 238-239; METSELAAR 2017, pp. 74-89.

<sup>60</sup> METSELAAR, 2017, p. 74; see also: BALDINGER, 2015, p. 40.

<sup>61</sup> WBV 2014/36, Stcrt. 36910, 24 December 2014; Explanatory memorandum of the amendment to the Aliens Act (memorie van toelichting), Kamerstukken II, 2015, 34088, 3, p. 16; IND, Werkinstructie 2014/10, para. 1, available at <https://ind.nl/over-ind/Cijferspublicaties/Paginas/Werkinstructies.aspx> [accessed 6 January 2017]; A.A.M.J. Smulders, 'De POK is dood, lang leve de integrale geloofwaardigheidsbeoordeling', *Journaal Vreemdelingenrecht* (2015) v8, p. 1.

series of credibility criteria, the applicant's credibility can actually arise and a correct weight can be given to gaps or inconsistencies in the applicant's storytelling.<sup>62</sup>

What is relevant about this case to me is that the political choice was to leave a subjective standard precisely because of its epistemological and justice inefficiencies. Indeed, in characterizing this new method and its objective nature Metselaar highlights:

The «integral credibility assessment» must be objective, structured and transparent, by using objective sources and transparently motivating how the balancing has taken place. The objectiveness refers to the requirement that the IND<sup>63</sup> should refrain from any subjective notions when assessing the credibility [...] under the doctrine of «positive persuasion» too many different situations required the same burden of proof and motivation<sup>64</sup>. In the «integral credibility assessment» all the arguments should be addressed in the motivation, but also the weight given to statements that influence the credibility of a material fact are motivated more transparently.<sup>65</sup>

This example gives us reasons to reflect on the adequacy of a more objective characterization of the evaluation tools and on the fact that Member States are concretely beginning to feel this need as a necessary (although not sufficient) basis for a greater correctness of international protection decisions.

### 3.5. Standards of proof as tools of evidential policy

Another variable to consider is a sometimes neglected feature of the standards of proof: that they are, essentially, instruments of evidential policy rather than instruments with an epistemic function (i.e. rules on how to know the facts or on the formation of the evidentiary *corpus*).

The setting of a proof standard is a political and functional choice for political purposes that mainly concern the allocation or «distribution of the risk of error».<sup>66</sup> They are adopted on the basis of a balance of the values at stake (in our case: human rights, ECHR fundamental rights, impact of the decision on the applicant's life, the applicant's limited probative capacity, etc.) and the objectives pursued by the State (needs budget, reception capacity, strategic purposes, etc.).<sup>67</sup>

In our field of interest, the standard of proof involves a choice on the acceptable proportions of protection of individuals who do not need it and of non-protection

<sup>62</sup> Werkinstructie 2014/10, para. 3.2.1.1e 3.2.1.2 available at: <https://ind.nl/over-ind/Cijferspublicaties/Paginas/Werkinstructies.aspx>; see also METSELAAR 2017, p. 76.

<sup>63</sup> Immigration and Naturalization Service, Minister of Justice and Security.

<sup>64</sup> This is an argument used also by Ferrer. He has frequently criticized the formulation of a «single standard» for each disciplinary sector: e.g. FERRER & TUZET, 2018, p. 6.

<sup>65</sup> METSELAAR, 2017, p. 78. See: Werkinstructie 2014/10, §3.2., available at <https://ind.nl/over-ind/Cijfers-publicaties/Paginas/Werkinstructies.aspx>; see also: Notes on the report (Nota naar aanleiding van het verslag), Kamerstukken I, 2015, 34088, E, p. 8.

<sup>66</sup> FERRER, 2021, p. 115-138; FERRER & TUZET, 2018, p. 458.

<sup>67</sup> In the case that interests us, of course, these ultimate values will be human rights and fundamental rights recognized by the ECHR.

of those who are entitled to it. By means of a standard, the ordering also expresses «the degree of probability from which [is] willing to take the hypothesis as proven» and that is «what degree of support [...] seems sufficient to accept as the factual hypothesis in question is true».<sup>68</sup> The lower the standard of proof, the more applicants we protect (assuming, as we said, the responsibility and costs of protecting even subjects who are not in real danger), the higher the standard —and therefore difficult to meet—, the fewer individuals we protect, assuming the risk of endangering a greater number of subjects who, while expressing a well-founded fear, have not reached the established threshold. In this second case, therefore, far fewer subjects will be protected who, however, will find themselves in situations of more glaring or exceptionally serious danger (as in the extreme case of Belgium until 2015).

Gama's proposals (especially those relating to a greater harmonization of the subjective and objective sphere) seem to me to go in the direction of not adequately considering this characteristic of the standards. By placing standards of proof as an essential element of what appears to be his (descriptive) proposal for a conception of proof,<sup>69</sup> he also seems to take for granted a certain role and weight of the standards in defining the conception of proof itself without, however, giving reasons. This is a point on which it seems to me that too little thought has still been given and which deserves more attention.

Having said that, I think we must continue to reflect on the issue of the «precision» of standards. Gama criticizes Ferrer's position by highlighting that the inductive nature of the assessment as well as the irrepressible vagueness of the language would make it impossible to create standards that set a sufficient evidential threshold to satisfy the precision criteria required by Ferrer.<sup>70</sup> Evidence would therefore be incompatible with this level of precision.<sup>71</sup> This seems to me to be a strong criticism that will have to be taken seriously by Ferrer at least as regards a greater clarification of the natural gradualness that the requirement of precision proposed by him presupposes. I believe, indeed, that even if absolute precision cannot be expected, one can continue to defend the need for an effort to seek a greater degree of precision compatible with probabilistic and inductive reality. This last theme is certainly relevant for this research area, but it would require an in-depth study that is not possible here, so I will leave it to future work. Now I will try to answer the questions that closed paragraph 3.3.

In questions 1) and 2) I asked whether the standards generally used by national jurisprudence on international protection are adequate or whether there is a need to modify or reformulate these decision rules.

---

<sup>68</sup> FERRER & TUZET, 2018, p. 457.

<sup>69</sup> GAMA, 2020, p. 17.

<sup>70</sup> GAMA, 2020, pp. 12-13. See DEI VECCHI, 2016 pp. 286-287.

<sup>71</sup> GAMA, 2020, p. 15; cf. FERRER, 2019, pp. 255-256 ff.

This sphere is certainly a *sui generis* field because, as I have said, it is characterized on the one hand by a peculiar lack of information and, on the other, it is marked by principles absent in canonical trial. If this suggests that the particularly low threshold envisaged by these standards is adequate, it is not sufficient to accept some actual inefficiencies of these standards: the case of Belgium mentioned above is a clear example of this.

The problems with these standards are two: the first in terms of the effective possibility of recourse by the applicants and the second in political terms. Actually, on one hand, they do not allow a distribution of the risk of error consistent with the principles that should govern such proceedings and, on the other hand, they also enter into conflict with current European law and supranational bodies guidelines.

This all generates, as I already said, a general worsening of conflicts of interest, differentiated levels of protection, the spread<sup>72</sup> of abusive practices at the individual and national level and discredit to the whole European asylum system. Having said this, it does not seem, and harmful nor dangerous to investigate standards that can provide greater democratic guarantees to applicants for international protection, and I believe this can also be pursued through standards free from terms for which a very limited justification can be given. The provision of standards that enhance the dimension of «acceptance» would therefore be desirable not only at the judicial level but also (and above all) at a first decision level.

In any case, the evidential issues (including those involving the formulation and application of standards) implied in the *evaluation* of international protection application have not yet been resolved by the European Court of Justice and this makes these issues still more interesting and open to discussion by the Justice, and this philosophers of law. Other issues that seem relevant to me and on which I would like to reflect more are: 1) whether the proof of «well-founded fear» referred to in Article 2, letter d), Directive 2011/95 diverges from the «proof of actual risk» referred to in Article 2, letter f), Directive 2011/95; 2) whether the approach to the evaluation of evidence diverges according to the different phraseology used by the *common* and *civil law* systems; 3) whether the creation of a single system of proof for the recognition of international protection can guarantee an equal level of protection throughout the Union and acceptable transparency of decisions. In this sense, the Union itself seems to be moving, working to transform all the directives that currently regulate this area into regulation.

---

<sup>72</sup> BLOMMAERT, 2001.

#### 4. THE CREDIBILITY OF THE APPLICANT AND THE RELIABILITY OF THE APPLICATION

The art. 4, §5, lett. e) of the Qualification Directive 2011/95/EU mentions the verification of the general reliability of the *applicant* as one of the necessary conditions for the application of the so-called benefit of the doubt, the rule according to which «when certain aspects of the applicant's declarations are not supported by documentary or other evidence, their confirmation is not [...] necessary».<sup>73</sup> This wording echoes the one expressed by UNHCR in its *Handbook* that the benefit of the doubt should be granted when «the applicant's statements [are] consistent and plausible and not [...] contrary to generally known facts», i.e. when the examiner is *convinced* of the «overall credibility of the applicant».<sup>74</sup>

The case law of the Member States, for its part, does not consider credibility as referring to the sincerity of the applicants, but to the overall reliability of their *report*, that is, to the statements and other evidence produced in support of an application.<sup>75</sup>

It seems, therefore, that the sources examined lack of clarity, confusing the reliability of the applicant, of the information material provided by him and that of the overall application for protection, which is the reason why it seems desirable to try to draw a distinction between «reliability» and «credibility». In this regard, a unitary evidential system should not be afraid of conceiving the evaluation of the application in two distinct moments: one relating to the judgment of the applicant's overall credibility and one (conclusive) on the general reliability of the application which would include the first.

The assessment of credibility should refer to the assessment of the reliability of the applicant's statements, while the second should be understood as a subsequent

---

<sup>73</sup> Art. 4 §5 of the Qualification Directive of 2011/95: the conditions necessary for the application of the benefit of the doubt (which presents itself as a relative legal presumption or a closing rule) are: a) that the applicant must have made «sincere efforts» to substantiate the request; b) that it has produced all the relevant elements in its possession and satisfactorily justified any lack of other significant elements; c) who has submitted the application as soon as possible or adequately justified not having done so; d) the decision maker will be able to apply this rule only when he is generally convinced of the consistency and plausibility of the applicant's statements (which must not contradict the general and specific information produced) and e) is therefore satisfied that the applicant is generally reliable. The benefit of the doubt raises a number of issues on which the CJEU has not yet provided guidance, including: 1) whether state decision-making bodies should (or should not) apply the benefit of the doubt as a rule or principle; 2) whether it can really be extrapolated from the provisions of Article 4, paragraph 5, of the Qualification Directive 2011/95; 3) whether it is a mere decision-making practice or whether it derives, as a consequence, from the application of a standard of evidence. See EASO 2018, p. 81. In the Italian legal system, this provision is present in art. 3 paragraph 5 of Legislative Decree 2007 n. 251: where it is stated that, in the absence of evidence, such statements «are considered true».

<sup>74</sup> UNHCR 2019, p. 44, §204. Italics mine.

<sup>75</sup> See EASO 2018a, p. 80.

and broader judgment (to which the assessment of credibility is functional) through which the overall assessment of the fulfilment of the criteria for the protection is carried out. Determining the general reliability of the application will therefore be the epistemic aim of the decision maker of an application for international protection.

Moving away from the overlapping use that courts, EU law and international sources make of these terms (distinguishing between credibility referring to the epistemic material provided by the subject and reliability of the question as a whole), would allow to recognize a specific decision-making relevance to both evaluation moments and to provide more publicity to the reasons underlying the decision.

Despite the terminology and conceptual confusion over the concepts of trustworthiness and credibility, the International Association of Refugee Law Judges (IARLJ), UNHCR and the European Asylum Support Office (EASO) are promoting increasingly in-depth studies that certainly support the clarifying efforts made by scholars and operators in the sector.

With reference to credibility, for example, three indicators have been developed: «coherence, «plausibility» and «sufficient detail».<sup>76</sup> Specifically, they facilitate the decision on the acceptability of facts<sup>77</sup> but they cannot be understood as necessary or sufficient conditions for the assessment of applications for international protection<sup>78</sup> or as binding or decisive elements.

The examining authority is therefore strongly invited to submit these indicators to a holistic judgment since none of them, in isolation, is considered decisive in establishing the general reliability of the application for protection. As part of this assessment, they will therefore be subjected to a double «balancing»: a) one between the indicators themselves (for example, between a plausibility judgment and one of insufficient details), and then b) between indicators and the so-called «specific circumstances of the applicant» (for example, between the judgment of inconsistency and the weight that elements like age, education, culture, religion, sex, sexual orientation, health, and vulnerability, have had on his report).<sup>79</sup>

#### 4.1. Credibility indicators<sup>80</sup>

Generally, the consistency judgment on the evidence provided by the applicant is divided into two: internal consistency and external consistency. By «internal consist-

---

<sup>76</sup> The assessment of the credibility of the applicant is considered by EASO and UNHCR one of the most important problems related to international protection application. See IARLJ 2013; IARLJ 2016; UNHCR 2013; UNHCR 2014; Hungarian Helsinki Committee 2013 and 2015.

<sup>77</sup> EASO 2018a, p. 88.

<sup>78</sup> EASO 2018a, p. 88.

<sup>79</sup> EASO 2018a, p. 88.

<sup>80</sup> EASO 2018a, p. 87.

ency» we basically mean the non-contradictory nature of the evidence (statements, documents and other) provided by the applicant. The sources recommend that operators should evaluate this indicator with extreme caution because, even in cases where the applicant tries to provide the best possible support to his application, it is rare that he is able to provide an absolutely coherent, detailed and precise report.<sup>81</sup> This is because the precarious condition in which those seeking international protection usually find themselves can often hinder assertiveness, narration and memory, influencing the accuracy, completeness or relevance of what is transmitted. There are many factors that can affect the quality of the evidence provided by the applicant: think of the incidence of factors such as: 1) excessive emotionality, trauma, cultural and linguistic differences<sup>82</sup> but also 2) the methods of enforcement and verbalization<sup>83</sup> and 3) the conditions of manifest and hidden vulnerability (cases of cognitive disability or immaturity, think of minors).

This does not mean that fraudulent or dishonest conduct cannot be revealed,<sup>84</sup> but only that internal consistency is not necessarily an index of credibility<sup>85</sup> because, often, a negative assessment of credibility can also be due to factors external to the applicant (think of the case of a *biased* or simply not properly trained examiner).<sup>86</sup>

By «external coherence», on the other hand, we generally mean the non-conflict of all the information provided by the applicant (in the interview and at other times) with the *body* of information already ascertained in the procedure (COI<sup>87</sup>, appraisals, travel itineraries, particular events, etc.) and those generally known through maxims of experience<sup>88</sup> and covering scientific laws (scientific laws concerning human biology, the onset of pathologies, etc.).<sup>89</sup> Also, with regard to this indicator, the impact of the personal and contextual circumstances mentioned above must be taken into account. Thus, a negative assessment of the coherence of the evidence adduced by the applicant cannot justify the refusal to grant protection unless these inconsistencies are of such importance as to significantly affect the judgment of the overall reliability of the application.

<sup>81</sup> EASO 2018a, p. 89. UNHCR 2013, chapt. 2 and 3.

<sup>82</sup> EASO 2018a, pp. 176-190. See HERLIHY & TURNER, 2009.

<sup>83</sup> Usually in the interview at the Territorial Commission the narrated is recorded but not transcribed. On the other hand, it is very rare that the interview is videotaped.

<sup>84</sup> EASO 2018a, pp. 89-90. On the contrast between applicants considered «deserving» and «abusive» underlying the legislation on special needs, see: COSTELLO & HANCOX 2016, p. 376.

<sup>85</sup> In the same way the applicant's demeanor (or attitude) during the examination is not.

<sup>86</sup> JANSEN & SPIJKERBOER 2011.

<sup>87</sup> EASO 2018a, p. 90.

<sup>88</sup> On the concept of experience maxim: CANALE & TUZET, 2019, pp. 46, 56; TUZET, 2016, pp. 181-193 and 174-179 but already: CALAMANDREI, 1965, p. 51, note 106: «they consist of definitions or hypothetical judgments of general content obtained from the experience of the facts but independent of the individual cases from whose experience they are drawn, and outside of which they claim to be valid also for further cases». My translation.

<sup>89</sup> EASO 2018a, p.100 et seq.; 107 and ff.

Article 4, §5, letter c) of the 2011 Qualification Directive juxtaposes the concept of external consistency with that of plausibility, conveying a certain overlap between the two concepts.<sup>90</sup> The CJEU, despite having cited the concept of plausibility,<sup>91</sup> has not clarified whether in the lexicon of the Directive it has a distinct space that the law does not value or whether the term «plausibility» is used in a purely rhetorical-reinforcing sense. Despite the lack of clarity of the European legislator wording, «plausibility» and «coherence» should not be overlapped since the latter would seem to have a narrower extension than the former: a report may, in fact, not be credible even if it is plausible.<sup>92</sup> For its part, UNHCR has instead stated that plausibility refers to «reasonableness» or «possibility» that is, how likely a hypothesis is in relation to common sense.<sup>93</sup>

Although the use of maxims of experience is valuable in any evidential context and even more in a context in which the examining authority is called to decide on the basis of a small or even non-existent set of evidence, it should be pointed out that here the use of such means can be riskier than in other procedural contexts.

This occurs essentially for two reasons: 1) the scarcity of evidence in these procedures and 2) the highly cultural connotation of the maxims of experience. The scarcity of the evidence is an additional risk factor to the use of the maxims (which, in themselves, already discount the typical difficulties of inductive laws) because they cannot be «supported» or «contradicted» by other data and will be used as a means to make up for this lack of information.

The second reason, on the other hand, relates to the fact that most of the maxims of experience develop (and represent) typical regularities of a determined and specific socio-cultural context and can, therefore, have little heuristic value (if not misleading) when applied to events that occurred in other contexts.

Consider, for example, a «intra-family punitive reaction [...] disproportionate according to our canons» but perfectly common in a certain cultural context or the «brutal reaction» of a social group due to the «transgression of religious rules»<sup>94</sup> that, in our context, would make such a reaction unlikely. In both cases, the risk is to use maxims developed—for example—in the current European cultural context to find an incorrect judgment of implausibility.

Two other possibly misleading «maxims» are those of «rarity» and that of «similarity». The first is based on the fact that if a phenomenon is attested by the sources as «very rare» (for example, the practice of female genital mutilation in those countries where this generally does not happen) then it must be considered implausible. The second

<sup>90</sup> MINNITI, 2020, §4.3.1., speaks of it in terms of «likelihood» or «logical likelihood».

<sup>91</sup> CJEU, judgment of 26 February 2015, case C-472/13, *Andre Lawrence Shepherd v. Bundesrepublik Deutschland*, EU: C: 2015: 117, §43.

<sup>92</sup> EASO 2018a, p. 92.

<sup>93</sup> See UNHCR 2013, pp. 176-177.

<sup>94</sup> MINNITI, 2020, §4.1.

requires considering implausible —and therefore lacking in authenticity— an evidential set considered stereotyped or too similar to other stories already heard.<sup>95</sup> These examples show that in contexts where the evidence is particularly incomplete, the use of maximum of experience is useful but at the same time requires higher standards of motivation, capable of counteracting the risks inherent in such generalizations.<sup>96</sup>

The third indicator concerns the «sufficiency of details». The lack of «relevant» or significant elements —pursuant to art. 4, §5, lett. b)— could be considered as a «surprising» element, if referred to the narration of a lived event. However, as previously mentioned, given the particular condition of the applicant, this event is frequent and should be assessed «in a balanced and objective way».<sup>97</sup> Moreover, an interesting question, but for now unanswered by jurisprudence and EU law, concerns the threshold relating to how many details can reasonably be expected from a subject who finds himself proposing an international protection application based on an experience referred as personal and authentic.<sup>98</sup> Also in this case, the indicator is highly sensitive to personal experiences and to the examination methods. For this reason, an adequate evaluation will have to carry out a judgment that concretely takes into account the impact of these variables.<sup>99</sup>

The goodness of the evaluation of these indicators will therefore also be a function of how well trained and competent of the authorities involved, which should guarantee high standards of flexibility to cultural diversity and the ability to build, with the applicant, a trusting relationship on the basis of which a serene and complete presentation of the relevant information may take place.<sup>100</sup>

#### 4.2. Credibility: special cases

As I have just mentioned, a relevant variable in the evaluation of applications for international protection is the negative impact that the individual and contextual

---

<sup>95</sup> UNHCR points out that this way of reasoning is very risky given the repetitiveness with which decision makers are subjected to examining and evaluating very similar causes: UNHCR 2013, pp. 79-80; EASO 2018a, p. 35. See also: Cass. Civ. Sez. VI-1 sent. 2014 n ° 22111, this is now generally recognized also by legal psychology.

<sup>96</sup> EASO 2018a, p. 93.

<sup>97</sup> EASO 2018a, p. 91.

<sup>98</sup> EASO 2018a, p. 91. See also the judgment of 27 May 2014 of the Administrative and Labor Court of Szeged (Hungary), 7.K.27.145/2014/9.

<sup>99</sup> Thus, an applicant will be able to provide vague answers, while reporting a true report and the lack of details must be considered «in broad terms», not as a decisive factor of non-credibility: EASO 2018a, p. 92. United Kingdom Immigration and Asylum Court, judgment of 12 June 2003, JB (Democratic Republic of the Congo), 2003, UKIAT 00012.

<sup>100</sup> This becomes essential especially on those occasions in which the applicant has been the victim of painful and traumatic experiences. EASO underlines that among the most damaging attitudes to the collection of evidence are attitudes of skepticism or frequent interruptions during the interview: EASO 2018a, p. 91.

circumstances of the applicant can have on the credibility and, consequently, on the general reliability of the application. Consider, for example, how much post-traumatic *stress* can influence the ability to remember events or the behaviour of the applicant during the interview, a moment which is, notoriously, the most delicate and decisive in the evaluation of the application for protection.<sup>101</sup> These difficulties have led to problematic enforcement scenarios in which applicants were subjected to inappropriate questioning or felt obliged to submit sensitive material to «prove» crucial elements of their application for protection (e.g. their sexual orientation).<sup>102</sup>

For example, cases of persecution linked to gender and sexual orientation, those involving victims of trafficking or at risk of being trafficked,<sup>103</sup> or disabled and minors, are particularly complex and harbingers of procedural challenges.<sup>104</sup> In the specific case of minors,<sup>105</sup> it is said that they must benefit of specific procedural and evidential precautions to guarantee the correctness and non-injurious nature of the decisions<sup>106</sup> because a stronger presumption of vulnerability applies to them.<sup>107</sup>

If they are accompanied by subjects (such as, for example, the family or others) who already integrate the protection conditions, the assessment is relatively simple as the fear is generally considered «founded» by presumption. Instead, in cases in which they are alone, the assessment must be carried out by experts and in compliance with the international legislation in force in their favour. However, the assessment of the validity of their fear must be calibrated according to the stage of psycho-cognitive development they have reached.<sup>108</sup>

The epistemological pitfall hidden in the evaluation of the fear of minors (as well as of subjects suffering from psychiatric pathologies) therefore lies in the fact that their perception of danger could, at times, not accurately reflect the reality of the facts.<sup>109</sup> Reflecting on these cases highlights in a particular way the problems inherent in an assessment system that conceives evidence in predominantly subjective terms.

<sup>101</sup> UNHCR 2013; EASO 2018. See BLACK, 1994.

<sup>102</sup> The practice has now been recognized as contrary to the fundamental right to private and family life and incompatible with human dignity under EU law. CJEU, judgment of 2 December 2014, joined cases C-148/13 to C-150/13, *A, B and C v. Staatssecretaris van Veiligheid en Justitie*, EU: C: 2014: 2406.

<sup>103</sup> UNHCR 2002, UNHCR 2012, UNHCR 2006.

<sup>104</sup> UNHCR 2009; EASO 2018a, p. 130 and ff. see also §29 of the Preamble of the Asylum Procedures Directive 2013/32/EU.

<sup>105</sup> UNHCR 2009, § 13.

<sup>106</sup> UNHCR 2009, § 65.

<sup>107</sup> Minors looking for international protection are considered vulnerable as they are particularly subject to being harmed by reason of their age. The presumed vulnerability increases if they are not «accompanied» by family or guardians.

<sup>108</sup> I am especially referring to the Convention on the Rights of the Child and its Protocols. See GAUDIO & BACCHIO, 2019.

<sup>109</sup> However, this problem would seem balanced by the standard of proof and by the humanitarian purpose which, in any case, tend to provide broad protection in this assessment procedure.

## 5. CONCLUSIONS

In the present article, I have analysed the main characteristics of the evidential assessment concerning the assignment of refugee *status*, underlining the complexity of this research area and its peculiarity with respect to the civil and criminal trial.

It has been noted how, in the European context, a unitary evidential system is lacking and how this constitutes a not negligible lacuna. Indeed, it removes effectiveness to the Common European Asylum System exposing applicants to differentiated and inadequate treatments and the States to a tightening of conflicts of interest with other Member States, with the Union and with individuals.

It has also been said that the contribution offered in evidential matters by bodies such as UNHCR, EASO and IARLJ, while relevant, is not enough to create a real culture of proof in this sector and a real harmonization of practices in European Union.

Subsequently, I have focused, more specifically, on some issues concerning the peculiar «subjectivity» that characterizes the proof required in this area and the probative standards. This reflection was, then, inserted into the debate that is touching the rationalist conception of legal proof promoted by Jordi Ferrer. It concluded by hoping for a more objective reformulation of the standards and a reflection on a unitary evidential asylum system.

Finally, some issues related to the evaluation of the applicant's storytelling were considered. Consideration was given, in particular, to the assessment of credibility and reliability, highlighting a certain confusion on this issue in international sources and in secondary EU law.

I hope that this text, with its merits and demerits, will bring scholars to this evidence area and offer them new tools for reflection.

## BIBLIOGRAPHY

- BALDINGER, D., 2015: *Vertical Judicial Dialogues in Asylum Cases. Standards on Judicial Scrutiny and Evidence in International and European Asylum Law*, Leiden-Boston, Brill NijHoff.
- BLACK, R., 1994: «Livelihoods under Stress: A Case Study of Refugee Vulnerability in Greece», *Journal of Refugee Studies*, vol. 7 (4), pp. 360-377.
- BLOMMAERT, J., 2001: «Investigating Narrative Inequality: African Asylum Seekers' Stories in Belgium», *Discourse & Society*, 12 (4), pp. 513-449.
- CALAMANDREI, P., 1965: *Opere giuridiche, vol. I: Problemi generali del diritto e del processo*, Roma, RomaTre Press, 2019, <http://romatpress.uniroma3.it/wp-content/uploads/2019/09/Opere-giuridiche-%E2%80%93-Volume-I-%E2%80%93-Problemi-general-del-diritto-e-del-processo.pdf>.
- CANALE, D. & TUZET, G., 2019: *La giustificazione della decisione giudiziale*, Turín, Giappichelli.
- COSTELLO, C. & HANCOX, E., 2016: «The Recast Asylum Procedures Directive 2013/32 / EU: Caught between the Stereotypes of the Abusive Asylum-Seeker and the Vulnerable Refugee», in Chetail, V., de Bruycker, P. & Maiani, F. (eds): *Reforming the Common European Asylum System—The New European Refugee Law*, Leiden, Brill-Nijhoff, p. 376.

- DEI VECCHI, D., 2016: «La prueba judicial como conocimiento: una caracterización poco persuasiva», in Ferrer Beltrán, J. & Vázquez, C. (eds.), *Debatiendo con Taruffo*, Barcelona, Marcial Pons.
- EASO, 2018a: *Evidence and credibility assessment in the context of the Common European Asylum System*, Luxemburgo, Publications Office of the European Union.
- , 2018b: *Conditions for the recognition of international protection (Directive 2011/95/EU)—A legal analysis*.
- EUROPEAN COMMISSION, 2020: *Amended proposal for a Regulation of the European Parliament and of the Council establishing a common procedure for international protection in the Union and repealing Directive 2013/32/EU*, 23.09.2020, COM (2020) 611 final.
- EUROPEAN MIGRATION NETWORK, 2014: *The Organization of Reception Facilities for Asylum Seekers in different Member States*,
- FERRER, J., 2005: *Prueba y verdad en el derecho*, Barcelona, Marcial Pons.
- , 2006: «Legal Proof and Fact Finders' Beliefs», *Legal Theory*, 12, pp. 293-314.
- , 2007: *La valoración racional de la prueba*, Madrid-Barcelona, Marcial Pons.
- , 2019: *Prueba y racionalidad de las decisiones judiciales*, Barcelona, Marcial Pons.
- , 2021: *Prueba sin convicción. Estándares de prueba y debido proceso*, Barcelona, Marcial Pons.
- FERRER, J. & TUZET, G., 2018: «Sulla necessità degli standard di prova per la giustificazione delle decisioni giudiziali», *Diritto & Questioni Pubbliche*, XVIII, 2, pp. 455-472.
- GAMA, R., 2020: «En búsqueda de El Dorado. La concepción racional de la prueba y la formulación de estándares de prueba precisos y objetivos», *Revus*, 43.
- GASCÓN ABELLÁN, M., 1999: *Los hechos en el derecho*, Barcelona, Marcial Pons.
- GAUDIO, R.M. & BACCHIO E. 2019: «La valutazione medico legale» in Bernardini, M. G. (a cura di), *Migranti con disabilità. Rappresentazioni, politiche, diritti*, Nápoles, Jovene, pp.147-159.
- GONZALEZ LAGIER, D., 2018: «Tres modos de razonar sobre hechos (y algunos problemas sobre la prueba judicial planteados a partir de ellos)» in Vazquez, C. (ed.): *Hechos y razonamiento evidencial*, Mexico, Editorial CEJI, pp. 17-43
- HERLIHY, J. & TURNER, S. W., 2009: «The Psychology of Seeking Protection», *International Journal of Refugee Law*, 21 (2), pp. 1-16.
- HUNGARIAN HELSINKI COMMITTEE, 2013: *Credibility Assessment in Asylum Procedure, A Multidisciplinary Training Manual*, vol. 1, <https://helsinki.hu/wp-content/uploads/Credibility-Assessment-in-Asylum-Procedures-CREDO-manual.pdf>.
- , 2015: *Credibility Assessment in Asylum Procedure, A Multidisciplinary Training Manual*, vol. 2, <https://helsinki.hu/wp-content/uploads/CREDO-training-manual-2nd-volume-online-final.pdf>.
- IARLJ, 2013: *Assessment of Credibility in Refugee and Subsidiary Protection Claims under the EU Qualification Directive, Judicial Criteria and Standards*, Progetto CREDO, [https://helsinki.hu/wp-content/uploads/Credo\\_Paper\\_March2013-rev1.pdf](https://helsinki.hu/wp-content/uploads/Credo_Paper_March2013-rev1.pdf)
- , 2014: *Due process standards for the use of country of origin information (COI) in administrative and judicial procedures*, [https://www.iarjl.org/images/stories/Tunis\\_conference/WPPapers/COI.pdf](https://www.iarjl.org/images/stories/Tunis_conference/WPPapers/COI.pdf).
- , 2016: *A Structured Approach to the Decision Making Process in Refugee and other International Protection Claims*, [https://www.iarmj.org/iarlj-documents/2017\\_publications/IARLJ\\_guidechartCOIJul17.pdf](https://www.iarmj.org/iarlj-documents/2017_publications/IARLJ_guidechartCOIJul17.pdf).
- LAUDAN, L. 2005: «Por qué un estándar de prueba subjetivo y ambiguo no es un estándar», *Doxa*, 28, pp. 95-113.
- , 2006: *Truth, Error, and Criminal Law. An Essay in Legal Epistemology*, Cambridge, Cambridge University Press, 2006.
- METSELAAR, E. J., 2017: «The motivation of the Dutch credibility assessment in return decisions and its judicial review in the light of relevant standards in international and EU law», Thesis for Master in law LLM, DOI:10.13140/RG.2.2.24323.91688.
- MINNITI, L., 2020: «La valutazione di credibilità del richiedente asilo tra diritto internazionale, dell'UE e nazionale», *Questione giustizia, Rubrica Diritti senza confini*, ASGI, 21 de enero de 2020, <https://www.asgi.it/notizie/la-valutazione-di-credibilita-richiedente-asilo-tra-diritto-internazionale-del-lue-e-nazionale/>.

- RICARTE, C., 2019: «Argumentación y teoría de la prueba en el mundo latino. Un punto de partida», Tesis doctoral, Universidad de Alicante.
- SUCAR, G. & CERDIO, J., 2017: *Derecho y verdad*, vol. 4, Valencia, Tirant.
- TARUFFO, M., 1992: *La prova dei fatti giuridici. Nozioni generali*, Milán, Giuffrè.
- , 2009: *La semplice verità*, Roma-Bari, Laterza.
- TUZET, G., 2006: *La prima inferenza, L'abduzione di C. S. Peirce fra scienza e diritto*, Turín, Giappichelli.
- , 2016, *Philosophy of Legal Proof*, Turín, Giappichelli.
- , 2020, «Assessment Criteria or Standards of Proof? An Effort in Clarification», *Artificial Intelligence & Law*, 28 (1), 99-109.
- TWINING, W., 1982: «The Rationalist Tradition of Evidence Scholarship», in Campell, E. & Waller, L. (eds.): *Well and Truly Tried*, Sídney, The Law Book Company, pp. 211-249.
- , 2020: «La prueba como un ámbito multidisciplinar: ¿de qué manera puede el derecho y su estudio contribuir a esta empresa?», in Ferrer Beltrán, J. & Vazquez, C. (eds.): 2020, *El razonamiento probatorio en el proceso judicial. Un encuentro entre diferentes tradiciones*, Barcelona, Marcial Pons.
- UNHCR, 1998: *Note on Burden and Standard of Proof in Refugee Claims*.
- , 2002: *Guidelines on International Protection No. 1: Gender-Related Persecution Within the Context of Article 1A (2) of the 1951 Convention and / or its 1967 Protocol Relating to the Status of Refugees*, 7 May 2002, HCR/GIP/02/01.
- , 2006: *Guidelines on International Protection No. 7: The Application of Article 1A (2) of the 1951 Convention and/or 1967 Protocol Relating to the Status of Refugees to Victims of Trafficking and Persons at Risk of Being Trafficked*, 7 April 2006, HCR/GIP/06/07.
- , 2009: *Guidelines on International Protection No. 8: Child Asylum Claims under Articles 1 (A) 2 and 1 (F) of the 1951 Convention and/or 1967 Protocol relating to the Status of Refugees*, 22 December 2009, HCR/GIP/09/08.
- , 2010: *The Heightened Risk Identification Tool (User Guide)*.
- , 2012: *Guidelines on International Protection No. 9: Claims to Refugee Status based on Sexual Orientation and/or Gender Identity within the context of Article 1A (2) of the 1951 Convention and/or its 1967 Protocol relating to the Status of Refugees*, 23 October 2012, HCR/GIP/12/01.
- , 2013: *Beyond proof. Credibility Assessment in EU Asylum Systems*, Brussels.
- , 2014: *The Heart of the Matter: Assessing Credibility when Children Apply for Asylum in the European Union*, CREDO report.
- , 2015: *Annotated Comments to Directive 2013/33/EU of the European Parliament and Council of 26 June 2013 laying down standards for the reception of applicants for international protection (recast) To assist with transposition and implementation*.
- , 2019: *Handbook on Procedures and Criteria for Determining Refugee Status*.



## «DÉJAME ATRAVESAR EL VIENTO CON DOCUMENTOS»\*: PROPUESTAS PARA REPLANTEAR LA PRUEBA DOCUMENTAL DEL SIGLO XXI\*\*

Vitor de Paula Ramos\*\*\*

UFRGS/UdG, abogado  
vitordepaularamos@hotmail.com

**RESUMEN:** En el presente artículo se defiende la necesidad de que las ideas sobre los documentos en el *civil law* consideren la realidad actual y se subraya la importancia de los procesos interpretativos para la reconstrucción de los sentidos. Así, en la primera parte se presentan y critican las premisas tradicionales, haciendo hincapié en los cambios culturales que se produjeron en la segunda mitad del siglo xx, que también deberían haber afectado a la teoría sobre los documentos. En la segunda, pensando en las fases de admisión, práctica y valoración de la prueba documental, se formulan críticas y propuestas para mejorar la forma de trabajar con documentos en los países de *civil law*.

**PALABRAS CLAVE:** Prueba documental; documentos; representación; signos; autenticidad; contradicción sobre documentos; medios de prueba.

---

\* El título parafrasea la canción *Sin documentos*, de Los Rodríguez. Composición de Andrés Calamaro, 1993.

\*\* Revisión: Laura CRIADO SÁNCHEZ.

\*\*\* Mi agradecimiento al Prof. Dr. Ronald J. ALLEN, por haberme recibido en la Northwestern University, en Chicago, como *Visiting Scholar*, donde pude investigar y debatir con él mis primeras ideas sobre la prueba documental. Agradezco, además, a las y los *peers* que revisaron el trabajo y me hicieron críticas muy importantes para mejorar su calidad. Por último, mi agradecimiento a Marina FISCHER y Achilles STEINHAUS.

**«LET ME CROSS THE WIND WITH DOCUMENTS»:  
PROPOSALS FOR RETHINKING DOCUMENTS USED  
AS EVIDENCE IN THE 21TH CENTURY**

**ABSTRACT:** In this article, the author argues that ideas about documents in civil law need to consider the current reality, and highlights the importance of interpretive procedures to reconstruct meanings. The first part offers an analysis of the traditional premises about documents in civil law countries, emphasizing the cultural changes in the second half of the 20th century, which should have had an impact on the theory about documents. In the second part, the author reflects on the phases of admission, production, and evaluation of documentary evidence, and presents ideas and proposals to rethink the way documents are processed in civil law countries.

**KEYWORDS:** Writings, recordings and Photographs; Documents; representation; signs; authenticity; contradictory about documents; means of proof.

**SUMARIO:** 1. INTRODUCCIÓN.—2. DESTRUYENDO CERTEZAS: LA FALSA IDEA DEL SIGNIFICADO AUTOEVIDENTE. LA NECESIDAD DE INTERPRETACIÓN Y DE RAZONAMIENTO PROBATORIO: 2.1. Los vientos: dos cambios significativos en la cultura jurídica y extrajurídica de la segunda mitad del siglo xx. 2.2. Cambian los vientos, se mantienen las ideas: 2.2.1. ¿Representación inmediata y objetiva?. 2.3. Algunas propuestas iniciales.—3. ATRAVESANDO EL VIENTO. LOS DOCUMENTOS EN EL PROCEDIMIENTO PROBATORIO: 3.1. La admisibilidad. Compartiendo responsabilidades epistémicas. 3.2. Contradicción y contenido: el alcance de la contradicción sobre documentos. 3.3. La valoración de los documentos: la necesidad de interpretación y el razonamiento probatorio.—4. CONCLUSIONES.

**RECOMMENDED CITATION:** VITOR DE PAULA RAMOS, 2022: «Déjame atravesar el viento con documentos»: propuestas para replantear la prueba documental del siglo XXI», in *Quaestio facti*, 3: 131-157. Madrid: Marcial Pons Ediciones Jurídicas y Sociales. DOI: [http://dx.doi.org/10.33115/udg\\_bib/qf.i3.22655](http://dx.doi.org/10.33115/udg_bib/qf.i3.22655)

## 1. INTRODUCCIÓN

En *O Tempo e o Vento*, una de las más importantes obras de la literatura de Rio Grande do Sul (sur de Brasil), Érico VERÍSSIMO narra 200 años de historia de una familia, donde el viento es el símbolo del paso del tiempo, de la transitoriedad. «Esta idea de transitoriedad se refleja en la propia narración, se trata de una saga de héroes del pasado, cuya importancia y nombres quedaron en la Historia, que ya forman parte del pasado y se han convertido en espectros, “fantasmas”»<sup>1</sup>.

<sup>1</sup> GUIA DO ESTUDANTE, 2018.

Con la prueba documental pasa algo parecido. Cuando, en la primera mitad del siglo xx, autores como CHIOVENDA (1923: 842) y CARNELUTTI<sup>2</sup> (1936: 690 y ss.; 1947: 122) escribían sobre la teoría de la prueba documental, la cámara Leica I, de 1925, empezaba a popularizar la fotografía.

A partir de ahí, la complejidad de la prueba documental solo ha aumentado. Primero, con la creciente popularización de la fotografía y los vídeos; y, después, en las décadas de 1990 y 2000, con la popularización, entre otras muchas cosas, de los ordenadores e internet; es decir, se añadieron a la realidad jurídica muchos tipos de documentos, cuya existencia CHIOVENDA y CARNELUTTI no podían ni siquiera imaginar.

En un mundo de *iPhones*, cámaras de seguridad, algoritmos, correos electrónicos y otros tantos documentos, cualquier lego puede suponer que debe quedar muy poco de las premisas de la teoría de la prueba documental de CHIOVENDA y CARNELUTTI. No obstante, las ideas más actuales de la doctrina y la jurisprudencia siguen trabajando, explícita o implícitamente, con muchas de las premisas de estos autores, lo que genera un escenario confuso e ineficiente. Es necesario, por ende, *atravesar el viento*, permitiendo que el paso del tiempo se refleje también en una forma de trabajar con documentos más actual.

Con este objetivo, en la primera parte del artículo, después de hacer un breve comentario histórico, presentaré algunas ideas de la doctrina clásica que, a mi juicio de forma injustificada, aún están presentes en los sistemas de *civil law*. En la segunda parte, intentaré desarrollar algunas propuestas para mejorar la forma de trabajar con documentos en el procedimiento probatorio, poniendo el acento más en el razonamiento probatorio y no tanto en el soporte o la autoría<sup>3</sup>.

## 2. DESTRUYENDO CERTEZAS: LA FALSA IDEA DEL SIGNIFICADO AUTOEVIDENTE. LA NECESIDAD DE INTERPRETACIÓN Y DE RAZONAMIENTO PROBATORIO

### 2.2. Los vientos: dos cambios significativos en la cultura jurídica y extrajurídica de la segunda mitad del siglo xx

Como he mencionado en la introducción, cuando CHIOVENDA (1923: 842) y CARNELUTTI<sup>4</sup> (1936: 690 y ss.; 1947: 122) escribieron sobre la prueba documental,

<sup>2</sup> Durante la primera mitad del siglo xx, otros autores también escribieron sobre el tema. Véase, por ejemplo, COUTURE, 1942: 218; GUIDI, 1950: 43. En los primeros años y décadas de la segunda mitad del siglo, DENTI, 1957: esp. 26-27; SANTOS, 1972: 60.

<sup>3</sup> Para un desarrollo más global del tema, véase DE PAULA RAMOS, 2021.

<sup>4</sup> Durante la primera mitad del siglo xx, otros autores también escribieron sobre el tema. Véase, por ejemplo, COUTURE, 1942: 218; GUIDI, 1950: 43. En los primeros años y décadas de la segunda mitad

hablar de documentos era sinónimo de hablar de documentos *escritos* y en *papel*. CARNELUTTI llegaba a señalar la posibilidad de que hubiera documentos en «tela, cera, metal, piedra y similares»<sup>5</sup>, pero afirmaba que el más difundido, entre los medios de elaboración documental, era el escrito, para concluir de forma contundente que:

[...] el documento es en papel [*cartaceo*]; por cierto, del mismo modo como hay un sinónimo entre documento y escrito, también lo hay entre documento y papel, en el sentido de que escrito y papel se relacionan, por excelencia, con el significado de documento<sup>6</sup>.

Dos aspectos muy relevantes para el estudio de los documentos deben, en mi opinión, extraerse de las ideas de CARNELUTTI.

El primero es que, como he mencionado, en 1936, CARNELUTTI ya hablaba de las llamadas «reproducciones mecánicas», aunque en aquella época la cámara Leica I, de 1925, estaba empezando a popularizar la fotografía. En las décadas posteriores, empezaron a surgir muchos otros tipos de documentos no escritos (o no necesariamente escritos) y sin soporte *cartaceo*.

En 1959, por ejemplo, la Xerox 914 empezó a hacer popular el uso de las fotocopias<sup>7</sup>; en 1970, la Sony Portapak, las cámaras de vídeo<sup>8</sup>, que pasaron a tener uso doméstico; en los años setenta, surgió el fax<sup>9</sup>. Y a partir de los noventa se produjeron algunos de los cambios más intensos, gracias a los correos electrónicos, internet y las cámaras digitales. Ya en los 2000, surgen los primeros teléfonos móviles con cámara fotográfica y, en 2007, se presentó el primer *iPhone*. Curiosamente, CARNELUTTI intentó definir las (primeras) «reproducciones mecánicas» como algo muy parecido a los documentos (escritos y en papel)<sup>10</sup>, pero su teoría general no cambió pese a la constatación de que no todos los documentos eran, ya en aquel momento, escritos y ni en papel.

El segundo aspecto importante es ver cómo, en la primera mitad del siglo xx, estaban presentes muchas de las ideas del siglo anterior, que explicaban que tanto la fotografía como los textos contendrían un sentido unívoco, que solo tenía que ser

---

del siglo, DENTI, 1957: esp. 26-27; SANTOS, 1972: 60.

<sup>5</sup> CARNELUTTI, 1936: 693.

<sup>6</sup> CARNELUTTI, 1936: 692-693.

<sup>7</sup> Sobre la historia de la Xerox 914, véase <https://es.digitaltrends.com/computadoras/xerox-prime-ra-fotocopiadora-historia-revolucion/>. Consultado por última vez el 28/10/2021.

<sup>8</sup> Para la historia sobre como la Portapak ha dado «voz a los *outsiders*» y cambió los *medios* para siempre, véase <https://www.atlasobscura.com/articles/revolution-televised-camera-portapak-sony-vi-deo>. Consultado por última vez el 28/10/2021.

<sup>9</sup> Sobre la historia del fax, véase <https://historiadel.com/fax/>. Consultado por última vez el 21/10/2021.

<sup>10</sup> CARNELUTTI, 1936: 694.

percibido. La fotografía, por ejemplo, era considerada, de hecho, como una mera reproducción: el «infalible lápiz del sol»<sup>11</sup>, donde el fotógrafo era un mero operador<sup>12</sup>.

Además, tras la forma de ver la interpretación de los textos que se contienen en documentos, probablemente se encontraba el mismo orden de ideas que defendía la doctrina clásica acerca de la interpretación de los textos legales. Es decir, que la actividad de interpretar constituiría una actividad de *descubrimiento* del sentido de la ley, que previamente le había dado el legislador (excepto en los casos de las leyes mal formuladas). CHIOVENDA, por ejemplo, defendía la idea de una *voluntad de la ley*, que ya se encontraba en el texto legal y que le correspondía al juez *descubrir*, razón por la que describía la sentencia como «la resolución del juez que afirma existente o inexistente la voluntad concreta de ley deducida en el pleito»<sup>13</sup>. Sobre la interpretación de los textos normativos procesales, CARNELUTTI aún es más contundente:

Por eso, ante cualquier declaración surge la pregunta: ¿Cuál es su verdadero significado? La pregunta se impone de verdad cuando el texto de la declaración es dudoso y también cuando las palabras puedan ser interpretadas con más de un significado; pero la verdad es que se presenta siempre, incluso cuando el texto es claro, pues siempre es posible el error de quien declara, es decir, el uso de palabras que no traducen exactamente el pensamiento y, por ello, siempre es oportuno verificar la correspondencia entre el pensamiento y el signo. Este es el problema de la interpretación<sup>14</sup>.

Por consiguiente, no es absurdo imaginar, respecto de la interpretación de los textos legales, que esa idea del juez que es la boca de la ley, que *descubre* el sentido ya listo y acabado del texto, se encontrara también en la interpretación de los textos presentes en los documentos.

En este momento histórico, fuera del campo de la prueba, algunas cosas que afectarían intensamente la forma de trabajar con documentos empezaban lentamente a cambiar. En otras palabras, los conocimientos que surgieron en la segunda mitad del siglo XX deberían, al menos en potencia, afectar sustancialmente la forma de trabajar con documentos.

En la primera mitad del siglo, SAUSSURE, considerado el padre de la lingüística moderna, ya había hecho importantes contribuciones y señalado, por ejemplo, el carácter arbitrario de los signos lingüísticos<sup>15</sup>, al defender que «el significado no tiene ninguna conexión [*attache*] natural con la realidad»<sup>16</sup> y que depende de «hábitos

<sup>11</sup> ROOT, 1864: 28.

<sup>12</sup> Sobre la historia de la fotografía y la figura del fotógrafo como operador, véase SONTAG, 1973: 67.

<sup>13</sup> CHIOVENDA, 1923: 803. No es el tema de este artículo, pero es interesante subrayar que esta tradición está dentro de un contexto cultural en el que las funciones del juez se acercaban a una burocratización del juez, «que había acabado homogeneizando la función judicial con la administrativa» (PICARDI, 2007: 170). Este modelo acababa justamente por replantear «en sustancia, la imagen del juez *bouche de la loi*» (PICARDI, 2007: 171).

<sup>14</sup> CARNELUTTI, 1936: 106.

<sup>15</sup> SAUSSURE, 1931: 75.

<sup>16</sup> SAUSSURE, 1931: 76.

colectivas»<sup>17</sup>. En la segunda mitad del siglo, importantes trabajos en el campo de la lingüística y la semiótica siguieron esta senda y, en la década de 1960, se publicaron algunas de las obras más importantes de HJELMSLEV (1961), MORRIS (1964) y ECO (1968).

Dichos trabajos, pese a sus peculiaridades, tenían en común, con relación a la lengua (y, por tanto, también a los textos), el desarrollo, sobre múltiples complejidades, de los procesos lingüísticos, incluso de atribución o reconstrucción de sentidos. Por tanto, ya quedaba muy claro que la interpretación estaba lejos de ser una mera operación de «descubrimiento» de sentidos, pues involucraba, entre otras cosas, sentidos literales y figurados (*denotación* y *connotación*)<sup>18</sup>, sentidos puramente contextuales<sup>19</sup>, diferencias entre los tipos de significación que los signos pueden tener en función del caso<sup>20</sup>, etc. Es decir, se entendía que se trataba de un proceso<sup>21</sup> que dependía no solo de los signos<sup>22</sup>, sino también de la *interpretación*.

En el campo de la teoría del Derecho, también se sintieron estas complejidades, pues, a lo largo del siglo pasado, la idea del *cognitivismo* interpretativo de los textos legales empezó a criticarse con dureza. Partiendo, en 1934, de KELSEN<sup>23</sup> —quien ha definido «el Derecho aplicable como un marco dentro del cual hay varias posibilidades de aplicación»—, en la segunda mitad del siglo pasado siguieron importantes trabajos, como, por poner solo dos ejemplos, los de HART<sup>24</sup> y TARELLO<sup>25</sup> (y después muchos otros)<sup>26</sup>. De modo que, hoy en día, se ha convertido casi en una obviedad afirmar que «cualquier texto normativo admite una pluralidad de interpretaciones» (GUASTINI, 2011: 39).

<sup>17</sup> SAUSSURE, 1931: 76.

<sup>18</sup> ECO, 1968: 33; MORRIS, 1964: 67.

<sup>19</sup> HJELMSLEV, 1961: 44-45.

<sup>20</sup> MORRIS, 1964: 4. El autor usa las palabras *negro*, *bueno* y *debe* como «ejemplos obvios». Sobre su teoría de la atribución de sentidos, véase HJELMSLEV, 1961: 56 y ss.

<sup>21</sup> ECO, 1968: 31.

<sup>22</sup> No es pertinente entrar ahora en un debate profundo sobre el significado de la palabra *signo*. Adoptaré la definición de PEIRCE (1873: 313), según la cual el signo es un objeto que «está por otro [*stands for another*] para alguna mente». En sentido análogo, HJELMSLEV (1969: 45), para quien el signo es «una expresión que señala hacia un contenido que hay fuera del signo mismo».

<sup>23</sup> Véase KELSEN, 1960: 390 y ss. Aquí citada la segunda edición. La primera edición es del 1934.

<sup>24</sup> HART, 1961: 124, para quien, por ejemplo, «[e]n todos los campos de la experiencia, no solo en lo de las reglas, hay un límite, inherente en la naturaleza del lenguaje, para la guía que el lenguaje general puede generar».

<sup>25</sup> TARELLO, 1980: 115, llega a hablar de «atribución» de sentidos por parte del intérprete: «[l]a atribución de significado a un texto escrito que se asume expresivo de normas jurídicas, consiste en una actividad de sujetos que solo ocasionalmente —y en el mundo moderno muy raramente— coinciden con los sujetos que han querido y redactado el texto; ello quiere decir que la atribución de significado a un documento legislativo es una operación que no tiene en cuenta necesariamente a quien ha querido y redactado el texto, es una operación en que los elementos necesarios son solo el texto y el sujeto que a él atribuye un significado».

<sup>26</sup> Véase, por ejemplo, CHIASSONI, 2007; GUASTINI, 2011. En Brasil, GRAU, 2002; ÁVILA, 2003; y entre la doctrina procesal, MITIDIERO, 2016: 65.

Asimismo, en la teoría de la fotografía, que antes se describía como el «infalible lápiz solar» (ROOT, 1864: 28), se empezó a subrayar la polisemia de las imágenes<sup>27</sup>, la importancia fundamental de los contextos<sup>28</sup> para su interpretación, etc. De modo contundente, se dijo que toda imagen tiene «subyacente a sus significantes, una “corriente flotante” de significados, de modo que el lector puede elegir algunos e ignorar a otros»<sup>29</sup>. Además, se pasó a demostrar cómo es que, lejos de ser una «reproducción mecánica», los aspectos técnicos y las elecciones del fotógrafo, por ejemplo, influyen en el resultado final y, por ende, en la representación e interpretación. La fotografía (y también los vídeos, etc.) dejan de verse como un proceso automático y se pasa a comprender la *expertise*, la habilidad y los valores artísticos del fotógrafo<sup>30</sup>. Una misma escena puede ser representada de formas muy distintas por diferentes fotógrafos, con diferentes técnicas y dispositivos, que cambian no solo la representación, sino también las posibles interpretaciones.

Por consiguiente, el desarrollo de las nuevas tecnologías y la percepción de la necesidad de procesos de interpretación en cualquier tipo de representación<sup>31</sup> —ya sea de textos, imágenes, vídeos, etc.—, significaron cambios radicales, que deberían haber determinado una revolución copernicana en el estudio de los documentos. No obstante, como intentaré demostrar en el próximo epígrafe, algunas de las ideas de la doctrina tradicional se mantuvieron, explícita o implícitamente, presentes y pasaron por diversas adaptaciones, que desafortunadamente hicieron que cada vez fueran más confusas.

## 2.2. Cambian los vientos, se mantienen las ideas

Pensando en documentos en papel y escritos (que, para CARNELUTTI, eran sinónimos), el pensamiento sobre la prueba documental giraba, en la doctrina procesal clásica, alrededor de saber, sobre todo, si el documento era público o privado. Por un lado, los documentos públicos, que llegaban a denominarse documentos auténticos<sup>32</sup>, provenían de un «funcionario público autorizado para dar fe pública»<sup>33</sup>. Por otro, los privados, no tendrían «por sí mismos fe plena»<sup>34</sup>, salvo mediante reconocimiento de la persona que los hubiera elaborado<sup>35</sup>. No es casualidad que tuviera

<sup>27</sup> BARTHES, 1964: 44.

<sup>28</sup> SONTAG, 1973: 82.

<sup>29</sup> BARTHES, 1964: 44. En el mismo sentido, SONTAG, 1973: 17 y SILBEX, 2004: 519.

<sup>30</sup> SONTAG, 1973: 8. MNOOKIN, 1998: 23; TAGG, 2005: 121-124.

<sup>31</sup> En los epígrafes siguientes, abordaré el significado de los *signos* y los procesos de *representación*.

<sup>32</sup> NEVES E CASTRO, 1917: 227 y 240.

<sup>33</sup> CHIOVENDA, 1923: 846.

<sup>34</sup> CHIOVENDA, 1923: 846.

<sup>35</sup> CHIOVENDA, 1923: 846.

mucha importancia la *sottoscrizione*<sup>36</sup>, es decir, saber si el documento estaba firmado o no por la persona que lo había elaborado y quien era su autor.

En 1936, cuando la cámara Leica I cumplía once años, CARNELUTTI hablaba de *documentos directos* para referirse, por ejemplo, a la fotografía, y afirmaba que «también para el empleo de estos documentos sirve la analogía de las disposiciones que la ley dicta para la escritura»<sup>37</sup>. Así, la estructura de los escritos doctrinales sobre la prueba documental seguía siendo, más o menos, la misma desde los tiempos de la doctrina clásica: *escritos* públicos y privados, firmas, autenticidad, etc. Se añadió un capítulo posterior<sup>38</sup>, acerca de las llamadas «reproducciones mecánicas». Se trataba de una forma de *cambiar sin cambiar*; hablar de las (en aquel momento) «nuevas» tecnologías sin replantearse la teoría general tradicional, que había sido pensada para los documentos en papel y escritos. La idea, sin embargo, no tiene ningún sentido, pues el hecho de que un documento contenga informaciones escritas se refiere al tipo de signo utilizado; el hecho de que el documento sea privado o público se refiere a la autoría, mientras que el hecho de que sea una «reproducción» mecánica se refiere al modo de generarse el documento. Por tanto, es evidente que se trata de planos distintos.

Aún hoy en día es común la comparación entre los documentos escritos y las «reproducciones mecánicas». Se dice, por ejemplo, que «el documento *stricto sensu* es el escrito»<sup>39</sup>; que los documentos escritos son «frutos de mimesis mecánicas»<sup>40</sup>; que los documentos pueden ser del tipo «escrito, [...] fonográfico, fotográfico o cinematográfico»<sup>41</sup>; que el Código de Proceso Penal italiano «igual a [parifica] los escritos a los otros documentos representativos de hechos»<sup>42</sup>, etc.

Una segunda «adaptación» se produjo, a partir de 1990, con lo que parte de la doctrina acordó en llamar «documentos informáticos»<sup>43</sup>. De nuevo, la doctrina añadió una nueva categoría, para, supuestamente, explicar las novedades tecnológicas, sin repensar la teoría general una vez más. En este caso, la diferencia radicaría en que el documento informático no tendría un «soporte *cartaceo*»<sup>44</sup>. Se llegó incluso a

<sup>36</sup> CARNELUTTI, 1936: 693; CHIOVENDA, 1923: 846.

<sup>37</sup> CARNELUTTI, 1936: 694.

<sup>38</sup> En este sentido, véase, por ejemplo, GUIDI, 1950: 56-59 (hablando de «documentos gráficos»); DENTI, 1957:74; MILHÔMENS,1982: 406; COMOGLIO, 2006: 419 y ROTA, 2012: 706 (en los que las reproducciones mecánicas están en un título sobre «figure particolari di prova documentale», en el primer caso, y en lo que llaman «gli altri documenti», en el segundo; en ambos casos después de un epígrafe dedicado a los documentos públicos y otro a los documentos privados); BONOMI, 2011: 177; MARINONI Y ARENHART, 2015: 717.

<sup>39</sup> ZANCHÈ, 2012: 7-8.

<sup>40</sup> ZANCHÈ, 2012: 40.

<sup>41</sup> CANTONE, 2004: 117 y 9.

<sup>42</sup> CARDINO, 2004: 16.

<sup>43</sup> En este sentido, véase, por ejemplo, TARUFFO, 2000: 98; COMOGLIO, 2006: 529; LUPANO, 2011: 269; ROTA, 2012: 728.

<sup>44</sup> CARDINO, 2004: 23.

hablar de «representación informática»<sup>45</sup> como si las palabras (u otros signos) usadas en un ordenador fueran distintas de las usadas en el papel<sup>46</sup>. Y, aún con mayor contundencia, más recientemente, se llega hasta a decir que el documento electrónico configuraría un «*tertium genus* con relación a la forma escrita y oral»<sup>47</sup>.

De nuevo, en vez de replantearse la teoría general, se añadió un nuevo capítulo. Sin embargo, en la idea de que puedan existir documentos escritos, reproducciones mecánicas y documentos electrónicos, a mi juicio, se mezclan distintos planos, pues la escritura se refiere al tipo de signos; la «reproducción» mecánica, a la forma de creación; y el hecho de ser electrónico, a su conservación.

Todo este contexto, aunque resumido, sirve para argumentar que algunas de las características señaladas por la teoría clásica —en un primer momento, con relación a los documentos escritos y en papel y en un contexto en el que la interpretación era un acto de *descubrimiento*— siguieron el camino de las teorías sobre la prueba documental. De la misma manera, de forma más explícita o implícita, más o menos contundente, las ideas de la doctrina clásica siguen presentes, aunque las premisas ya no sean las mismas.

### 2.2.1. ¿Representación inmediata y objetiva?

La representación, para CARNELUTTI, era un «subrogado de la percepción»<sup>48</sup>, un «equivalente sensible», que tendría condiciones de «despertar [*risvegliare*]» la idea que determinaría la percepción de un hecho<sup>49</sup>. Una primera característica que, en la doctrina clásica, diferenciaba al documento de otros medios de prueba fue su supuesta capacidad de *representación inmediata*<sup>50</sup>, en la medida que «la individualidad del hecho que se representa se traduce inmediatamente en un objeto exterior»<sup>51</sup>. Existirían, entonces, documentos *directos* (como una fotografía) y documentos *indirectos* (como un dibujo)<sup>52</sup>, de acuerdo con la «mayor o menor inmediatez»<sup>53</sup>. Para CARNELUTTI, la representación sería *mediata* cuando el hecho primero se fijara en la memoria y después lo «reprodujera» el testigo<sup>54</sup>.

<sup>45</sup> CARDINO, 2004: 23; GRAZIOSI, 1998: 481. Sobre documentos informáticos, también BONOMI, 2011: 188.

<sup>46</sup> En sentido contrario, defendiendo que la «scrittura rimane scrittura, anche se mutano e si rinnovano i supporti tecnici con i quali essa viene confezionata e comunicata o riprodotta», véase TARUFFO, 2000: 95.

<sup>47</sup> SOUSA, 2017: 56.

<sup>48</sup> CARNELUTTI, 1947: 122.

<sup>49</sup> CARNELUTTI, 1947: 122.

<sup>50</sup> CARNELUTTI, 1947: 140; GUIDI, 1950: 19.

<sup>51</sup> CARNELUTTI, 1947: 140.

<sup>52</sup> CARNELUTTI, 1947: 110; GUIDI, 1950: 58-59.

<sup>53</sup> CARNELUTTI, 1947: 110

<sup>54</sup> CARNELUTTI, 1947: 140.

La idea de documentos directos e indirectos se sigue aceptando hasta hoy en día, por ejemplo, cuando se dice que «de un lado [está] la fotografía (documento directo), de otro la escritura (documento indirecto)»<sup>55</sup>; que «el documento indirecto requiere siempre una aportación humana [...]. Por otro lado, el documento directo se forma automáticamente»; que la «representación [...] puede darse indirectamente, a través de la narración del autor del documento, o directamente, mediante la observación directa del propio documento»<sup>56</sup>; que el «original es el documento [...] que representa directa e inmediatamente las declaraciones de voluntad»<sup>57</sup>; que las llamadas reproducciones mecánicas «son capaces de reproducir, directa e inmediatamente, el hecho objeto de prueba»<sup>58</sup>.

Esa es, también, la idea de la doctrina y las legislaciones que adoptan la idea de «prueba plena» para algunos tipos de documentos o para documentos cuando se dan determinados requisitos<sup>59</sup>. Por ejemplo, en el caso de los documentos privados, de acuerdo con el art. 2702 del CC italiano y del art. 408 del CPC brasileño, estos harían «prueba plena» de las *declaraciones* de quienes los hubieran firmado. De nuevo aquí se está imaginando que las declaraciones, por sí solas, permiten un acceso directo a su contenido, que tiene un único sentido que podría descubrir quien debe juzgar (y, en realidad, considerar algunos hechos probados)<sup>60</sup>.

En primer lugar, es necesario señalar que no hay representaciones *directas* ni *inmediatas*. Hablar de *representación* es distinto de hablar de *presentación*, es decir, cuándo un objeto es solo un objeto. Puedo *presentar* un perro, pero el mismo perro, en principio, nada *representa*. Los documentos utilizan *signos*. Y justamente que el objeto esté «por otro [*stands for another*] para alguna mente» (PEIRCE, 1873: 313), puede considerarse un *signo*. Los signos, por ende, justamente señalan algún «contenido fuera del signo» (HJELMSLEV, 1969: 45), que no es algo unívoco.

En segundo lugar, la dicotomía entre documentos directos e indirectos oculta un aspecto muy importante: lo que la doctrina considera relevante para saber si la representación es *directa* o *indirecta* es saber si los hechos «pasan» o no por la mente humana. Es decir, el documento directo permitiría un acceso directo a los hechos y el indirecto, no; pero este carácter indirecto no se refiere a los signos, sino a la idea de pasar por la mente humana. En ambos casos, por ende, el proceso de interpretación de los signos queda oculto o directamente se considera (implícita o explícitamente) que no existe. Cuando una jueza mira una fotografía, por la supuesta condición de «reproducción mecánica» de esta «prueba directa», tendría un acceso directo a los hechos. En el caso de una declaración por escrito, no, pues los hechos se habrían almacenado en la memoria de quien escribe y solo después, de forma mediata, en el

<sup>55</sup> CANTONE, 2004: 5.

<sup>56</sup> BERTOLINO, 2011: 7-8.

<sup>57</sup> DI IOIRIO, 2011: 132.

<sup>58</sup> MARINONI y ARENHART, 2015: 717.

<sup>59</sup> A este tema me referiré en el epígrafe 3.3, *infra*.

<sup>60</sup> Volveré sobre la cuestión de la prueba plena en el epígrafe específico a la valoración, 3.3, *infra*.

papel. En ambos casos, sin embargo, «indirecto» sería el camino de los hechos hasta el documento; el camino del documento a la jueza parecería, en ambos los casos (de forma errónea), como directo.

Todos estos problemas quedan claros, en mi opinión, cuando se ve que, en la mayor parte de los escritos sobre documentos de las últimas décadas, es muy raro encontrar algún comentario<sup>61</sup>, aunque sea al margen, sobre la interpretación de los signos o la valoración de los documentos<sup>62</sup>. Por otro lado, en este sentido de imaginar una equivalencia sensible *inmediata*, directa, ya se decía en el pasado, por ejemplo, que la prueba documental permitiría una «deducción» (CARNELUTTI, 1947: 110); más recientemente, que un documento contendría una «representación real u objetiva» (REICHEL, 2009: 254); que el documento podría «hacer que de nuevo estuvieran presentes hechos naturales» (ZANCHÈ, 2012: 9); que «el documento [...] es todo lo que represente objetivamente un hecho»<sup>63</sup> (DIDIER JR. y BRAGA, 2014: 3); que el documento sería capaz de «evocar en juicio un fragmento de realidad»<sup>64</sup> (CARDINO, 2004: 60); o que «el documento es capaz, por sí solo, de representar el hecho» (MARINONI, ARENHART Y MITIDIERO, 2015: 359). Todo ello brinda una idea de *objetividad* del documento que considero que no es correcta<sup>65</sup>.

Es importante entender, además, que no hay una única forma de representar un hecho<sup>66</sup> y que no necesariamente la representación tiene algún tipo de equivalencia con la percepción. Un átomo puede ser representado por la palabra «átomo», por una fotografía hecha con una cámara de alta tecnología o mediante un dibujo. Y una persona como yo, que nunca ha visto un átomo, puede asimismo imaginar que la representación que en el colegio aprendemos para identificar un átomo no tiene «equivalencia sensible» con un átomo real. Del mismo modo, una fotografía de un átomo jamás podría ser un «equivalente sensible» de la percepción, pues nuestros ojos no tienen capacidad de ver algo tan pequeño; es decir, la mayor parte de las personas jamás ha visto un átomo.

La semiótica es el campo del conocimiento que estudia la teoría general de los signos, y señala que, pese a que estos sean «transportadores de sentidos» (HJELMSLEV,

<sup>61</sup> Es excepción, por ejemplo, NIEVA FENOLL, 2010: 316. Hablando sobre que los documentos reproducen solo una parte y también es necesario respecto a los «instrumentos mecánicos (...) en algún modo creer», ZANCHÈ, 2012: 25.

<sup>62</sup> Véase, por ejemplo, CANTONE, 2004; RONCO, 2011; ZANCHÈ, 2012; ROTA, 2012 y SOUSA, 2017

<sup>63</sup> DIDIER JR. y BRAGA, 2014: 3.

<sup>64</sup> CARDINO, 2004: 60.

<sup>65</sup> La *Corte di Cassazione* italiana, solo para poner otro ejemplo, ya decidió que los documentos «figurativos» y directos ofrecen una «descripción inmediata de los hechos». *Corte di Cassazione* Sez. V, 18 ottobre 1993, n. 10309, Fumero, rv. 195556. Más recientemente, el mismo tribunal reafirmó su jurisprudencia, que mantiene, al menos, desde los años 1990, en el sentido de que «la fotografía constituye prueba preconstituida de su conformidad con las cosas y los lugares representados». *Corte di Cassazione*, Civile, Rel. Ragonesi Vittorio, n. 308/2020.

<sup>66</sup> TARUFFO, 1992: 439 hablaba de una equivalencia que podría ser solo parcial, pero que también podría ser suficiente para conocer el hecho.

1969: 44), todo el proceso de significación es polisémico y depende de unos procesos complejos de *interpretación*. Es cierto que, en algunos casos, la interpretación será más fácil —como cuando veo escrita la palabra «agua» en un vaso lleno de un líquido incoloro—; no obstante, para alguien que no hable castellano no le será posible interpretar ni siquiera una palabra tan sencilla como esa.

El proceso de interpretación, además, no tiene una única dirección o resultado necesario, que permitiera una «equivalencia», menos aún directa. A una persona que lea la palabra «perro» le puede venir a la cabeza un pitbull; a otra, un caniche. Las dos personas estarán en lo cierto, pues el pitbull y el caniche son perros. ¿Pero cuál sería, entonces, la equivalencia sensible? La idea de «despertar» una idea, por tanto, me parece que ofrece poca información y, en todo caso, es insuficiente para explicar la complejidad de la representación. Otro factor que introduce complejidad en la interpretación es que los signos —sean los de una fotografía, un dibujo o un texto— pueden tener sentidos *denotados* y *connotados*. Los primeros son los sentidos literales, concretos, mientras que los segundos, son los sentidos figurados, no literales. El dibujo de un cubo de hielo tiene el sentido *denotado* —sentido literal, concreto— de un cubo de hielo: agua congelada. También puede tener el sentido *connotado* —figurado, no literal— de frío o, en otro ejemplo, de una persona sin emociones<sup>67</sup>. La clave estará en el contexto. Un vídeo puede contener imágenes de una persona recibiendo un disparo y, si el vídeo fue grabado por una cámara de seguridad, la respuesta será llamar a la policía; no obstante, si la misma escena aparece en una película de Netflix, ya no será necesario avisar a las autoridades.

Por último, no todos los documentos funcionan del mismo modo, algo que afecta, también, a la representación y sus contenidos: ni los documentos «en papel», ni las llamadas «reproducciones mecánicas», ni los referidos «documentos informáticos», que, como hemos visto, son categorías poco informativas. Para entender la representación de un documento y analizar su valor probatorio es necesario ir más allá del papel (o de cualquiera que sea el soporte) y entender la forma de inclusión, el funcionamiento y la conservación de los signos.

Estos son solo algunos ejemplos que, a mi juicio, demuestran que la idea de documentos directos o bien de representación documental como equivalencia sensible inmediata, lo que hace, en realidad, es ensombrecer la necesidad y la importancia de llevar a cabo procesos interpretativos sobre los documentos. Y de esa forma, lo único que sucede es que las interpretaciones de nuestros jueces y de nuestras juezas se mantienen ocultas y no se pueden controlar intersubjetivamente.

---

<sup>67</sup> Imagínese que, en una conversación de *WhatsApp*, una persona pregunta: «¿Cómo ha reaccionado tu novio(a) a la noticia?» y que la respuesta sea el dibujo de una piedra de hielo.

### 2.3. Algunas propuestas iniciales

La doctrina procesalista más moderna sostiene que el documento es «una cosa cualquiera (escrita o figurativa) que —como *vox mortua*, traducida de forma duradera en signos estables, [...]— sirve para describir, representar y hacer cognoscible un hecho ocurrido en la realidad o incluso para reproducir una determinada manifestación del pensamiento y la voluntad» (COMOGLIO, 2010: 427)<sup>68</sup>.

La primera cautela que debemos tener es que decir que el documento «sirve para», que tiene el objetivo de *hacer cognoscible un hecho* no significa, no puede confundirse con el hecho de que en efecto así lo haga. Un texto escrito confuso no permitirá que se conozca ni se infiera nada de los hechos, pero seguirá siendo un documento. Otro aspecto es que no veo la necesidad de que los signos sean estables o duraderos, incluso porque no está claro a partir de que momento podría considerarse que los signos son duraderos. Un documento cuyos signos duren cinco segundos será, en mi opinión, durante esos cinco segundos, un documento. Las cámaras de seguridad que graben vídeos que se borren en 24 horas, seguirán generando documentos. Documentos cuyos signos se actualicen como forma ordinaria de funcionamiento también seguirán siendo documentos.

Por tanto, lo que me parece claro es que una definición de documento que pretenda abarcar todos los tipos de documentos actuales debe, necesariamente, contar con un grado alto de generalidad. En este sentido, puede decirse que el documento es un *objeto, corpóreo o no, que contiene signos añadidos por humanos o máquinas*. Y, una vez criticadas las premisas de la doctrina tradicional, ¿cuáles serían, entonces, los factores relevantes a tener en cuenta para mejorar el trabajo con documentos?

*En primer lugar*, los tipos de signos presentes en el documento, que pueden ser esencialmente tres: índices (*indexes*), íconos o símbolos<sup>69</sup>.

Los índices tienen una relación causal automática con la realidad y cambian de acuerdo con ella. Es el caso de una radiografía, que cambia según el hueso esté o no fracturado. Del mismo modo, la fotografía de una persona sonriendo es diferente de otra tomada cuando la persona no sonrío.

Los íconos son aquellos signos que tienen alguna relación de semejanza con la idea u objeto y no cambian automáticamente con la realidad. Por ejemplo, un dibujo o una representación gráfica tridimensional. Una persona que hace un dibujo o un muñeco en 3D de otra puede elegir dibujarle sonriendo, aunque en el momento en

---

<sup>68</sup> En el mismo sentido, PROTO PISANI, 2006: 494. TARUFFO, 1992: 439 subraya que el equivalente al hecho podría ser un equivalente *parcial*, ya que un hecho no podría reproducirse de forma integral; en todo caso, podría ser suficiente para conocer el hecho.

<sup>69</sup> PEIRCE, 1901: 519.

el que hizo el dibujo estuviera seria<sup>70</sup>. El dibujo o la simulación serán «buenos» si guardan una semejanza considerable con la persona dibujada.

Por último, los símbolos tienen un significado arbitrario, que se atribuye de acuerdo con hábitos colectivos. Es el caso de las palabras, las frases, los textos, como el de una señal de *prohibida la entrada*<sup>71</sup>: en España, este símbolo se representa mediante un círculo rojo con un rectángulo blanco horizontal en el centro. En Brasil, donde hay otros hábitos colectivos, se representa mediante una flecha negra cortada por una línea roja diagonal.

Un documento podrá contener, por ende, distintos tipos de signos. Índices, como fotografías, imágenes, radiografías; íconos, como el retrato de la Mona Lisa, un esquema 3D, un dibujo, un croquis, etc; y símbolos, como palabras, frases o textos. Un mismo documento podrá incluso contener muchos tipos de signos, por lo que, para realizar una interpretación adecuada, será imprescindible entender cómo fueron generados o añadidos.

Si se parte de esta idea se superan las confusiones tradicionales, que harían que un correo electrónico que tuviera texto y una fotografía fuese, al mismo tiempo, un escrito privado, una reproducción mecánica y un documento informático. Pasaríamos a decir que se trata de un documento que contiene símbolos e índices.

*En segundo lugar*, es necesario entender si el documento está sujeto a cambios *fisiológicos* o *patológicos*. Un historial médico funciona todo el tiempo con cambios que introducen los médicos y médicas en los signos del documento, circunstancia que, al ser parte del funcionamiento natural de este tipo de documento, no será un problema ni una patología. Por otro lado, si el historial médico permite que un anestesista acusado de mala praxis cambie «70 mg» por «7 mg», este será un cambio *patológico* —que sería aún más grave si el *software* no permitiera que quedara constancia de los cambios—.

Así, por ejemplo, ante índices, será imprescindible entender los límites naturales de lo que generó el signo: ¿cuál era el alcance del micrófono? ¿Qué frecuencia de sonido tiene? ¿Dónde estaba ubicada la cámara de seguridad? ¿Cuál es la definición de las imágenes? Con los íconos, por ejemplo, será importante entender: ¿qué datos se utilizaron para hacer la simulación 3D? ¿Cómo se eligieron las proporciones? ¿Cómo se pueden confirmar esos datos? Y, por último, en el caso de símbolos, como los textos: ¿cómo se puede conocer la colectividad a la que esa comunicación pertenece? ¿A brasileños? ¿A brasileños que viven en la ciudad de Porto Alegre? ¿A brasileños que viven en la ciudad de Porto Alegre y son abogados? ¿A brasileños que viven en la ciudad de Porto Alegre, son abogados y trabajan en una misma oficina? ¿Qué significan estas palabras? ¿Qué significan las frases usadas en el contexto de la comunicación? *So on and so forth...*

<sup>70</sup> Eco, 1968: 108, hablando de la Mona Lisa y la Sra. Giocondo.

<sup>71</sup> Eco, 1968: 108.

Por último, *en tercer lugar*, el origen del documento, en seres humanos o máquinas, también será importante para entender su funcionamiento. Por ejemplo, los documentos generados directamente por seres humanos estarán sujetos no solo a *errores honestos*, sino también a *mentiras*<sup>72</sup>. Y, de acuerdo con lo que sabemos, la percepción y la memoria humanas tienen múltiples límites, que pueden determinar que se introduzcan en los documentos, de modo genuinamente *sincero*, signos que no se correspondan con la realidad. O puede darse el caso de documentos creados de formas *mentirosas*, con intención de engañar.

A su vez, las máquinas no mienten, pero han de estar calibradas e, incluso cuando lo están, tienen límites, tasas de error, puntos ciegos, sesgos debidos a la programación, etc. Algunas máquinas no hacen razonamientos propios, como, por ejemplo, un termómetro<sup>73</sup>, que genera un documento sobre la temperatura medida. Otras máquinas utilizan algoritmos, cálculos, etc. para llegar a una conclusión<sup>74</sup>. En ambos casos, será imprescindible entender como se llevaron a cabo los razonamientos, para comprender los documentos generados.

Por todo ello, en resumen, analizar un documento significa entender su «funcionamiento»; sin examinar el razonamiento que subyace a su creación y funcionamiento, dará lugar a un análisis incompleto, cuyos resultados muchas veces serán desastrosos y equivocados. Si los documentos contienen *signos*, que están por otras cosas (representan), es imprescindible conocer o investigar las realidades que representan e intentar conocer en la mayor medida posible el «mundo de fuera».

En el siguiente apartado, intentaré esbozar algunas formas de trabajar con documentos más actuales, respetando y considerando las nuevas realidades y, sobre todo, el razonamiento probatorio que está por detrás de los documentos.

### 3. ATRAVESANDO EL VIENTO. LOS DOCUMENTOS EN EL PROCEDIMIENTO PROBATORIO

Con la estructura pensada para los documentos escritos y en papel adaptada, durante décadas, a otros tipos de documentos, muchos de nuestros sistemas procesales de *civil law* acaban adoptando prácticas sobre la prueba documental que, hoy en día, parecen incompletas o, en ocasiones, directamente equivocadas.

El planteamiento de este epígrafe no es exhaustivo, pues solo pretendo presentar algunas ideas sobre la admisión, la práctica y la valoración de la prueba documental, teniendo en cuenta su relación con el razonamiento probatorio que está por detrás.

---

<sup>72</sup> DE PAULA RAMOS, 2019: 84, incluso con indicaciones bibliográficas específicas.

<sup>73</sup> LEHRER, 1995: 160; HARRÉ, 2010: 32.

<sup>74</sup> ROTH, 2017: 2019; VÁZQUEZ, 2015: 267.

### 3.1. La admisibilidad. Compartiendo responsabilidades epistémicas

Pensando en un documento *escrito* y en *papel*, la doctrina procesal tradicional desarrolló, en la primera mitad del siglo xx, las ideas de *falsedad material* e *inautenticidad*, factores que podrían determinar la inadmisibilidad de un documento<sup>75</sup>.

La falsedad material sería un cambio en el *elemento extrínseco del documento*<sup>76</sup>, por ejemplo, en una hoja o en una palabra de un contrato; en el caso de vídeos, una edición de las imágenes. En este sentido, en el pasado, se describía la falsedad como un problema en la «integridad física del papel escrito» (AMARAL, 1958: 43-44). Esta idea está presente, de forma similar hasta nuestros días<sup>77</sup>, por ejemplo, en los arts. 141 del CPC mexicano, 242 del CPC peruano, 346.3 del CPC chileno y 412 del CPC brasileño y, obviamente, ya no se aplica solo con relación al papel o a los llamados documentos escritos (*rectius*: documentos con símbolos).

Dicho tipo de falsedad no tendría relación directa con los hechos o informaciones del documento<sup>78</sup>, sino, más bien, con el llamado *soporte*. Un documento materialmente falso en el que se dice que Edson Arantes do Nascimento, Pelé, jugó el Mundial de 1970, seguiría siendo materialmente falso, pero la información (que Pelé jugó en el Mundial de 1970) sería, en cualquier caso, verdadera<sup>79</sup>. Del mismo modo, una fotografía editada seguiría siendo materialmente falsa, pero la información (por ejemplo, la existencia de un accidente de tráfico) también podría, en cualquier caso, ser verdadera.

Además de la *veracidad material*, la doctrina procesal defendió la necesidad de analizar la *autenticidad* del documento, que tradicionalmente se interpreta con relación casi directa con el hecho de conocer su *autoría*<sup>80</sup>. Esta idea está presente en muchos de nuestros códigos, como, por ejemplo, los arts. 412 del CPC brasileño, 138 del CPC mexicano, 247 del CPC peruano y 2703 del CC italiano<sup>81</sup>. Así, por ejemplo, los documentos públicos podrían considerarse pruebas de su *autenticidad*,

<sup>75</sup> Pese a que en las legislaciones no siempre esté clara la consecuencia de reconocer la falsedad o la inautenticidad.

<sup>76</sup> CHIOVENDA, 1923: 845.

<sup>77</sup> En la doctrina, véase, por ejemplo, BERTOLINO, 2011: 12; GRAZIOSI, 2000: 167; COMOGLIO, 2010: 442.

<sup>78</sup> MARINONI Y ARENHART, 2015: 743.

<sup>79</sup> En el CPC peruano, de hecho, en su art. 237, se prevé: «Son distintos el documento y su contenido. Puede subsistir este aunque el primero sea declarado nulo».

<sup>80</sup> ALVARO DE OLIVEIRA Y MITIDIERO, 2012: 115; PROTO PISANI, 2006: 425; MILHOMENS, 1982: 353; ROTA, 2012: 615-616; MARINONI, ARENHART Y MITIDIERO, 2015: 365. Se debatió, por ejemplo, la necesidad de tratar de la autoría en el caso de las llamadas «reproducciones mecánicas». En este sentido, véase, por ejemplo, CARDINO, 2004: 51 y ZANCHÈ, 2012: 39. Afirmando una utilidad *in re ipsa*, con independencia de quien formó las fotografías o vídeos, véase CANTONE, 2004: 122-123.

<sup>81</sup> NEVES E CASTRO, 1917: 254. CHIOVENDA, 1923: 843-844.

mientras que en el caso de documentos privados sería necesario reconocer la firma [*sottoscrizione*]<sup>82</sup>.

Considero que, sobre todo hoy en día, no pueden servir de parámetro de admisibilidad de un documento ni una cosa ni la otra.

Como ya he señalado, hay documentos en los que su propio funcionamiento normal presupone la existencia de cambios *fisiológicos*. Es decir, el documento será diferente el 12 de abril y el 14 de abril. Por ejemplo, los profesionales médicos *deben* alterar a diario un historial médico para reflejar los cambios fisiológicos, sin que ello disminuya en nada la calidad de su contenido. Del mismo modo, los dispositivos que registran la velocidad de los vehículos, los llamados tacógrafos, o los dispositivos de ultrasonido, funcionan justamente «cambiando» el documento continuamente.

Además, la idea de «paternidad» de un documento, muchas veces, se ve ensombrecida o resulta poco útil. Una doctora puede pedir a su ayudante que introduzca y actualice en el sistema del hospital el historial del paciente. El registro, pese a que no sea «auténtico» en el sentido tradicional, sin lugar a dudas podrá ser admitido como prueba de los hechos. Del mismo modo, un vídeo hecho por una persona desconocida puede ser relevante y admisible para probar unos hechos.

Con todo ello, no quiero decir, como es obvio, que esas dos cuestiones no sean importantes. Lo que sí defiendo es que tales factores son solo una parte del conocimiento del funcionamiento de los documentos; e, incluso, en algunos casos, una parte que puede no tener importancia. En mi opinión, lo que verdaderamente importa es entender el «funcionamiento» del documento. Es decir, los procesos de creación y conservación. Por ejemplo, para averiguar el funcionamiento, la existencia de cambios patológicos o fisiológicos, los tipos de signos, las formas de creación, los límites naturales, etc.

El problema es que, en la época de la doctrina procesal clásica, solo podía existir un original de una carta escrita en un papel. Al garantizar que el documento presentado era *el original*, y que había sido producido por el autor que se declaraba, solo faltaba averiguar si existía algún cambio o falsificación en el soporte<sup>83</sup>. Hoy, empero, las modernas tecnologías prácticamente «difuminaron la línea entre el original y la

---

<sup>82</sup> CHIOVENDA, 1923: 844. Más recientemente véase, por ejemplo, GRAZIOSI, 2000: 165; RUSSO, 2011: 87. Afirmando la necesidad de conocer la procedencia, una versión más actual de la *sottoscrizione*, CARDINO, 2004: 51. Hablando de la importancia del conocimiento de la «paternidad», ZANCHÈ, 2012: 39.

<sup>83</sup> McCORMICK *et al.*, 2020: § 232. En este sentido, es interesante la afirmación de que «*la stampa cartacea di un file memorizzato nell'hard disk di un computer rappresenterà pur sempre una copia dell'unico originale, costituito dall'hard disk stesso e della sequenza di magnetizzazione impressavi*» (CARDINO, 2004: 25), que me parece un intento de adaptar la lógica del papel al «no papel», en mi opinión, sin éxito. Al fin y al cabo, basta pensar en un «documento informático» compuesto por una fotografía hecha con un iPhone, un texto retirado de una página web y un texto mecanografiado por el propio autor. ¿Cuál sería, en este caso, el original?

copia» (LILLY, 1996: 615)<sup>84</sup>. En el caso de un documento manuscrito que se mecanografía y guarda como documento «.doc» y, más tarde, se convierte en documento «.pdf», que se envía por correo electrónico y, después, se imprime en papel, ¿cuál de los cinco sería el documento original? Mucho más importante que analizar el soporte, por ende, es analizar el documento en su complejidad, entendiendo, de hecho, su funcionamiento.

Ante esta cuestión, en los Estados Unidos, la llamada *best evidence rule* ha sufrido muchos cambios en los últimos años. Si antes le bastaba a la parte presentar el ejemplar original para demostrar la autenticidad del documento, ahora, el proceso de autenticación es mucho más complejo: no hay «un único y bien definido sentido» (ALLEN *et al*, 2011: 187), lo que da lugar a que los abogados den rienda suelta a su creatividad<sup>85</sup>. La idea, pese a la complejidad de su ejecución en algunos casos concretos, es sencilla y está entre las 901 *Federal Rules of Evidence*: sobre la parte que presenta la prueba recae la carga de aportar pruebas suficientes que fundamenten la decisión [*to support a finding*] de que el elemento de prueba es «lo que la parte afirma que es»<sup>86</sup>. La regla coloca, en general, casi toda la responsabilidad epistémica, en la parte que presenta el documento.

Por otro lado, los países de *civil law*, cuando una parte aporta un documento, sobre la parte contraria suele recaer la carga de alegar, en un determinado plazo, la inautenticidad o la falsedad (por ejemplo, 6 días, de acuerdo con el artículo 346.3 del CPC chileno, 15 días, según el art. 437.1 del CPC brasileño, etc.). No siempre queda claro cuál sería la consecuencia de no tachar el documento, pero, en general, puede entenderse que daría lugar a una presunción de autenticidad y veracidad<sup>87</sup>. En este caso, la regla hace recaer toda la responsabilidad epistémica en la parte contraria. En mi opinión, la regla estadounidense puede ser una buena indicación sobre qué debería ser la *autenticidad*, con el cuidado de no confundir conocer el *proceso de creación del documento* (y la forma de conservación de la información) con el «contenido» o el fondo de los hechos que se pretenden probar. Cuando una parte aporta, por ejemplo, un vídeo de una cirugía como prueba de la mala praxis del médico, a los efectos de conocer la autenticidad en este sentido más amplio que propongo, interesará saber si el vídeo, de hecho, fue grabado con una cámara automática durante la cirugía, si es posible apagar durante unos momentos ese tipo de cámaras, si se puede borrar información, si queda constancia de la edición del contenido del vídeo, etc. No obstante, en ese momento, no será relevante saber si, de hecho, el médico

<sup>84</sup> En sentido análogo, MURPHY Y GLOVER, 2015: 709

<sup>85</sup> LILLY, CAPRA Y SALTZBURG, 2019: 55.

<sup>86</sup> Sobre el tema, véase ALLEN *et al*, 2011: 187; MUELLER Y KIRKPATRICK, 2012: 1083; MURPHY Y GLOVER, 2015: 709; MCCORMICK *et al*, 2020: § 221.

<sup>87</sup> Lo que, a su vez, abriría la puerta a otras preguntas: ¿debe el juez, ante un documento que la parte contraria no haya tachado, si considera que existe la posibilidad de que haya sido falseado, no actuar de oficio?

ha cometido o no mala praxis, algo que se decidirá cuando se valore esta prueba y el conjunto probatorio.

Sin embargo, no me parece que la mejor alternativa sea hacer recaer la responsabilidad epistémica solo en una u otra parte; más bien, debería hacerse que ambas partes la compartan. Sobre la parte que aporta la prueba debería recaer la carga de demostrar el «funcionamiento» del documento, es decir, el proceso de creación y conservación de la información que contiene. Sobre la parte contraria, ante esta información y pruebas y teniendo acceso a todo lo necesario para la generación del documento (tema que trataré más a fondo en el próximo epígrafe), la de demostrar posibles incompatibilidades.

Así, por ejemplo, en el caso del vídeo antes mencionado, la parte que aporta el vídeo debería demostrar como fue creado, el tipo de cámara que se utilizó, la forma como se conservan los archivos (por ejemplo, con metadatos, etc.). La parte contraria, en esta situación, podría desarrollar su propia labor epistémica, evaluando la calidad de la información brindada e, incluso, solicitar a su propio perito que haga un análisis preventivo, a fin de que, siendo el caso, se concluya que la prueba aportada no presenta problemas. Además, en ese momento, el juez (o jueza) podría evaluar la fiabilidad de la prueba aportada e intentar saber si la información facilitada, de hecho, demuestra el «funcionamiento» del documento en cuestión.

El documento deberá inadmitirse solo cuando sea irrelevante. Por ejemplo, cuando se demuestra que las fotografías no son del lugar del accidente o los análisis médicos no tienen relación con lo que se esté debatiendo en el caso concreto. La autenticidad no es necesariamente una cuestión de *todo o nada*. Es posible, por ejemplo, que haya pruebas sobre el funcionamiento de un documento y de un cambio hecho de forma patológica. En este sentido, al ser posible reconocer este cambio por metadatos, por ejemplo, puede darse el caso de que, si se quita la información añadida mediante el cambio patológico, el documento sea auténtico en todo lo demás. En otras palabras, todo el contenido del documento podría considerarse auténtico si se elimina el cambio patológico realizado. No hace falta que sea una relación de *todo o nada*<sup>88</sup>.

De esta forma, al entender el «funcionamiento» del documento y compartir las responsabilidades epistémicas, la discusión sobre la *autenticidad* sería mucho más completa y actual y permitiría no solo resolver los problemas sobre la inadmisión de la prueba —cuando la irrelevancia quede probada—, sino también preparar las discusiones que se entablarán en las etapas siguientes del procedimiento.

---

<sup>88</sup> En Brasil y Chile, por ejemplo, la consecuencia de reconocer la inautenticidad o la falsedad no está clara en la ley. En Perú, el CPC habla de *ineficacia* (art. 242) del documento considerado falso; en México, el CPC también considera que el juez civil deberá decidir sobre el *valor* del documento tachado por la parte contraria (art. 141). Sea como fuere, en general, parece que se trata de una relación de *todo o nada*. O bien el documento es verdadero y auténtico, con lo que tendría valor probatorio, o es falso o no auténtico, con lo que perdería su valor probatorio. En mi opinión, la propuesta defendida en el texto aporta una solución más actual y con mejores posibles resultados.

### 3.2. Contradicción y contenido: el alcance de la contradicción sobre documentos

La doctrina clásica subrayaba que, además de la *falsedad material* y la *inautenticidad*, el documento podría ser *falso ideológicamente*. En este caso, el problema no sería una intervención en el «material», sino en su contenido<sup>89</sup>. Ya no se trataría de una cuestión de admisibilidad, sino de una cuestión de fondo, que debería resolverse en la sentencia. El ejemplo clásico es el de un funcionario público que declare cosas que no se correspondan con la verdad<sup>90</sup>, pero es posible imaginar también un vídeo hecho en estudio, que muestre un asesinato que jamás se cometió. El documento sería auténtico (en el sentido tradicional), pues lo emite el funcionario o el estudio; es decir, sería materialmente verdadero, pero sería usado como prueba, por ejemplo, de un delito que nunca existió.

En mi opinión, la idea de *falsedad ideológica* no parece útil y brinda de nuevo una visión equivocada de objetividad o un único sentido a la prueba documental que elimina indebidamente el trabajo de conocer el funcionamiento del documento y su interpretación. Y ello, en mi opinión se debe a que se parte de la idea tradicional de que el documento ya tendría en sí mismo toda la información necesaria para su interpretación, de modo que la contradicción respecto de los documentos podría ser algo meramente formal. Si una parte aporta un documento, la otra parte tiene un plazo para manifestarse sobre este y, con ello, sería suficiente.

Solo para poner un ejemplo, en este sentido, el art. 437.1 del CPC brasileño prevé la necesidad de que cuando una parte aporte documentos, la parte contraria tenga un plazo de 15 días<sup>91</sup>. Si el actor, en una demanda de daños contra una constructora, presenta fotografías del baño de su casa, la constructora tendrá, por tanto, 15 días para manifestarse. Pero en estos 15 días, no le permitirá ningún tipo de acceso a la residencia, por lo que la discusión sobre la fotografía se dará exclusivamente con relación al «papel» (o archivo).

Pasando a considerar el documento como algo que contiene signos, que puede tener distintos funcionamientos, orígenes, límites, etc., los significados se presentan ya no de forma meramente objetiva o automática, sino de una forma que depende de un complejo proceso de interpretación. Y, naturalmente, si una parte tiene acceso al baño, para tomar sus fotografías, y la otra no, la contradicción de la parte que tiene

<sup>89</sup> AMARAL, 1958: 46. Recientemente, PROTO PISANI, 2006: 425; MARINONI Y ARENHART, 2015: 731.

<sup>90</sup> CHIOVENDA, 1923: 845-846.

<sup>91</sup> La Corte Constitucional italiana, por ejemplo, en un interesante fallo que «*l'allegata difficoltà di esaminare le prove documentali di cui è chiesta l'ammissione*» puede ser «*agevolmente superata dalla concessione di un termine*» (C. Cost. Italiana, 6 de julio de 1994, n. 284, disponible en <https://www.giurcost.org/decisioni/1994/0284s-94.html>, consultado por última vez el 27/10/2021). Para un comentario sobre esta decisión, véase CANTONE, 2004: 137.

acceso será más amplia que la de la que no tiene acceso. De esta forma, un juez o jueza, en la fase de valoración, tendrá que interpretar y valorar un documento sin que una de las partes haya tenido condiciones de ejercer la contradicción en el sentido fuerte<sup>92</sup> —de modo que permita que las partes influyan en el rumbo del razonamiento—; algo que, sin lugar a dudas, vulnera el derecho a la contradicción.

Para que esto sea posible y que las partes estén realmente en condiciones de igualdad, ambas deben tener acceso no solo a los documentos aportados por la parte contraria, sino también a todo lo que es necesario para generarlos y conservarlos. Así, por ejemplo, si una de las partes aporta una fotografía de un edificio con intención de probar que la constructora causó daños porque lo entregó con patologías constructivas, el contradictorio solo podrá tener lugar, de forma efectiva, si la constructora también tuviera acceso al edificio. No solo para poder tomar sus propias fotografías del lugar, sino para poder denunciar ante el juez los problemas en la representación de la realidad que detecte, etc. —como sucedería si la fotografía hubiera sido tomada con un plano contrapicado, que produce una distorsión y hace que los objetos cercanos parezcan más grandes (por ejemplo, un agujero en la pared)—.

Del mismo modo, si una parte aporta un documento con símbolos generado por una máquina, el contradictorio deberá incluir el acceso a la máquina, para que la parte contraria también pueda hacer sus «test». En este contexto, por ejemplo, la parte «contra» quien se produjo un documento podrá demostrar un error en el algoritmo de la máquina o que, en un tipo determinado de hechos, la máquina tiene una programación sesgada. Si la parte solo presenta en juicio el «resultado final», el documento generado por la máquina (un examen de COVID «positivo», por ejemplo), la parte contraria no podrá realmente ejercer su derecho a al contradictorio, puesto que no tendrá acceso a ninguna información, con lo que no podrá contradecirle.

En resumen, a sabiendas de que el documento no da un «acceso directo» u objetivo a la realidad y que el documento contiene signos, que representan de distintos modos aspectos de la realidad, el principio del contradictorio debe ampliarse para ir más allá del documento.

### 3.3. La valoración de los documentos: la necesidad de interpretación y el razonamiento probatorio

Muchos de los sistemas procesales del *civil law* crearon sistemas de valoración de la prueba documental que incluyen jerarquías y reglas de prueba legal<sup>93</sup>. El Código Civil italiano, por ejemplo, en el art. 2700 prevé que el documento público hace «piena prova» (prueba plena) de los hechos certificados por el funcionario público, la

---

<sup>92</sup> Sobre la contradicción como derecho de influencia sobre las posibles decisiones del juez (o jueza), MITIDIERO, 2015: 66; ALVARO DE OLIVEIRA, 2009: 131-132.

<sup>93</sup> ROTA, 2012: 598.

misma previsión que el art. 215 del Código Civil brasileño (y el art. 202 del Código Federal de Procedimientos Civiles mexicano). La idea, de hecho, estaba ya presente en CHIOVENDA (1923: 844) y en CARNELUTTI (1936: 691) y sigue presente en la doctrina<sup>94</sup> hasta nuestros días.

En el caso de los documentos privados, harían «prueba plena» solo de las *declaraciones* de quien los hubieran firmado (art. 2702 del CC italiano, art. 408 del CPC brasileño)<sup>95</sup>. La razón, de acuerdo con la doctrina clásica, sería que los documentos privados no darían «por sí solos *fe plena*» (CHIOVENDA, 1923: 846), excepto cuando los reconoce la persona «contra» quien se hayan presentado<sup>96</sup>. La misma idea de prueba «plena» aparece, en ocasiones, en las llamadas «reproducciones mecánicas» (art. 2712 del CC italiano y art. 225 del CC brasileño).

Dejando de lado las confusas y ya criticadas categorías doctrinales y legales —de documentos públicos, privados, escritos, reproducciones mecánicas, etc.—, en primer lugar debe decirse que la idea de que algo pueda ser una «prueba plena» plantea muchos problemas. En este sentido, la doctrina ya intentó aclarar que la prueba plena de la que habla, por ejemplo, la norma italiana, sería la eficacia «de una prueba por sí sola suficiente para fundar la declaración [*accertamento*] del hecho» (COMOGLIO, FERRI Y TARUFFO, 2005: 481)<sup>97</sup>, algo que, sin embargo, no podría excluir «automáticamente la admisión de pruebas diferentes o contrarias, ni la eventualidad de que el juez realizara una valoración diferente» (COMOGLIO, FERRI Y TARUFFO, 2005: 481)<sup>98</sup>. Pero, entonces, ¿cuál sería la ventaja de este tipo de previsiones legales? En mi opinión, ninguna.

Como ya he tenido la oportunidad de mencionar en los epígrafes anteriores, es necesario analizar el funcionamiento de un documento para poder conocer su valor probatorio. Un documento escrito, creado a partir de métodos muy fiables y mantenido de forma muy segura, puede tener mucho más valor probatorio que otro; un documento privado con códigos complejos que garanticen la inalterabilidad de su contenido puede tener mucho más valor probatorio que un documento público producido sin cuidados; y una fotografía tomada con una cámara profesional puede tener mucho más valor probatorio que otra fotografía hecha con un iPhone. La idea de que el legislador deba atribuir los valores probatorios me parece una reliquia histórica, que, en la actualidad, carece de significado<sup>99</sup>.

En segundo lugar, suponiendo que fuera posible considerar un documento como «prueba plena», sería necesario saber qué hechos deberían considerarse «plenamente» probados. Y ello, en la práctica, es mucho más complejo de lo que parece, pues, como

<sup>94</sup> Véase, por ejemplo, PROTO PISANI, 2006: 425; BIORCI, 2011: 43; COMOGLIO, 2010: 419; ROTA, 2012: 598.

<sup>95</sup> MARINONI, ARENHART Y MITIDIERO, 2015: 378; SANTOS, 1972: 54.

<sup>96</sup> CHIOVENDA, 1923: 846; RUSSO, 2011: 104;

<sup>97</sup> Sobre las diferencias generales entre documentos públicos y privados, véase TARUFFO, 2014: 74.

<sup>98</sup> En el mismo sentido, MILHOMENS, 1982: 353; ROTA, 2012: 644.

<sup>99</sup> En el mismo sentido, NIEVA FENOLL, 2010: 312.

ya he demostrado, el documento no contiene una única posibilidad interpretativa o un «hecho» que esté guardado en el documento.

En el caso de un vídeo de un accidente de tráfico, por ejemplo, ¿qué hechos quedarían probados «plenamente»? ¿Juan mató a María? ¿Juan se saltó el semáforo en rojo? Distintos ángulos, o incluso otros vídeos de la misma escena, podrían generar percepciones e interpretaciones distintas; el tipo de dispositivos, las lentes, etc., también. Además, el vídeo, por mejor que sea, nunca podrá significar una representación total: muchas informaciones relevantes para la prueba de los hechos que componen el *thema probandum* no se contendrán en el vídeo, como, por ejemplo, la velocidad de los coches, si alguno de los conductores estaba o no bajo los efectos del alcohol, etc. Y pese a no estar presentes en el vídeo, poniendo en jaque la idea de «plenitud», son extremadamente relevantes, por ejemplo, para definir la responsabilidad civil de una u otra persona.

Lo mismo se puede decir de los documentos con símbolos (públicos o privados). En un acta notarial, en la que el notario describe un diálogo entre dos personas, ¿qué hechos quedarían probados «plenamente»? Se podría pensar en la existencia de la conversación, en cada palabra de la conversación, en la presencia de determinadas personas y en muchas otras posibilidades. En cualquier caso, sin embargo, aunque la prueba pudiera considerarse «prueba plena», siempre sería posible que, con una grabación de audio, por ejemplo, una persona pretendiera demostrar la equivocación del notario, que habría entendido mal alguna palabra o frase.

La cuestión es que, como he mencionado en los epígrafes anteriores, no se puede acceder directamente al sentido de un documento, por lo que este ha de ser siempre objeto de interpretación. Y, además de la interpretación del documento, este deberá ser valorado, más tarde, en su conjunto, es decir, con las demás pruebas, en el momento de la valoración.

En cualquier caso, el valor probatorio de un documento dependerá siempre de muchos factores, como, por ejemplo, la calidad de la información, la calidad de la conservación de los signos, la forma de creación, las posibilidades de efectuar cambios fisiológicos y patológicos, los límites naturales de ese tipo de creación de documentos. Y un solo documento puede aportar un peso mucho menor cuando se analiza de forma aislada que cuando se considera dentro de un conjunto probatorio, en el que se combinan las pruebas<sup>100</sup>.

Por tanto, la valoración de la prueba documental debería efectuarse no con reglas de prueba legal o tasadas, sino mediante una valoración racional<sup>101</sup>. Después de que las partes hayan aportado toda la información del documento, en la fase de admisión, considerando los requisitos de autenticación propuestos y tras la etapa de contradictorio, de un efectivo diálogo entre las partes, que considere no solo el documento,

<sup>100</sup> Sobre el peso de la prueba combinada, véase HAACK, 2014: 208-238.

<sup>101</sup> Sobre el tema véase FERRER BELTRÁN, 2007; TARUFFO, 1992: 395.

sino también todo lo necesario para su formación (lugar, instrumentos, sustancias, algoritmos, etc.). Cuando llegue la fase de valoración, en consecuencia, la decisión tendrá mejores condiciones de considerar la prueba documental más allá de los signos e interpretar los documentos considerando el razonamiento que subyace.

#### 4. CONCLUSIONES

A lo largo del presente artículo, he presentado críticas y propuestas para trabajar con documentos considerando la evolución de las nuevas tecnologías y ámbitos, como la semiótica y la teoría de la fotografía. De todo lo expuesto, se pueden extraer las siguientes conclusiones:

1) Muchas de las teorías que sirvieron como base para la estructuración de la forma de trabajar con documentos en los ordenamientos de *civil law* parten de las ideas de la doctrina procesal italiana de primera mitad del siglo XX, sobre todo, CHIOVENDA y CARNELUTTI.

2) Dicha doctrina se gestó en un momento en el que el documento era sinónimo de ser en *papel y escrito* y en el que gran parte de las tecnologías que hoy utilizamos no existían.

3) A partir de la segunda mitad del siglo XX, muchos estudios demostraron que los signos no son objetivos, ya que todo proceso de significación depende de la interpretación.

4) En el caso de los documentos jurídicos, por ende, considerar una supuesta objetividad o representación inmediata de un documento sobre los hechos, acaba ocultando los procesos interpretativos, lo que impide su control intersubjetivo.

5) Partiendo de factores como el tipo de signo, el modo de formación y la manera de conservación de los documentos, se puede crear una teoría sobre la prueba documental más actual, que considere los diferentes tipos de documentos que existen hoy en día.

6) Partiendo de estas nuevas premisas, debe replantearse la idea de autenticidad para dejar de analizar solo los cambios y la autoría y pasar a entender el documento en su complejidad. Además, las responsabilidades epistémicas no deben recaer solo en una u otra parte, sino que estas han de ser compartidas. Deben inadmitirse los documentos que se demuestren irrelevantes y permitir que los análisis no sean del tipo *todo o nada*.

7) Dado que el documento no es un equivalente de los hechos, sino que solo representa realidades en sus distintas formas, para que haya contradictorio es imprescindible que el sistema obligue a la parte que aporta el documento a (o tenga la carga de) dar acceso a la parte contraria a todo lo necesario para la formación y la conservación del documento (lugares, máquinas, objetos, etc.).

8) En la etapa de valoración, los sistemas deben abandonar las reglas de prueba legal, habida cuenta de que, tras el proceso de autenticación y contradictorio relativo al documento, el juez deberá valorar la prueba en su singularidad, considerando todos los elementos del documento aportado que sean relevantes para determinar su valor individual, así como su valor en el contexto del conjunto probatorio.

## BIBLIOGRAFÍA

- ALLEN, R. J.; KUHN, R. B.; SWIFT, E.; SCHWARTZ, D. S.; PARDO, M. S., 2011. *Evidence. Text, problems, and cases*. [1989]. 5.ª ed. Frederick: Kluwer.
- ALVARO DE OLIVEIRA, C. A.; MITIDIERO, D., 2012. *Curso de Processo Civil. Vol. 2. Processo de Conhecimento*. São Paulo: Atlas.
- ALVARO DE OLIVEIRA, C. A., 2009. *Do Formalismo no Processo Civil*. 3.ª ed. São Paulo: Saraiva.
- AMARAL, S., 1958. *Falsidade Documental*. São Paulo: RT.
- BARTHES, R., 1964. «La Rhétorique de l'image». En *Communications*, 4, pp. 32-51.
- BERTOLINO, G., 2011. «Documento e Rappresentazione della Realtà». En RONCO, A. (org.). *Il Documento nel Processo Civile*, pp. 1-32. Bologna: Zanichelli.
- BONOMI, E., 2011. «Le riproduzioni meccaniche ed i documenti che non riflettono una scrittura». En RONCO, A. (org.). *Il Documento nel Processo Civile*. Bologna: Zanichelli.
- CANTONE, R., 2004. *La Prova Documentale*, pp. 177-200. Milano: Giuffrè.
- CARDINO, A., 2004. «La nozione di documento nel Sistema Processuale Penale». En CARDINO, A.; GUIDA, R.; RANALDI, A. *Processo Penale e Prove Documentali*, p. 2-144. Padova: CEDAM.
- CARDINO, A., 2004a. «La valutazione della prova documentale». En CARDINO, A.; GUIDA, R.; RANALDI, A. *Processo Penale e Prove Documentali*, pp. 2-144. Padova: CEDAM.
- CARNELUTTI, F., 1936. *Sistema del Diritto Processuale Civile. Vol. I. Funzione e Composizione del Processo*. Padova: CEDAM.
- , 1947, *La Prova Civile*. 2.ª Ed. Roma: Ateneo.
- CHIOVENDA, G., 1923. *Principii di Diritto Processuale Civile*. 3.ª ed. Napoli: Jovene. (Reimpresión: 1965).
- COMOGLIO, L. P.; FERRI, C.; TARUFFO, M., 2005. *Lezioni sul Processo Civile. I. Il Processo Ordinario di Cognizione*. 4.ª Ed. Bologna: Il Mulino.
- COMOGLIO, L. P., 2010. *Le Prove Civili*. 3.ª Ed. Bologna: UTET Giuridica.
- COUTURE, E., 1942. *Fundamentos del Derecho Procesal Civil*. 4.ª ed. (póstuma). Montevideo-Buenos Aires: B de F. Imp. en 2010.
- CUNHA, L. C., 2006. «Ação Rescisória Fundada Em Documento Novo». En: *Revista de Processo*, vol. 134. São Paulo: RT, p. 7-29.
- DENTI, V., 1957. *La Verificazione delle Prove Documentali*. Torino: Editrice Torinese.
- DE PAULA RAMOS, V., 2019. *Prueba testifical*. Trad. Laura Criado Sánchez. Madrid: Marcial Pons.
- , 2021. *Prova Documental*. Salvador: JusPodivm, 2021.
- DI IORIO, R., 2011. «Le altre scritture private, le copie e il telefax». En RONCO, A. (org.). *Il Documento nel Processo Civile*, pp. 117-176. Bologna: Zanichelli.
- DI SABATO, D. «Scrittura e documento scritto». En: *Scrittura e Diritto. Quaderni della Rivista Trimestrale di Diritto e Procedura Civile*, n. 3, pp. 217-224. Milano: Giuffrè.
- DIDIER JR., F.; BRAGA, P. S.; OLIVEIRA, R. A., 2015. *Curso de Direito Processual Civil - v. 2*. 10.ª ed. Salvador: Editora Jus Podivm.
- DIDIER JR., F.; BRAGA, P. S., 2014. «Carta Psicografada Como Fonte De Prova No Processo Civil». En *Revista de Processo*, vol. 234, Agosto, pp. 33-61.
- ECO, U., 1968. *La Struttura Assente. Introduzione alla Ricerca Semiologica*. Milano: Ed. Bompiani.

- FERNANDES, A. S., 2007. «Prova e Sucedâneos da Prova no Processo Penal». En: *Revista Brasileira de Ciências Criminais*, vol. 66/2007, pp. 193-236, mayo-junio.
- GRAZIOSI, A., 1998: «Premesse ad una teoria probatoria del documento informatico». En *Rivista Trimestrale di Diritto e Procedura Civile*, n. 2.
- , 2000. «Parola detta, parola scritta e parola telemática, questioni in tema di prova e di provenienza». En: *Scrittura e Diritto. Quaderni della Rivista Trimestrale di Diritto e Procedura Civile*, n. 3, pp. 163-176. Milano: Giuffrè.
- GUASTINI, R., 2011. *Interpretare e Argomentare*. Milano: Giuffrè.
- GUIA DO ESTUDANTE, 2018. «O tempo e o vento». *Análise da Obra de Érico Veríssimo*. Disponible en: <https://guiadoestudante.abril.com.br/estudo/o-tempo-e-o-vento-analise-da-obra-de-erico-verissimo/>. Consultado por última vez el 25/10/2021, a las 10:56.
- GUIDI, P., 1950. *Teoria Giuridica del Documento*. Milano: Giuffrè.
- HARRÉ, R. «Equipment for an Experiment». En: *Spontaneous Generations: A Journal for the History and Philosophy of Science*, Vol. 4, No. 1 (2010), 30-38.
- HART, H. L., 1961. *The Concept of Law*. 1a. ed., 10a. Reimp. Oxford: Oxford, University Press.
- HJELMSLEV, L., 1969. *Prologomena to a Theory of Language*. Trad. Fracis Whitfield. Londres: The University of Wisconsin Press.
- KELSEN, H., 1960. *Teoria Pura do Direito*. 2.ª Ed. Cit. en la traducción de João Baptista Machado. São Paulo: Martins Fontes (2012).
- LILLY, G. C.; CAPRA, D. J.; SALTZBURG, S. A., 2019. *Principles of Evidence*. 8a. Ed. St. Paul: West Academic Publishing.
- MACCORMICK, N.; SUMMERS, R., 1991. «Interpretation and Justification». En *Id.* (orgs.). *Interpreting Statutes. A Comparative Study*. Surrey: Ashgate Publishing.
- LEHRER, K. 1995. «Knowledge and the Trustworthiness of Instruments». En: *The Monist*, 78, n.º 2: 156-170.
- LUPANO, M., 2011. «L'informatica e le nuove frontiere della prova documentale». En RONCO, A. (org.). *Il Documento nel Processo Civile*, p.p. 249-280. Bologna: Zanichelli.
- MARINONI, L. G.; ARENHART, S. C., 2015. *Prova e Convicção*. 3.ª Ed. São Paulo: RT, 2015.
- MARINONI, L. G.; ARENHART, S. C.; MITIDIERO, D., 2015. *Novo Curso de Processo Civil. Vol. 2. Tutela dos Direitos Mediante Procedimento Comum*. São Paulo: RT.
- MCCORMICK, C.; BROWN, K.; DIX, G.; IMWINKELRIED, E.; KAYE, D.; MOSTELLER, R.; SWIFT, E., 2020. *McCormick on Evidence*. 8.ª Ed. Eagen: Thomson Reuters, 2020, versión en e-book.
- MILHÔMENS, J., 1982. *A Prova no Processo*. Rio de Janeiro: Forense.
- MITIDIERO, D., 2015. *Colaboração no Processo Civil*. São Paulo: RT.
- MNOOKIN, J. L., 1998. «The Image of Truth: Photographic Evidence and the Power of Analogy». En: *10 Yale J. L. & Human.* 1, 17.
- MUELLER, C. B.; KIRKPATRICK, L. C., 2012. *Evidence*. 5a. ed. Frederick: Kluwer.
- NEVES E CASTRO, F. A., 1880. *Teoria das Provas e Suas Aplicações aos Atos Cíveis*. Reimp. 2000, a partir de la 2.ª Ed. Campinas: Servanda.
- NIEVA FENOLL, J., 2010. *La Valoración de la Prueba*. Madrid: Marcial Pons.
- PEIRCE, C., 1873. «On the nature of Signs». En HOOPES, J. (org.), 1991. *Peirce on Signs*. Chapel Hill y Londres: The University of Carolina Pressa.
- , 1901. «Sign». En HOOPES, J. (org.), 1991. *Peirce on Signs*. Chapel Hill y Londres: The University of Carolina Pressa, pp. 239-259.
- PROTO PISANI, A., 2006. *Lezioni di Diritto Processuale Civile*. [1999]. 5.ª ed. Napoli: Jovene.
- REICHELT, L. A., 2009. *A Prova no Direito Processual Civil*. Porto Alegre: Livraria do Advogado.
- ROOT, M. A., 1864. *The Camera and the Pencil. Or the Heliographic Art its theory and practice in all its various branches*. New York: Appleton and Co.
- ROTA, F. «I Documenti». En: TARUFFO, M. (org.), 2012. *La Prova nel Processo Civile*. Milano: Giuffrè.
- RUSO, M., 2011. «La Scrittura Privata in Senso Stretto». En RONCO, A. (org.). *Il Documento nel Processo Civile*, pp. 73-116. Bologna: Zanichelli.

- SANTOS, M. A., 1972. *Prova Judiciária no Cível e Comercial. Vol. IV. Prova Documental*. 4.<sup>a</sup> Ed. São Paulo: Max Limonad.
- SILBEY, J. M., 2004. «Judges as Film Critics: New Approaches to Filmic Evidence». En: *U. Mich. J.L. Reform*, 37, 493, 508 n. 62.
- SOSA, L. F., 2017. *O Valor Probatório do Documento Eletrónico no Processo Civil*. 2.<sup>a</sup> Ed. Coimbra: Almedina.
- SONTAG, S., 1973. *On Photography*. New York: Farrar, Straus & Giroux.
- TAGG, J., 2005 «In the Valley of the Blind». En: KELSEY, R.; STIMSON, B. *The Meaning of Photography*. Williamston: Clark Studies in the Visual Arts, 2005, pp. 118-129.
- TARELLO, G., 1980. *L'interpretazione della legge*. Milano: Giuffrè.
- TARUFFO, M., 1992. *La Prova dei Fatti Giuridici*. Milano: Giuffrè, 1992.
- , 2000. «Parola Scritta e Parola Informatica nel Processo Civile». En *Scrittura e Diritto. Quaderni della Rivista Trimestrale di Diritto e Procedura Civile*, n. 3, pp. 91-106. Milano: Giuffrè.
- , 2014. *A Prova*. Trad. João Gabriel Krás Couto. São Paulo: Marcial Pons.
- VÁZQUEZ, Carmen. *De la prueba científica a la prueba pericial*. Madrid: Marcial Pons, 2015.
- ZACCHÈ, F., 2012. *La Prova Documentale*. Milano: Giuffrè.



## LA DISTINCIÓN ENTRE ESCLARECIMIENTO PROBATORIO Y ESCLARECIMIENTO PREPARATORIO EN EL PROCESO CIVIL

Thomas Vogt Geisse

Profesor de Derecho Procesal de la Universidad de Chile  
thomasvogt@derecho.uchile.cl

**RESUMEN:** Este trabajo propone adoptar la distinción entre esclarecimiento probatorio y esclarecimiento preparatorio en la reflexión comparada sobre el proceso civil. Se trata de una distinción funcional de normas procesales dirigidas al esclarecimiento de los hechos. Permite identificar normas dirigidas a formar la decisión fáctica del juzgador (esclarecimiento probatorio) y distinguirlas de normas dirigidas a proporcionar a las partes el conocimiento suficiente para comprender la fuerza de sus posiciones y tomar decisiones informadas antes de avanzar a la etapa probatoria (esclarecimiento preparatorio). Se demostrará que la confusión de estas categorías podría llevar una comprensión defectuosa del esclarecimiento preliminar del pre-trial anglosajón. A modo de ilustración, se mostrará que la *Rule 102* de las Reglas Europeas de Proceso Civil ofrece un ejemplo de cómo conservar los beneficios del intercambio de información del *pre-trial* de una forma compatible con la estructura procesal continental.

**PALABRAS CLAVE:** derecho comparado; proceso civil; acceso a la prueba; Discovery; exhibición de documentos.

## THE DISTINCTION BETWEEN PROOF AND PREPARATORY ACCESS TO INFORMATION IN CIVIL PROCEDURE

**ABSTRACT:** This paper proposes to adopt the distinction between proof and preparatory access to information when taking a comparative approach to civil procedure. Such a distinction provides means for functional differentiation of procedural rules regarding factual decision-making. It also allows identifying rules aimed at the factual decision making by a judge or a jury (proof) as distinctive from rules aimed at providing information to the parties and enabling them to understand the strength of their positions and to make informed decisions on how to further proceed with the case (preparatory access to information). This paper will show how failing to distinguish these two categories can lead to faulty comparative reflections, such as equating the measures of the pre-trial of the common law countries (preparatory tools) and the documentary exhibits of the continental civil procedure (evidentiary tools). By way of illustration, this paper will show that Rule 102 of the European Rules of Civil Procedure is an example of how to preserve the benefits of pre-trial information exchange in a way that is compatible with the continental procedural structure.

**KEYWORDS:** comparative law; civil procedure; access to evidence; discovery; disclosure of documents.

**SUMARIO:** INTRODUCCIÓN.— 1. DIFICULTADES DEL ANÁLISIS COMPARADO DEL ESCLARECIMIENTO PROCESAL.— 2. FUNCIONALISMO COMO MÉTODO DE COMPARACIÓN.— 3. LA DISTINCIÓN ENTRE ESCLARECIMIENTO «PROBATORIO» Y ESCLARECIMIENTO «PREPARATORIO»: 3.1. El esclarecimiento probatorio.- 3.2. El esclarecimiento preparatorio.— 4. BENEFICIOS DE DISTINGUIR ENTRE PREPARACIÓN Y PRUEBA.— 5. UN ESCLARECIMIENTO PREPARATORIO CONSISTENTE CON LA TRADICIÓN CONTINENTAL.— CONCLUSIONES.— BIBLIOGRAFÍA.

**RECOMMENDED CITATION:** THOMAS VOGT GEISSE, 2022: «La distinción entre esclarecimiento probatorio y esclarecimiento preparatorio en el proceso civil», in *Quaestio facti*, 3: 159-173. Madrid: Marcial Pons Ediciones Jurídicas y Sociales. DOI: [http://dx.doi.org/10.33115/udg\\_bib/qf.i3.22711](http://dx.doi.org/10.33115/udg_bib/qf.i3.22711)

## INTRODUCCIÓN

En los países de tradición continental, las reformas al proceso civil suelen discutirse con una mirada de reojo hacia el proceso anglosajón. Sus figuras aparentemente novedosas y exitosas —la “teoría del caso”, el “*case management*”, o la “*alternativel/appropriate dispute resolution*”— brillan como posibles remedios contra los males que aquejan a procesos continentales en crisis. En materia de determinación de los hechos, el amplio y temprano acceso a la información que permite el *pre-trial discovery* (o instituciones parecidas) seduce con una promesa de eficaz resolución de disputas y el acertado establecimiento de los hechos. Es frecuente que se resalten las bondades del acceso a la información que ofrece el *pre-trial* en contraste con el carácter limitado de procedimientos probatorios continentales.<sup>1</sup> Este artículo

<sup>1</sup> Con distintos matices: en Chile, PEÑA (2017): 79-120; en España, CERRATO (2021): 3-16; en Alemania; KOCH (2013): 224 ss.

mostrará —desde el derecho comparado— que estas comparaciones son apresuradas y una mejor comprensión comparada del esclarecimiento de los hechos en el proceso civil requiere un mayor grado de diferenciación.

El mensaje central de la ciencia del derecho comparado es la necesidad de fijar parámetros de comparabilidad respecto a las normas jurídicas de distintos ordenamientos. Para el derecho *procesal* comparado, esto es de esencial importancia, ya que las normas procesales solo pueden revelar su sentido si se considera el contexto procedimental en el que operan. Por eso, sin resolver la pregunta sobre la comparabilidad, es imposible evaluar, de forma controlada, las ventajas y desventajas de una regla de esclarecimiento.<sup>2</sup>

Como aporte a este desafío de transparentar la comparabilidad de normas procesales, ofrezco en este artículo la distinción entre los conceptos de esclarecimiento procesal “probatorio” y “preparatorio”, que considero esencial para el análisis del *discovery* estadounidense, del *disclosure* inglés o de figuras parecidas del *pre-trial* anglosajón. Procederé de la siguiente forma: Primero describiré, a partir de un ejemplo, las dificultades que surgen al comparar normas de esclarecimiento procesal (1). Luego, abordaré el funcionalismo como base metodológica de la comparación en el derecho procesal (2). A continuación, explicaré la diferencia entre esclarecimiento procesal probatorio y preparatorio (3) y los beneficios de esta distinción para la reflexión comparada (4). Finalmente, presentaré la Rule 120 de las European Rules of Civil Procedure de 2021 (ERCP)<sup>3</sup> como un modelo de esclarecimiento preparatorio compatible con la tradición continental (5). El artículo cierra con un conjunto de conclusiones.

## 1. DIFICULTADES DEL ANÁLISIS COMPARADO DEL ESCLARECIMIENTO PROCESAL

La dificultad de comparar normas jurídicas, incluidas aquellas dirigidas al esclarecimiento de los hechos, es que su significado varía según donde el observador fije la mirada. Dependerá de su enfoque si normas de distintos ordenamientos se pueden considerar semejantes o diferentes. Para ejemplificar esto, miremos dos normas específicas provenientes de distintas tradiciones procesales: por un lado, el art. 328 1 de la Ley de Enjuiciamiento Civil española (LEC), por el otro, la *Rule 34(a)(1)* de las *Federal Rules of Civil Procedure* que rigen el proceso civil federal en los Estados Unidos (FRCP):

Artículo 328 I LEC

---

<sup>2</sup> VOGT (2020): 4.

<sup>3</sup> ELI-UNIDROIT Model European Rules of Civil Procedure (2021), [https://www.europeanlawinstitute.eu/fileadmin/user\\_upload/p\\_eli/Publications/200925-eli-unidroit-rules-e.pdf](https://www.europeanlawinstitute.eu/fileadmin/user_upload/p_eli/Publications/200925-eli-unidroit-rules-e.pdf). [consulta: 31 de octubre de 2021]

“1. Cada parte *podrá* solicitar de las demás la exhibición de documentos *que no se hallen a disposición de ella* y que se refieran al objeto del proceso o a la eficacia de los medios de prueba.

2. A la solicitud de exhibición deberá acompañarse copia simple del documento y, si no existiere o no se dispusiere de ella, se indicará *en los términos más exactos posibles el contenido de aquel*”

FRCP 34(a)(1)(A)

“A party *may* serve on any other party a request within the scope of Rule 26(b): *to produce* and permit the requesting party or its representative to inspect, copy, test, or sample the following *items in the responding party’s possession, custody, or control*: any designated documents or electronically stored information ...”

En cuanto a la extensión de la facultad de exhibición la *Rule 26(b)*:

(1) Scope in General. Unless otherwise limited by court order, the scope of discovery is as follows: Parties may obtain discovery regarding any nonprivileged matter *that is relevant* to any party’s claim or defense and *proportional* to the needs of the case, considering ...”.

Estas normas procesales, ¿son similares o diferentes? Ambas están dirigidas a lograr algún tipo de esclarecimiento de hechos relevantes para el proceso. Ambas consagran la facultad de una parte (“podrá”, “*may*”) de solicitar acceso a una fuente de información en control de la contraparte. ¿Esas similitudes bastan para afirmar que pertenecen a una misma categoría?

Si tomamos distancia de esa similitud inmediata y miramos su función dentro del contexto procesal, aparece una diferencia determinante. El art. 328 I se ubica en el capítulo VI del Libro II de la LEC que trata los “medios de prueba”. La exhibición ahí regulada se entiende como operación probatoria, en el sentido que permite suministrar una base para la decisión del tribunal sobre la ocurrencia de los hechos. La FRCP 34, en cambio, opera antes de cualquier etapa probatoria – en general, antes siquiera de designarse el órgano que decidirá sobre la *quaestio facti*. Esto demuestra que tiene una función completamente distinta a la función probatoria del art. 328 I LEC. Es parte de una fase preliminar (el “*pre-trial discovery*”) de intercambio de información con el fin de que las *partes*, y no un tribunal, adquieran un conocimiento más acabado sobre la fuerza relativa de sus posiciones en miras a una eventual etapa probatoria. Su función no es aportar a una cognición autoritativa por parte del tribunal, sino a una prognosis de las partes sobre el futuro desarrollo del proceso.<sup>4</sup> A partir de aquí ya podemos apreciar que nos falta una distinción en lo que respecta al esclarecimiento procesal de los hechos. Pero antes de desarrollarla, conviene exponer algunos aspectos metodológicos.

<sup>4</sup> Detalles sobre esta función de esclarecimiento preparatorio en los procesos civiles del *common law*, VOGT (2020): 42-79.

## 2. FUNCIONALISMO COMO MÉTODO DE COMPARACIÓN

La reflexión precedente nos lleva a preguntas sobre el método funcional de derecho comparado —en la versión desarrollada por ZWEIGERT y KÖTZ en los años noventa<sup>5</sup>— aplicado al ámbito del derecho procesal. Según esta doctrina, el trabajo de comparación presupone la definición de un problema común a comparar. Este problema funge como *tertium comparationis*.<sup>6</sup> Permite identificar las normas que los ordenamientos prevén para regularlo y estas serán comparables en la medida que cumplen esa función, en calidad de equivalentes funcionales.<sup>7</sup> La idea central es que es preferible que figuras jurídicas de distintos ordenamientos no se comparen como abstracciones, sino que en función de un problema previamente definido.<sup>8</sup> Para identificar funciones comunes, estos autores recomiendan presumir la similitud de problemas en los distintos ordenamientos.<sup>9</sup> Este funcionalismo también ha sido considerado adecuado para la comparación de derecho procesal.<sup>10</sup>

La amplia acogida de este método funcional generó críticas que, en lo más relevante, denuncian el sesgo nacional o eurocentrista que promueve la presunción de similitud que los autores proponen para identificar problemas comunes.<sup>11</sup> Estas críticas solo prosperaron en el sentido que impulsaron correcciones al método. Según versiones actuales del funcionalismo, la identificación del problema y de la función debe ser un resultado, no de presumir similitud, sino de operaciones más complejas. Según una opinión, se debe seguir un conjunto de pasos hermenéuticos que permitan un ir y venir de la mirada comparativa.<sup>12</sup> Según otra, se debe resaltar que cada función que se fije da cuenta de solo una de varias posibilidades de entender una norma, resaltando el método como una propuesta de interpretación funcional, sin pretensión de verdad.<sup>13</sup>

Según este funcionalismo interpretativo, se sugiere proceder según los siguientes pasos.<sup>14</sup> En primer lugar, se debe interpretar el derecho desde una perspectiva funcional, esto es, como respuesta a un problema definido con más o menos precisión. A partir de esta función será posible, en un segundo paso, “generar comparabilidad”

---

<sup>5</sup> ZWEIGERT y KÖTZ (1998).

<sup>6</sup> Categoría indispensable en toda ciencia comparada, JANSEN (2019): 295.

<sup>7</sup> ZWEIGERT y KÖTZ (1998): 40.

<sup>8</sup> ZWEIGERT y KÖTZ (1998): 38 s.

<sup>9</sup> ZWEIGERT y KÖTZ (1998): 45.

<sup>10</sup> GOTTWALD (2005): 231; STÜRNER/STADLER (1995): 265.

<sup>11</sup> Una exposición detallada de las múltiples críticas en KISCHEL (2015): 95-108. Uno de los críticos más destacados es Pierre Legrand, en cuanto resalta que el poder discursivo del observador convierte el ejercicio de comparación en una imposición etnocentrista, LEGRAND (2003): 255.

<sup>12</sup> KISCHEL (2015): 164 ss.

<sup>13</sup> MICHAELS (2019): 386.

<sup>14</sup> Siguiendo a MICHAELS (2019): 386-388.

identificando equivalentes funcionales. La identificación de equivalentes funcionales envuelve una propuesta de tratar normas como comparables para aportar a su mejor comprensión.<sup>15</sup> Este paso es importante, porque es sobre esta base de equivalencia funcional que posteriormente, en un tercer paso, los equivalentes se comparan según un criterio adicional, que puede ser histórico, económico, filosófico o institucional, según el enfoque que le quiera dar cada autor.<sup>16</sup> En este último paso se pueden hacer juicios de calidad, por ejemplo, sostener que una norma es mejor que otra desde un punto de vista económico o filosófico.

La distinción que propongo pretende facilitar el segundo de estos tres pasos al momento de comparar normas de esclarecimiento procesal. Sirve para identificar equivalentes funcionales de forma que la comparación permita conclusiones más precisas. Se basa en la premisa que en materia procesal la comparación es más reveladora si los equivalentes funcionales se definen incorporando, en la mayor medida posible, el contexto procesal. Para comparar el proceso como “experiencia situativa”<sup>17</sup> es necesario, tal vez más que en otras materias, ir más allá de los conceptos extraídos del tenor de las normas y mirar, a la vez, la estructura particular de los procesos en los que se aplican. Como veremos, la distinción entre preparación y prueba que propongo a continuación facilita la integración del contexto al definir los equivalentes funcionales.

### 3. LA DISTINCIÓN ENTRE ESCLARECIMIENTO «PROBATORIO» Y ESCLARECIMIENTO «PREPARATORIO»

Ahora, con estas ideas sobre el método funcional en mente, podemos volver a la comparación del art. 328 I LEC y la FRCP 34 considerando los pasos explicados arriba: (i) proponer una función desde un problema común, (ii) identificar de equivalentes funcionales, (iii) comparar normas equivalentes y, eventualmente, hacer un juicio de calidad.

Entonces, en primer lugar, todo proceso civil se enfrenta a la necesidad de esclarecer hechos inciertos para una acertada aplicación del derecho civil (problema común) y, por lo tanto, podemos identificar una función de normas procesales que buscan el esclarecimiento de los hechos (función común). En segundo lugar: una forma de cumplir esta función es mediante normas que obligan a revelar (esto es, individualizar y, eventualmente, producir) documentos. Por tanto, podemos interpretar, tanto el art. 328 I LEC como la FRCP 34 en clave de dicha función de esclarecer los hechos en el proceso y concluir que son comparables, es decir, equivalentes fun-

---

<sup>15</sup> MICHAELS (2019): 387.

<sup>16</sup> *Loc. cit.*

<sup>17</sup> STÜRNER y STADLER (1995): 281.

cionales como herramientas de esclarecimiento. Desde aquí se podría, en un tercer paso, hacer un juicio de valor respecto a su calidad, por ejemplo, como herramienta para el establecimiento de la verdad (si interesa la calidad epistémica de la sentencia), o como herramienta para promover acuerdos (si interesa la aptitud del proceso de promover la autocomposición), o desde el punto de vista de los costos, etc.

Como ya adelanté bajo I., considero que aquí el segundo paso expuesto —el establecimiento de la equivalencia— resulta algo apresurado. Como mostraré a continuación, es conveniente incluir el contexto procesal antes de fijar la equivalencia de una norma y proceder a la comparación. La distinción que propongo, entre esclarecimiento probatorio y preparatorio, facilita la incorporación de ese contexto. Permite hilar más fino, ya que subdivide la función genérica “esclarecer hechos en el proceso” en dos categorías funcionales más específicas que permiten identificar con una mayor precisión normas de esclarecimiento equivalentes.

### 3.1. El esclarecimiento probatorio

Por un lado, todo proceso civil prevé una etapa de esclarecimiento *probatorio* identificable por su orientación hacia la comprobación de los hechos por parte del juzgador. Esta abarca tanto la producción de fuentes de información necesarias para que el juzgador pueda determinar lo que efectivamente ocurrió, como también información sobre la confiabilidad de estas fuentes.<sup>18</sup> A esta categoría pertenece art. 328 I LEC recién citado. Estas normas pueden variar considerablemente según tradición jurídica: en el *common law* se desenvuelven en el marco de un *trial* – una audiencia pública de funcionamiento bipartido y estrictamente concentrada; en la tradición continental, en cambio, la actividad probatoria se produce en una audiencia de juicio unipolar y con flexibilidad para reagendar sesiones y profundizar en la averiguación de lo ocurrido.<sup>19</sup> Por lo mismo, las normas probatorias de distintas tradiciones pueden tener un contenido muy diferente a primera vista, como demuestra la práctica del *cross-examination*, muy distinta a la interrogación judicial típica de los procesos de tradición continental. No obstante, se deben mirar como equivalentes funcionales en cuanto cumplen la misma función probatoria, esto es, una acertada cognición judicial.

### 3.2. El esclarecimiento preparatorio

El esclarecimiento *preparatorio* tiene una función distinta. No va dirigido al tribunal, sino a las partes. Su finalidad es que estas puedan tomar decisiones informadas

<sup>18</sup> Considero que esto es lugar común. Mayor detalle en VöGT (2020): 18 ss.

<sup>19</sup> P. ej., DAMAŠKA (1997) y LANGBEIN (1985): 823-866. Todo esto sin perjuicio de tendencias convergentes, vid. párrafos 12 y 13 del preámbulo de las ERCP.

sobre la disputa, con anterioridad a la fase probatoria del proceso. El acceso a información puede cobrar relevancia antes de la iniciación del proceso (esclarecimiento prejudicial) o después de iniciado el proceso, durante la etapa de intercambio de los escritos de discusión, pero antes de la fase probatoria (esclarecimiento pre-probatorio). En todo caso, este tipo de esclarecimiento no es funcional a la toma de decisión del juzgador, sino a una mayor autonomía de las partes en el ejercicio de sus facultades. Está dirigido a facilitar su decisión sobre si conviene seguir adelante con el proceso y, en caso afirmativo, la decisión sobre la extensión y los términos en los cuales se someterá a prueba la disputa.

Estos “términos” no son definibles en abstracto. Depende de cómo está regulada la prueba en un ordenamiento específico. Serán diferentes, según si la prueba se desarrolla en un trial o en una audiencia de estilo continental. En el primer caso, el esclarecimiento preparatorio está dirigido a generar un *prima facie case* susceptible de ventilarse en un trial de funcionamiento bipolar y concentrado. Cada parte debe estar en condiciones de elaborar una *teoría del caso* con antelación al trial, en que incorpore el *case law* que considera aplicable y contenga un contraste persuasivo frente al caso de la contraparte.<sup>20</sup> Esto presupone que las partes puedan acceder a todos los antecedentes del conflicto, lo que a su vez exige amplios derechos de esclarecimiento preparatorio en la fase del *pre-trial (discovery)*. Esta necesidad de diseñar un caso *prima facie* para optar a un trial explica la existencia de un conjunto de facultades de investigación privadas (FRCP 30-36) destinadas, no solo a revelar fuentes relevantes de información, sino a averiguar la existencia de fuentes potencialmente relevantes en el control de la contraparte. Solo así se asegura que cada parte pueda construir, con autonomía, una versión del caso con expectativa de éxito en un trial. Esto explica también que el acceso proceda según un amplio criterio de relevancia, limitado solamente por consideraciones de proporcionalidad a las necesidades de la causa (ver FRCP 26(b), transcrita arriba). Esta etapa de esclarecimiento privado pre-probatorio corresponde a una decisión del sistema que apuesta a que, bajo los incentivos adecuados, las partes están en mejores condiciones de seleccionar el nivel óptimo de información necesaria para resolver el asunto.<sup>21</sup>

En el modelo de tradición continental, en cambio, el esclarecimiento preparatorio tiende a ser prescindible. En el proceso continental, la fase probatoria no se orienta en dos versiones opuestas del caso, sino en una controversia respecto de *afirmaciones* sobre hechos jurídicamente relevantes aportadas por las partes en sus escritos de discusión (demanda, contestación, etc.). Esto se debe a que la decisión sobre los hechos se orienta en las normas jurídicas abstractas de los códigos, y no, como en el *common law*, en una selección de precedentes propuesta por cada una de las partes para su

---

<sup>20</sup> ANDERSON, SCHUM y TWINING (2005): 153.

<sup>21</sup> NANCE (2016): 190-192.

versión del caso.<sup>22</sup> En efecto, en el continente no tiene sentido desarrollar “teorías del caso” (en su sentido propio). El litigante continental persuade —no destilando argumentos desde precedentes— sino demostrando que las afirmaciones sobre hechos que alega permiten subsumirse bajo las normas legales que invoca a su favor.

No cabe duda de que también para el litigante del continente es deseable tener toda la información disponible y así estar en condiciones de evaluar sus expectativas de éxito antes de recibirse la causa a prueba, especialmente en los ordenamientos que prevén la condena en costas al litigante derrotado. Pero el acceso pre-probatorio a la información no es, como en el common law, condición indispensable para que la parte pueda acceder a la etapa probatoria y comprobar su versión de los hechos.<sup>23</sup> En un proceso civil continental, la parte puede solicitar diligencias probatorias sin saber con exactitud qué resultado van a arrojar; y si su práctica da cuenta de la necesidad de someter a prueba nuevos hechos relevantes (por ejemplo, tras conocer el resultado de una pericia o de una declaración de testigo), en general se permite agendar una sesión adicional y continuar la audiencia de prueba en una fecha posterior si es necesario para una acertada decisión judicial.<sup>24</sup> Eso puede explicar que el derecho procesal continental tradicionalmente solo ofrezca escasas oportunidades, reguladas como medidas prejudiciales, para que los litigantes puedan acceder a información para su propia toma de decisión.

#### 4. BENEFICIOS DE DISTINGUIR ENTRE PREPARACIÓN Y PRUEBA

Aplicando la distinción recién desarrollada, cambia la forma de comprender el art. 328 I LEC y la FRCP 34(a)(1). Estas normas, que a primera vista eran equivalentes funcionales, dejan de serlo. Vistas desde nuestra óptica, la primera resulta ser una norma de esclarecimiento probatorio, mientras que la segunda es de esclarecimiento preparatorio. Para comparativistas continentales, la distinción exige dejar de ver el *discovery* y otros sistemas de esclarecimiento preliminar del *pre-trial* anglosajón como figuras relacionadas con el derecho probatorio y la cognición judicial.

---

<sup>22</sup> Al respecto VOGT (2020): 42 ss. Obviamente se pueden hacer matices en este punto, resaltando que también el “precedente” continental influye en las decisiones de los litigantes y que también el derecho anglosajón está regulado por normas abstractas consagradas en *statutes*. Para mayor claridad omito las discusiones sobre una supuesta convergencia entre las tradiciones jurídicas cuanto al sistema de fuentes. En todo caso, creo que las diferencias en regulación y funcionamiento de los procesos civiles que describo demuestran las diferencias que aún subsisten en la argumentación jurídica entre un sistema de *case law* y otro de derecho codificado.

<sup>23</sup> En los procesos civiles de Inglaterra y los EE.UU. el litigante que no logra armar un caso *prima facie* a favor de su versión de los hechos puede ser derrotado sin necesidad de un *trial*, como consecuencia de un *summary judgement*.

<sup>24</sup> Al menos en su versión tradicional, VOGT (2020): 33 s., sin perjuicio de esfuerzos para una mayor concentración en varios ordenamientos nacionales de tradición continental.

En esta línea, el beneficio de la distinción es que previene la confusión del *pre-trial discovery* o *disclosure* con normas probatorias. Esto es importante, porque la confusión podría llevar a entender que el intercambio de información del *pre-trial* es, a la vez, información que se exhibe al tribunal y aporta así a una mejor toma de decisión por el tribunal. Esto no es así. A pesar de ordenar la producción de fuentes de información, el *discovery* opera en un ámbito funcionalmente distinto al de la rendición de pruebas. El acceso a información que concede el *discovery* está dirigido exclusivamente a la cognición privada preparatoria; las exhibiciones documentales del proceso de tradición continental, en cambio, están reguladas como aportes probatorios, es decir, aportes a la cognición del tribunal.

La distinción explica que el *discovery* pueda producir un volumen de información que, como producción probatoria, sería impensable. Según datos de 2008, en las causas de grandes empresas que llegan a juicio se intercambia un promedio de aproximadamente 5 millones de páginas por causa,<sup>25</sup> y los volúmenes van en aumento exponencial.<sup>26</sup> Un tribunal no estaría en condiciones de valorar racionalmente la cantidad de información que ingresaría al juicio en virtud del amplio criterio de relevancia que admite el *discovery*.<sup>27</sup> Por cierto, estos números astronómicos —hay que imaginarse una biblioteca de 5000 libros de 1000 páginas cada uno— se refieren a casos complejos y muestran el lado más extremo de la litigación civil. Pero deja de manifiesto la función que cumple la figura. Al tratarse un instrumento de esclarecimiento preparatorio, la información revelada en el *discovery* queda a disposición de las partes, con el fin que estas efectúen una selección de lo relevante y diseñen su versión del caso. De hecho, el mismo estudio indica que de los millones de páginas exhibidas en el *discovery* solo 0,1% se utilizan para efectos de probar los hechos en juicio.<sup>28</sup> Solo esa pequeña fracción se presenta al juzgador en el trial para su valoración.

La función preparatoria del *discovery* es consistente con cómo se organiza y financia la abogacía en los EE.UU. Es común que los abogados se asocien como grandes empresas con amplio personal de apoyo. A modo de referencia: en los EE.UU. existen 94 estudios jurídicos que emplean a más de 500 abogados (año 2018).<sup>29</sup> En Alemania, solo tres estudios llegan a semejantes dimensiones (año 2019/2020).<sup>30</sup> En casos complejos, la labor de análisis del producto del *discovery* puede llegar a costar

<sup>25</sup> La cifra exacta es 4980441 páginas, *vid.* Litigation Cost Survey of Major Companies (2010): 3. [https://www.uscourts.gov/sites/default/files/litigation\\_cost\\_survey\\_of\\_major\\_companies\\_0.pdf](https://www.uscourts.gov/sites/default/files/litigation_cost_survey_of_major_companies_0.pdf). [consulta: 31 de octubre de 2021]

<sup>26</sup> ARMOUR, PARNHAM y SAKO (2020): 20.

<sup>27</sup> Una explicación considerando las normas de exhibición del proceso civil alemán, VOGT (2020): 71.

<sup>28</sup> Litigation Cost Survey of Major Companies (2010), *loc. cit.*: “The ratio of pages discovered to pages entered as exhibits is as high as 1000/1. In 2008, on average, 4,980,441 pages of documents were produced in discovery in major cases that went to trial – but only 4,772 exhibit pages actually were marked.”

<sup>29</sup> <https://www.ilrg.com/nlj250/attorneys/desc/2>. [consulta: 31 de octubre de 2021]

<sup>30</sup> <https://de.statista.com/statistik/daten/studie/191422/umfrage/die-groessten-kanzleien-in-deutschland-nach-anzahl-der-anwaelte/>. [consulta: 31 de octubre de 2021]

varios millones de dólares.<sup>31</sup> ¿Quién la financia? Los clientes corporativos pueden internalizar estos costos, los demás deben buscar financiamiento en el mercado crediticio o, si se trata de un demandante, delegar la búsqueda de financiamiento al estudio de abogados prometiéndole a cambio un porcentaje de la eventual ganancia obtenida en virtud de un acuerdo o veredicto favorable. En este escenario, los abogados asumen el rol de agentes comerciales en búsqueda de terceros dispuestos a cofinanciar su cartera de causas (“*portfolio funding*”).<sup>32</sup>

Estas reflexiones ilustran por qué en países de tradición continental, en general ajenos a la idea que el levantamiento privado de información es parte estructural de la litigación, es difícil de concebir un *discovery* u otro modelo de facultades privadas de esclarecimiento con un amplio criterio de acceso. Incluso cuando el proceso civil continental contempla normas de esclarecimiento preparatorio, lo hace regulando medidas prejudiciales concretas, limitadas a fines específicos (prueba anticipada y exhibición de ciertas categorías de documentos para facilitar la preparación de la demanda).

Las diferencias entre el esclarecimiento preparatorio del *pre-trial* y el procedimiento probatorio del continente parecen no haber permeado aún en el pensamiento procesal hispanoparlante. A modo de ejemplo, un reciente artículo<sup>33</sup> considera equivalentes funcionales la exhibición documental de terceros de la LEC con las reglas de *disclosure* de las Civil Procedure Rules (CPR) que rigen el proceso civil de Inglaterra y Gales. Si incorporamos nuestra distinción, vemos que esta equivalencia no puede ser muy fructífera, ya que las normas probatorias de la LEC están dirigidas a la cognición judicial, mientras que las normas de las CPR están dirigidas a dar acceso a información a las partes en el *pre-trial*, antes de una etapa probatoria. De haberse valido de la distinción aquí presentada, la autora se hubiese abstenido de afirmaciones del tipo “[l]as dos instituciones procesales [la exhibición documental de la LEC y el *disclosure* de las CPR] se desarrollan estrictamente en sede de prueba documental”.<sup>34</sup> Aunque sus descripciones de las normas del *disclosure* son rigurosas, la autora no repara en que tienen una función distinta a las normas de prueba. Al igual que el *discovery*, el *disclosure* no busca una comprobación judicial de lo ocurrido, sino facilitar decisiones preparatorias de los litigantes. La confusión del ámbito preparatorio y probatorio la hace llegar a un juicio de valor muy divulgado en el derecho procesal comparado, pero, en mi opinión, apresurado:<sup>35</sup> “De ello se desprende que el sistema anglosajón es más garantista en este extremo por cuanto apuesta por la incorporación al proceso de todos los elementos probatorios”.<sup>36</sup> Esta afirmación es, en el mejor de los casos, incompleta. No aclara el fin para el cual el sistema an-

---

<sup>31</sup> Litigation Cost Survey of Major Companies (2010): 15.

<sup>32</sup> STEINITZ (2019): 1085 s.

<sup>33</sup> CERRATO (2021): 3-16.

<sup>34</sup> CERRATO (2021): 13.

<sup>35</sup> Para críticas a académicos alemanes que incurren en el mismo error, VOGT (2020): 132-136.

<sup>36</sup> CERRATO (2021): 14.

glosajón permite una mayor amplitud en el esclarecimiento. Las normas de *disclosure* están dirigidas a asegurar que *las partes* conozcan la fuerza relativa de su postura y de la postura de la contraria y puedan así seleccionar los elementos probatorios que les conviene usar frente al juzgador en un eventual trial. Sin embargo, un amplio esclarecimiento preparatorio no necesariamente garantiza una cognición judicial acertada, conforme a la verdad. La selección de las partes tenderá a ser incompleta y solo capaz de reflejar dos versiones extremas de lo ocurrido, ya que se adapta a la teoría del caso que cada parte formulará en contraste con la postura contraria. TARUFFO reconoció este punto con notable agudeza al advertir que los procesos adversariales anglosajones no son adecuados para la averiguación de la verdad. Refiriéndose a los instrumentos del *discovery*, explica:

Es necesario tener en cuenta que estos instrumentos están dirigidos a hacer que *las partes* (no el juez) puedan tomar conocimiento de las pruebas de las que disponen las contrapartes; sin embargo, no necesariamente todas estas pruebas son efectivamente practicadas en el *trial* (entre otras razones, porque también pueden ser objeto de *discovery* pruebas inadmisibles). Por consiguiente, el *discovery* está dirigido a maximizar las oportunidades defensivas de las partes, pero no a maximizar la posibilidad de que la decisión final se funde en la determinación de la verdad de los hechos.<sup>37</sup>

La necesidad de distinguir entre esclarecimiento probatorio y preparatorio en el derecho procesal comparado también se puede ver como una consecuencia de esta observación.<sup>38</sup>

## 5. UN ESCLARECIMIENTO PREPARATORIO CONSISTENTE CON LA TRADICIÓN CONTINENTAL

Como ya se mencionó, en el proceso de tradición continental la etapa probatoria se activa controvirtiendo afirmaciones sobre hechos jurídicamente relevantes. A partir de estas afirmaciones controvertidas se puede practicar un esclarecimiento “profundo”, mediante diversas diligencias probatorias. Esto torna estructuralmente prescindible un esclarecimiento pre-probatorio. Sin embargo, también en estos procesos puede ser beneficioso que las partes, antes de la etapa probatoria, puedan acceder a las fuentes en manos de la contraparte para un conocimiento acabado de los hechos. Este conocimiento podría evitar procesos, promover salidas mediante acuerdo evitando la costosa etapa probatoria o, al menos, limitar la actividad probatoria mediante una mayor precisión de los hechos controvertidos. La distinción entre preparación y prueba nos permite orientar cómo incorporar al proceso civil continental los beneficios del *discovery* —esto es, la promoción de acuerdos, delimitación eficaz de la disputa y eliminación de sorpresas que puedan alargar un juicio— sin

<sup>37</sup> TARUFFO (2010): 129, nota 206.

<sup>38</sup> Esto independientemente de la discusión sobre la aptitud de cada tipo de esclarecimiento de promover una decisión judicial conforme a la verdad.

alterar el derecho probatorio ni anteponer una etapa de esclarecimiento sin bordes, incompatible con la tradición continental.

Pese a su antigüedad, sigue siendo referente en esta materia la propuesta de *Rolf Stürner* de 1976. Propone consagrar un deber de esclarecimiento prejudicial de ambas partes, fundado en el postulado que cada parte de un proceso futuro tiene el deber de revelar a la contraria la información y las fuentes en su poder, en la medida que la contraparte se encuentre en desconocimiento inimputable.<sup>39</sup> Su base dogmática estaría en la idea que —como efecto anticipado— la relación jurídico procesal impone a cada persona un deber de hacer lo posible para evitar procesos innecesarios.<sup>40</sup> Dado que el acceso a información tiene la aptitud de hacer prescindible un proceso, sería razonable asumir un deber de esclarecimiento prejudicial dentro de ciertos márgenes. Para evitar el costo excesivo y uso abusivo de las solicitudes de información, STÜRNER propone delimitarlas exigiendo subsidiaridad y plausibilidad.

La subsidiaridad dispone que si el interesado está en condiciones de obtener por su cuenta la información, no tiene derecho a obtenerla de otra persona. Esto evita que se use solo para liberarse de los costos de levantar la información.

La plausibilidad busca evitar el uso frívolo de este derecho, exigiendo cierto grado de verosimilitud del relato que se acompaña a la solicitud. Este debe contener indicios que muestren una razonable posibilidad de que se haya configurado el derecho que se pretende hacer valer.<sup>41</sup>

El requerido que no cumple con este deber de información se arriesga, como sanción, a soportar las costas en un proceso posterior, si aparece que este se podría haber evitado transparentando las posiciones (independientemente del resultado del proceso).<sup>42</sup>

Desarrollos posteriores han precisado algunos de estos elementos para un acceso temprano a la información, buscando una fórmula compatible con la estructura del proceso continental y sus instituciones. El resultado de estos esfuerzos ha quedado plasmado en la ERCP 102.

*Rule 102. Relevant Criteria where an application for access to evidence is made*

*(1) A party or prospective party applying for an order for access to evidence must*

*(a) identify, as accurately as possible in the light of the circumstances of the case, the specific sources of evidence to which access is sought, or alternatively*

*(b) identify closely defined categories of evidence by reference to their nature, content, or date.*

<sup>39</sup> STÜRNER (1976): 274.

<sup>40</sup> STÜRNER (1976): 269 s., generalizando la idea detrás del § 93 de la ZPO que, en caso de allanarse el demandado, transfiere al demandante las costas del litigio en la medida que el demandado “no haya provocado, mediante su conducta, la necesidad de presentar la demanda”. Considera que aquí hay una sanción al demandante que interpone demanda sin informar previamente al demandado la razón de su pretensión, generando gastos innecesarios (269).

<sup>41</sup> STÜRNER (1976): 277, 123.

<sup>42</sup> STÜRNER (1976): 284.

(2) *An application must satisfy the court of the plausibility of the merits of the applicant's claim or defence by demonstrating that*

(a) *the requested evidence is necessary for the proof or proposed proof of issues in dispute in proceedings or in contemplated proceedings;*

(b) *the applicant cannot otherwise gain access to this evidence without the court's assistance; and*

(c) *the nature and amount of evidence subject to the application is reasonable and proportionate. For this purpose the court will take into account the legitimate interests of all parties and all interested non-parties.*

(3) *If an application for access to evidence is made prior to the commencement of proceedings, the applicant must indicate with sufficient precision all elements necessary to enable the court to identify the claim for relief which the applicant intends to make.*

(4) *The court may not grant any application under this Rule which involves a vague, speculative, or unjustifiably wide-ranging search for information.*

Esta regla no está dirigida a la toma de decisión del tribunal (prueba), sino que de las partes (preparación). Propone una regulación de esclarecimiento preparatorio, tanto pre-probatorio como prejudicial.

Para evitar los abusos y costos excesivos del *discovery* y *disclosure*, requiere la máxima especificidad posible de los ítems solicitados (ERCP 102(1)), además de plausibilidad (ERCP 102(2)) demostrando necesidad (1.), subsidiaridad (2.) y proporcionalidad (3.) de la solicitud. En caso de una solicitud prejudicial, esto es, anterior a la demanda y, por tanto, a los relatos de los escritos de la fase de discusión, la norma exige suficiente precisión del derecho que el solicitante pretende hacer valer (ERCP 102(3)). La norma cierra con un requisito negativo: la ausencia de vaguedad, especulación o amplitud injustificada de la solicitud (ERCP 102(4)). Vemos aquí una propuesta de un modelo de esclarecimiento preparatorio continental. No va dirigido a la prueba judicial de los hechos, sino a una toma de decisión de las partes. En este sentido es funcionalmente equivalente al *discovery*, pero presenta una regulación acorde con la estructura de procesos de tradición continental.

## CONCLUSIONES

En este trabajo he propuesto la distinción entre esclarecimiento probatorio y esclarecimiento preparatorio para enriquecer los debates comparados sobre el acceso a la información y a la prueba en el proceso civil. Su base metodológica es la comparación funcional interpretativa, ajustada a las exigencias del derecho procesal como objeto de estudio comparado. La disciplina del derecho procesal comparado exige considerar el contexto procesal de cada ordenamiento o tradición jurídica. Para incorporar este contexto en la comparación de normas de esclarecimiento procesal debemos diferenciar dos categorías funcionales. Por un lado, se deben identificar aquellas dirigidas a una acertada decisión del juzgador de hechos; las llamé normas de esclarecimiento probatorio. Estas se diferencian de normas dirigidas a las partes y a sus abogados para un acabado conocimiento de su posición con anterioridad a la prueba y una mejor toma de decisiones respecto a cómo seguir adelante con la

causa. Las llamé normas esclarecimiento preparatorio. Esta diferencia funcional permite una mejor comprensión de las normas de esclarecimiento procesal en cuanto evita una inerte comparación entre normas probatorias continentales con normas de esclarecimiento del *pre-trial* anglosajón, que son propias de una cultura procesal y organización de la abogacía no replicable en el continente. Sus beneficios, sin embargo, se pueden asegurar mediante un modelo de esclarecimiento preparatorio compatible con los procesos civiles de tradición continental, para lo que la ERCP 102 ofrece un buen punto de referencia.

## BIBLIOGRAFÍA

- ANDERSON, T., SCHUM, D., TWINING, W., 2005: *Analysis of Evidence*, 2a Ed., Cambridge/New York: Cambridge University Press..
- ARMOUR, J.; PARNHAM, R.; SAKO, M., 2020, "Augmented Lawyering", en: *European Corporate Governance Institute - Law Working Paper 558/2020*. [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=3688896](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3688896)
- CERRATO, E., 2021, "Análisis de la exhibición documental de terceros y la *disclosure against a person not a party* en el proceso civil", *Revista Ius et Praxis*, 27, N. 1: 3-16.
- DAMAŠKA, M., 1997, *Evidence Law Adrift*, New Haven/London: Yale University Press.
- GOTTWALD, P., 2005, "Zum Stand der Zivilprozessrechtsvergleichung", en: Bachmann, B. et al. (Eds.), *Grenzüberschreitungen – Beiträge zum Internationalen Verfahrensrecht und Schiedsgerichtsbarkeit. Festschrift für Peter Schlosser zum 70. Geburtstag*, Tübingen: Mohr Siebeck: 227-245.
- JANSEN N., 2019, "Comparative Law and Comparative Knowledge", en: REINMANN M.; ZIMMERMAN R. (Eds.), *The Oxford Handbook of Comparative Law*, 2a Ed., Oxford/New York: Oxford University Press.
- KISCHEL, U., 2015, *Rechtsvergleichung*, München: C.H.Beck.
- LANGBEIN, J., 1985, "The German Advantage in Civil Procedure", en: *The University of Chicago Law Review*, Vol. 52, N° 4, 823-866.
- LEGRAND, P., 2003, "The Same and the Different", en: LEGRAND, P., MUNDAY, R. (Eds.), *Comparative Legal Studies: Traditions and Transitions*, Cambridge: Cambridge University Press, 240 – 311.
- MICHAELS, R., 2019, "The Functional Method of Comparative Law", en: REINMANN M.; ZIMMERMAN R. (Eds.), *The Oxford Handbook of Comparative Law*, 2a Ed., Oxford/New York: Oxford University Press.
- NANCE, D., 2016, *The Burdens of Proof – Discriminatory Power, Weight of Evidence, and Tenacity of Belief*, Cambridge University Press, Cambridge.
- PEÑA MARDONES, C., 2017, "Traduciendo el *discovery* al *civil law* chileno: Su aporte a los procesos de reforma procesal civil", en: *Revista Ius et Praxis*, 23, N. 2: 79-120.
- STEINITZ, M., 2019, "Follow the Money? A Proposed Approach for Disclosure of Litigation Finance Agreements", *UC Davis Law Review*, 53: 1073-1116.
- STÜRNER, R., 1976, *Die Aufklärungspflicht der Parteien des Zivilprozesses*, Tübingen: JBC Mohr.
- STÜRNER, R., STADLER, A., 1995, "Eigenarten der Prozeßrechtsvergleichung", en: GILLES, P. (Ed.), *Transnationales Prozessrecht*, Baden-Baden, 263-292.
- TARUFFO, M., 2010: *Simplemente la verdad – el juez y la construcción de los hechos* (trad. Accatino, D.), Madrid: Marcial Pons.
- VOGT GEISSE, T., 2020, *Aufklärung und Informationskontrolle im Zivilprozess. Eine vergleichende Studie zum deutschen, englischen und US-amerikanischen Recht*, Tübingen: Mohr Siebeck.
- ZWEIGERT, K.; KÖTZ, H., 1998, *Introducción al derecho comparado*, Oxford/México: Oxford University Press.



## LOS PODERES PROBATORIOS DEL JUEZ DE AMPARO. UN GARANTE DE LA CORRECCIÓN EPISTÉMICA DE LOS HECHOS

Verónica Lorena Osornio Plata

Abogada postulante y docente de la cátedra de amparo  
en la Universidad Panamericana

**RESUMEN:** El artículo tiene como objetivo visibilizar la importancia de potenciar los poderes probatorios del juez en un medio de control constitucional, como lo es el juicio de amparo en México, en específico el relativo a la incorporación oficiosa de pruebas, en aras de fortalecer la averiguación de la verdad en el proceso. Así mismo, se pretende mostrar la insuficiencia de algunos argumentos que han obstaculizado el otorgamiento o el ejercicio de los poderes probatorios: la pérdida de imparcialidad judicial y la violación al principio de aportación de parte. Se advierte que dicho poder se prevé en la Ley de Amparo en supuestos limitados y que se tiene que recurrir a la codificación civil supletoria para regular su ejercicio, cuyo texto, en la práctica, restringe el principio de contradicción, en detrimento del derecho de defensa de las partes, por lo que es necesaria una reforma que subsane tales deficiencias e imponga límites a su actuación.

**PALABRAS CLAVE:** juez; poderes probatorios; prueba de oficio; prueba; verdad; amparo.

## EVIDENTIAL POWERS OF AMPARO JUDGES A GUARANTOR OF THE EPISTEMIC CORRECTNESS OF THE FACTS

**ABSTRACT:** This article aims to show the importance of enhancing evidentiary powers of judges in processes of constitutional control, such as the amparo trial in Mexico, specifically the ones related to the mandatory incorporation of evidence, in order to strengthen the investigation of the truth

in the process. This article also subverts the ideas that have hindered its exercise: loss of impartiality or transgression of the part contribution principle. It points out two issues: first, this powers are provided by the Amparo Law in limited cases and second, additional civil codification must be used to regulate the exercise of the Amparo Law, whose text, in practice, restricts the contradiction principle, to the detriment of the defense rights of both parties. This is the reason why it is necessary to implement an amendment that remedy such deficiencies and impose limits on the behavior of the judges.

**KEYWORDS:** Judge; evidencial faculties; ex officio evidence; evidence; truth; amparo.

**SUMARIO:** 1. INTRODUCCIÓN.—2. LAS PARTES: SUJETOS EN QUIENES DESCANSA, DE MANERA PRIORITARIA, LA ACTIVIDAD PROBATORIA EN EL JUICIO DE AMPARO.— 3. LOS PODERES DE INSTRUCCIÓN DEL JUEZ DE AMPARO.— 4. ALGUNAS IMPLICACIONES PRÁCTICAS EN EL DESPLIEGUE DE LOS PODERES PROBATORIOS OFICIOSOS.— 5. DECONSTRUCCIÓN DE LOS PREJUICIOS PROBATORIOS QUE LIMITAN LA ACTIVIDAD PROBATORIA DEL JUZGADOR.— 6. CONFIGURACIÓN Y DELIMITACIÓN DE LA PRUEBA DE OFICIO.— 7. CONCLUSIÓN.- 8. BIBLIOGRAFÍA.

**RECOMMENDED CITATION:** VERÓNICA LORENA OSORNIO PLATA, 2021: «Juez de amparo un garante de la corrección epistémica de los hechos», in *Quaestio facti*, 3: 175-215. Madrid: Marcial Pons Ediciones Jurídicas y Sociales. DOI: [http://dx.doi.org/10.33115/udg\\_bib/qf.i3.22727](http://dx.doi.org/10.33115/udg_bib/qf.i3.22727)

*«Ningún ordenamiento procesal moderno renuncia a orientarse hacia la búsqueda de la verdad de los hechos, y ningún ordenamiento procesal moderno confía exclusivamente en la iniciativa de las partes para conseguir ese objetivo».*

TARUFFO<sup>1</sup>

## 1. INTRODUCCIÓN

Una de las premisas esenciales en que se sustenta la concepción racionalista de la prueba es la que reconoce una relación teleológica entre prueba y verdad, que permite afirmar que la función de la prueba es la determinación de la verdad sobre los hechos<sup>2</sup>.

Al respecto, FERRER (2021:18) sostiene que el concepto de verdad que resulta útil para efectos de tal relación, es el de *verdad como correspondencia* que permite sostener que un enunciado sobre los hechos (formulado en el proceso y sujeto a prueba) es verdadero si y solo si, se corresponde con el mundo (externo del proceso)<sup>3</sup>

<sup>1</sup> TARUFFO, 2008: 179.

<sup>2</sup> En cuanto a la relación entre prueba y verdad véase FERRER, 2005: 54-78.

<sup>3</sup> TARUFFO (2010:99-100) sostiene que un enunciado es verdadero o no es verdadero, que no puede ser «más o menos verdadero»; que aquello que puede variar es el grado de confirmación que se puede atribuir a un enunciado sobre la base de los conocimientos disponibles y que, en ese sentido, podrá

De modo que sostener que un enunciado sobre los hechos está probado es afirmar que es probablemente verdadero, en razón de las pruebas disponibles, pues no se desconoce que, tal como lo afirma el referido autor, otro lugar común de la concepción racionalista de la prueba, es el relativo a que el razonamiento probatorio es probabilístico y, en consecuencia, no es posible alcanzar certezas racionales acerca de la verdad de los hechos.

En consecuencia, existen razones válidas para considerar que uno de los fines esenciales del proceso es el descubrimiento de la verdad, así como que es una condición necesaria para la corrección jurídica de la decisión, como afirma TARUFFO (2010: 132 ss.), en tanto se debe fundar en una determinación verdadera de los hechos del caso; de suerte que el proceso es el medio, no el fin y por lo cual

La cuestión no consiste, entonces, en excluir en absoluto cualquier límite o reserva a la búsqueda de la verdad en el ámbito del proceso. Se trata más bien, de evaluar qué límites y reservas están justificadas y cuáles en cambio —además de resultar epistemológicamente contraindicadas— carecen de una justificación ideológica aceptable (TARUFFO, 2010: 158).

En ese sentido, TARUFFO (2008: 176) añade que si no se reconoce importancia tanto a la determinación de la verdad en el proceso como a la calidad de la decisión que en este se adopta, el proceso estaría reducido a un mera “representación ritual” en el que se haría creer a las partes y a la sociedad que se ha hecho justicia a fin de legitimar la decisión. Por otra parte, sostiene que si la finalidad del proceso se reduce a la mera solución de la controversia, existen otros caminos más rápidos y eficaces para alcanzar ese objetivo.

De modo que es factible asumir que el objetivo fundamental de la actividad probatoria en el proceso judicial es la averiguación de la verdad de los enunciados fácticos. Sin embargo, tanto FERRER (2017: 94-95) como TARUFFO (2010: 156) afirman que no es la única finalidad del proceso, pues en este también cobran importancia otras funciones tales como la celeridad en la toma de las decisiones, el secreto de determinadas comunicaciones, el respeto por ciertos valores, y se efectúan elecciones económicas, se enfrentan problemas sociales, entre otros.

---

existir, por tanto, una mayor o menor aproximación a la verdad, de acuerdo al contexto y según las circunstancias. De ahí que el conocimiento de la verdad es el que puede calificarse como relativo y no la verdad en sí misma. Lo anterior a propósito de algunas distinciones formuladas entre verdad absoluta y relativa en el proceso.

Asimismo, detalla que carece de sustento la distinción entre verdad formal y verdad procesal que radica en que existen normas que limitan o condicionan de alguna forma la determinación de los hechos o bien que ponen fin a la búsqueda de la verdad y que, por tanto, fuera del proceso la búsqueda de la verdad “verdadera” puede efectuarse libremente. Ello, sobre la idea de que no existen varias especies de verdad según si estamos dentro o fuera del proceso, pues la verdad de los enunciados sobre los hechos depende de la realidad de estos hechos, lo que puede acontecer tanto dentro como fuera del proceso. Sumado a que la existencia de reglas que limiten el descubrimiento de la verdad, solo producen un déficit en la determinación de la verdad, pero ello no trae consigo que exista una verdad procesal, sino solo una verdad limitada o incompleta o bien que no se determine ninguna verdad; y que, por tanto, el problema no es la verdad, sino los límites que la propia regulación del proceso impide que sea determinada.

Al respecto, cabe aclarar que, tal como afirma VAN FRASSEN, «decir que algo es la finalidad de una institución o actividad, no excluye que existan otras finalidades o propósitos» (FERRER, 2017: 31).

Asumido que la averiguación de la verdad es objetivo central de la actividad probatoria, se deben extraer las consecuencias de esa asunción para cada uno de los momentos de esta actividad<sup>4</sup>. En este sentido, como sostiene FERRER (2007: 68) el primer momento de dicha actividad probatoria, esto es, « la conformación del conjunto de elementos de juicio o pruebas» habrá de tener como meta el obtener o adquirir el mayor número de pruebas relevantes posible y de la mayor fiabilidad, sobre la idea (retomando a LAUDAN) de que los errores judiciales en los que se pudiera incurrir respecto de los hechos disminuirán en razón de que el juez disponga de la mayor cantidad y calidad posible de elementos de juicio relevantes del caso. Al respecto, FERRER (2007: 68) añade que constituye un principio epistemológico indiscutible que «cuanta más información relevante esté a disposición de quien debe decidir, mayor probabilidad de acierto en la decisión», lo que lleva a considerar que el proceso ha de ser diseñado de modo que facilite la incorporación del mayor número de pruebas relevantes y, por tanto, debemos esclarecer y/o determinar cuáles son aquellos mecanismos procesales que sirven de herramienta para ello.

Frente a tal escenario se plantea el problema de determinar qué es aquello que el juez puede o no hacer para allegarse del mayor número de elementos de juicio (pruebas) que le permitan alcanzar la determinación de la verdad de los hechos (de manera supletoria y complementaria respecto de la actividad probatoria de las partes) en donde el tema de los poderes probatorios para incorporar pruebas no ofrecidas por las partes juega un rol epistémico importante.

Es por todo ello que el presente trabajo tiene como objetivo central analizar si en México el “juicio de amparo” se erige como un modelo de proceso judicial que exige de la institución probatoria el que se dirija hacia la averiguación de la verdad, para lo cual será necesario: analizar cuáles son los mecanismos procesales con los que cuenta el juez de amparo a efecto de averiguar la verdad (de manera supletoria y complementaria a la actividad probatoria de las partes) sobre los enunciados fácticos en que se sustenta la contienda constitucional; identificar las implicaciones prácticas que presenta su ejercicio; mostrar la insuficiencia de algunos argumentos que han obstaculizado su despliegue, como son la pérdida de imparcialidad y violación al principio de aportación de parte; y, finalmente, sugerir una modificación a la ley de la materia, a fin de garantizar y fortalecer su debida práctica, a través de la configuración de los límites respectivos.

Es importante precisar que, en México, el “juicio de amparo” es «un medio de control de la constitucionalidad, convencionalidad y legalidad de los actos, normas y omisiones que provienen de las autoridades y de los particulares que actúan como

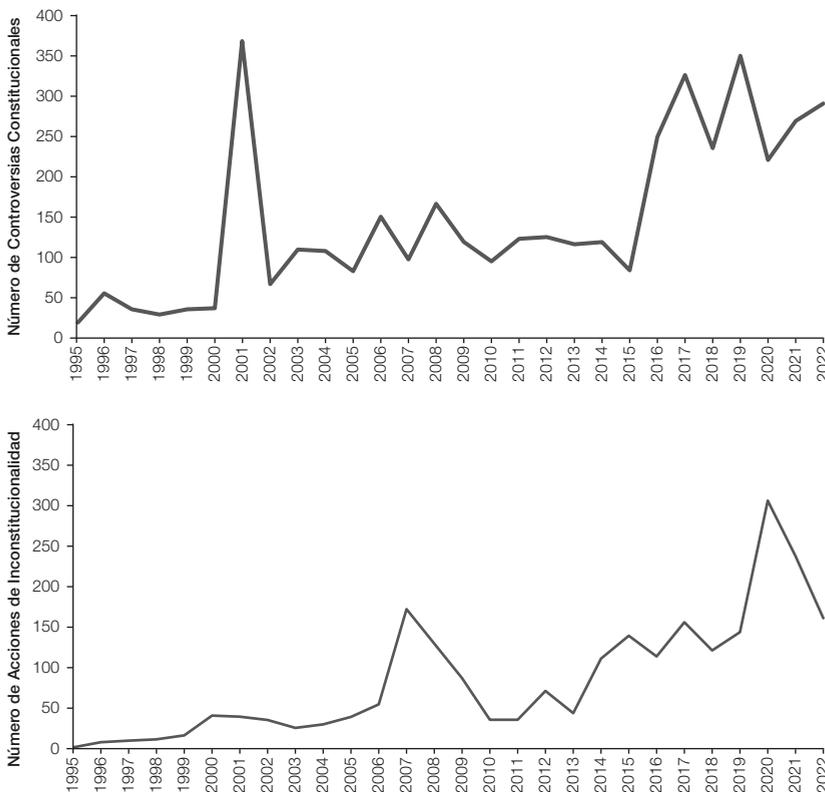
---

<sup>4</sup> Según FERRER (2021: 22-23) son tres los momentos de la actividad probatoria: 1) El momento de la conformación del conjunto de elementos del juicio o acervo probatorio; 2) El momento de la valoración de la prueba y 3) El momento de la decisión sobre la prueba.

autoridades, el cual se sigue en forma de juicio ante los tribunales federales» (CAMPUZANO, 2015: 1).

El juicio de amparo es el medio de control constitucional más utilizado en México en comparación con diversos medios de control constitucional. Para dar cuenta de este aspecto, a modo de ejemplo, se señala que como resultado de una solicitud de información pública formulada al Consejo de la Judicatura Federal se reportó que en 2019, ingresaron 556.928 demandas de amparo indirecto; mientras que, durante ese mismo año, lo hicieron 350 controversias constitucionales y 150 acciones de inconstitucionalidad.

Es más, cabe destacar que en 2019 ha sido el año en que el número de controversias constitucionales tramitadas ha sido el más alto de los últimos tiempos, información que se concluye de las siguientes gráficas elaboradas por la Unidad General de Transparencia y Sistematización de Información Judicial de la SCJN con información del Sistema de Control y Seguimiento de Expedientes<sup>5</sup>:



<sup>5</sup> Consultables en: <https://estadisticajudicial.scjn.gob.mx/alex/analisisControversias.aspx>

Una vez visibilizada la importancia que cobra la tramitación del juicio de amparo en México debido al alto número de personas que recurren a su promoción ante una eventual vulneración de derechos fundamentales, es que se considera necesario, tal como se adelantó, verificar si de conformidad con el marco legal vigente, el juicio de amparo se erige como un modelo de proceso judicial que exige de la institución probatoria el que se dirija hacia la averiguación de la verdad, para lo cual el tema de los poderes probatorios del juez juega un papel importante en tal cometido.

Se aclara que, para efectos del presente trabajo, el acento se pone en el “juicio de amparo indirecto” y no así en el “juicio de amparo directo”, toda vez que en este último el despliegue probatorio, por regla general, es nulo<sup>6</sup>.

## 2. LAS PARTES: SUJETOS EN QUIENES DESCANSA, DE MANERA PRIORITARIA, LA ACTIVIDAD PROBATORIA EN EL JUICIO DE AMPARO

El juicio de amparo indirecto, por regla general, se inicia con la presentación de la demanda, en la que, entre otras cuestiones, de conformidad con lo dispuesto en el artículo 108 de la Ley de Amparo (en adelante LA) se habrá de señalar: el acto reclamado, la autoridad responsable a la que se atribuye su emisión, «los hechos o abstenciones que constituyan los antecedentes del acto reclamado o que sirvan de fundamento a los conceptos de violación». y los preceptos legales en los que se contengan los derechos o garantías cuya violación se reclama.

El Pleno de la Suprema Corte de Justicia de la Nación (en adelante SCJN) ha establecido que quien promueva juicio de amparo indirecto podría ubicarse en dos supuestos:

### 1. Aducir un *interés jurídico*, para lo cual el promovente tendría que demostrar<sup>7</sup>:

---

<sup>6</sup> En el amparo directo, el acto reclamado solo podrá ser una sentencia definitiva, laudo o resolución que pone fin a un juicio, de modo que el tribunal colegiado que conoce de este, únicamente analiza su constitucionalidad a la luz de las pruebas que ya fueron desahogadas en el juicio de origen ante la autoridad jurisdiccional que dictó dicha sentencia, laudo o resolución, de modo que el ofrecimiento probatorio a cargo de las partes es nulo. Ello ha producido incluso que algunos procesalistas lo consideren, más que un juicio constitucional, una tercera instancia.

Por otra parte, en el juicio de amparo indirecto el ofrecimiento probatorio, por regla general, no está limitado, ya que el acto reclamado en este medio de control constitucional, a diferencia del amparo directo, no se ciñe a meras resoluciones jurisdiccionales, pues entre los múltiples supuestos de procedencia se encuentra, por ejemplo, la posibilidad de reclamar la inconstitucionalidad de una norma general, como lo sería una ley federal, caso en el cual el quejoso habría de demostrar, mediante las pruebas que considere conducentes para ello, la titularidad de un derecho y la afectación a este, ya sea con motivo de la mera entrada en vigor de dicha ley (autoaplicativa) o en razón de un acto de aplicación de la ley en cuestión (heteroaplicativa).

<sup>7</sup> Así, por ejemplo, una persona que reclama la falta o ilegal emplazamiento a un juicio (tercero extraño por equiparación) en el que se le ha condenado a la entrega o devolución de un bien inmueble

- a) Ser titular de un *derecho subjetivo*; y,
- b) Una afectación inmediata y directa en dicho derecho con motivo del acto de autoridad.

2. Alegar un *interés legítimo*, para lo cual el promovente tendría que acreditar<sup>8</sup>:

- a) La existencia de un *derecho objetivo*; y,
- b) La existencia de una afectación cierta a la esfera jurídica del promovente (afectación real y actual en virtud de su especial situación frente al orden jurídico).

Ahora bien, una vez admitida a trámite la demanda, el juez señalará fecha y hora para la audiencia constitucional y pedirá a la autoridad responsable que rinda *informe con justificación*, en el que, entre otras cuestiones, se habrá de manifestar en relación con la existencia o no del acto reclamado; las razones y fundamentos para sostener la improcedencia del juicio y acompañará a este las constancias necesarias para apoyar tales aspectos<sup>9</sup>.

Incluso, la LA dispone que si la autoridad responsable no rinde el informe respectivo, *se presumirá* cierto el acto reclamado (salvo prueba en contrario) quedando a cargo del quejoso acreditar su inconstitucionalidad, cuando dicho acto no sea por sí mismo violatorio de derechos humanos.

Cabe destacar que la parte quejosa no podría estar liberada de demostrar la afectación que dice sufrir con motivo del acto de autoridad que reclama por el hecho de que la autoridad responsable reconociera, al rendir su informe justificado, la existencia del acto reclamado, en virtud de que una cosa es la *existencia del acto reclamado en sí mismo* y otra el *perjuicio o agravio que le depara a la persona en concreto*.

Del marco legal antes expuesto es dable sostener que son las partes en quienes descansa, de manera prioritaria, la actividad probatoria en el juicio de amparo, pues tienen la carga de demostrar o desvirtuar tanto la existencia o no del acto reclamado, como la afectación al derecho fundamental que se dice vulnerado.

---

que habita, tendría que demostrar para efectos del juicio de amparo, por una parte, la existencia de un derecho real a su favor en relación con dicho inmueble, como podría ser la propiedad, a través de la escritura pública en la que se contenga el contrato de compraventa de este; y, por otra parte, la afectación que está sufriendo al ser desposeído de un bien de su propiedad sin haber sido oído y vencido en juicio.

<sup>8</sup> Por ejemplo, si una persona se dice afectada en su derecho a un ambiente sano con la construcción de un complejo hotelero en una zona de manglares, cuya licencia de impacto ambiental considera fue emitida sin observar los requisitos legales a los que está sujeta su expedición, tendría que demostrar el interés diferenciado (cualificado) que guarda en relación con el resto de la población (quienes también están interesados en la protección del medio ambiente) y acreditar que habita o es usuaria de los servicios ambientales en la zona de influencia de ese ecosistema. En casos análogos, la SCJN ha determinado que no sería necesario demostrar el daño al medio ambiente o los servicios ambientales pues, en todo caso, y atendiendo al principio de precaución, tal daño constituirá la materia de fondo del juicio del amparo. Véase amparo en revisión 307/2016 de la Primera Sala de la SCJN.

<sup>9</sup> Art. 117 LA.

Ahora bien, resulta interesante cuestionarse *si la actividad probatoria que despliegan las partes en el amparo tiene la finalidad de la búsqueda de la verdad, por ser uno de los fines esenciales del proceso.*

La respuesta es que no necesariamente y ello obedece a que las partes, como en cualquier otro proceso, defienden sus propios intereses, los cuales no siempre coinciden con la verdad. En la defensa de tales intereses, como afirma FERRER (2007: 39), las partes pueden manipular el material probatorio (ejemplo: introduciendo un testimonio falso) o dejar de presentar medios de prueba que, aun cuando pudieran ser útiles para la búsqueda de la verdad, pondrían resultar perjudiciales para tales intereses. Lo que permite afirmar que, al menos en ciertos casos<sup>10</sup>, la actividad de las partes en el proceso podría generar una cierta limitación al conocimiento sobre la verdad de los hechos.

En esa misma línea, TARUFFO (2010: 192 ss.) explica que usualmente el fin de las partes no es otro que ganar el juicio, sirviéndose del “derecho de prueba”, conforme al cual pueden valerse de todos los medios posibles para intentar demostrar el fundamento de sus alegaciones y enfocando su interés en demostrar lo que tienen la carga de probar; fin que se persigue, según detalla, con independencia de que la decisión se funde en una determinación verdadera de los hechos, sumado a que no tienen obligación o deber de respetar la verdad en el ámbito de dicha actividad. Así pues, se puede descartar que las partes tengan un interés común en que la verdad sea determinada.

Ante esta situación, la posibilidad de que el juez ordene de forma oficiosa pruebas no aportadas por las partes parece constituir una herramienta necesaria (aunque no suficiente) para el conocimiento de la verdad, en tanto permitiría contrarrestar, o al menos minimizar, los estragos que genera la defensa de los intereses de las partes durante la sustanciación del juicio, la cual, como se ha visto, en algunos casos, lleva a manipular o a omitir medios de prueba que ponen al descubierto situaciones que generan un perjuicio a las partes, con la consecuencia de distorsionar la verdad.

En ese sentido, resulta necesario identificar, como se hará en el próximo apartado, *con qué poderes probatorios cuenta el juez de amparo que le permitan maximizar las probabilidades de averiguar la verdad, frente a la eventualidad de que las partes, en defensa de sus intereses, intenten distorsionarla.*

---

<sup>10</sup> Los casos a los que se hace referencia en este apartado son aquellos juicios de amparo en los que el acto reclamado no tiene como origen o antecedente algún procedimiento (en el que las partes ya han ejercido su derecho de ofrecer y desahogar pruebas ante la autoridad responsable) y, en consecuencia, la posibilidad de ofrecer y desahogar pruebas se actualiza de manera plena en el propio juicio de amparo, como por ejemplo cuando se reclama una desaparición forzada de personas, de modo que existe el riesgo de que se pueda manipular el material probatorio o se pueda dejar de presentar alguno.

### 3. LOS PODERES DE INSTRUCCIÓN DEL JUEZ DE AMPARO

En este apartado se dará cuenta pormenorizada de todos los poderes probatorios con los que legal y jurisprudencialmente cuenta el juez de amparo, entre ellos, el de incorporar pruebas no solicitadas por las partes como un medio para maximizar la averiguación de la verdad, lo que permitirá identificar más adelante, tanto las implicaciones prácticas de dicho poder, como justificar que no tiene sentido que, por un parte, se pretenda negar dicha potestad y, por otra, se reconozca a su favor diversos poderes en materia de prueba, en el entendido de que cualquier potestad probatoria puede dar lugar a pérdida de imparcialidad.

Para dar cuenta de tales poderes se atenderá a la clasificación (no necesariamente exhaustiva) que ha formulado FERRER (2017: 97 y ss.) y conforme a la cual identifica entre los principales poderes probatorios de un juzgador los siguientes seis:

#### 3.1. Potestad de admitir o desechar pruebas

Dicha potestad está prevista a favor del juez de amparo, sin embargo, no es irrestricta, pues está sujeta a dos reglas: 1) Que la prueba ofrecida por alguna de las partes esté reconocida en la ley (a excepción de la confesional por posiciones<sup>11</sup>) y 2) que la prueba «necesariamente tenga relación inmediata con los hechos controvertidos, que no es otra cosa que el principio de idoneidad de la prueba.»<sup>12</sup>. Reglas cuya inobservancia permiten al juzgador desechar pruebas desde su anuncio y no reservarse hasta la celebración de la audiencia constitucional.

#### 3.2. La capacidad de intervenir en la práctica de la prueba (distinta a la de incorporación oficiosa de prueba)

El artículo 80 de la codificación supletoria faculta al juez a decretar, en todo tiempo, la repetición o ampliación de cualquier diligencia probatoria, siempre que se estime necesaria y sea conducente para *el conocimiento de la verdad*.

Así, por ejemplo, tratándose de la prueba pericial, el juez podrá formular todas las aclaraciones que estime conducentes y podrá exigirles la práctica de nuevas diligencias. Lo que en términos análogos se dispone para el caso de la prueba testimonial<sup>13</sup>.

Asimismo, cuenta con la potestad para requerir a alguna autoridad omisa la aportación de documentos que las partes le hubieran solicitado con anterioridad

---

<sup>11</sup> Art. 119 LA.

<sup>12</sup> Jurisprudencia P./J. 41/2001

<sup>13</sup> Art. 148 y 79 de la codificación supletoria.

para exhibirlos como prueba en el juicio de amparo<sup>14</sup>; potestad que puede ejercer también tratándose de particulares (ajenos a la *litis* constitucional) para que exhiban documentos que obren en su poder<sup>15</sup>.

### 3.3. La capacidad de indicar a las partes lagunas probatorias que estas deberían integrar, pudiendo incluso determinar qué pruebas concretas deberían aportar al proceso

En la LA no se prevé algún precepto que faculte al juez a indicar a las partes la posibilidad de que estas colmen lagunas probatorias. Sin embargo, por vía jurisprudencial<sup>16</sup> se ha establecido que si el quejoso omite señalar la fecha en que tuvo conocimiento del acto reclamado, el juez debe prevenirlo para que subsane tal deficiencia. Esta potestad contribuye a integrar una laguna probatoria, en la medida en que tal dato está vinculado con un hecho que el quejoso afirma, concerniente al conocimiento del acto reclamado y que permitirá establecer, en su momento, la oportunidad de la acción constitucional.

### 3.4. La capacidad del juez de disponer la incorporación de pruebas no solicitadas por las partes

El desarrollo de este punto implicará un análisis más minucioso, en tanto pretende ser la columna vertebral de la postura crítica que se desarrollará más adelante.

La LA no contiene disposición normativa alguna que prevea expresamente la capacidad del juez de incorporar a juicio pruebas no aportadas por las partes, salvo en dos supuestos específicos:

a) Cuando se reclamen actos que tengan o puedan tener como consecuencia privar de la propiedad o de la posesión y disfrute de sus tierras, aguas y pastos y montes a los ejidos o a los núcleos de población que de hecho o por derecho guarden el estado comunal, o a los ejidatarios o comuneros (artículo 75, último párrafo).

b) Cuando resuelva sobre la suspensión definitiva del acto reclamado, pues el artículo 143 lo faculta expresamente para solicitar documentos y ordenar las diligencias que considere necesarias.

Asimismo, de forma implícita, tratándose de *desaparición forzada de personas* al disponerse que el juez «requerirá a las autoridades correspondientes toda la información que pueda resultar conducente para la localización y liberación de la probable víctima»<sup>17</sup>.

<sup>14</sup> Art. 121 LA.

<sup>15</sup> Jurisprudencia 1a. /J. 28/2009.

<sup>16</sup> Jurisprudencia 2a. /J. 161/2013 (10a.).

<sup>17</sup> Art. 15 LA.

Vinculado con tal potestad, la LA también prevé que frente a la eventualidad de que alguna o todas las constancias del expediente hayan sido extraviadas (incidente de reposición de constancias de autos) «el juzgador está facultado para investigar de oficio la existencia de las piezas de autos desaparecidas, valiéndose para ello de todos los medios de prueba admisibles en el juicio de amparo y ley supletoria»<sup>18</sup>.

No pasa inadvertido que en relación con el tema de la incorporación oficiosa de pruebas el art. 75 de la LA también dispone:

“Artículo 75. En las sentencias que se dicten en los juicios de amparo el acto reclamado se apreciará tal y como aparezca probado ante la autoridad responsable. No se admitirán ni se tomarán en consideración las pruebas que no se hubiesen rendido ante dicha autoridad.

No obstante lo dispuesto en el párrafo anterior, en el amparo indirecto el quejoso podrá ofrecer pruebas cuando no hubiere tenido oportunidad de hacerlo ante la autoridad responsable. Adicionalmente, en materia penal, el juez de distrito deberá cerciorarse de que este ofrecimiento en el amparo no implique una violación a la oralidad o a los principios que rigen en el proceso penal acusatorio.

*El Órgano jurisdiccional deberá recabar oficiosamente las pruebas rendidas ante la responsable y las actuaciones que estime necesarias para la resolución del asunto.* En materia penal, se estará a lo dispuesto en la última parte del párrafo anterior” (Lo destacado es propio).

En cuanto al sentido y alcance de dicho numeral, la Primera Sala de la SCJN al resolver la contradicción de tesis 171/2020, determinó que su primer párrafo contiene el denominado “*principio de limitación de la prueba*”, el cual entraña dos obligaciones a cargo del juzgador: 1) lo obliga a apreciar el acto reclamado en la forma misma en que lo haya tenido por apreciado la autoridad responsable; y, 2) obliga al juzgador a no admitir ni tomar en consideración pruebas que no se hayan rendido ante dicha autoridad. Lo anterior, con la finalidad de «evitar que la jurisdicción de amparo se sustituya en la autoridad responsable y, en su lugar, garantizar que solo realice una revisión de la constitucionalidad del acto reclamado»<sup>19</sup>, pues el juicio de amparo no es una extensión del juicio de origen, ni debe confundirse con este.

De modo que «la regla y no la excepción, es la imposibilidad del ofrecimiento de pruebas para alterar la percepción de la responsable del acto reclamado»<sup>20</sup> y la única excepción se actualizará en «aquellos casos en que el oferente no haya tenido oportunidad de ofrecer pruebas ante la responsable»<sup>21</sup>, esto es, cuando, por ejemplo,

---

<sup>18</sup> Art. 71 LA.

<sup>19</sup> Párrafo 43 de la ejecutoria de la CT 171/2020.

<sup>20</sup> *Ibid.*, párrafo 53.

<sup>21</sup> Esta excepción, según aclaró la Primera Sala, tiene una prohibición tajante en materia penal y se actualiza cuando tal ofrecimiento probatorio vulnera la oralidad o los principios del procedimiento acusatorio, lo que acontece, por ejemplo, cuando en un juicio de amparo indirecto promovido por el probable responsable de la comisión de un delito se reclama una orden de aprehensión dictada bajo el sistema procesal penal acusatorio y se pretende, durante la sustanciación del juicio de amparo, ofrecer pruebas tendentes a demostrar su inconstitucionalidad, las cuales no fueron rendidas ni examinadas por el juez de control.

se trate de una persona ajena al procedimiento de creación del acto reclamado, o cuando la ley que lo rija no establezca la posibilidad de ofrecer y desahogar pruebas.

Lo que permite sostener que si bien el numeral en comento prevé expresamente que el juez de amparo puede recabar pruebas de forma oficiosa, lo cierto es que su ejercicio, como se vio, es excepcional e incluso está vedado (salvo las excepciones previstas en el propio numeral) en aquellos casos en los que el juicio de amparo tenga como antecedente u origen un procedimiento de corte jurisdiccional o un procedimiento seguido en forma de juicio, esto es, procedimientos cuya nota distintiva es el respeto a las formalidades esenciales del procedimiento, entre las que destaca, el derecho de ofrecer y desahogar pruebas.

Limitación probatoria a cargo del juez que, cabe aclarar, no se extiende a aquellos otros casos en los que el acto reclamado no tiene como antecedente u origen algún juicio o procedimiento, como cuando se reclama, por ejemplo, una desaparición forzada de personas o bien el reclamo de una ley por considerarse inconstitucional el precepto en el que se funda el cobro de alguna contribución.

De ahí que no es factible considerar que el art. 75 de la LA dispone de una facultad plena a favor del juez de amparo a fin de incorporar pruebas distintas a las ya ofrecidas por las partes (salvo por la limitada excepción que se menciona). No obstante dicha omisión, los arts. 79 y 80 de la codificación civil supletoria prevén lo siguiente:

ARTÍCULO 79. Para conocer la verdad, puede el juzgador valerse de cualquier persona, sea parte o tercero, y de cualquier cosa o documento, ya sea que pertenezca a las partes o a un tercero, sin más limitaciones que las de que las pruebas estén reconocidas por la ley y tengan relación inmediata con los hechos controvertidos.

Los tribunales no tienen límites temporales para ordenar la aportación de las pruebas que juzguen indispensables para formar su convicción respecto del contenido de la litis, ni rigen para ellos las limitaciones y prohibiciones, en materia de prueba, establecidas en relación con las partes.

ARTÍCULO 80. Los tribunales podrán decretar, en todo tiempo, sea cual fuere la naturaleza del negocio, la práctica, repetición o ampliación de cualquier diligencia probatoria, siempre que se estime necesaria y sea conducente para el conocimiento de la verdad sobre los puntos controvertidos. En la práctica de esas diligencias, obrarán como lo estimen procedente, para obtener el mejor resultado de ellas, sin lesionar los derechos de las partes, y procurando en todo su igualdad.

Sumado a ello, en criterios jurisprudenciales la SCJN ha establecido que en algunos asuntos el juez debe allegarse de pruebas oficiosamente, tales como:

a) Ante la existencia de algún indicio de que en el juicio se actualiza alguna causal de improcedencia<sup>22</sup>;

b) Asuntos cuyo planteamiento exija juzgar con perspectiva de género<sup>23</sup>;

---

Sin embargo, cabe aclarar que en el presente trabajo no será materia de análisis tal prohibición en materia penal.

<sup>22</sup> Jurisprudencia 1a./J. 163/2005

<sup>23</sup> Jurisprudencia 1a./J. 22/2016

- c) Menores de edad o incapaces<sup>24</sup>;
- d) Actos de tortura<sup>25</sup>;
- e) Para proveer respecto de la suspensión provisional<sup>26</sup>.

### 3.5 La capacidad de alterar la carga de la prueba durante el desarrollo del proceso (carga dinámica de la prueba<sup>27</sup>)

En la LA no se prevé alguna potestad a cargo del juez de alterar, en el juicio de amparo indirecto, la carga de la prueba caso a caso (carga dinámica de la prueba). Sin embargo, jurisprudencialmente sí se han establecido ciertas reglas generales, en cuanto a la “*inversión de la carga de la prueba*”. Ejemplo de esto último es lo resuelto en el amparo en revisión 307/2016 por la Primera Sala de la SCJN, en el que se determinó que cuando en un juicio de amparo se reclamen violaciones al derecho a un ambiente sano causadas por la afectación de los recursos naturales, en atención al «principio de precaución<sup>28</sup>», puede revertirse la carga de la prueba a cargo del agente potencialmente responsable y, en consecuencia, el juez puede ordenar todos los medios de prueba necesarios para identificar el riesgo o daño ambiental.

Por otra parte, no se desconoce que jurisprudencialmente, pero tratándose d el “juicio de amparo directo”, también se han establecido ciertas reglas en cuanto a la “*inversión de la carga de la prueba*”, a la luz de los «principios de disponibilidad y facilidad probatoria», concebidos solo como un «deber de aportar prueba» tratándose, por ejemplo, de casos como el bullying en los centros escolares en los que existe dificultad de probar un hecho negativo a cargo de la víctima y son tales centros quienes tienen mejor forma de demostrar que se ha cumplido con los deberes que demanda tener menores bajo su cuidado<sup>29</sup>.

<sup>24</sup> Jurisprudencia 1a./J. 191/2005

<sup>25</sup> Tesis 1a. CCCXXVI/2018

<sup>26</sup> Jurisprudencia 2a. /J. 27/2004

<sup>27</sup> FERRER (2019: 63 ss.) sostiene que la concepción dinámica de la carga de la prueba tiene sus antecedentes doctrinales y jurisprudenciales en Argentina con Peyrano y que en atención a ella el juez tiene la potestad de alterar la carga de la prueba en función de su evaluación, en el caso concreto, sobre cuál de las partes tiene mejor acceso a la prueba o mayor facilidad para producirla; y que esa posibilidad en ocasiones es utilizada en el momento de dictar sentencia, esto es, cuando la parte perjudicada ya no tiene la posibilidad de aportar pruebas que le permitan satisfacer la carga, con una grave afectación al principio de seguridad jurídica.

<sup>28</sup> “En términos del artículo 15 de la Convención de Río sobre el Medio Ambiente y el Desarrollo, conforme al principio de precaución, cuando la experiencia empírica refleja que una actividad es riesgosa para el medio ambiente, resulta necesario adoptar todas las medidas indispensables para evitarla o mitigarla, aun cuando no exista certidumbre sobre el daño ambiental. Este principio demanda una actuación estatal ante la duda de que una actividad pueda ser riesgosa” (Tesis 1a. CCXCIII/2018 (10a.).

<sup>29</sup> Tesis 1a. CCCXXXI/2015

Reglas de “inversión de la carga de la prueba” que si bien rigen para el juicio ordinario que constituye el origen o antecedente del acto reclamado en el juicio de amparo directo, se estima, pueden ser adoptadas, por identidad de razón, tratándose del “juicio de amparo indirecto” frente algún problema probatorio que verse exclusivamente sobre la *litis* constitucional, como podría ser la existencia del acto reclamado o la improcedencia del juicio, en tanto que no se debe soslayar que en ciertos casos<sup>30</sup> se debe garantizar la estricta aplicación del «principio de limitación probatoria» que impide al juzgador incorporar pruebas que no fueron ofrecidas y desahogadas ante la autoridad responsable.

### 3.6. La imposición al juez del deber de decidir qué hipótesis fácticas se consideran probadas, unida a la ausencia de un estándar de prueba

A los efectos de este apartado es menester retomar que uno de los lugares comunes del razonamiento probatorio es que es probabilístico, de modo que no es factible alcanzar certezas racionales sobre la verdad de los hechos, por lo que se torna imprescindible determinar cuál es el grado de probabilidad suficiente para aceptar como probada una hipótesis (FERRER, 2021: 26), esto es, un estándar de prueba.

En relación con los estándares de prueba, FERRER (2021: 109) precisa que

Son reglas que determinan el nivel de suficiencia probatoria para que una hipótesis pueda considerarse probada (o suficientemente corroborada) a los efectos de una decisión sobre los hechos. Al realizar esa determinación, cumplen tres funciones de la máxima importancia en el marco del proceso de decisión probatoria: 1) aportan los criterios imprescindibles para la justificación de la decisión misma, por lo que hace a la suficiencia probatoria; 2) sirven de garantía para las partes, pues les permitirán tomar sus propias decisiones respecto de la estrategia probatoria y controlar la corrección de la decisión sobre los hechos, y 3) distribuyen el riesgo del error entre las partes.

El nivel de exigencia o severidad de un estándar de prueba es una *decisión política*, que habrá de establecer el legislador. Esa decisión, como afirma LAUDAN (2013: 105) «no baja del cielo, no está de modo alguno implícito en la naturaleza del concepto de estándar de prueba. Más bien, su exigencia refleja una decisión colectiva por parte de la sociedad para ubicar el umbral referido en un punto y no en otro».

Ahora bien, tratándose del juicio de amparo, el legislador no ha establecido (al menos de manera expresa) ningún estándar de prueba, pues en la LA no se prevé algún precepto normativo que disponga el nivel de suficiencia probatoria que como tal se debe colmar.

<sup>30</sup> En aquellos en que el acto reclamado tiene como origen o antecedente un juicio o procedimiento seguido en forma de juicio, en el que haya tenido la oportunidad de ofrecer y desahogar pruebas.

No se desconoce que en la codificación supletoria<sup>31</sup> el legislador estableció, en cuanto a la «valoración probatoria», que el juez goza de la más amplia libertad para hacer análisis de las pruebas rendidas, para determinar el valor de estas y para fijar el resultado final de dicha valoración, a no ser que la ley fije las reglas para efectuar tal valuación en relación con cada especie de prueba. Sin embargo, la labor de un estándar de prueba no la puede suplantar la valoración probatoria que de las pruebas se haga, en tanto que, como señala FERRER (2021: 26), esta determina el grado de probabilidad o de corroboración de una hipótesis a la luz de las pruebas presentadas, pero nada dice sobre su «suficiencia».

De modo que lo más parecido (aunque incipientemente) al establecimiento de un grado de prueba mínimo para efectos probatorios en la LA está en su art. 113, al prever que el juez que conozca del juicio de amparo indirecto examinará el escrito de demanda y si existiera causa «manifiesta» e «indudable»<sup>32</sup> de improcedencia la desechará de plano; así como con lo complementado jurisprudencialmente al respecto, en el sentido de que la actualización de la causal de improcedencia debe probarse «plenamente» y no inferirse con base en presunciones<sup>33</sup>.

Al respecto, cabe aclarar que desde el momento de la presentación de la demanda el juez de amparo efectúa valoración probatoria, pues la prueba documental (a diferencia de las pruebas pericial, testimonial e inspección judicial que deben ser ofrecidas con 5 días hábiles de anterioridad a la fecha fijada para la audiencia constitucional) puede ofrecerse y exhibirse con motivo de ese primer escrito. De modo que es factible que el juez advierta la actualización de alguna causal de improcedencia de la lectura del escrito de demanda, de los escritos aclaratorios de esta o de los documentos que se anexan a esas promociones.

Por otra parte, también está vinculado (aunque incipientemente) con el establecimiento de un grado de prueba mínimo para efectos probatorios, lo que la SCJN ha establecido en relación con estos dos temas:

*a) Interés jurídico /legítimo*

En relación con la demostración a cargo del quejoso de ser titular de un derecho subjetivo y la afectación a este por un acto de autoridad (interés jurídico) o bien la acreditación de la titularidad de un derecho objetivo y la afectación a este por un acto de autoridad (interés legítimo) la SCJN ha referido que tales aspectos deben acreditarse «fehacientemente» y no inferirse con base en presunciones<sup>34</sup>.

*b) Interés suspensional*

<sup>31</sup> Art. 197.

<sup>32</sup> Por «manifiesto» debe entenderse lo que se advierte en forma patente, notoria y absolutamente clara, y por «indudable», que se tiene la certeza y plena convicción de algún hecho, esto es, que no puede ponerse en duda por lo claro, seguro y evidente que es. (CT 24/2005-PS).

<sup>33</sup> Tesis sin número, de rubro: «IMPROCEDENCIA DEL AMPARO. DEBE PROBARSE PLENAMENTE Y NO APOYARSE EN PRESUNCIONES».

<sup>34</sup> Jurisprudencia 1a./J.168/2007

La SCJN ha establecido que la suspensión provisional de los actos reclamados, solicitada por la parte quejosa, supone la demostración de su interés (jurídico o legítimo<sup>35</sup>) aunque sea «de forma indiciaria», a fin de establecer con suficiente garantía de acierto que realmente es titular de un derecho.

En estos supuestos, como se observa, se ha pretendido establecer un umbral que determine el grado de prueba mínimo haciendo uso de vocablos tales como: «manifiesto e indudable» para el caso de las causales de improcedencia que den lugar al desechamiento de la demanda; «fehacientemente» para la acreditación de la titularidad tanto de derechos subjetivos como objetivos y la afectación a estos (interés jurídico/legítimo); e «indiciariamente» para efectos del otorgamiento de medidas cautelares.

De manera que es factible sostener que se ha intentado instruir, en términos muy generales e imprecisos, al juez de amparo en cuanto al grado de corroboración que deben alcanzar las premisas fácticas. Sin embargo, con el uso de tales vocablos lo único que se puede colegir, sin mayor certeza, es la exigencia de un umbral “alto” tratándose de interés jurídico/legítimo y “bajo” tratándose del interés suspensorial para la suspensión del acto reclamado.

Por tanto, dada la falta de precisión en dichos umbrales, el juez de amparo tiene hasta cierto punto posibilidad de decidir en cada caso cuán alta debe ser la corroboración de una hipótesis fáctica para considerarla probada.

#### 4. ALGUNAS IMPLICACIONES PRÁCTICAS EN EL DESPLIEGUE DE LOS PODERES PROBATORIOS OFICIOSOS

Como ha quedado precisado, entre los poderes probatorios con los que cuenta el juez de amparo está el relativo a incorporar en el juicio pruebas no solicitadas por las partes, el cual será materia de análisis pormenorizado en el presente apartado, por dos razones:

a) Una, ya que es un tema que ha empezado a ser analizado más allá del derecho probatorio, con una perspectiva epistemológica, la cual no se limita a describir cómo deciden los juzgadores, sino, como afirma VÁZQUEZ (2015: 59) propone cómo hacerlo si lo que se pretende es alcanzar decisiones racionales sobre las controversias judiciales<sup>36</sup>.

<sup>35</sup> Jurisprudencia 2a./J.61/2016

<sup>36</sup> La actividad probatoria «es un evento fundamentalmente epistemológico, pues en dicho razonamiento están continuamente en juego cuestiones relacionadas con creencias, el conocimiento, las pruebas, los estándares de prueba, etc» de modo que la epistemología tiene mucho que decir sobre la prueba en tanto que «su objeto de estudio son precisamente las creencias, la justificación de las creencias, la verdad de las creencias y el conocimiento» VÁZQUEZ (2015: 59).

b) Dos, porque, pese a tener fundamento legal, el despliegue de los poderes probatorios del juez, en la práctica, encuentra muchas limitantes, entre estas, dos centrales:

b.1 El que no se les reconozca, ni llame por su nombre cuando son desplegadas y como consecuencia de ello se vulnere el derecho de defensa de las partes; y,

b.2 El que, vía jurisprudencial, se intenten restringir dicho poder probatorio a cargo del juzgador sobre el argumento toral de que la actividad probatoria debe centrarse solo en las partes.

Al respecto, es importante retomar que las actividades probatorias a cargo del juez de amparo están previstas en los arts. 79 y 80 de la codificación supletoria, numerales de cuyo contenido se pueden extraer las siguientes notas distintivas:

(i) *Facultad*: decretar la práctica de cualquier diligencia probatoria.

(ii) *Condición*: siempre que se estime necesaria y conducente para el conocimiento de la verdad sobre los puntos controvertidos.

(iii) *Oportunidad*: En todo tiempo.

(iv) *Límite a la incorporación probatoria*: siempre que no se lesionen los derechos de las partes y se procure en todo su igualdad.

(v) *Tipo de pruebas que puede ser incorporadas*: cualquiera, siempre que esté reconocida en la ley y tenga relación inmediata con los hechos controvertidos.

(vi) *Medios para ejercer tal potestad*: el juzgador puede valerse de cualquier persona, sea parte o tercero, y de cualquier cosa o documento, ya sea que pertenezca a las partes o a un tercero.

(vii) *Excepciones*: no rigen para el juez las limitaciones y prohibiciones que en materia de prueba están establecidas en relación con las partes.

(viii) *Finalidad*: Conocer la verdad sobre los puntos controvertidos (dimensión epistémica del proceso).

En la doctrina mexicana, como vía jurisprudencial, al ejercicio o despliegue de tales facultades a cargo del juez de amparo, se le ha denominado: *diligencias para mejor proveer*<sup>37</sup>.

En torno a dichas diligencias, vía jurisprudencia, la SCJN estableció que el juez no viola garantías individuales [*sic*] en perjuicio de las partes, cuando omite ordenar diligencias para mejor proveer, porque es una facultad que la ley le confiere, pero que una vez que ejerza dicha facultad, genera un derecho en favor de las partes, no solo

<sup>37</sup> Al respecto, el ex ministro MEDINA MORA (2017: 419) sostiene que: «las partes no son los únicos facultados para allegarse de material probatorio, ya que en ciertos casos, el legislador ha autorizado vía normativa un papel activo al juzgador... En ese tenor, las diligencias para mejor proveer... se traducen en la herramienta idónea con la que cuenta el juzgador para el esclarecimiento de los hechos controvertidos, para despejar, en mayor medida, las dudas acerca de la verdad objetiva que pretende alcanzar y plasmar en su sentencia al resolver el litigio que es sometido a su conocimiento».

para que las diligencias tengan lugar, sino para que su eficacia probatoria sea calificada en sentencia<sup>38</sup>.

Incluso, en relación con tales diligencias, pero vinculadas a diverso medio de control constitucional (controversias constitucionales) la SCJN ha sostenido dos cuestiones: 1) que el juzgador podrá ordenarlas “*en todo tiempo*”, es decir, desde la admisión de la demanda hasta el momento en que el ministro somete a consideración del Pleno el proyecto de resolución respectivo<sup>39</sup>; y 2) que el hecho de que un tercero aporte pruebas a la controversia puede tenerse como una auténtica diligencia para mejor proveer<sup>40</sup>.

En relación con el criterio previsto en el punto dos que antecede, LARA (2011: 92) afirma que tal pronunciamiento da lugar a que, en el marco de las diligencias para mejor proveer, los terceros al juicio puedan aportar pruebas a la manera de *amicus curie*. Esta afirmación resulta debatible si se toma en consideración que al menos en dos de los asuntos en los que el Máximo Tribunal del País ha desplegado sus poderes probatorios de manera oficiosa, de los cuales enseguida se dará cuenta, el tratamiento que se le ha dado a la incorporación de pruebas no ordenadas por las partes justo ha sido el de un *amicus curie* y no como diligencias para mejor proveer, lo que ha generado que sean excluidas de ciertas reglas probatorias, en detrimento del derecho de defensa de las partes.

a) Caso Militar con VIH (amparo en revisión 510/2004<sup>41</sup>).

Resuelto el 6 marzo 2007 y cuyos antecedentes son, en esencia: un militar que acude a la realización de un examen médico con el objeto de obtener una promoción al grado de Capitán Segundo de Infantería y con motivo de este se le detecta VIH. Derivado de ello, se emite un certificado médico en el que se le declaró “*inútil*” para el servicio, por lo que fue dado de baja, lo que a su vez generó privarlo de atención médica.

Inconforme con tal determinación, el quejoso promovió juicio de amparo, el cual fue sobreseído; sin embargo, en el amparo en revisión, se determinó que el juez había atribuido erróneamente los actos reclamados a autoridades diversas y en cuanto al problema de constitucionalidad remitió los autos a la SCJN. La *litis* se centró en determinar si es admisible dar de baja a un miembro del ejército por motivos de inutilidad, cuando es portador del VIH.

Durante la discusión del asunto varios ministros manifestaron haberse allegado de información científica para poder resolver el asunto. Así, por ejemplo el ministro Salvador Aguirre Anguiano mandó hacer un estudio médico para su información personal; el ministro Valls consultó un libro del doctor Jaime Sepúlveda sobre SIDA.

<sup>38</sup> Tesis de rubro: «DILIGENCIAS PARA MEJOR PROVEER».

<sup>39</sup> Jurisprudencia P. /J. 37/2002.

<sup>40</sup> Tesis P.CIX/95

<sup>41</sup> Véase COSSÍO, MEJÍA GARZA, CABAÑAS RIVERO, 2017.

Incluso, el ministro José Ramón Cossío, aprovechando un convenio de colaboración entre la SCJN y la Academia Mexicana de Ciencias, envió un cuestionario a efecto de que fuera respondido por especialistas en el tema, entre cuyas preguntas a manera ilustrativa se citan las siguientes:

¿Puede desempeñar un trabajo una persona que padece el Síndrome de la Inmunodeficiencia Humana, SIDA? ¿Existen limitaciones en el tipo de trabajo que puede desempeñar?

Sin embargo, el ministro Cossío también precisó que no estaba trayendo “*pruebas*” a juicio, sino solo allegándose de información para formarse “*convicción*” en el caso; sumado a que él no tenía el carácter de parte y que, por tanto, según los arts. 79 y 80 de la codificación supletoria no le eran aplicables las reglas de prueba que sí le resultaban aplicables a las partes.

Ello, frente al argumento de otros ministros tales como: Sergio Salvador Aguirre Anguiano, Marino Azuela Huitrón y Genaro Góngora Pimentel, quienes sostuvieron la necesidad de ordenar el desahogo de una prueba pericial, en forma, esto es, que siguiera las reglas procesales que rigen la referida prueba, a efecto de no violar el derecho de defensa de las partes; sin embargo, la mayoría de los ministros consideró que no era necesario y resolvieron lo siguiente:

No es la ausencia de salud lo que faculta al empleador para separar de sus funciones al subordinado, sino la incapacidad para poder llevar a cabo las tareas encomendadas que esa carencia genera. Esto último es a lo que, en todo caso, razonablemente debe atenderse para que el individuo abandone sus labores [...] <sup>42</sup>.

El sustento de esta determinación, según se lee en la propia sentencia, obedeció a la información médica allegada por miembros del Máximo Tribunal del país con base en el art. 79 de la codificación civil supletoria.

b) Caso “*El Encino*” (incidente de inejecución de sentencia 40/2003) <sup>43</sup>.

En el año dos mil se ordenó la expropiación de dos fracciones de un predio a fin de construir dos vialidades importantes (las cuales darían acceso a un hospital y a una universidad).

Inconforme, el propietario de dicho predio promovió juicio de amparo, el cual fue concedido para el efecto de que dichas fracciones le fueran devueltas, pues la autoridad responsable no logró demostrar que estas fueran la única opción para concluir y/o construir tales avenidas.

Ante el incumplimiento de la sentencia, el juez de distrito abrió un incidente de inejecución de sentencia, el cual fue declarado fundado. Determinación que fue confirmada por un tribunal colegiado de circuito y, en consecuencia, el asunto fue remitido a la SCJN, para efectos de determinar si en el caso era procedente el cumplimiento sustituto de la sentencia de amparo.

<sup>42</sup> Amparo en Revisión 510/2004.

<sup>43</sup> Véase OSORNIO, 2018: 327-345.

La Segunda Sala ordenó que los autos fueran devueltos al juez a fin de que ordenara la apertura de un incidente innominado, a fin de determinar si existía o no imposibilidad material o jurídica para devolver los terrenos. Al respecto, se concluyó que era factible su devolución, pues las avenidas que se pretendía construir aún no estaban concluidas.

El ministro Cossío Díaz concluyó que el análisis efectuado por el juez se ciñó al hecho de si las obras estaban concluidas o no, no obstante que debía hacerse una valoración más amplia, en la que se determinara si la ejecución de la sentencia afectaba gravemente a la sociedad o a terceros.

Atendiendo a ello, el ministro consideró necesario allegarse de lo que denominó: “*elementos de mayor convicción*” (no pruebas, según aclaró), en términos de lo previsto en el art. 79 del código supletorio. Asimismo, requirió al Rector de la UNAM para que propusiera tres expertos (en economía, arquitectura y urbanismo) a fin de que emitieran un “*estudio*” (no dictamen pericial), el cual, una vez elaborado, se ordenó agregar a los autos del incidente, sin que al respecto se corriera vista a las partes.

Cabe añadir que a tales expertos no se les pidió que acreditaran su profesión, ni que aceptaran ni protestaran el cargo en su calidad de expertos (peritos), así como tampoco ratificaron sus conclusiones (reglas a las que usualmente está sujeto un dictamen pericial)<sup>44</sup>.

Con base en lo expuesto en dicho estudio (porque se prescindió del resto del material probatorio que ya obraba en autos), se concluyó que en el caso, lo procedente era decretar el cumplimiento sustituto del fallo constitucional y que, en consecuencia, la autoridad debía efectuar el pago de daños y perjuicios o bien celebrar un convenio con la parte quejosa.

Como se observa, en los asuntos que han sido relatados con anterioridad, el ejercicio práctico de lo dispuesto en los artículos que regulan las diligencias para mejor proveer han traído como consecuencia los siguientes aspectos:

a) Que se distinga entre “*elementos de mayor convicción*” (entendidos como elementos con carácter meramente informativo que generan mayor convicción al juzgador para decidir) y “*pruebas*” (que serían las propiamente aportadas por las partes);

b) Que la inclusión de los “*elementos de convicción*” no esté sujeta a ninguna regla (tales como exigir que los profesionales que participaron en su elaboración acrediten su profesión);

c) Que no se da vista a las partes con las pruebas incorporadas a juicio;

d) La inutilización del resto del material probatorio.

---

<sup>44</sup> Durante los debates del asunto, el ministro ponente sostuvo: «con toda franqueza si me hubiera parecido una falta de respeto pedirle, en este carácter, al señor Rector que me acreditara la personalidad o la profesión de estas personas... primero, por la calidad de las instituciones con las que estamos interactuando; segundo, porque no estoy desahogando pruebas».

A partir de lo dicho es dable concluir que si bien se ha hecho un ejercicio práctico de las diligencias para mejor proveer, su despliegue no ha sido reconocido abierta ni expresamente, tan es así que se utilizan denominaciones como “*elementos de mayor convicción*”, lo que ha ocasionado que se les exente de sujetarse a las reglas procesales que salvaguardan los derechos de defensa de las partes.

Al respecto, cabe aclarar que aun cuando las diligencias para mejor proveer no constituyen en sí mismo una prueba, sino el mecanismo mediante el cual el juez ordena la incorporación de una prueba al proceso, lo cierto es que todas aquellas pruebas que se incorporan con motivo de su despliegue no están exentas de las reglas probatorias que estén previstas en la LA, en tanto que la única diferencia con las pruebas ofrecidas por las partes es solo el sujeto que las propone.

No pasa inadvertido que en términos del art. 79 de la codificación supletoria no rigen para el juzgador las *limitaciones y prohibiciones* que en materia de prueba están establecidas para las partes; sin embargo, ello, por sí mismo, no lo exime de atender las *reglas procesales* que regulan su desahogo, cuando ordena su incorporación de forma oficiosa.

Piénsese, por ejemplo, en la prueba testimonial. En relación con esta existen en la legislación de la materia, o bien se han establecido por vía jurisprudencial, reglas, limitantes y prohibiciones, a saber:

LIMITANTE	PROHIBICIÓN	REGLA
Una <i>limitante</i> que se ha establecido es la relativa a que cuando el oferente la anuncie y exceda el número de 3 testigos para un mismo hecho, el juzgador, oficiosamente, debe decretar su reducción <sup>45</sup> .	El plazo para el ofrecimiento de la testimonial <i>no podrá ampliarse</i> con motivo del diferimiento de la audiencia constitucional <sup>46</sup> .  *Salvo que se trate de hechos supervenientes.	El oferente de la prueba testimonial deberá, previamente a la celebración de la audiencia constitucional, proporcionar el nombre de las personas que rendirán su testimonio, a fin de que las partes en litigio constitucional, así como el Juez Federal, tengan conocimiento, con precisión, de las personas que lo prestarán y de esta manera estén en aptitud de determinar su idoneidad, permitiendo que se formulen y preparen oportunamente, por escrito o verbalmente, las repreguntas y, en su caso, la posible impugnación de dicha idoneidad <sup>47</sup> .

<sup>45</sup> Jurisprudencia 2a./J. 159/2011

<sup>46</sup> Art. 119 LA y jurisprudencia 2a. /J. 109/2016 (10a.)

<sup>47</sup> Jurisprudencia P./J. 75/2001

Por lo que si en términos de lo previsto en la codificación civil supletoria no rigen para el juzgador las limitaciones y prohibiciones que en materia de prueba están establecidas para las partes, tratándose de la testimonial, el juez solo estaría exento de limitar el número de testigos para corroborar un hecho, pero no así respecto de la obligación de dar a conocer las partes el nombre de quien daría su testimonio.

En ese sentido, se considera aun cuando la SCJN, en los ejemplos antes citados, ha hecho un despliegue de las diligencias para mejor proveer, no ha acatado las reglas procesales que rigen el desahogo de las pruebas que ha incorporado, ni tampoco ha observado, en detrimento del derecho de defensa de las partes, el *principio de contradicción*.

Por otra parte, tal como se adelantó, algunos tribunales, vía jurisprudencial, han intentado limitar los poderes probatorios a cargo del juez para lo cual se citarán dos ejemplos:

a) Tesis I.15o.C.9 K (10a.) del Décimo Quinto Tribunal Colegiado en Materia Civil del Primer Circuito, relativo a que los tribunales no deben invocar oficiosamente, con el carácter de “*hecho notorio*” las sentencias almacenadas y capturadas en el Sistema Integral de Seguimiento de Expedientes (SISE) a fin de constituir o perfeccionar una prueba deficientemente ofrecida para demostrar ser titular de un derecho y que este ha sido afectado (interés jurídico),

...so pena de violentar los principios de igualdad procesal y de contradicción, que posibilitan debatir sobre la prueba de la parte contraria, al trastocar la distribución de las cargas probatorias y, además, generaría inseguridad jurídica, por permitir que el juzgador, discrecionalmente, allegue a los autos datos no invocados ni aportados en el juicio natural para constituir o perfeccionar una prueba y resolver con base en esta sobre un presupuesto procesal *cuya demostración corresponde exclusivamente al promovente*.

En relación con este tema, es importante retomar que el Pleno de la SCJN<sup>48</sup> ya ha establecido que las sentencias almacenadas y capturadas en el Sistema Integral de Seguimiento de Expedientes (SISE) tienen el carácter de hechos notorios, de modo que en términos de lo previsto en el art. 88 de la codificación civil supletoria pueden ser invocados, aunque no hayan sido alegados ni probados por las partes. Lo cual obedece a que tal como afirma GASCÓN (2005: 104) se trata de «hechos conocidos (o que pueden ser conocidos) por todos, y por tanto, también por el juez; es decir, aquellos cuya existencia está plenamente acreditada y es de -dominio público-».

Como se observa, el criterio en mención desconoce lo resuelto por la SCJN, pues impide al juez de amparo citar con la calidad de “*hecho notorio*” el contenido de alguna sentencia de la que haya tenido conocimiento con motivo de utilizar el SISE.

Ahora bien, podría considerarse el supuesto de que ante tal imposibilidad el juez ordenara una diligencia para mejor proveer a fin de incorporar a juicio el contenido de dicha sentencia; sin embargo, ello no puede ser entendido de tal forma, si se toma en cuenta que el tribunal colegiado precisó que la demostración de la titularidad de un derecho y la afectación a este (interés jurídico) corresponde en exclusiva al pro-

<sup>48</sup> Jurisprudencia P. / J. 16/2018 (10a.).

movente del juicio, consideración con la cual excluyó al juez de cualquier despliegue de sus facultades probatorias.

Al respecto, es importante aclarar que si bien el criterio en comento no establece, al menos de forma expresa, una prohibición al despliegue oficioso de los poderes probatorios del juez, sí lo hace de forma implícita, al considerar que se trastocan las cargas probatorias si eventualmente se incorporan datos que perfeccionen una prueba deficientemente ofrecida por las partes, soslayando que precisamente esa es una de las finalidades de las que gozan las diligencias para mejor proveer, pues tal como lo afirma PABÓN (2019: 233) los poderes que se confieren al juez son “*eventuales*” y “*por vía de integración*”, ya que si las partes incorporaron al proceso todas las pruebas disponibles, el juez ya no ejercería ninguna iniciativa probatoria.

b) Tesis III.5o.A.14 K (10a.) del Quinto Tribunal Colegiado en Materia Administrativa del Tercer Circuito, en la que se determina que el juez de amparo no puede invocar, oficiosamente, como hecho notorio, la información obtenida de la consulta a una página de internet para constituir o perfeccionar una prueba deficientemente ofrecida para demostrar la titularidad de un derecho y su afectación a este (interés jurídico), pues ello corresponde exclusivamente al quejoso.

Criterio del que nuevamente se advierte la imposición de una limitante a los poderes probatorios del juzgador, pues se excluye la posibilidad de que ordene alguna diligencia para mejor proveer, sobre el argumento toral de que solo corresponde al promovente demostrar (en exclusiva) la titularidad de un derecho y su afectación a este.

Sumado a que se pretende condicionar la incorporación oficiosa de una prueba al hecho de que no se beneficie con ello a la parte quejosa, desconociendo, tal como sostiene TARUFFO (2010: 200) que la disponibilidad de la prueba pertenece a la dimensión epistémica del proceso, esto es, a la técnica del proceso y no al principio dispositivo en estricto sentido (cuestión que será analizada a detalle en el siguiente apartado).

Por tanto, los precedentes antes enunciados permiten concluir tres cuestiones: 1) que cuando se han ejercido los poderes probatorios del juez no se ha garantizado el derecho de defensa que tienen las partes, pues se ha hecho caso omiso a lo que implica el principio de contradicción; y 2) que cuando se intenta poner límites a su ejercicio, se hace sobre la base de que su despliegue altera las cargas procesales (en detrimento de los principios dispositivos y de aportación de parte) y 3) que cuando se afirma que con la actividad oficiosa a cargo del juez se estaría beneficiando a alguna de las partes, de forma implícita se estaría aludiendo a la pérdida de imparcialidad de aquel. Prejuicios procesales que se intentarán desvanecer en el siguiente apartado.

#### 4. DECONSTRUCCIÓN DE LOS PREJUICIOS PROBATORIOS QUE LIMITAN LA ACTIVIDAD PROBATORIA DEL JUZGADOR

Uno los temas más álgidos que se presentan cuando se debate en torno a las facultades probatorias del juez de amparo es el relativo a que su despliegue vulnera el principio dispositivo, así como el de aportación de parte; incluso, se llega atribuir al juzgador un carácter inquisitivo que tiende a provocar su parcialidad en el asunto.

A efecto de disipar lo que denominaremos “*prejuicios probatorios*” se tendrá como eje toral de este apartado el trabajo de dos grandes procesalistas, a saber: Jordi Nieva Fenoll y Joan Picó i Junoy, quienes han contribuido con un estudio histórico relevante en torno al principio inquisitivo y de aportación de parte, respectivamente, que permite dar una nueva mirada a las viejas concepciones formuladas en torno a dichos principios.

Así es, en el trabajo intitulado: “*El mal nombre del principio inquisitivo*” NIEVA (2014: 616 ss.) afirma que, tradicionalmente, el principio dispositivo había incluido también en su ámbito el principio de aportación de parte, pero que hoy en día, se tiende a separar ambos principios de la siguiente forma:

PRINCIPIO DISPOSITIVO	PRINCIPIO DE APORTACIÓN DE PARTE
Supone que queda en manos del particular la decisión de acudir al proceso, de iniciarlo y de continuarlo en sus distintos grados de jurisdicción, así como de concretar el objeto del juicio y, finalmente, disponer sobre el mismo <sup>49</sup> .	Son los litigantes, y no el juez, los que tienen que procurar los hechos y los medios de prueba.

El principio de “aportación de parte” estaría fundado, desde el punto de vista histórico, por el brocardo: “*iudex iudicare debet secundum allegata et probata, et non secundum conscientiam*” el cual, según destaca PICÓ (2007: 161 ss.) ha sido erróneamente modificado, tanto por cierta parte de la doctrina alemana de finales del siglo XIX, como de la italiana de principios del siglo XX, para quedar de la siguiente forma: “*iudex iudicare debet secundum allegata et probata partium*”, lo que ha impedido al juez resolver más allá de los hechos y pruebas aportados por las partes.

En efecto, como resultado de un extenso análisis a las fuentes históricas de la doctrina del *ius commune*, PICÓ refiere que el brocardo original

<sup>49</sup> NIEVA (2017: 3) afirma que su primera formulación hizo referencia, más que al propio principio dispositivo, al que desde hace ya más de un siglo se llama principio de aportación de parte, aunque no dejaba de referirse al primero.

Únicamente pretendía destacar los límites cognoscitivos del juez que deben reflejarse en su sentencia (*iudex iudicare debet*), a saber, los hechos según fueron alegados y probados en el proceso (*secundum allegata et probata*), debiendo rechazarse cualquier conocimiento privado (*conscientiam*) que el juez pudiera tener sobre estos<sup>50</sup>.

Y que, en ese sentido, la formulación del brocardo por parte de la doctrina procesalista en materia civil incurre en una doble y errónea modificación, pues

- a) Por un lado, se añade la palabra «*partium*» o «*partibus*»; y
- b) Por otro lado, se omite la expresión «*et non secundum conscientiam*».

Con base en lo cual concluye que si bien existe evidencia histórica para sostener que se impide al juez resolver más allá de los hechos alegados por las partes, lo cierto es que ello no sucede respecto de las pruebas. Al respecto, PICÓ aclara lo siguiente:

El brocardo no tiene por objeto indicar la mayor o menor facultad de iniciativa probatoria del juez, sino cosa bien distinta, que hace referencia a la imposibilidad de permitir el conocimiento privado del juez y a la congruencia de la sentencia<sup>51</sup>.

Por otra parte, NIEVA (2014: 617-619) agrega que el principio de aportación de parte habría excluido la prueba de oficio en el proceso civil, al menos desde el siglo XVI, no obstante que sin el *partium* la actividad probatoria del juez no tendría por qué descartarse.

Asimismo, explica que entre algunos doctrinarios es usual identificar como antagónico tanto del principio dispositivo como de aportación de parte, al *principio inquisitivo*. Principio este último, conforme al cual, el juez oficiosamente debe aportar los hechos y los medios de prueba. Con lo cual, afirma, se estaría identificando solo el antagónico del principio de aportación de parte, no así el relativo al principio dispositivo (NIEVA, 2014: 619 ss.).

A efecto de solventar tal cuestión, NIEVA intenta separar las dos manifestaciones del principio inquisitivo para quedar como siguen:

PRINCIPIO DE OFICIALIDAD	PRINCIPIO INQUISITIVO
El principio de oficialidad incluiría la decisión oficial de inicio del proceso, su continuación en ulteriores instancias, y la concreción y disposición sobre el objeto del juicio.	Se referiría solamente a la prueba de oficio, lo que además es perfectamente coherente con la terminología latina de la palabra empleada ( <i>inquirere</i> ) cuyo significado literal es “ <i>buscar</i> ”.

De modo que los principios y sus respectivos antagónicos quedarían de la siguiente forma:

- Principio dispositivo
- Principio de oficialidad
- Principio inquisitivo
- Principio de aportación de parte

<sup>50</sup> PICÓ, 2011: 296.

<sup>51</sup> PICÓ, 2011: 296.

En cuanto al principio inquisitivo, NIEVA (2014: 623-624) añade que es conveniente superar la confusión histórica que sucedió en el momento de creación del sistema procesal penal inquisitivo canónico en 1215, que llevó a considerar que la *inquisitio* era un modo de comenzar el proceso, como la *accusatio* o la *denunciatio*; no obstante que la *inquisitio* era la prueba de oficio con la cual se inicia el proceso penal, confundándose ambas actividades; y que, por tanto, es conveniente reducir el contenido del referido principio a la prueba de oficio, tanto por razones etimológicas (inquire= buscar) como teóricas, para encontrar un principio antagónico tanto al dispositivo como al de aportación de parte respectivamente.

En relación con ello, agrega que es factible visualizar un proceso dispositivo que tenga pruebas de oficio; o bien un proceso en el que rija el principio de oficialidad, pero que respete el principio de aportación de parte.

Lo hasta aquí expuesto permite concluir que cuando un juez de amparo practica prueba de oficio no dispone del objeto del juicio y, por tanto, no está vulnerando el principio dispositivo, pues solo se podría considerar que dispone de este si, además de ejercer esos poderes probatorios de forma oficiosa, tuviera a su disposición todo el contenido del principio de oficialidad, lo que en el caso no acontece. Por el contrario, en la materia rige lo dispuesto en el art. 107, fracción I, constitucional, el principio de iniciativa o instancia de parte, conforme al cual «el juicio (de amparo) jamás pueda operar oficiosamente y, por lo mismo, que para que nazca sea indispensable que lo promueva alguien» (SCJN, 1994: 31). Este principio, en palabras de BURGOA (2004: 269) entraña que el juicio de amparo: «nunca procede oficiosamente, es decir, sin que haya un interesado legítimo en provocar su actividad tuteladora [...]».

Todo ello, sumado al hecho de que si bien el principio de aportación de parte impide al juez resolver más allá de los hechos alegados por las partes, no lo hace respecto de la iniciativa probatoria que el juez pudiera asumir de oficio, pues como se vio, en sus orígenes, solo pretendía evitar que el juez recurriera a su propio conocimiento privado.

Otro de los aspectos que limita el actuar del juzgador en materia de prueba es el vinculado con la pérdida de imparcialidad, con la consecuencia de terminar favoreciendo a una de las partes, pues no estaría en condiciones de valorar de forma objetiva la prueba cuya incorporación ordenó de forma oficiosa.

Al respecto, el art. 17 constitucional establece que toda persona tiene derecho a que se le administre justicia por tribunales que emitan sus resoluciones de manera pronta, completa e «imparcial». Asimismo, el diverso art. 100 constitucional, establece que la carrera judicial se regirá por diversos principios, entre estos, el de «imparcialidad», el cual es definido por el Código de Ética del Poder Judicial de Federación, como

La actitud del juzgador frente a influencias extrañas al Derecho, provenientes de las partes en los procesos sometidos a su potestad. Consiste en juzgar, con ausencia absoluta de designio anticipado o de prevención a favor o en contra de alguno de los justiciables.

Como se observa, en el sistema judicial mexicano la imparcialidad tiene un carácter dual, pues, por un lado, constituye un “derecho” en favor de las partes, que habrá de materializarse en función de la forma en que se administre justicia por parte de los tribunales ante quienes se somete un conflicto (derecho a un juez o tribunal imparcial); y, por otro, es una “obligación” que ha de guiar el desempeño de la función judicial.

De modo que la imparcialidad judicial, como afirma PICÓ, se configura como un elemento consustancial a todo el proceso, el cual comporta

el derecho de las partes de pretender y esperar que el juez les trate de igual modo, bajo el mismo plano de igualdad... la existencia de cualquier tipo de prejuicio o interés del juzgador a favor de una parte, o en perjuicio de la otra, supone conculcar el principio básico de la igualdad de armas procesales, provocando que el juez deje de estar legitimado para resolverlo<sup>52</sup>.

Cabe agregar que, en 1981, México ratificó la Convención Americana sobre Derechos Humanos, en cuyo art. 8.1 se prevé lo siguiente: «Toda persona tiene derecho a ser oída, con las debidas garantías y dentro de un plazo razonable, por un juez o tribunal competente, independiente e *imparcial*... ».

La Corte Interamericana de Derechos Humanos (en adelante CIDH), al dar sentido y alcance a dicho numeral ha considerado que la imparcialidad «exige que el juez que interviene en una contienda particular se aproxime a los hechos de la causa careciendo, de manera subjetiva, de todo prejuicio y, asimismo, ofreciendo garantías suficientes de índole objetiva que permitan desterrar toda duda que el justiciable o la comunidad puedan albergar respecto de la ausencia de imparcialidad»<sup>53</sup>.

Asimismo, retomando lo sostenido por la Corte Europea de Derechos Humanos, la CIDH precisó:

... la *imparcialidad personal o subjetiva* se presume a menos que exista prueba en contrario, consistente, por ejemplo, en la demostración de que algún miembro de un tribunal o juez guarda prejuicios o parcialidades de índole personal contra los litigantes. Por su parte, la denominada *prueba objetiva* consiste en determinar si el juez cuestionado brindó elementos convincentes que permitan eliminar temores legítimos o fundadas sospechas de parcialidad sobre su persona. Ello puesto que el juez debe aparecer como actuando sin estar sujeto a influencia, aliciente, presión, amenaza o intromisión, directa o indirecta, sino única y exclusivamente conforme a -y movido por- el Derecho<sup>54</sup>

Cuando, en una parte del precedente interamericano, se hace referencia a que con motivo de la imparcialidad, el juez debe actuar exclusivamente conforme a -y movido por- el derecho, se equipara a la «imparcialidad» con el diverso valor/deber de «neutralidad» no obstante que en la doctrina se suelen distinguir entre estos, en atención a las actitudes (factores de actitud) de los jueces, tal como señala PAPAYANNIS (2016: párrafo 44) en la siguiente clasificación:

<sup>52</sup> PICÓ, 1998: 25-27.

<sup>53</sup> Caso Apitz Barbera y otros vs. Venezuela Sentencia de 5 de agosto de 2008, párrafo 56; y Urrutia Laubreaux vs. Chile sentencia de 27 de agosto de 2020, párrafo 118.

<sup>54</sup> Caso Apitz Barbera y otros vs. Venezuela de 5 de agosto de 2008, párrafo 56; y Urrutia Laubreaux vs. Chile de 27 de agosto de 2020, párrafo 118.

— Independencia: «Es una actitud hacia toda presión o influencia externa. Los jueces independientes resisten o rechazan toda esa presión o influencia».

— Imparcialidad: «Es una actitud hacia las partes involucradas y el tema de la disputa. Los jueces imparciales tienen una visión sin prejuicios de todas las partes y no tienen ningún interés personal en el resultado de la disputa».

— Neutralidad: «Es una actitud hacia la ley. Los jueces neutrales están comprometidos con el punto de vista legal».

En cuanto a este tercer punto, el autor en comentario refiere:

el árbitro puede estar en total desacuerdo con el contenido que identifica como legal. Sin embargo, la neutralidad, precisamente, consiste en razonar con los valores que nos proporciona la ley y no con los propios. Un juez neutral es aquel que se compromete a analizar cualquier asunto que sea sometido a su consideración desde el punto de vista jurídico<sup>55</sup>.

Lo hasta aquí expuesto permite sostener que se ha reconocido que la imparcialidad judicial tiene dos dimensiones:

una de carácter personal vinculada con las circunstancias del juzgador, con la formación de su convicción personal en su fuero interno en un caso concreto; y otra funcional, predicable de las garantías que debe ofrecer el órgano encargado de juzgar y que se establece desde consideraciones orgánicas y funcionales. La primera debe ser presumida mientras no se demuestre lo contrario. La segunda reclama garantías suficientes para excluir cualquier duda legítima sobre su imparcialidad<sup>56</sup>.

En cuanto al carácter personal de la imparcialidad, FERRER MAC-GREGOR (2014,15) sostiene que esta implica que «el juez tenga la capacidad de tomar la distancia necesaria y que se resista a sucumbir a cualquier influencia de carácter subjetivo» mientras que respecto al carácter funcional de la imparcialidad, dice «hay que verificar si, con independencia de la actitud personal del juez, existen circunstancias objetivas verificables que pueden hacer sospechar de su imparcialidad». En cuanto a este último punto el autor de referencia aclara que en el tema de la imparcialidad, “las apariencias” pueden revestir cierta importancia y por consiguiente se debe inhibir todo juez respecto de quien exista una razón legítima para temer su falta de imparcialidad.

En esa misma línea, el Comité de Derechos Humanos, en su Observación General sobre El derecho a un juicio imparcial y a la igualdad ante los tribunales y cortes de justicia, sostuvo

El requisito de imparcialidad tiene dos aspectos. En primer lugar, los jueces no deben permitir que su fallo esté influenciado por sesgos o prejuicios personales, ni tener ideas preconcebidas en cuanto al asunto sometido a su estudio, ni actuar de manera que indebidamente promueva los intereses de una de las partes en detrimento de los de la otra. En segundo lugar, el tribunal también debe parecer imparcial a un observador razonable. Por ejemplo, normalmente no puede ser considerado imparcial un juicio afectado por la participación de un juez que, conforme a los estatutos internos, debería haber sido recusado<sup>57</sup>.

<sup>55</sup> PAPAYANNIS, 2016: 38.

<sup>56</sup> FERRER, E., 2014: 14-15.

<sup>57</sup> Comité de Derechos Humanos, 90 período de sesiones Ginebra, Observación General núm. 32.

Una vez precisado lo anterior, surge el siguiente interrogante: ¿El despliegue oficioso de los poderes probatorios del juez afecta su imparcialidad en alguno de sus sentidos?

De forma previa a dar respuesta a dicho interrogante, se estima que se deben tomarse como base las siguientes consideraciones:

a) Durante la sustanciación del juicio de amparo el juez dispone del proceso, con la consecuencia de poder beneficiar o perjudicar a alguna de las partes con la toma de alguna determinación, por lo que no es factible considerar que ello solo acontecerá cuando ordena prueba de oficio.

Al respecto, destaca lo sostenido por TARUFFO (2008: 182) en el sentido de que en todas las ocasiones en que el juez dispone o decide alguna cuestión relativa al procedimiento (por ejemplo, cuando admite o excluye una prueba; reduce la lista de testigos o cierra la instrucción probatoria) toma partido a favor de una parte y en contra de la otra. Pese a ello, nadie piensa que por esa razón, en cada uno de esos momentos procesales, el juez ha perdido su imparcialidad.

b) Antes de ordenar una diligencia para mejor proveer no se sabe cuál va ser el resultado y, en consecuencia, a cuál de las partes puede beneficiar o perjudicar su incorporación.

Al respecto, sostiene PICÓ (2007: 112) que si, por ejemplo se ordenara la prueba pericial, no conoce cuál va ser su resultado.

c) Con independencia del origen de la prueba (aportada por las partes o por el juez) una vez incorporada al juicio, en virtud del principio de adquisición procesal, cualquiera de las partes puede valerse de esta, de modo que no es factible sostener que la prueba de oficio solo favorece a una de las partes.

5) No tendría sentido que, por una parte, se niegue al juez de amparo la capacidad de disponer la incorporación de pruebas no solicitadas por las partes y, por otra, tal como se vio en el apartado que antecede, se reconozcan a su favor amplios poderes en materia de prueba como, por ejemplo: la potestad de admitir y desechar pruebas; la capacidad de intervenir en la práctica de las pruebas (al formular preguntas a los peritos o testigos); la potestad de indicar alguna laguna probatoria o, incluso, la capacidad de decidir en cada caso cuán alta debe ser la corroboración de una hipótesis fáctica para considerarla probada, ante la ausencia de un estándar de prueba como tal en la LA.

Lo anterior en el entendido de que el despliegue de cualquier potestad probatoria puede dar lugar a la pérdida de imparcialidad del juzgador. De modo que no se podría caer en el extremo de suprimir todos los poderes con que cuenta con tal de evitar ese peligro. Al respecto ilustra lo sostenido por Barbosa Moreira.

llevada la suposición (de pérdida de imparcialidad del juzgador) hasta las últimas consecuencias lógicas, las leyes deberían prohibir de modo absoluto cualesquiera iniciativas oficiales en materia

de prueba, lo cual probablemente jamás ocurrió y no es propugnado siquiera por los más radicales representantes del 'dispositivismo' en la ciencia procesal civil<sup>58</sup>.

A la luz de estos primeros elementos parece que otorgar al juez el poder de ordenar pruebas de oficio no afecta necesariamente a la imparcialidad que debe observar como juzgador. Sin embargo, es necesario también preguntarse si el ejercicio de ese poder por el juez pudiera dar lugar a sesgos cognitivos que pongan en riesgo la imparcialidad que debe observar.

La noción de sesgo, sostiene BUSTAMANTE, da cuenta

de la inclinación o predisposición hacia determinada dirección suponiendo diversos procedimientos mentales, usualmente inconscientes, no de tipo aleatorio, sino sistemático, que producen juicios o decisiones que en ocasiones pueden alejarse o apartarse de la verdad, de una posición acordada o de determinados estándares o modelos de comportamiento humano<sup>59</sup>.

Entre los diversos tipos de sesgos se ubican «los sesgos cognitivos», que son definidos por PÁEZ como:

Una desviación sistemática, involuntaria e inconsciente de una norma o de un estándar de racionalidad al emitir un juicio perceptual o conceptual, al recordar un evento o al hacer una predicción. No se trata de un simple error, sino de comportamientos que ocurren consistentemente bajo circunstancias similares, y que por lo tanto son predecibles y replicables. Son causadas principalmente por el uso de heurísticas, atajos que utilizamos en el procesamiento de información, pero también son producto de las limitaciones de nuestra capacidad cerebral de procesar información, de influencias emocionales, morales y sociales, y de distorsionadores en el proceso de almacenamiento y búsqueda de información de la memoria<sup>60</sup>.

Por otra parte, entre los sesgos cognitivos, se identifica «el sesgo por confirmación» que es definido por BUSTAMANTE como:

La tendencia a buscar, interpretar o recordar información de manera que apoye las creencias, expectativas o hipótesis, independientemente de que estas sean verdaderas o no. Lo que resalta en este sesgo es que la búsqueda, interpretación o recuerdo se hace de modo unilateral, sin atender otras posibles alternativas<sup>61</sup>.

En cuanto a los tipos de «sesgos por confirmación», para efectos de este trabajo, se destacan dos (BUSTAMANTE, 2021:9):

a) Búsqueda de pruebas: «el sujeto buscará información que confirme sus creencias o hipótesis de forma unilateral y con ello evitará o dejará de lado aquellas que las refuten... se producirá una restricción de la atención a una hipótesis, lo que conllevará a excluir la posibilidad de interpretar los datos que apoyan cualquier otra explicación alternativa»; y,

<sup>58</sup> Cfr. en PICÓ, 2007: 113.

<sup>59</sup> BUSTAMANTE, 2021: 6.

<sup>60</sup> PÁEZ, 2019: 3.

<sup>61</sup> BUSTAMANTE, 2021: 9.

b) Interpretación de las pruebas: «en las que el sujeto tiende a interpretar la información de forma selectiva, sobre todo aquella que es ambigua con la intención de favorecer sus creencias».

Una vez visibilizado el papel que juegan los «sesgos cognitivos» en la toma de decisiones o formulación de juicios, es factible considerar que la decisión judicial no está exenta de ellos y, en ese sentido, el juez, al momento de ordenar la incorporación oficiosa de alguna prueba, podría verse vinculado por «el sesgo de confirmación» en al menos dos momentos centrales del juicio de amparo, a saber:

1) En el momento de la conformación del conjunto de elementos de juicio o pruebas, toda vez que el juez iniciará la búsqueda de aquellas que confirmen la hipótesis de los hechos que crea o tenga la expectativa de ser más conveniente (sin atender otras alternativas) de modo que pudieran presentarse los siguientes escenarios<sup>62</sup>:

a) Que el juez ignore la existencia de las pruebas que son contrarias a la hipótesis que él cree o considera más conveniente y, por tanto, para confirmar esta, estime necesario la incorporación oficiosa de diversas pruebas; o bien

b) Que el juez sí examine las pruebas que son contrarias a la hipótesis que él cree o considera más conveniente, pero lo haga «críticamente» de manera que concluya que son contradictorias y/o defectuosas y, en consecuencia, intente justificar la incorporación de otras que sí confirmen su hipótesis.

2) En el momento en que se proceda a la valoración de los elementos del juicio o pruebas disponibles en el proceso (tanto las incorporados por las partes como de ordenados de forma oficiosa por el juzgador), donde nuevamente pueden plantearse dos escenarios:

a) Que se niegue valor probatorio a las pruebas contrarias a su hipótesis, ya sea porque las ignora por completo o porque considerándolas se muestre menos perceptivo con estas. Lo que podría hacerse evidente, por ejemplo, cuando el juez efectúa un escueto o insuficiente relato de su contenido, de modo que las considera irrelevantes o poco confiables.

b) Que, a pesar de que no haya pruebas que confirmen su hipótesis, intente dar interpretaciones alternativas, incluso podría considerar que los defectos o ambigüedades en el resto del material probatorio de cierta forma confirman su hipótesis.

---

<sup>62</sup> Para poder describir los eventuales escenarios a los que se podrá enfrentar el juzgador tanto al momento de buscar prueba como de interpretarla se atendió al fenómeno de la «asimilación sesgada» estudiado a detalle por Lord, Ross y Lepper, entre cuyas conclusiones destacan: «1) que a veces no se ignora la información contraria, sino que se le examina críticamente y se concluye que es contradictoria o defectuosa y 2) a pesar de que la prueba no confirma la hipótesis, el sujeto ensaya cualquier interpretación alternativa, incluso podrá considerar que las ambigüedades y defectos en los datos sugieren de alguna forma la corrección de su hipótesis, manteniendo o incluso reforzando sus ideas preconcebidas». Cfr. en BUSTAMANTE, 2021: 11 y 29.

Lo antes expuesto se plantea sobre la aclaración de que no siempre que se efectúa una búsqueda y una interpretación de pruebas para confirmar una hipótesis, se configura un sesgo confirmatorio, en tanto que esas pruebas podrían ser claramente adecuadas, de modo que la confirmación de una hipótesis de la que se habla en este apartado linda con la «selectividad deliberada de la evidencia», con un «moldeo involuntario» y hasta con el «deseo de defender las creencias que se desean mantener» y que, por tanto, resulta más factible hablar de «una tendencia general a un inapropiado refuerzo de hipótesis o creencias cuya verdad está en cuestión» (BUSTAMANTE, 2021: 12).

De cara al tema de los sesgos, la respuesta a la pregunta de si *el despliegue oficioso de los poderes probatorios del juez afecta su imparcialidad en alguno de sus ámbitos* pareciera no tener que ser negativa en sentido contundente, o al menos no en todos los casos, pues, como se ha visto, es factible que en algunos de esos ámbitos los sesgos bloqueen en mayor medida el análisis de fiabilidad de las pruebas y justifiquen de forma irracional la incorporación de una prueba al proceso. En estos casos la imparcialidad podría verse afectada, tanto en su aspecto personal como funcional, puesto que se vinculan las circunstancias personales del juzgador en la formulación de su convicción en el caso concreto y se niega la posibilidad de brindar elementos convincentes que permitan eliminar sospechas de parcialidad. Sin embargo, los potenciales sesgos cognitivos que pudiera generar el ejercicio del poder de ordenar pruebas de oficio se pueden limitar si se adoptan mecanismos que los contrarresten o disminuyan y, en ese sentido, eviten la parcialidad involuntaria que puedan generar en el juzgador.

A fin de clarificar este último aspecto, se torna imprescindible lo sostenido por PÁEZ (2019:10-14) en cuanto a que las decisiones judiciales están afectas de principio a fin por sesgos cognitivos y que la posibilidad de eliminarlos por completo es nula, de modo que «un juez responsable intentará, en la medida de lo posible, llegar a ser consciente de sus sesgos [de carácter individual] y anular su efecto a través de este autoconocimiento», pero sin que ello garantice plenamente su no aparición.

Lo anterior permite sostener que el acento debe ponerse en algún mecanismo que permita restringir, disminuir o limitar el efecto negativo de los sesgos, como elemento deslegitimador en la toma de las decisiones.

Al respecto, PÁEZ (2019: 9-29) afirma que la exigencia de motivación como remedio para limitar los efectos negativos de los sesgos es insuficiente, ya que difícilmente describe el razonamiento que llevó el juez en cada etapa del proceso. Por ello, se torna difícil identificar si la toma de decisión se vio afectada por algún sesgo en específico. En ese sentido, sugiere que es desde las reglas procesales donde se deben intentar subsanar las limitaciones que imprimen los sesgos al razonamiento práctico, esto es, la «eliminación de sesgos a través del derecho», lo que implica una intervención directa sobre la normativa y el diseño institucional.

La propuesta de PÁEZ en ese sentido es crear guías de práctica judicial, análogas a las guías de práctica clínica a partir de las cuales se puedan desarrollar protocolos

de actuación y listas de verificación que deban ser observadas sin excepción. Lo que aterrizado en el campo de la incorporación oficiosa de pruebas a cargo del juzgador podría equiparse con la configuración de los límites que deben guardar su despliegue en el proceso, en aras de evitar los efectos negativos de los sesgos de confirmación que pudieran actualizarse en detrimento de la imparcialidad del juzgador.

En cuanto a la configuración de tales límites, PICO (1998: 106-107) ha sostenido en esencia los tres siguientes:

a) La prueba debe limitarse a los hechos discutidos por las partes en el proceso; de modo que el juez no puede realizar una actividad tendente a investigar o aportar otros hechos, ni fallar alterándolos;

b) En los autos deben constar las «fuentes de prueba» sobre las cuales tendrá lugar la actividad probatoria a cargo del juez; así por ejemplo, tratándose de la prueba testimonial podría ordenar la presencia de un tercero que pueda tener conocimiento de datos relevantes para la solución del asunto siempre que su identidad haya sido relevada durante el juicio; o bien ordenar alguna prueba documental, citada por alguna de las partes; y para la prueba pericial, la advertencia de hechos cuya corroboración requieren de conocimientos técnicos especializados.

c) Respeto irrestricto al *principio de contradicción*, con la posibilidad de admitir pruebas que le permitan a las partes confrontar la prueba incorporada por el juez.

Estos límites (sobre todo los primeros dos) tienen como finalidad esencial evitar que el juzgador se lance en una búsqueda irrestricta de pruebas que puedan confirmar la hipótesis sobre los hechos que asuma de forma preconcebida. Sin embargo, no impiden que la incorporación oficiosa de pruebas, aun sujeta a tales límites, se restrinja, exclusivamente, a la búsqueda de pruebas que tengan por objeto confirmar la hipótesis preconcebida y, por tanto, omita ordenar la incorporación de aquellas que resulten contrarias a la hipótesis de que se trata.

En relación con este tipo de sesgo, es importante aclarar que no es propio o exclusivo de la incorporación oficiosa de pruebas a cargo del juzgador, sino que puede presentarse en cualquier otra etapa del juicio, por ejemplo, cuando el juez efectúa la valoración probatoria de las pruebas aportadas exclusivamente por las partes, en tanto podría intentar justificar aquella hipótesis que asume de forma preconcebida sobre los hechos y, en consecuencia, restar valor probatorio a aquellas otras que sean contrarias a dicha hipótesis.

Ante la insuficiencia de los límites antes referidos para prevenir este tipo de sesgos, se estima que será el tribunal colegiado de circuito, que eventualmente conozca del recurso de revisión que se interponga contra la sentencia dictada en el juicio de amparo, el que habrá de jugar un papel trascendental a fin de advertir y contrarrestar si el despliegue probatorio a cargo del juzgador atendió a aspectos objetivos y no a una mera selectividad deliberada de evidencia. Para ello, se estima, adquiere una alta relevancia el aspecto de las “apariencias”. En relación con estas apariencias, el TEDH

ha acuñado la famosa expresión de «no solo debe hacerse justicia, sino advertirse que se hace»<sup>63</sup> y respecto de lo cual ha sostenido

que la imparcialidad personal de un juez puede determinarse, según las circunstancias concretas del caso, con base en el comportamiento del juez durante el procedimiento, el contenido, los argumentos y el lenguaje utilizado o los motivos para llevar a cabo la investigación, que indiquen una falta de distancia profesional frente a la decisión<sup>64</sup>.

Aspectos que de ser tomados en consideración, al analizar el comportamiento procesal del juzgador, permitirían al tribunal colegiado estar en aptitud de verificar, por ejemplo, si tenía sustento conducir la investigación de los hechos hacia cierta dirección y no optar por otra, en un momento procesal determinado, en el que ya obraban ciertas pruebas en el expediente.

Por lo que si bien es cierto no existe un mecanismo que logre erradicar por completo los sesgos cognitivos que pueden afectar la decisión judicial, también lo es que, como se ha explicado, existen instrumentos que permiten disminuir el efecto negativo de estos. Entre ellos, la configuración de límites al despliegue de los poderes probatorios a cargo del juzgador e incorporar su revisión por un órgano colegiado. Es por ello que no se puede alegar, sin más, pérdida de imparcialidad, mientras existan parámetros objetivos que permitan verificar si tiene justificación el ejercicio de tales poderes y no obedecen a una mera selectividad deliberada de evidencia.

Por otra parte, no se debe olvidar que el concepto de imparcialidad, como señala TARUFFO (2010: 138) presenta diversos niveles de variabilidad semántica, al tratarse de un concepto que está presente no solo en el ámbito procesal sino también en el de la ética y política; que dentro del contexto procesal se ha visto descuidada la conexión entre «imparcialidad y verdad» con base en la cual el concepto de imparcialidad del juez no se agota en las condiciones que la ley establezca para salvaguardar que no sea parcial (impedimentos o recusación -que son las figuras que se prevén en la LA-), sino que se dirige a un fin determinado: la búsqueda de la verdad; y que desde esa perspectiva «el juez sería realmente imparcial en tanto busca de forma objetiva la verdad de los hechos, haciendo de ella el verdadero y exclusivo fundamento racional de la decisión». Ello se corrobora con lo dispuesto en el 10 del Código de Modelo Iberoamericano de Ética, al prever: «*El juez imparcial es aquel que persigue con objetividad y con fundamento en la prueba la verdad de los hechos...*».

Por tanto, si se pugna por un juez de amparo activo (que no autoritario) para quien la búsqueda de la verdad sea el fin del proceso, se debe abandonar la idea de que la imparcialidad es sinónimo de pasividad procesal y tener presente que lo afirmado por TARUFFO (2010: 139) en cuanto a que «la pasividad del juez terminaría por jugar un rol contra-epistémico, pues el juez pasivo dejaría que el proceso estuviera dominado por sujetos que, en la mayoría de los casos, son indiferentes (si es que no hostiles) con respecto al descubrimiento de la verdad».

<sup>63</sup> Caso Maurice vs. France, Sentencia de 11 de julio de 2013, párr. 78.

<sup>64</sup> Cfr. en FERRER, E., 2014: 20 (Caso Kyprianou vs. Cyprus).

Incluso, TARUFFO (2010: 139) añade que la determinación de los hechos realizada de forma imparcial constituye una condición necesaria del debido proceso, en tanto que este se encuentra dirigido hacia la consecución de decisiones justas. En donde, cabría agregar, el significado de dicho valor (justicia), no sea otro que: aquel orden social bajo cuya protección puede progresar la búsqueda de la verdad<sup>65</sup>.

Por otra parte, no sería congruente que mientras en el derecho privado (derecho civil), a partir de una visión publicista del proceso<sup>66</sup>, se ha atendido a la necesidad de dotar al juez de los poderes probatorios necesarios para contribuir activamente a la satisfacción del interés público en cuanto a la recta aplicación de la ley; en cambio, en el juicio de amparo, que no es un juicio ordinario, sino un medio de control constitucional, se niegue al juez la capacidad de disponer la incorporación de pruebas no solicitadas por las partes.

De ahí que la relevancia de reconocer poderes probatorios al juez de amparo, específicamente en lo que hace a la incorporación oficiosa de prueba, tenga un impacto diferenciado con el análisis que habitualmente se realiza frente a procedimientos de naturaleza ordinaria. En efecto, mientras en estos se declara o se reconoce un derecho en favor de alguien, en el amparo se salvaguarda la titularidad de un derecho fundamental ya constituido frente a la afectación que sufre con motivo de un acto de autoridad (o de uno equivalente al de autoridad). En ese sentido, los mecanismos para una mejor integración de la premisa fáctica, por mayoría de razón, tendrían que ser potencializados, en aras de preservar el orden constitucional.

A estos efectos, se estima que el antídoto central para contrarrestar los eventuales actos en que pudiera incurrir el juez en detrimento de los principios dispositivo, de aportación de parte e imparcialidad judicial, no consiste en erradicar los poderes probatorios con los que cuenta, sino en esclarecer los límites dentro de los cuales se pueden ejercer, lo que será motivo de análisis pormenorizado en el siguiente apartado.

---

<sup>65</sup> Palabras expresadas por Kelsen al dar una conferencia de despedida como profesor de la Universidad de California. Cfr. en PICÓ, 2007: 99.

<sup>66</sup> «Algunos autores suelen argumentar la imposibilidad de atribuir iniciativa probatoria al juez civil debido a la naturaleza privada del interés discutido en el proceso [...]. Si bien es cierto que lo discutido en el proceso civil tiene, por regla general, un carácter disponible o privado, ello no comporta que tales características puedan igualmente predicarse del proceso, pues el modo de desarrollarse el mismo no pertenece a los litigantes sino al Estado, único titular de la función jurisdiccional. [...] Como ya destacó en 1943 Calamandrei *«de la consideración de la jurisdicción, también en materia civil, como una función pública, se deriva la necesidad técnica de dar al juez todos poderes necesarios para poder cooperar activamente a la satisfacción del interés público que también en el proceso civil está en juego; y basta reconocer el carácter público de la función jurisdiccional para deber considerar técnicamente inadecuado a los fines de la justicia un sistema en el que el juez asiste como mero espectador»* PICÓ, 2007: 105.

## 5. CONFIGURACIÓN Y DELIMITACIÓN DE LA PRUEBA DE OFICIO

El despliegue práctico de las diligencias para mejor proveer, a cargo del juez de amparo, constituye un poder discrecional y no propiamente un deber<sup>67</sup>, en el entendido de que será potestad del juzgador determinar, en cada caso concreto, si ordena o no la incorporación de pruebas no aportadas por las partes.

Sin embargo, debe entenderse que ese poder discrecional no es sinónimo de arbitrariedad y que como cualquier otro poder está limitado o sujeto a controles que garanticen su debido ejercicio, los cuales son:

1) Las diligencias para mejor proveer tienen una naturaleza *complementaria* y *supletoria* respecto de las actividades probatorias de las partes. Lo que implica que su despliegue acontecerá una vez que las partes hayan ejercido su derecho de ofrecer todas las pruebas disponibles para la determinación de los hechos y sean necesarios mayores datos para la construcción de la premisa fáctica.

2) Las diligencias para mejor proveer tienen el carácter de *subsidiarias*. De modo que no en todos los casos es indispensable su despliegue. Estimar lo contrario implicaría soslayar que las partes pueden haber ejercido completamente su derecho a proponer todas las pruebas disponibles (como ocurre en la práctica en la mayoría de los casos) y no sean necesarios mayores datos para la toma de la decisión.

3) Las diligencias para mejor proveer tienen un carácter inquisitivo (en su connotación de búsqueda) que no autoritario, que no es incompatible con los principios dispositivos ni de aportación de parte, según se explicó.

4) No rempazan la carga probatoria de las partes, en tanto que su despliegue no obedece necesariamente a una deficiencia en la actividad probatoria de estas, sino a la necesidad de complementar el acervo probatorio, una vez que las partes ya han ejercido su derecho de ofrecer todas las pruebas disponibles, en tanto resulta relevante una prueba o varias a fin de que el juez pueda tomar una decisión más acertada sobre los hechos.

5) Su despliegue debe circunscribirse a los hechos expuestos por las partes y que han identificado como objeto del proceso, con independencia de si han sido objetos o no por estas o se ha establecido algún acuerdo probatorio en relación con ellos.

6) Es necesario que consten en el proceso las fuentes de prueba sobre las cuales tiene lugar la posterior actividad probatoria del juez.

---

<sup>67</sup> Solo se consideraría un “deber” hacer uso de ese poder discrecional cuando así se requiera para la integración de la premisa fáctica. Lo que en palabras de LARA (2011:110) se explica así: «Desde mi punto de vista, el uso discrecional de cualquier facultad discrecional debe ser visto como un deber: *el deber consiste precisamente en el uso (adecuado al caso concreto) de la facultad discrecional*, y su incumplimiento derivaría del no ejercicio de la facultad en aquellos casos en que se hacía necesario» (Lo destacado es propio).

7) Es necesario que su despliegue respete, en todo momento, el *principio de contradicción* y el *derecho de defensa* de las partes, a fin de que puedan «controlar la relevancia y admisibilidad de las pruebas que el juez dispone de oficio... que estén en condiciones de aportar pruebas contrarias o diferentes respecto de las dispuestas por el juez, y que tengan por último la posibilidad de discutir el resultado y valor de estas pruebas» (TARUFFO, 2010: 203).

8) Es indispensable que su despliegue respete el “*principio de limitación de la prueba*” conforme al cual, según se vio, cuando en el juicio de amparo se reclamen actos que tengan como origen o antecedente algún procedimiento de carácter jurisdiccional o procedimiento seguido en forma de juicio, el juez está impedido para incorporar pruebas que no haya sido rendidas ante la autoridad responsable, con la única excepción de que no se hubiera tenido la oportunidad de hacerlo (excepción que, cabe aclarar, tiene una prohibición tajante en materia penal si se vulnera la oralidad o los principios que rigen el proceso penal acusatorio). Lo que permite garantizar que el juicio de amparo no se convierta en una extensión de aquellos procedimientos que constituyen el origen del acto reclamado y, en consecuencia, que la incorporación oficiosa de pruebas a cargo del juez de amparo verse propia y exclusivamente sobre la *litis* constitucional.

La configuración de tales límites<sup>68</sup> plantea el siguiente interrogante: ¿Frente a qué circunstancias el juez estará en aptitud de ordenar la incorporación oficiosa de una prueba?

La respuesta a dicha pregunta, se propone, debería atender a ciertas eventualidades probatorias -que surgen una vez ofrecidas y desahogadas las pruebas de las partes-. Estas eventualidades pueden mermar la completitud del acervo probatorio y, con ello, disminuir la probabilidad de tomar una decisión acertada sobre los hechos. Esto puede darse al menos en tres situaciones en el juicio de amparo (sin pretensión de exhaustividad):

a) Cuando una de las partes alegó y demostró un hecho y la contraparte no formuló objeción alguna (entendida como la negación explícita de la verdad del enunciado) o incluso manifestó su conformidad con aquel por así convenir a sus intereses; por lo que para efectos procesales se tendría que tener por verdadero. Sin embargo, pudiera acontecer que como resultado de un contraste con el resto del

<sup>68</sup> Límites que resultan congruentes con la línea jurisprudencial sostenida por la Primera Sala de la SCJN, al resolver el amparo directo en revisión 3596/2016, en el que, al pronunciarse sobre la facultad de los jueces penales de recabar prueba de oficio en el sistema mixto o tradicional, estableció que su despliegue está sujeta a ciertos límites, a saber:

<sup>a)</sup> La prueba solicitada por el juez debe ceñirse en todo momento a los hechos que son objeto del proceso penal;

<sup>b)</sup> Es indispensable que existan fuentes de prueba en el proceso a partir de las cuales se desprenda la necesidad de la práctica u obtención de nuevos medios probatorios;

<sup>c)</sup> Es imprescindible que la autoridad jurisdiccional garantice en todo momento el principio de contradicción y el derecho de defensa de las partes.

material probatorio al juzgador le generara incertidumbre el acontecimiento de tal hecho o incluso le resultara inverosímil.

Por ello, atendiendo a una condición de justicia de la decisión, el enunciado que en relación con el hecho no se hubiere objetado, solo podrá ser tenido por verdadero siempre que, como afirma TARUFFO (2010: 152-153), el juez haya madurado un juicio sobre la verdad del enunciado que lo describe. A estos efectos, debe entenderse que el juez puede extraer elementos de juicio sobre la verdad o falsedad de aquel de cualquier fuente de información, entre ellas, claro está, las que ordene de forma oficiosa para esclarecer tal aspecto.

b) Cuando el juzgador carece de los conocimientos necesarios para evaluar la fiabilidad y la validez de las pruebas basadas tanto en testimonio científico como técnico. El juez ha de ejercer esa facultad una vez que ya ha planteado y agotado todas las aclaraciones pertinentes que pudiera formular a los expertos (a través, por ejemplo, de haber ampliado, de forma oficiosa, el cuestionario que habrán de responder). Es decir, ejercería sus facultades probatorias no porque el dictamen, por ejemplo, resulte contradictorio o con falta de coherencia, sino porque, quizá, no tenga los conocimientos suficientes para corroborar la científicidad de la técnica o método empleado.

Negar el despliegue de poderes probatorios officiosos tratándose de este supuesto implicaría desconocer que, como sostiene NIEVA (2010: 285), lo que suele encontrarse en la práctica es «un juez que no posee la información necesaria ni siquiera para entender debidamente el dictamen» y que el problema que se deriva de ello es que «el juez se limite a transcribir las razones del perito sin entenderlas realmente».

Al respecto, como afirma VÁZQUEZ<sup>69</sup>, resulta importante que el juez se convierta en un consumidor sofisticado de conocimiento experto; o como sostiene GASCÓN (2016: 359 ss.) es necesario aumentar o reforzar la base cognitiva del juzgador para hacer frente a este tipo de asuntos. Para ello es que legalmente existe una facultad a cargo del juez para conseguir esos objetivos.

c) Cuando el objeto de la controversia involucre una vulneración a los derechos fundamentales de ciertos grupos vulnerables, tales como: (i) menores de edad o incapaces; (ii) trabajadores; (iii) ejidatarios y comuneros o bien núcleos de población ejidal o comunal; (iv) el inculpado o el sentenciado o bien la víctima en su carácter de quejoso; (v) quienes por sus condiciones de pobreza o marginación estén en clara desventaja social para su defensa en el juicio. Supuestos todos ellos, que si bien son previstos en la LA (art. 79) como aquellos en los que el juez debe “suplir la deficiencia de la queja”<sup>70</sup>, sirven de parámetro para identificar las temáticas que han de activar

<sup>69</sup> Cfr. en «La prueba pericial en la experiencia estadounidense», Cátedra de Cultura Jurídica, disponible en: <https://youtu.be/IBJ-bhYVyA8>

<sup>70</sup> En este apartado, la referencia a la figura de la “suplencia de la deficiencia de la queja” debe ser entendida no en su concepción tradicional (limitada a subsanar o integrar conceptos de violación o agravios) sino en toda su amplitud, como ha sido entendida por la SCJN, por ejemplo, tratándose de menores (1a./J.191/2005), pues a nada práctico conduce limitar la actuación a subsanar meras alegacio-

los poderes probatorios del juez. Esto último si se considera que su incorporación en la ley de la materia atiende al hecho de que se trata de ciertos sectores que histórica y socialmente presentan posiciones asimétricas durante el proceso, dadas sus condiciones educativas, sociales, económicas, o de cualquier otra naturaleza, ocasionándose una disparidad que repercute en su derecho de acceso a la justicia<sup>71</sup>.

Todo ello permitiría acotar, al menos es estos casos, que el poder probatorio del juez no es discrecional, sino obligatorio.

Incluso, recientemente, vía jurisprudencial, se ha incluido como parte de los supuestos respecto de los que procede la suplencia de la queja, lo relativo a las personas con discapacidad<sup>72</sup>. Sin embargo, se estima que el despliegue probatorio a cargo del juez tratándose de tal supuesto, se debe atemperar en cada caso concreto, pues como lo señaló la Primera Sala<sup>73</sup> podría fundarse en la indebida suposición de que una persona, por el solo hecho de tener una deficiencia física, mental, intelectual o sensorial, no está en posición de defenderse en igualdad de condiciones.

d) En los supuestos legales y jurisprudenciales específicamente establecidos y que han quedado precisados en el punto 3 de este trabajo.

## 6. CONCLUSIÓN

El marco legal, jurisprudencial y doctrinal antes expuesto permite concluir que en el juicio de amparo la actividad probatoria que se despliega sí tiene como uno de sus objetivos centrales el conocimiento de la verdad. Tan es así que se reconocen una serie de caminos que conducen a ella, entre ellos, claro está, el que coloca al juez como el *garante de la corrección epistémica de los hechos*, al facultarlo con un cúmulo de poderes en materia de prueba. Entre tales poderes destaca el relativo a la incorporación de prueba no aportada por las partes, en relación con el cual en este trabajo se sugiere que, cada vez que sea ejercido, debe ser llamado por su nombre, en pleno respeto y potencialización del principio de contradicción, para no dejar en estado de indefensión a las partes.

Al respecto, cabe destacar que en LA no existe una regulación explícita sobre el despliegue de la incorporación oficiosa de prueba a cargo del juzgador, pues, como se visibilizó, su ejercicio solo se prevé en relación con un número limitado de supuestos. Ante esa ausencia, el juez tiene que recurrir a la codificación civil supletoria, en la

---

nes sí, finalmente, no se cuenta con los medios de prueba para acreditar aquellas. Lo que es acorde con lo expuesto por la Primera Sala del Máximo Tribunal (1a./J. 65/2013 (10a.), en cuanto a que «el referido derecho implícito (a ofrecer pruebas), por sí mismo, sin vinculación alguna con argumentos tendentes a oponer excepciones y defensas, carecería de toda lógica y a nada práctico llevaría procesalmente».

<sup>71</sup> Tesis: 1a. CCI/2018

<sup>72</sup> Tesis 2a. L/2020 (10a.)

<sup>73</sup> Tesis 1a. CCXV/2018 (10a.)

que si bien en los arts. 79 y 80 se establece un límite importante para el ejercicio de tal potestad, al señalar que su despliegue se debe circunscribir a los hechos expuestos por las partes, lo cierto es que el resto de su contenido restringe el principio de contradicción, en detrimento de las partes, quienes no pueden confrontar la prueba incorporada por el juez.

En consecuencia, se estima necesaria una reforma legislativa que, se sugiere, debiera tener lugar sobre la propia LA y no así en la codificación civil supletoria, en la que no solo se establezca expresamente tal potestad a cargo del juzgador, sino que también se delimite su ejercicio, para lo cual se torna necesario identificar en qué supuestos puede ser desplegada y, sobre todo, se garantice el derecho de las partes a confrontar la prueba que eventualmente sea incorporada de forma oficiosa por el juzgador.

Por otra parte, cabe destacar que si bien el Código Federal de Procedimientos Civiles es aplicable supletoriamente a cualquier tipo de procedimiento federal, el amparo no puede ni debe ser considerado como uno más de aquellos, no al menos para la regulación de la incorporación oficiosa de prueba a cargo del juzgador, pues como se destacó, constituye un medio de control constitucional, cuyo objeto es la defensa de los derechos fundamentales y, por tanto, exige una regulación propia, específica e independiente de la materia civil que es sobre la que versa el código supletorio.

A todo ello se suma el hecho de que si en la LA se prevé una regulación detallada del despliegue de otros poderes probatorios a cargo del juez, no existe razón que justifique la exclusión de esa regulación tratándose de la incorporación oficiosa de prueba como un poder más. Aunado a que no habría razón que explique por qué tratándose de la suspensión provisional del acto reclamado, por ejemplo, sí se hace una referencia destacada en la ley de la materia en cuanto al tema de la incorporación oficiosa de prueba, y tratándose del juicio en lo principal se incurre en una omisión sobre ese tema.

## BIBLIOGRAFÍA

- BURGOA, I., 2004: *El juicio de amparo*. México: Porrúa.
- BUSTAMANTE, J. F., 2021: «Sesgos cognitivos en la valoración individual de la prueba» en <<https://dugi-doc.udg.edu/handle/10256/19419>>.
- CAMPUZANO, A., 2015: *Manual para entender el juicio de amparo Teórico-Práctico*: Thomson Reuters.
- COSSÍO, J. R., MEJÍA, R., CABAÑAS, J., 2017, *El uso de evidencia científica y opinión experta en las sentencias de la SCJN*. México: El Colegio Nacional de México-Tirant lo Blanch.
- FERRER, J., 2005: *Prueba y verdad en el derecho*, Madrid: Marcial Pons.
- , 2007: *La valoración racional de la prueba*. Madrid: Marcial Pons.
- , 2017: «Los poderes probatorios del juez y el modelo de proceso», en *Cuadernos Electrónicos de Filosofía del Derecho*, 36: 88-108, disponible en: <<https://ojs.uv.es/index.php/CEFD/article/view/10632>>.
- , 2019: «La carga dinámica de la prueba. Entre la confusión y lo innecesario», en *Contra la carga de la prueba*. Madrid: Marcial Pons, 53-87.

- , 2021: *Prueba sin convicción. Estándares de prueba y debido proceso*. Madrid: Marcial Pons.
- FERRER, E., VENTURA, M., 2014: «El derecho a un juez o tribunal imparcial: Análisis del artículo 8.1 de la Convención Americana sobre Derechos Humanos (a la luz del caso del Pueblo Indígena Mapuche vs Chile)», en <https://www.corteidh.or.cr/tablas/r33505.pdf>
- GASCÓN M., 2016: «Conocimientos expertos y deferencia del juez» en *Revista Doxa. Cuadernos de Filosofía del Derecho*, 39: 359 y ss.
- , (2005) «La actividad judicial: problemas interpretativos» en *La argumentación en el derecho*. Perú: Palestra, 101-130.
- LARA, R., 2011: «Motivación de los hechos: Reflexiones sobre las diligencias para mejor proveer», en *Revista Isonomía*, 35:84-118.
- LAUDAN, L., 2012: *Verdad, Error y Proceso Penal*, Marcial Pons: Madrid.
- MEDINA, E., 2017: «Análisis de la Naturaleza Jurídica y los Límites de las facultades para mejor proveer en el juicio de amparo» en *El Juicio de Amparo en el Centenario de la Constitución de 1917, pasado, presente y futuro*. México: UNAM, 417-430.
- NIEVA, J., 2010: *La valoración de la prueba*, Marcial Pons: Madrid.
- , 2014: «El mal nombre del principio inquisitivo», en *Justicia. Revista de derecho procesal*, 1: 131-166.
- , 2017: «La actuación de oficio del Juez Nacional Europeo», en *Diario la Ley*, 9000, sección doctrina. Ed. Wolters Kluwer.
- OSORNIO, V., 2018: «De las facultades de un juez de amparo para verificar la validez científica o técnica de un dictamen pericial», en *Revista del Instituto de la Judicatura Federal*, 46: 327-345.
- , 2019: «El interés legítimo y sus notas distintivas a partir de lo resuelto en la contradicción de tesis 111/201», en *Revista del Instituto de la Judicatura Federal*, 48: 143-169.
- PABÓN, L., 2021: «Imparcialidad y prueba de oficio ¿Entre la búsqueda de la verdad o la suplantación de las partes», en *Garantías Procesales y Poderes del Juez*, México: Zela, 223-245.
- PÁEZ, A., 2019: «Los sesgos cognitivos y la legitimidad racional de las decisiones judiciales», en: <https://philpapers.org/archive/PEZLSC.pdf>
- PAPAYANIS, D., 2016: «Independencia, imparcialidad y neutralidad en la adjudicación judicial», en *Revista de Teoría Constitucional y Filosofía del Derecho*, 28, consultable en: <https://journals.openedition.org/revus/3547>.
- PICÓ J., 1998: *La imparcialidad judicial y sus garantías: la abstención y la recusación*. Barcelona: Bosch.
- , 2007: *El juez y la prueba. Estudio de la errónea recepción del brocardo iudex iudicare debet secundum allegata et probata, non secundum conscientiam y su repercusión actual*. Barcelona: Bosch.
- , 2011: «El juez civil y la prueba: una historia mal contada», en *Boletín de la Academia de Ciencias Políticas y Sociales*, consultable en [http://acienpol.msinfo.info/bases/biblo/texto/boletin/2011/Bol-ACPS\\_2011\\_150\\_293-332.pdf](http://acienpol.msinfo.info/bases/biblo/texto/boletin/2011/Bol-ACPS_2011_150_293-332.pdf)
- SUPREMA CORTE DE JUSTICIA DE LA NACIÓN, 1994: *Manual del Juicio de Amparo*, México: Themis.
- TARUFFO, M., 2010: *Simply la verdad*, Madrid: Marcial Pons.
- , (2008): *La prueba*, Madrid: Marcial Pons.
- VÁZQUEZ, C., 2015: *De la prueba científica a la prueba pericial*, Madrid: Marcial Pons.



## **SOBRE LA OBLIGACIÓN DE APLICAR LOS PROTOCOLOS DE ESTAMBUL Y MINNESOTA EN LA JURISPRUDENCIA DE LA CORTE INTERAMERICANA DE DERECHOS HUMANOS**

María Elena Lumiento  
Universidad de Buenos Aires  
melumiento@derecho.uba.ar

**RESUMEN:** La autora analiza la posición de la Corte Interamericana de Derechos Humanos sobre el deber de aplicar los Protocolos de Estambul y de Minnesota en las investigaciones penales que se llevan a cabo en el ámbito interno de los Estados bajo su jurisdicción para determinar la responsabilidad de hechos de ejecuciones ilegales y de tortura. Señalando algunos de los problemas, normativos y epistemológicos, que conlleva esa posición.

**PALABRAS CLAVE:** derechos humanos; derecho internacional; soft law; derecho penal; prueba.

## **ON THE DUTY TO OBEY THE ISTAMBUL AND MINNESOTA PROTOCOLS IN THE JURISPRUDENCE OF THE INTER-AMERICAN COURT OF HUMAN RIGHTS**

**ABSTRACT:** The author analyzes the position of the Inter-American court in relation to the duty to apply the Istanbul and Minnesota protocols in judicial investigations carried out in the domestic

---

\* Quiero agradecer especialmente a Paula Rodríguez, Pablo Jalil, Julieta Rábanos, Carmen Vázquez y Daniela Accatino por la lectura y comentarios críticos de los borradores del presente al presente trabajo. El agradecimiento se extiende a dos dictaminadores/as anónimos/as. Los errores, como siempre, a mi exclusiva cuenta deben imputarse.

sphere to clear up acts of torture and illegal executions. It also points out some normative and epistemological problems derived from accepting that position.

**KEYWORDS:** human rights; soft law; internacional law; criminal law; evidence.

**SUMARIO:** PLANTEO DEL PROBLEMA.— 1. LA FUERZA NORMATIVA DE LOS PROTOCOLOS DE ESTAMBUL Y MINNESOTA SEGÚN LA JURISPRUDENCIA DE LA CORTE IDH: 1.1. El deber de realizar una investigación eficaz. 1.2. La Obligatoriedad de los protocolos inferida a partir del deber (implícito) de investigar las violaciones a los derechos humanos en el ámbito interno. 1.3. Condena al Estado de adoptar las normas de los protocolos como algún tipo de norma jurídica en el ámbito interno. 1.4. Algunas conclusiones del apartado.— 2. LOS PROTOCOLOS DE ESTAMBUL Y MINNESOTA: 2.1. Orígenes, finalidades y características. 2.2. Estatus normativo de los protocolos según el derecho internacional. 2.3. Los protocolos contienen «reglas técnicas».— 3. Una revisión crítica: 3.1. Estándar mínimo y eficacia. 3.2. Las normas de los protocolos no pueden reemplazar a las normas internas obligatorias.— 4. A modo de conclusión.— Bibliografía.

**RECOMMENDED CITATION:** MARÍA ELENA LUMIENTO, 2022: «Sobre la obligación de aplicar los Protocolos de Estambul y Minnesota en la jurisprudencia de la Corte Interamericana de Derechos Humanos», in *Quaestio facti*, 3: 217-249. Madrid: Marcial Pons Ediciones Jurídicas y Sociales. DOI: [http://dx.doi.org/10.33115/udg\\_bib/qf.i3.22702](http://dx.doi.org/10.33115/udg_bib/qf.i3.22702)

## PLANTEO DEL PROBLEMA

La Corte Interamericana de Derechos Humanos (en adelante, “Corte IDH”) ha afirmado, en numerosos casos, que los Protocolos conocidos comúnmente como “Protocolo de Estambul” (en adelante, “PE”) y “Protocolo de Minnesota” (en adelante, “PM”) deben ser aplicados por los Estados, miembros del Sistema Latinoamericano de protección de los derechos humanos, cuando estos realicen investigaciones para determinar la responsabilidad penal de aquellos imputados, en los ámbitos internos, de hechos de tortura o de ejecuciones ilegales.

Los protocolos regulan los estándares mínimos con los que, según el Derecho Internacional de los Derechos Humanos, deben obtenerse y producirse las pruebas para que pueda predicarse que una investigación es *eficaz*<sup>1</sup>.

Sin embargo, esas normas, por la forma en que fueron emitidas en el sistema internacional, no resultan obligatorias para los Estados. En otros términos, los proto-

<sup>1</sup> Una aclaración terminológica: el concepto de «eficacia» no es utilizado en los Protocolos —ni en la jurisprudencia de la Corte IDH— en su faz jurídica, sino que remite a su definición lexicográfica. Concretamente, la RAE define «eficaz» como «eficacia» y a la eficacia como «1. f. Capacidad de lograr el efecto que se desea o se espera». Mostraré *infra* (3) que, según la jurisprudencia de la Corte IDH, una investigación eficaz es aquella que permite llegar a un resultado determinado. El resultado esperado es la imposición de un castigo a los culpables de las violaciones a los derechos humanos.

colos no fueron adoptados mediante la firma de un tratado ni pueden ser calificados como una costumbre internacional (*i.e.* de una práctica generalizada aceptada como derecho) y tampoco dan cuenta de principios generales del derecho aceptados por la comunidad internacional.

La falta de encuadre en alguna de las fuentes del derecho internacional determina que esos documentos tengan el carácter de *soft law*, es decir, de criterios o estándares que no resultan, en principio, vinculantes para los Estados.

Para fundamentar su aplicabilidad, en el ámbito de los procesos penales internos de los Estados, la Corte IDH sostuvo que esas normas instancian ciertos deberes contenidos en la Convención Americana sobre Derechos Humanos (en adelante, CADH). Específicamente, el organismo internacional entendió que la aplicación de los Protocolos en las investigaciones penales internas que realicen los Estados de graves violaciones a los derechos humanos es una consecuencia derivada del compromiso que asumen estos de garantizar derechos (art. 1.1. CADH), las garantías judiciales (art. 8), de la garantía del acceso a la justicia (art. 25 CADH) y del derecho a la verdad. O, en otros términos, al interpretar el alcance de esas obligaciones estatales citadas la Corte crea – o induce – un nuevo deber: de realizar una investigación eficaz, el que solo puede ser satisfecho mediante el uso de los protocolos de Estambul y de Minnesota.

En ese contexto, el trabajo analiza las implicancias de la posición de la Corte IDH de sostener que los Estados tienen el deber de aplicar los protocolos en su ámbito interno. Puntualmente, se examinará si entender que esas normas obligan a los Estados conlleva contradicciones normativas con las normas internas que regulan la prueba (reglas de adquisición, producción y de valoración de la prueba) y con ciertos principios catalogados como “fundamentales” (*i.e.* normas constitucionales). Por último, se señalará que, aun cuando se entiendan que las normas resultan aplicables, estas presentan ciertas deficiencias técnicas que desaconsejan su uso.

## 1. LA FUERZA NORMATIVA DE LOS PROTOCOLOS DE ESTAMBUL Y MINNESOTA SEGÚN LA JURISPRUDENCIA DE LA CORTE IDH

La Corte IDH demanda el cumplimiento de los estándares contenidos en los Protocolos de Estambul y de Minnesota en los procesos judiciales penales que se lleven adelante en los ámbitos internos de los Estados para determinar la responsabilidad de los presuntos autores de las violaciones al derecho a la vida y a la prohibición de la tortura<sup>2</sup>. Asimismo, en la posición del organismo, el cumplimiento de esos estándares son los que, hacen posible, en otros derechos, al «derecho a la verdad».

---

<sup>2</sup> Aunque, claro está, este deber no se agota en el cumplimiento de esos estándares, para un panorama completo de la posición de la Corte IDH, véase CEJIL 2010: 38.

El derecho a la verdad<sup>3</sup> y el deber de realizar una investigación eficaz presentan una característica común: son inducidos de los mismos derechos explícitos. Es decir, son entendidos como consecuencia de la existencia y aplicación del derecho a las garantías judiciales (art. 8 CADH), el derecho al acceso a la justicia (art. 25 CADH) y del deber del Estado de respetar los derechos establecidos en la Convención Americana de Derechos Humanos (art. 1.1 CADH)<sup>4</sup>.

El razonamiento de la Corte se estructura sobre la base de las siguientes premisas: a) con la ratificación de la CADH, los Estados se convierten en garantes de los derechos protegidos por la Convención (art. 1.1. CADH) y; b) para no transgredir ese deber de garantía no deben lesionar ni tolerar que alguien lesione alguno de esos derechos, así como tampoco pueden «proteger» a los responsables de esas lesiones. Sobre esa base concluye que, cuando se viola alguno de esos derechos protegidos, el deber de garantía del Estado se «activa» e incluye la investigación y el castigo de quienes son responsables de esas violaciones, así como la reparación integral del daño causado a las víctimas<sup>5</sup>.

En ese contexto, en la posición de la Corte IDH, las reglas forenses contenidas en los Protocolos permiten efectivizar esa garantía, en tanto, al establecer los criterios de investigación, posibilitarían, —de observarse muy estrictamente—, el castigo de los responsables. Esa posición hoy puede considerarse una jurisprudencia consolidada.

La Corte IDH, en significativos y numerosos precedentes, ha sostenido que esas reglas *deben* aplicarse al momento de producir la prueba en el ámbito interno, puesto que son las únicas que instancian el derecho implícito de las víctimas a una investigación eficaz. Ello implica —según lo dicho— que la estricta observancia de los estándares fijados en los Protocolos por parte de los Estados miembros implica una mayor posibilidad de determinar a los eventuales responsables de los delitos.

El castigo, nos agrega un problema que parecería convertir a la obligación de realizar una investigación eficaz, típicamente concebida como una obligación de medios, en una obligación de «resultado»; aunque esa conclusión no es reconocida abiertamente por la Corte IDH. En sus sentencias se da un reconocimiento —pacífico y uniforme— sobre que del deber de realizar una investigación da origen a una obligación de medios. Pero, no es menos cierto, que, frente a la falta de castigo, como se mostrará luego, se ha ordenado la reapertura de las actuaciones,

<sup>3</sup> El derecho a la verdad no es un derecho consagrado en la CADH pero es reconocido como derecho implícito en la jurisprudencia de la Corte IDH a partir del caso “Velázquez Rodríguez” (sent. del 29/07/1988, párr. 174). Luego, volvió sobre la cuestión en los casos “Bámaca Velázquez” (sent. del 25/10/2000), “Barrios Altos” (sent. del 25/11/2003) y en todos los casos que aquí se analizan). Actualmente, su existencia se encuentra “consolidada”. Una visión crítica de la jurisprudencia de la Corte IDH y sus consecuencias para los ámbitos penales internos de los Estados puede verse: PASTOR y MACULAN, 2013.

<sup>4</sup> Sobre las formas de inferir principios de normas positivas, véase ALONSO 2018.

<sup>5</sup> Véase Corte IDH en el caso “Velázquez Rodríguez vs. Honduras”, párr. 166.

así como la aplicación de los estándares omitidos, para obtener ese resultado dentro del proceso interno.

Resumidamente la posición de la Corte IDH podría reconstruirse de la siguiente manera: la obligación de investigar graves violaciones a los derechos humanos es uno de los deberes elementales del Estado para garantizar la tutela de otros derechos fundamentales. Correlativamente, y como parte de un recurso efectivo contra esas violaciones, el Estado en el que hayan tenido lugar esos acontecimientos debe garantizar a las víctimas un adecuado acceso a la justicia que permita asegurar una investigación seria sobre los hechos denunciados.

Seguidamente, mostraré qué alcance tuvo en la jurisprudencia de la Corte IDH el deber de realizar una investigación eficaz<sup>6</sup> y su vinculación con la «obligación» de aplicar los protocolos en los procesos penales internos en los que se investiguen graves violaciones a los Derechos Humanos.

### 1.1. El deber de realizar una investigación eficaz

La Corte IDH ha establecido que en los casos de graves violaciones a los derechos humanos existe un *deber de diligencia* que debe observarse en el ámbito interno de los Estados que investiguen esos hechos para determinar la responsabilidad penal de los responsables de esas violaciones.

Esa obligación estatal de investigar diligentemente es inducida por el organismo a partir del reconocimiento del derecho a la vida y la prohibición de la tortura (art. 3 CADH) y de la garantía del acceso a la justicia (art. 25 CADH). Además, el organismo ha sostenido que debe ser *necesariamente* efectiva para tener un efecto tutelar, aleccionador y disuasivo<sup>7</sup>.

En efecto, desde su primer caso contencioso “*Velázquez Rodríguez*”, la Corte IDH sostuvo que el Estado tiene el deber jurídico: a) de prevenir, razonablemente, las violaciones de los derechos humanos; b) de *investigar seriamente* con los medios a su alcance las violaciones que se hayan cometido dentro del ámbito de su jurisdicción a

<sup>6</sup> El término eficacia es ambiguo y admite, por ello, más de un significado en sentido jurídico. Siguiendo a Guastini en el lenguaje doctrinal, el vocablo «eficacia» —en tanto predicado de normas o actos jurídicos— es usado en no menos de tres sentidos distintos. En un primer sentido, eficacia significa capacidad de producir efectos jurídicos. En segundo sentido eficacia significa aplicabilidad. Así se considera eficaz la norma que puede o debe ser aplicada. En un tercer sentido, eficacia designa la peculiar fuerza de una norma, es decir, su ubicación en la jerarquía del ordenamiento jurídico (GUASTINI, 2017: 399-400). También, según el autor, cabría agregar un cuarto sentido no estrictamente jurídico, sino sociológico-jurídico, donde se denomina «eficaz» a toda que sea efectivamente usada, observada o aplicada.

<sup>7</sup> Corte IDH. *Caso de los Hermanos Gómez Paquiyauri Vs. Perú*. Fondo, Reparaciones y Costas. Sentencia de 8 de julio de 2004. Serie C No. 110, párr. 130; *Caso Myrna Mack Chang Vs. Guatemala*. Fondo, Reparaciones y Costas. Sentencia de 25 de noviembre de 2003. Serie C No. 101, párr. 156.

fin de identificar a los responsables; c) de imponerles las sanciones pertinentes y; d) de asegurar a la víctima una adecuada reparación<sup>8</sup>. Asimismo, entendió a la obligación de investigar, en sí misma, como una forma de reparación, ante la necesidad de remediar la violación del derecho a conocer la verdad en el caso concreto<sup>9</sup>.

Respecto de las consecuencia del deber de diligencia, la Corte IDH ha sostenido que la investigación que realice el Estado para aclarar los hechos que constituyen graves violaciones a los derechos humanos en su ámbito interno debe ser oportuna<sup>10</sup> (*i.e.* debe ser iniciada de manera inmediata para impedir la pérdida de pruebas que puedan resultar fundamentales para la determinación de responsabilidades); debe realizarse en un plazo razonable<sup>11</sup> (*i.e.* sin demoras prolongadas e innecesarias) y ‘propositiva’<sup>12</sup> (*i.e.* debe ser impulsada de oficio por las autoridades judiciales locales, las que deben proponer medidas de investigación pertinentes para el esclarecimiento de los hechos).

A su turno, respecto al tipo de obligación a que da lugar el deber de diligencia, sostuvo que da cuenta o responde a la estructura de una obligación de medios. En términos de la Corte IDH: «[E]l deber de investigar es una obligación de medios, y no de resultados. Sin embargo, debe ser asumida por el Estado como un deber jurídico propio y no como una simple formalidad condenada de antemano a ser infructuosa»<sup>13</sup>. Sin embargo, a la par ha exigido el castigo y frente a la falta de ese resultado ha ordenado la apertura de las actuaciones en el ámbito interno<sup>14</sup>.

Puntualmente, en el ya citado caso “*Velázquez Rodríguez*”, la Corte IDH entendió que el fin explícito de la investigación debe ser «identificar a los responsables»; imponerle las «sanciones pertinentes» y «asegurar a la víctima una adecuada reparación».

También, en el caso “*Barrios Altos*” la Corte se ha expedido en un sentido similar, e incluso, el Juez Sergio García Ramírez ha expresado que: «[E]s inadmisibles la impunidad de las conductas que afectan más gravemente los principales bienes

<sup>8</sup> Corte IDH. *Caso Velázquez Rodríguez Vs. Honduras*. Fondo, Reparaciones y Costas. Sentencia de 29 de julio de 1988. Serie C No. 4, párr. 174.

<sup>9</sup> Corte IDH. *Caso Anzualdo Castro Vs. Perú*. Excepción Preliminar, Fondo, Reparaciones y Costas. Sentencia de 22 de septiembre 2009. Serie C No. 202, párr. 118. Cfr. Corte IDH. *Caso de la Masacre de Pueblo Bello Vs. Colombia*, *supra* nota 22, párr. 266.

<sup>10</sup> Véase Corte IDH. *Caso Ximenes Lopes Vs. Brasil*, *supra* nota 22, párr. 189.

<sup>11</sup> Ya que ello por sí solo, en los términos de la Corte IDH, implica una violación a la garantía del acceso a la justicia (art. 25 CADH). Para determinar la razonabilidad del plazo ha sostenido que es preciso tomar en cuenta tres elementos: a) complejidad del asunto, b) actividad procesal del interesado y c) conducta de las autoridades judiciales. En tal sentido, véase Corte IDH. *Caso de las Hermanas Serrano Cruz Vs. El Salvador*, *supra* nota 63, párr. 69.

<sup>12</sup> Corte IDH. *Caso de la Masacre de Pueblo Bello Vs. Colombia*, *supra* nota 22, párr. 143; *Caso de la Masacre de Mapiripán Vs. Colombia*, *supra* nota 25, párr. 219 y 223; *Caso de la Comunidad Moiwana Vs. Suriname*, *supra* nota 1, párr. 145 y *Caso Juan Humberto Sánchez Vs. Honduras*, *supra* nota 1, párr. 132.

<sup>13</sup> Véase Corte IDH. *Caso Heliodoro Portugal Vs. Panamá*, *supra* nota 49, párr. 144.

<sup>14</sup> GALAIN PALERMO sostiene que incluso que existe un derecho absoluto a la verdad que implica el castigo de todos los responsables de las graves violaciones a los derechos humanos.

jurídicos sujetos a la tutela de ambas manifestaciones del Derecho internacional. La tipificación de esas conductas y el procesamiento y sanción de sus autores —así como de otros participantes— constituye una obligación de los Estados, que no puede eludirse a través de medidas tales como la amnistía, la prescripción, la admisión de causas excluyentes de incriminación y otras que pudieran llevar a los mismos resultados y determinar la impunidad de actos que ofenden gravemente esos bienes jurídicos primordiales»<sup>15</sup>.

En el mismo sentido, en el caso “*La cantuta vs. Perú*” ha sostenido que los esfuerzos para conocer la verdad son «inútiles» si no vienen acompañados del castigo de los responsables<sup>16</sup>. El castigo, además, debe incluir a todos los responsables y, por ende, no alcanza con que se sancione a alguno de ellos. En palabras de la Corte IDH:

[...] si bien han sido condenados algunos de los responsables de la masacre, subsiste una impunidad generalizada en el presente caso, en la medida en que no ha sido determinada toda la verdad de los hechos ni la totalidad de las responsabilidades intelectuales y materiales por los mismos.<sup>17</sup>

También, la Corte IDH ha sostenido que se encuentra prohibida la aplicación en el ámbito interno de los Estados, a los casos en que se investiguen graves violaciones de derechos humanos, ciertos institutos constitucionales y/o procesales, *v.gr.*, las amnistías, indultos e incluso la prescripción, cuando estos resulten obstáculos para la identificación y castigo de los responsables de los hechos<sup>18</sup>.

De lo anterior se sigue que, pese a que la Corte IDH sostiene que el deber de realizar una investigación eficaz da cuenta de una obligación de medios, lo cierto es que la exigencia de un «castigo» invierte la calificación de ese derecho dando lugar a una verdadera obligación de resultado’ (*i.e.* una sanción a una persona o personas).

Por último, respecto a los medios que permiten alcanzar ese resultado, y que a la vez sirven para evaluar si un Estado ha cumplido con el deber de realizar una investigación eficaz, el organismo internacional ha utilizado como referencia —casi de modo exclusivo— a los Protocolos de Estambul y de Minnesota, perdiendo de vista que existen numerosos documentos que establecen reglas similares a las contenidas en ellos y, por ello, también podrían resultar aplicables en tanto pretender ser referencias del estado de la *lex artis*<sup>19</sup>.

---

<sup>15</sup> Véase Corte IDH. *Caso Barrios Altos vs. Perú*, voto voto concurrente del juez Sergio García Ramírez, párr. 13.

<sup>16</sup> Véase Corte IDH. *Caso La Cantatuta vs. Perú*, párr. 161.

<sup>17</sup> Véase Corte IDH. *Caso de la Masacre de Mapiripán Vs. Colombia*, *supra* nota 25, párr. 236.

<sup>18</sup> Véase Corte IDH. *Caso Barrios Altos vs. Perú*, párr. 41 y ss.

<sup>19</sup> Tales como, las Observaciones y recomendaciones aprobadas por consenso el 21 de febrero de 2003 en la Conferencia Internacional de Expertos gubernamentales y no gubernamentales, en el marco del Proyecto Las personas desaparecidas y sus familiares del Comité Internacional de la Cruz Roja; reconociendo la importancia de sus recomendaciones en el tema de la desaparición forzada; El Protocolo Modelo para la investigación forense de muertes sospechosas de haberse producido por violación de los derechos humanos de la Oficina del Alto Comisionado de las Naciones Unidas para los Derechos Humanos; El Informe sobre mujeres no identificadas o de identificación dudosa, muertas en las ciudades

Observo que la exigencia de castigo, tiene como consecuencia transformar a las normas contenidas en los Protocolos en los medios para cumplir con ese deber. A la vez, esos medios son calificados, según mostraré infra 1.2 y 1.3, como «necesarios», puesto que son los únicos que posibilitarían la aplicación de un castigo a los responsables de los hechos. Incluso, no solo la Corte IDH sostuvo que los Estados deben tener en cuenta esas reglas en los ámbitos internos, sino que, en los casos en los que fueron omitidos, ordenó la reapertura del proceso a efectos de su aplicación.

## 1.2. La Obligatoriedad de los protocolos inferida a partir del deber (implícito) de investigar las violaciones a los derechos humanos en el ámbito interno

En un gran grupo de casos la Corte IDH sostuvo que la aplicación de los protocolos resultaba una consecuencia derivada del deber de realizar una investigación eficaz de las violaciones a los derechos humanos fundamentales ocurridas en el ámbito interno de los Estados.

Ese criterio ha sido establecido por la Corte IDH en los casos: “*Juan Humberto Sánchez Vs. Honduras*”<sup>20</sup>; “*Caso de la Masacre de Pueblo Bello vs. Colombia*”<sup>21</sup>; “*Caso Baldeón García Vs. Perú*”<sup>22</sup>; “*Ximenes López vs. Brasil*”<sup>23</sup>; “*Caso Vargas Areco vs. Paraguay*”<sup>24</sup>; y “*Espinoza Gonzáles Vs. Perú*”<sup>25</sup>.

---

de Juárez y Chihuahua del Equipo Argentino de Antropología Forense –*Mass Fatality Incidents: A Guide for Human Forensic Identification del Departamento de Justicia de los Estados Unidos de América; Lessons Learned From 9/11: DNA Identification in Mass Fatality Incidents del US Department of Justice*; La gestión de cadáveres en situaciones de desastre: guía práctica para equipos de respuesta de la Organización Panamericana de la Salud, Organización Mundial de la Salud, el Comité Internacional de la Cruz Roja y la Federación Internacional de Sociedades de la Cruz Roja y de la Media Luna Roja100; Manejo de cadáveres en situaciones de desastre de la Organización Panamericana de la Salud; Normas Directivas para Médicos con respecto a la Tortura y a otros Tratos o Castigos Crueles, Inhumanos y Degradantes, Impuestos sobre Personas Detenidas o Encarceladas (Declaración de Tokio de la Asociación Médica Mundial); Principios de ética médica aplicables a la función del personal de salud, especialmente los médicos, en la protección de personas presas y detenidas contra la tortura y otros tratos o pena crueles, inhumanos y degradantes; Principios y directrices básicos sobre el derecho de las víctimas de violaciones manifiestas de las normas internacionales de derechos humanos y de violaciones graves del derecho internacional humanitario a interponer recursos y obtener reparaciones, entre otros documentos internacionales.

<sup>20</sup> Corte IDH. *Juan Humberto Sánchez Vs. Honduras*. Excepción Preliminar, Fondo, Reparaciones y Costas. Sentencia de 7 de junio de 2003. Serie C No. 99.

<sup>21</sup> Corte IDH, *Caso de la masacre de Pueblo Bello vs. Colombia*. Fondo, reparaciones y costas. Sentencia de 31 de enero de 2006. Serie C. 140.

<sup>22</sup> Corte IDH. *Caso Baldeón García Vs. Perú*. Excepciones preliminares, Fondo, Reparaciones y Costas. Sentencia de 6 de abril de 2006. Serie C No. 147.

<sup>23</sup> Corte IDH. *Caso Ximenes López vs. Brasil*. Excepciones preliminares, Fondo, reparaciones y Costas. Sentencia de 4 de julio de 2006. Serie C No. 149.

<sup>24</sup> Corte IDH. *Caso Vargas Areco vs. Paraguay*. Excepciones preliminares, Fondo, reparaciones y Costas. Sentencia de 30 de octubre de 2008. Serie C No. 155.

Nota 25 en página siguiente

El argumento esgrimido por la Corte IDH para sostener la aplicación de esas normas fue, sistemáticamente, el mismo: en los casos que tenían por objeto la investigación de hechos de tortura o de ejecuciones ilegales, el Estado tenía la obligación de realizar una investigación seria, imparcial y efectiva de lo acaecido; siendo los protocolos documentos que señalan los requisitos mínimos de la investigación y, por tanto, debieron ser utilizados por los Estados en esas investigaciones.

Asimismo, el Tribunal Internacional afirmó que cualquier carencia o defecto en la investigación que perjudique la eficacia para establecer la causa de la muerte o identificar a los responsables materiales o intelectuales, implica una violación a la obligación de respetar y proteger los derechos humanos consagrados en la Convención Americana de Derechos Humanos<sup>26</sup>.

En términos de la Corte IDH:

«El Protocolo (de Minnesota) ha señalado como requisitos mínimos de la investigación: la identificación de la víctima, la recolección y preservación de pruebas relacionadas con la muerte con el fin de ayudar en el potencial procesamiento de los responsables, la identificación de posibles testigos y la obtención de sus declaraciones en relación con la muerte, la determinación de la causa, manera, lugar y tiempo de la muerte, así como cualquier patrón o práctica que pueda haber provocado la muerte, la distinción entre muerte natural, suicidio y homicidio, la identificación y aprehensión de la o las personas involucradas en la muerte y la presentación de los presuntos perpetradores ante un tribunal competente establecido por ley. *En este caso no se cumplieron dichos parámetros*» (Corte IDH, caso *Humberto Sanchez vs. Honduras*, párr. 127) (el destacado no es original). «La Corte ha constatado que en Honduras existió y existe un estado de *impunidad* respecto de los hechos del presente caso (supra 135), cuando el Estado tiene la obligación de prevenir e investigar lo sucedido. La Corte entiende como impunidad *la falta en su conjunto de investigación, persecución, captura, enjuiciamiento y condena de los responsables de las violaciones de los derechos protegidos por la Convención Americana*, toda vez que el Estado tiene la obligación de combatir tal situación por todos los medios legales disponibles ya que la impunidad propicia la repetición crónica de las violaciones de derechos humanos y la total indefensión de las víctimas y de sus familiares» (Corte IDH, caso *Humberto Sanchez vs. Honduras*, párr. 149) (el destacado no es original).

En el caso *Vargas Areco* la Corte IDH sostuvo que los procedimientos (se refería a las medidas de prueba y entre ellas, a la autopsia realizada en la investigación judicial del hecho llevada a cabo por la República de Paraguay) deben tomar en consideración las normas internacionales de documentación e interpretación de los elementos de prueba forense respecto de la comisión de actos de tortura y, particularmente, las definidas en el Protocolo de Estambul (párr. 93).

Sobre esa base concluyó la Corte que la investigación de la ejecución extrajudicial de Gerardo Vargas Areco, así como de su supuesta tortura, no se llevó a cabo de

<sup>25</sup> Corte IDH. *Espinoza González Vs. Perú*. Excepciones preliminares, Fondo, Reparaciones y Costas. Sentencia de 20 de noviembre de 2014. Serie C No. 289.

<sup>26</sup> Cfr., *Caso Humberto Sánchez*, párr. 127; *Caso Baldeón García*, párr. 96; *Caso de la Masacre de Pueblo Bello*, párr. 177; y *Caso de la «Masacre de Mapiripán»*, supra nota 8, párr. 224; *Caso Ximenes López*, párr. 179; CASO VARGAS ARECO, párr. 91).

manera tal que garantizara los derechos reconocidos en los artículos 4 y 5.1 de la Convención, en relación con los artículos 1.1 de la misma, y 6 y 8 de la Convención Interamericana para Prevenir y Sancionar la Tortura, en perjuicio de los familiares del niño Vargas Areco (párr. 94).

Merece ser destacado que la Corte IHD analiza infracciones evidentes al deber de investigar dado que, en algunos casos, ni siquiera se realizaban autopsias de los cuerpos encontrados (por ejemplo, en los casos “*Humberto Sánchez*”; “*Vargas Areco*”; “*Baldeón García*”; “*Masacres de Ituango*”; “*Masacre de Pueblo Bello*”). Sin embargo, también hubo otros casos en los que se habían realizado medidas de pruebas en el ámbito interno y la Corte IDH entendió que estas no cumplían requisitos mínimos de fiabilidad dado que no respetaban los parámetros establecidos, *v.gr.*, en el Protocolo de Minnesota<sup>27</sup>.

Asimismo, todos los casos tienen en común que la Corte IDH ordenó re abrir la investigación del hecho y aplicar los estándares omitidos. Esas investigaciones habían sido ‘cerradas’ en los ámbitos internos por falta de prueba en la gran mayoría de los casos

### 1.3. Condena al Estado de adoptar las normas de los protocolos como algún tipo de norma jurídica en el ámbito interno

En otro segundo grupo de casos, la Corte, además de señalar que deben observarse las normas sobre prueba contenidas en los protocolos, le impone al Estado, cuya responsabilidad internacional declara, el deber de dictar una norma que los incorpore como normas jurídicas internas.

Los casos en los que la Corte IDH adoptó esa posición son dos: “*González y otras (campo algodón) vs. México*”<sup>28</sup> y “*Rosendo Cantú vs. México*”<sup>29</sup>.

Ambos casos tienen en común, además del país demandado, un contexto de violencia generalizada contra grupos vulnerables, sea por una condición de género o bien por tratarse de un integrante de una comunidad originaria.

Al igual que en el otro grupo de casos analizados, la Corte IDH hizo especial hincapié en que las falencias de la investigación, por haber omitido la aplicación de

<sup>27</sup> Dado que se contaba con una descripción completa de las lesiones que presentaba el cuerpo o porque existían discrepancias entre peritos que no se habían salvado en el contexto de la investigación solicitando un perito tercero en discordia, etc. Tal es el caso, por ejemplo, de las falencias señaladas por la Corte IDH en la sentencia conocida como “*Campo Algodonero*”.

<sup>28</sup> Corte IDH. *Caso González y otras (“Campo Algodonero”) Vs. México*. Excepción Preliminar, Fondo, Reparaciones y Costas. Sentencia de 16 de noviembre de 2009. Serie C No. 205.

<sup>29</sup> Corte IDH. *Caso Rosendo Cantú y otra Vs. México*. Excepción Preliminar, Fondo, Reparaciones y Costas. Sentencia de 31 de agosto de 2010. Serie C No. 216.

los estándares internacionales, impidieron arribar a conclusiones sobre los hechos en el ámbito interno lo que, en definitiva, determinó que no pudiera castigarse a los responsables.

Ahora bien, si bien los dos casos presentan un sinfín de aristas interesantes para analizar, por razones de espacio realizaré solo una breve referencia al caso “ROSENDO CANTÚ”. El caso trataba de un hecho de violación perpetrado por dos militares a una mujer campesina. El hecho habría tenido lugar en el marco de un interrogatorio realizado en una comunidad originaria de México.

La Corte IDH —en el proceso internacional y a los efectos de declarar la infracción del deber de realizar una investigación eficaz— entendió que la violación se encontraba probada, sobre la base de la declaración de una única prueba de cargo: la declaración de la víctima<sup>30</sup>. En consecuencia, el organismo declaró la responsabilidad del Estado México porque en la investigación no se habían respetado los estándares establecidos en el PE, ni para la toma de testimonio de víctimas de tortura, ni tampoco los establecidos para efectuar los peritajes médicos y psiquiátricos de constatación de lesiones<sup>31</sup>.

En concreto, sostuvo que el informe fiel del examen físico y ginecológico no fue detallado y completo, como recomienda el Protocolo de Estambul, y iv) los métodos de investigación tampoco cumplieron con los parámetros establecidos en el PE (párr.168). Teniendo en consideración los estándares que habían sido omitidos por el Estado ordenó, además, la reapertura del proceso interno (Corte IDH, Caso “*Rosendo Cantú vs. México*”, punto 1 del resolutorio).

Asimismo, la Corte IDH ordenó que México continúe con el proceso de estandarización de un protocolo de actuación para el ámbito federal conforme a los parámetros establecidos en el Protocolo de Estambul (Corte IDH, Caso “*Rosendo Cantú vs. México*”, punto 16 del resolutorio).

---

<sup>30</sup> Aunque también valoró: 2) la credibilidad del relato de la víctima (la que encontraba respaldo en el dictamen psiquiátrico, efectuado sobre la víctima en el ámbito internacional y varios años después de los hechos y daba cuenta de que esta habría sufrido estrés postraumático agudo y depresión); 3) la declaración de testigos de oídas (quienes habrían visto a la víctima llorando y con su cara ensangrentada) y; 4) un informe del médico legista que atendió a y que sostuvo que la víctima presentaba una *escoriación* en el ojo derecho. Ahora bien, las pruebas no eran independientes de la declaración de la víctima y tampoco daban cuenta de un hecho de violación más allá del propio relato de la víctima.

<sup>31</sup> México alegó encontrarse en un proceso de estandarización de sus normas internas a lo dispuesto en el Protocolo de Estambul, reconoció parcialmente su responsabilidad por la falta de atención a la salud de la Sra. Rosendo Cantú (párr. 240). Además, respecto de la valoración de la declaración de la víctima, sostuvo que no podía ser considerada prueba plena sin otros elementos independientes corroborantes, y agregó que no podía existir una violación al deber de investigar porque la investigación de los hechos se encontraba en curso (párr.73 y 158).

#### 1.4. Algunas conclusiones del apartado

De la reconstrucción de las sentencias de la Corte IDH surge que el organismo entiende que los estándares de investigación establecidos en los Protocolos de Estambul y de Minnesota constituyen la única forma válida para cumplir con ciertos deberes derivados, a su vez, de ciertas obligaciones ‘implícitas’ contenidas en los arts. 1.1., 8 y 25 de la CADH.

Esas normas, en el sistema de derecho internacional, no son consideradas obligatorias porque no constituyen una fuente del derecho internacional. Aunque tampoco podrían ser consideradas fuente en virtud de la aplicación de las normas de la Convención Americana de Derechos Humanos que regulan la competencia del organismo internacional<sup>32</sup>.

En otro orden de cosas, cuando la Corte IDH resolvió los casos mencionados no analizó si las normas forenses que resultaron aplicadas en el ámbito interno, efectivamente daban o no cuenta de parámetros forenses adecuados. Es decir, no se analizó si el estándar aplicado en el ámbito interno era insuficiente o inadecuado o si, por el contrario, en esos casos puntuales el órgano de investigación estatal interno había fallado en su aplicación. Pero, en cualquier caso, según lo entiende la Corte IDH, frente a la existencia de normas aplicables en el ámbito interno, esas reglas se vieron *desplazadas* en su aplicación por las normas de los protocolos, dado que estas no se adecuan a los estándares mínimos establecidos en ellos.

Según he mostrado, en los casos que tuvo a decisión la Corte IDH, entendió que las normas contenidas en los protocolos deben ser aplicadas en los procesos penales internos. En otros términos, afirmó que existe el deber de aplicar de las reglas contenidas en el PE y el PM. De otro modo, el organismo debería haber utilizado los protocolos como “parámetro comparativo” para determinar si el Estado había incurrido en responsabilidad internacional, pero no hizo esto. Fue más lejos y ordenó reabrir las investigaciones y aplicar los estándares internacionales omitidos. De este modo, sentó el criterio de que esas reglas no pueden ser omitidas en las investigaciones internas que se

---

<sup>32</sup> A diferencia de otros organismos internacionales, no tiene competencia para la aplicación e interpretación de otras normas de derecho internacional que las contenidas en la CADH. Un ejemplo de un organismo que tiene una competencia más amplia es la Corte Internacional de Justicia. Así, lo prevén expresamente los arts. 34 a 38 del Estatuto de la Corte Internacional de Justicia. Asimismo, el estatuto establece qué normas deben aplicarse para resolver las cuestiones sometidas a su jurisdicción. Así ese organismo deberá aplicar: a. las convenciones internacionales, sean generales o particulares, que establecen reglas expresamente reconocidas por los Estados litigantes; b. la costumbre internacional como prueba de una práctica generalmente aceptada como derecho; c. los principios generales de derecho reconocidos por las naciones civilizadas; d. las decisiones judiciales y las doctrinas de los publicistas de mayor competencia de las distintas naciones, como medio auxiliar para la determinación de las reglas de derecho, sin perjuicio de lo dispuesto en el artículo 59 (véase art. 38 ECIJ). Asimismo, esa disposición no restringe la facultad de la Corte para decidir un litigio ex aequo et bono, si las partes así lo conviniere.

realicen para el juzgamiento de esos hechos. Así, la Corte IDH le reconoció un carácter vinculante a los Protocolos: los Estados que hayan ratificado la Convención Americana de Derechos Humanos no solo deben tener en cuenta lo dispuesto en los protocolos, sino que tienen el deber de usar, aplicar y observar efectivamente esas normas.

Además, para que una investigación sea eficaz, según los términos de las sentencias analizadas, debe obtenerse un resultado: el castigo de los responsables de los hechos. Es decir, la investigación interna debe concluir afirmando la responsabilidad de una 'persona' como responsable de los hechos violatorios de derechos humanos fundamentales imponiéndole un castigo. A la vez, la Corte IDH sugiere que ese resultado se obtiene necesariamente a través del estricto cumplimiento de las normas contenidas en los Protocolos.

## 2. LOS PROTOCOLOS DE ESTAMBUL Y MINNESOTA

Según se mostró, cuando la Corte IDH se refirió en sus sentencias a las normas que instancian el deber de realizar una investigación eficaz, hizo mención a que en los procesos internos en los que juzgaron ciertos hechos graves que constituyen delitos y violaciones a la Convención Americana de Derechos Humanos, no se aplicaron las reglas contenidas en los Protocolos de Estambul y de Minnesota.

Los protocolos fueron emitidos en el ámbito de la Organización de las Naciones Unidas y, en esencia, son dos instrumentos internacionales que tienen por objeto establecer los estándares mínimos de diligencia para la investigación de los delitos de tortura —y otros tratos crueles, inhumanos y degradantes— (Protocolo de Estambul), y de desaparición forzada de personas y/o ejecuciones ilegales (Protocolo de Minnesota).

Sin embargo, no fueron adoptados a través de tratados internacionales, por lo que esas normas, en principio, no son normas obligatorias y, por tanto, no resultan exigibles ni crean deberes jurídicos exigibles de manera inmediata, así como tampoco de su incumplimiento se debería seguir la responsabilidad internacional del Estado.

### 2.1. Orígenes, finalidades y características

Los documentos tienen origen en la preocupación de la comunidad internacional sobre la falta de efectividad en la investigación y castigo de hechos de tortura y de ejecuciones ilegales<sup>33</sup>. Esa falencia, encontraba origen en la falta de estándares científicos unificados, lo representaba el principal obstáculo para juzgar y prevenir esos hechos<sup>34</sup>.

<sup>33</sup> Cfr. CEJIL, 2010: 46 y ss.

<sup>34</sup> Concretamente, en la versión original del PM se señalaba que ya desde 1979 se venía manifestando la necesidad de establecer normas científicas de uso común para la investigación de muertes.

El primer documento emitido en el ámbito internacional es el conocido como Protocolo de Minnesota, cuya primera versión llevó el nombre de «Manual de la ONU para la prevención efectiva y la investigación de ejecuciones extralegales, arbitrarias y sumarias» (1991)<sup>35</sup>. Ese documento fue elaborado como norma complementaria de los Principios de Naciones Unidas relativos a una prevención efectiva y la investigación de ejecuciones extralegales, arbitrarias y sumarias, aprobado dos años antes (1989)<sup>36</sup>.

La versión original del Protocolo de Minnesota se elaboró con la intervención de grupos de investigación «expertos», dirigido por el *Minnesota Lawyers International Human Rights Committee* e integrado por profesionales forenses de diversas áreas, organizaciones no gubernamentales y otros especialistas. Su principal objetivo era contribuir a superar los inconvenientes técnicos, que se observaban en las investigaciones penales internas, y, en consecuencia hacer más probable que las investigaciones penales prosperen, (PM 1991).

A su turno en 1999, se crea el Protocolo de Estambul cuyo nombre oficial es «Manual para la investigación y documentación eficaces de la tortura y otros tratos o penas crueles, inhumanas o degradantes»<sup>37</sup>.

Ese manual fue elaborado por más de 75 «expertos» representando a más de 40 organizaciones de 15 países. Entre los que se encontraban: médicos forenses, médicos, psicólogos, observadores de derechos humanos y juristas de Alemania, Chile, Costa Rica, Dinamarca, Estados Unidos de América, Francia, India, Israel, Países Bajos, Reino Unido, Sri Lanka, Sudáfrica, Suiza y Turquía, así como de los territorios

---

Así, se sostiene que: «En 1979 el Grupo Médico Danés de Amnesty International expresó el deseo de adoptar normas internacionales establecidas para llenar certificados de defunción. En 1984 J.L. THOMSEN observó divergencias en el ejercicio de la medicina forense y que las comunicaciones se facilitarían al contar con directrices y definiciones comunes. Las organizaciones no gubernamentales destacaron la necesidad de elaborar y aprobar normas internacionales como resultado práctico de sus misiones en países donde se denunciaba que habían tenido lugar ejecuciones extralegales, arbitrarias o sumarias. Por ejemplo, una misión de Amnesty International en un país llegó a la conclusión en 1983 de que los procedimientos de las autoridades para registrar e investigar muertes violentas eran del todo inadecuados para determinar la causa de más de 40.000 muertes ocurridas entre 1979 y 1984, o para identificar a los responsables. Los procedimientos eran incluso inadecuados para determinar el número preciso de esas muertes. Análogamente, una delegación de la Asociación Estadounidense para el Progreso de la ciencia enviada a otro país para secundar en la identificación de miles de personas, secuestradas o asesinadas entre 1976 y 1983 llegó a la conclusión de que la identificación de los restos rebasaba la capacidad de las instituciones locales, y recomendó la creación de un centro investigador nacional con especialistas forenses idóneos y un director con facultades independientes de investigación. Con todo, la delegación opinaba que aún la identificación de un número reducido de los “desaparecidos” y la determinación de la causa de sus muertes podría tener un efecto disuasivo importante si cabía identificar y enjuiciar a los responsables» (PM, 1991: introducción). Es decir, claramente, el origen de las normas es la necesidad práctica de contar con estándares unificados de investigación.

<sup>35</sup> Aprobado por la Res. U.N. Doc. E/ST/CSDHA/12 (1991).

<sup>36</sup> <https://www.ohchr.org/EN/Issues/Executions/Pages/RevisionoftheUNManualPreventionExtra-LegalArbitrary.aspx?platform=hootsuite>

<sup>37</sup> Aprobado por la Res. U.N. Doc. HR/P/PT/8 (1999).

palestinos ocupados (PE, 2001: 2). Asimismo, y al igual que el PM, el PE fue elaborado para facilitar a los Estados la investigación penal a través del establecimiento de reglas para la documentación eficaz de las pruebas de la tortura (PE 1999).

Los documentos tienen por objetivo establecer reglas que, de cumplirse integralmente, permitirían que se realice una investigación eficaz de las violaciones graves a los derechos humanos. Así el PM se sostiene que:

«El objeto del Protocolo de Minnesota es proteger el derecho a la vida y promover la justicia, la rendición de cuentas y el derecho a una reparación mediante la promoción de una investigación eficaz de toda muerte potencialmente ilícita o sospecha de desaparición forzada. El Protocolo establece una norma común de desempeño en la investigación de una muerte potencialmente ilícita o una sospecha de desaparición forzada, así como un conjunto común de principios y directrices para los Estados, las instituciones y las personas que participen en la investigación» (PM, §1).

En similar sentido en el PE puede leerse que:

«El presente manual se ha preparado para contribuir a que los Estados utilicen uno de los medios fundamentales para la protección de los individuos contra la tortura: una documentación eficaz. Esta documentación saca a la luz las pruebas de torturas y malos tratos de manera que se pueda exigir a los torturadores que den cuenta de sus actos y permitir que se haga justicia» (PE, p. 3).

Los protocolos sugieren a los Estados su adopción como normas de documentación de pruebas internas o, cuanto menos, que sus estándares, pese a no ser normas internas, se apliquen a los procesos penales en los que se investiguen hechos de homicidio o tortura (PM 2016: párr. 4 / PE ). Su fundamento radica en que sus normas son una «garantía» en caso de violación a un derecho fundamental a la vida (PM 2016, párr. 8).

Algo similar surge de la lectura del PE dado que, si bien no afirma que sus disposiciones resultan vinculantes para los Estados, sí se sostiene que concretan el contenido del deber de investigar todo hecho de tortura. En este sentido, el PE realiza una extensa recopilación de las normas jurídicas de derecho internacional sobre derechos humanos que resultan vinculantes para los Estados miembros del Sistema de las Naciones Unidas y en ese marco, especifica que existe un deber de realizar una investigación eficaz para cumplir con esas normas (PE, cap. 1). Es decir, pese a reconocer que el documento en sí no es vinculante, sí lo es el deber de investigar, y castigar a los culpables, conforme a las normas del derecho internacional de los derechos humanos.

En efecto, la forma en que se justifica la fuerza normativa de los protocolos respecto de los países miembros de las Naciones Unidas es que instancian ciertos deberes cuyo respeto resulta obligatorio para la comunidad internacional por resultar derechos contenidos en normas de *Hard Law*. De este modo, según esa reconstrucción normativa, las normas de los protocolos resultarían complementarias de las normas vinculantes puesto que definirían el estándar de diligencia en la investigación que requieren esos derechos humanos vulnerados.

En lo que respecta a las medidas de prueba, si lo que se investiga es una ejecución ilegal, deben producirse medidas para: a) identificar a la víctima o las víctimas; b)

recuperar y preservar todo material probatorio de la causa y las circunstancias de la muerte, y de la identidad del autor o los autores del delito; c) identificar posibles testigos y obtener sus testimonios en relación con la muerte y las circunstancias que la rodearon; d) determinar la causa, la manera en que se produjo, el lugar y el momento de la muerte, y todas las circunstancias del caso y; e) determinar quién estuvo involucrado en la muerte y su responsabilidad individual en ella. Con respecto al punto d), es decir, al determinar el modo en que se produjo la muerte, se aclara que en la investigación se debe distinguir entre: la muerte por causas naturales, la muerte por accidente, el suicidio y el homicidio (PM 2016: párr. 25). Asimismo, respecto de cada uno de esos tópicos, el PM reglamenta, en forma exhaustiva, qué pruebas concretas deben producirse para alcanzar esos resultados. Esas medidas explican, por ej., cómo debe realizarse una autopsia y cómo deben documentarse los resultados, así como también como deben trabajar los investigadores en la escena del crimen. En cambio, si lo que se investiga es un hecho de tortura, el Protocolo dispone que se debe tratar de: 1) obtener las declaraciones de las víctimas de la presunta tortura; 2) recuperar y preservar las pruebas, incluidas médicas pero también otros tipos de pruebas, como recolección de fibras, etc.; 3) identificar a posibles testigos y obtener sus declaraciones; 4) realizar informes médicos y psicológicos sobre la presunta víctima de tortura y; 5) recolectar toda otra prueba que coadyuve a determinar cómo, cuándo y dónde se han producido los presuntos hechos de tortura, así como cualquier tipo de pauta o práctica que pueda haber dado lugar a la tortura (PE 2001:párr. 77).

En los que respecta a los estándares propios de la investigación forense, los Protocolos regulan qué pruebas son relevantes con relación al fin de esclarecer los hechos y cómo deben ser producidas. O, en otros términos, determinan qué pericias deben realizarse y cómo deben documentarse los resultados obtenidos.

En resumidas cuentas, ambos documentos establecen un conjunto de normas de común desempeño, con miras a resguardar y asegurar la prueba para que pueda ser presentada en los procesos penales internos (cfr. PM 2016: párr. 8 y PE 2001:1-2, 92; 102). Asimismo, se caracterizan principalmente: 1) por su carácter forense, fundamentalmente, en virtud de su contenido y porque pertenecen a la serie de «entrenamiento profesional» del Alto Comisionado de los Derechos Humanos y; 2) porque pretenden regular los estándares mínimos de investigación para la adecuada documentación de la prueba de hechos de tortura y de ejecuciones ilegales. Por ello, sus normas se encuentran dirigidas a establecer las condiciones del estado del arte en la materia. Concretamente, pretende instituir cómo debe llevarse la investigación (qué órgano debe llevarla adelante y con qué facultades), y también qué hechos deben investigarse (*v.gr.*, se establecen los hechos que configuran formas típicas de tortura), qué prueba/s resultarían relevante/s, cómo deberían producirse y documentarse sus resultados para luego poder ser utilizados en el ámbito de los procesos judiciales internos.

## 2.2. Estatus normativo de los protocolos según el derecho internacional

Como se señaló al comienzo de este apartado, las normas de los protocolos no son consideradas *hard law*, es decir, no producen obligaciones jurídicas exigibles de manera inmediata. Esto implica que sus disposiciones no son, en principio, obligatorias para los Estados y, por ende, en caso de incumplimiento no podrían seguirse consecuencias jurídicas.

Por ello, una de las falencias que tiene la postura adoptada por la Corte IDH respecto del deber de aplicar las normas contenidas en los protocolos, es que esas directrices no resultan obligatorias para los Estados y, por ello, tienen el estatus de *soft law* (derecho blando).

Concretamente, las normas de *soft law* no tienen origen en una de las fuentes del derecho internacional. Siguiendo en este punto a GUASTINI, existen dos tipos de fuentes de las normas de derecho internacional: 1) la convención (adopción de tratados celebrados entre Estados y que vinculan a quienes los han suscripto) y; 2) la costumbre internacional (que vincula a la generalidad de Estados) (2016: 661). Las normas de *soft law* no se subsumen en ninguna de esas dos estructuras: no tienen origen en un tratado internacional ni tampoco evidencian una costumbre internacional<sup>38</sup>. Por ello, en el sistema internacional no existe acuerdo sobre qué es el *soft law* ni sobre el alcance de su fuerza normativa. Algún autor, incluso, ha sugerido que existen tantas definiciones del *soft law* como escritores sobre el tema (GOLD, 1996).

Ahora bien, sobre lo que sí hay consenso es que cuando este tipo de disposiciones se emite es porque no hay acuerdo en la comunidad internacional sobre que regulación deba ser vinculante para los Estados (GUZMÁN y MAYER, 2010: 188)<sup>39</sup>. O dicho de otro modo, solo hay acuerdo en que las normas de *soft law* carecen de obligatoriedad y, por ello, resulta difícil sostener que resulten vinculantes o de aplicación obligatoria en los ámbitos internos.

La posición dominante en derecho internacional entiende que, dado que el *soft law* carece de obligatoriedad, las normas que reciben esa calificación serían «normas morales»<sup>40</sup>. Ese carácter, además, sería congruente con la explicación de los efectos de su incumplimiento, dado que de ello no debería derivarse, aunque de hecho su-

---

<sup>38</sup> Existen algunos autores que propician que las normas de *soft law* dan cuenta de una costumbre, lo cierto es que, si necesitan especificarse las conductas en documentos que las contengan van a ser en el futuro costumbre internacional. Así, eventualmente, este tipo de «normas» puede ser un presupuesto para la conformación de una práctica, pero no dan cuenta de una práctica actualmente existente.

<sup>39</sup> Aunque, no es menos cierto que tampoco hay una explicación satisfactoria sobre cómo deben interactuar las normas pertenecientes al sistema internacional con las normas pertenecientes a los sistemas normativos de los Estados. Algunos de los problemas que se generan pueden verse en RODRÍGUEZ y VICENTE, 2009 y; en RODRÍGUEZ y ORONESU, 2017.

<sup>40</sup> Véase la reconstrucción de las teorías sobre el *soft law* descritas por COLMEGNA (2012: 30-31).

cede, la declaración de responsabilidad internacional del Estado incumplidor. Esta posición tiene en contra la opinión de HART.

El mentado autor demostró que las reglas del derecho internacional, a menudo, son totalmente indiferentes desde el punto de vista moral. Así, puede existir una regla porque es conveniente o necesario tener una pauta clara y fija sobre el punto en cuestión, pero no porque se atribuya ninguna importancia moral a la norma particular. Bien puede tratarse de un conjunto importante reglas posibles, cada una de las cuales habría estado igualmente bien. De este modo, concluye el autor que las reglas jurídicas contienen muchos detalles específicos, y hacen distinciones arbitrarias, que serían ininteligibles como elementos de reglas o normas morales (HART, 1998: 282).

Ejemplo de ello son las reglas contenidas en el protocolo de Minnesota. Por ejemplo, la norma que determina cómo debe realizarse una necropsia, si bien esa previsión es conveniente y necesaria —porque sirve para facilitar la prueba de un hecho (*i.e.* como ocurrió la muerte de la víctima)— no podría calificarse como moralmente comprometida. Por ello, la asimilación de las normas del derecho internacional a normas morales, no sería una correcta explicación del contenido de esas normas.

También, se sostiene que el *soft law* constituye una nueva fuente del derecho internacional y que, de reputarse obligatorias para los Estados miembros por su contenido, deben reconducirse a algunas de las «fuentes» reconocidas del Derecho internacional, principalmente los tratados o la costumbre internacional (ASAMOAH, 1996: 70 y SALMON, 1999: 18). Así, las normas de *soft law* resultarían normas ‘complementarias’ de las normas obligatorias (según, GIOVALI citado por LICHTENSTEIN, 2001: 1440/1).

Sin embargo, esa posición no explica qué estatus normativo propio tiene el *soft law*. Por ende, su obligatoriedad sería dependiente —y, por tanto, condicionada— a la existencia de una norma obligatoria de la que la norma de *soft law* sea accesoria. Esa posición presenta el inconveniente de que debería probarse que toda norma *soft law* podría reconducirse como una instancia de una norma de *hard law*. A la vez, no habla a favor de la obligatoriedad del *soft law*, sino de que un intérprete le asigne ese carácter al momento de aplicarla. También, deja sin explicar cómo un órgano que no tiene facultades para emitir normas vinculantes termina creando normas vinculantes.

Otra posición es la del ya citado prof. GOLD, quién entiende que los documentos catalogados como *soft law* serían preferencias expresadas por un organismo internacional sobre un determinado comportamiento que deberían adoptar los Estados, pero de su cumplimiento no se derivan consecuencias jurídicas, como sería la declaración de la responsabilidad internacional del Estado incumplidor (GOLD, 1996: 301). En otros términos, se tratarían de normas no vinculantes de organizaciones internacionales que despliegan una actividad normativa indirecta o exhortativa por cuanto proponen una conducta no obligatoria, pero deseable.

Esta posición presenta la ventaja de que puede explicarse sobre la base de las normas de competencia de los órganos del sistema internacional que emiten disposi-

ciones de *soft law*. Por ejemplo, el PM fue emitido por el Alto Comisionado para los Derechos Humanos (en adelante, por sus siglas, 'ACDH'). El ACDH es un órgano que es auxiliar a las Naciones Unidas creado por la Asamblea General de la Naciones Unidas a través de la Res. A/RES/48/141 (de fecha 7 de enero de 1994) y ratificada su creación en la Conferencia Mundial de Derechos Humanos de Viena (1993). Su ámbito de competencia se encuentra prevista en el ap. 4 de la resolución A/RES/141 y, en lo que aquí interesa, no tiene prevista específicamente la emisión de normas vinculantes para los Estados, si bien, es cierto que en la Conferencia se recomendó que «se haga un esfuerzo concertado para alentar y facilitar la ratificación de los tratados y protocolos de derechos humanos adoptados en el marco del sistema de las Naciones Unidas, y la adhesión a ellos, o la sucesión en los mismos, con el propósito de *conseguir su aceptación universal*» (A/CONF.157/23, p. 15).

Sin perjuicio de que aquí no se efectuará una reconstrucción de la competencia de todos los órganos del sistema internacional que emiten este tipo de normas puede asegurarse que, en general, el órgano que emite normas de *soft law* no tiene competencia para dictar normas vinculantes para los Estados miembros de ese sistema. O, en otros términos, las normas jurídicas que le confieren potestades no incluyen la potestad de dictar normas obligatorias.

En suma, de las formas de reconstruir el estatus normativo del *soft law* la teoría que tiene mayor poder explicativo es la que considera que esas normas no pueden ser obligatorias y, pese a estar escritas en términos exhortativos, expresan preferencias sobre cursos de acción. Dicho de otro modo, una exhortación da cuenta de las preferencias de quien la emite y su sentido normativo es el de crear pautas para guiar la acción de aquellos a quienes está dirigida (i.e. cumplen una función directiva).

### 2.3. Los protocolos contienen «reglas técnicas»

Entiendo que hay buenas razones para sostener que las normas contenidas en los Protocolos tienen la pretensión de ser normas técnicas; es decir, que este fue el objetivo principal que tuvo en miras el organismo que emitió la regulación de los protocolos.

Básicamente, como se mostró muy brevemente en el punto 2.1. del presente trabajo, ambos conjuntos de normas regulan diversos aspectos forenses. Es decir, regulan qué pruebas exige la investigación de hechos de tortura u homicidios y cómo deben producirse, i.e. qué profesionales deben hacer y qué estándares deben aplicar estos para la producción de esas pruebas. Esas reglas, en tanto estándares científicos, son los que la Corte IDH entiende que los Estados deben aplicar en sus investigaciones internas para determinar la responsabilidad penal de quienes sean indicados como autores de los hechos.

En este sentido, resulta pertinente recordar que la finalidad de las normas fue contribuir a generar estándares de procedimiento frente al desconocimiento, la ausencia

de reglas o criterios uniformes aplicables en los distintos países del Sistema de Naciones Unidas. Por ello, esos documentos fueron redactados, por reconocidos expertos de diversas áreas forenses<sup>41</sup> y pretenden concretar el ‘estado del arte’ en la materia.

Es útil aclarar que el concepto de *lex artis* se utiliza, en general, en el ámbito jurídico con dos alcances. En un primer sentido, se quiere hacer referencia a las reglas técnicas de actuación de cualquier profesión. Esa *lex artis* suele aplicarse a aquellas profesiones que utilizan técnicas especializadas y que tienen resultados constatables por los sentidos (empíricos), como por ejemplo la actividad médica. Además, también se tiene en cuenta la evolución de las técnicas, por lo que un profesional para cumplir *la lex artis*, debe acomodar su actuación a los conocimientos científicos y técnicos aceptados al tiempo de llevar a cabo su función (ROMERO FLORES, 2001: 272 con cita de RUIZ VADILLO, 1994: 72). Por ello su función principal es determinar si una acción ha sido correctamente ejecutada con relación a resultado. Sin embargo, esas reglas no suelen ni estar escritas ni sistematizadas. En segundo lugar, se utiliza la noción de *lex artis* como estándar o arquetipo de conducta de un agente que sirve para enjuiciar aquellas acciones u omisiones profesionales en orden a determinar su responsabilidad (DE LAS HERAS GARCÍA, 2005: 19).

Los Protocolos se inscriben en el primero de los sentidos detallados, al hacer explícitos el conjunto de elementos necesarios y suficientes para cumplir con un estándar de debida diligencia en la investigación de las violaciones al derecho a la vida y la integridad física. Por ello entiendo que las normas contenidas en los documentos pueden entenderse en el sentido de que establecen el conjunto sistematizado de reglas técnicas aplicables a las investigaciones sobre ejecuciones ilegales y tortura.

Las normas técnicas son aquellas reglas que ‘deben’ cumplirse si se quiere alcanzar un fin determinado. Según la tradicional definición de VON WRIGHT una regla técnica es la que establece que «debe hacerse A si quiere obtener el resultado B» (1970: 29-30). No nos obligan a asumir determinados fines, pero – de alguna manera – nos imponen los medios para conseguir tales fines (GONZALEZ LAGIER, 1993: 474). De allí que la voluntariedad es un elemento fundamental de este tipo de reglas: las reglas técnicas no establecen deberes en sentido estricto, no son prescripciones, puesto que la realización de la acción A depende de la *voluntad* del sujeto de obtener el resultado B (FERRER BELTRÁN, 2019: 61/2).

En la noción VON WRIGHT una regla técnica es una condición necesaria para obtener un determinado resultado. De esto se sigue que si A no es necesaria para obtener B no podría afirmarse que exista un «deber» de realizar A. Lo que implicaría que el medio tiene que ser efectivamente necesario para alcanzar el fin. O en otros términos, dado que las reglas técnicas pertenecen al ámbito del razonamiento práctico y sugieren un curso de acción indicando un deber técnico (GONZÁLEZ LAGIER, 1993:

---

<sup>41</sup> En este sentido véase p. VII del PM y pp. X-XII del PE.

488), presuponen también una proposición anankástica. Es decir, una descripción verdadera de que con cierta acción se alcanza un determinado fin.

En efecto, se sostiene que entre las reglas técnicas y las proposiciones anankásticas existe una conexión lógica: «Cuando se da la directriz ‘Si quieres hacer la cabaña habitable, tienes que instalar calefacción’ se presupone (lógicamente) que si la cabaña no es acondicionada con calefacción no llegará a ser habitable». De manera que las normas técnicas se fundan siempre en proposiciones anankásticas 42 (GONZÁLEZ LAGIER, 1993:475).

Sin embargo, la exigencia de la necesidad del medio, para que la regla sea calificada como técnica, ha sido morigerada por GONZÁLEZ LAGIER. Ese autor sostiene que para que exista una regla técnica basta con que el medio puede ser considerado una condición suficiente como una condición contribuyente. En sus términos:

*[C]iertas condiciones suficientes pueden convertirse en necesarias dependiendo del contexto; la disyunción de todas las condiciones suficientes de un fin es una condición necesaria (y suficiente) de ese fin; una condición contribuyente es todo lo que podemos hacer en un momento determinado; puede ser absurdo realizar un medio que es condición necesaria de un fin si no puede completarse el contexto causal para convertirla en necesaria y suficiente, etc. [...] Quizás podría establecerse una graduación dentro de las reglas según las razones sean más o menos compulsivas. Si el medio es condición necesaria y suficiente del fin las razones son mucho más compulsivas que si el medio es suficiente o contribuyente (1993: 493).*

Lo interesante de la posición de VON WRIGHT es que un deber técnico puede estar enlazado genuinamente a un deber normativo. Así, el citado autor señala que un deber técnico puede respaldar a una norma jurídica o a una orden, puede ser parte de un razonamiento que lleve al legislador a emitir una ley u ordenar algo. Por otro lado, dado un deber deóntico puede que consideremos que si no hacemos *x* o *y* no podremos cumplir con ese deber; entonces surge un deber técnico, una necesidad práctica, derivado de la existencia de un deber deóntico (GONZÁLEZ LAGIER, 1994: 478).

No obstante, las reglas técnicas se diferenciarían de las reglas deónticas, en que estas últimas, reemplazan al razonamiento práctico dando, en terminología de Raz, razones excluyentes para la acción. Por el contrario, las reglas técnicas muestran —con el fin de dirigir la conducta del agente— que si es cierto que se tiene el deseo de determinado fin (razón operativa), y dado que cierta acción es condición de ese fin (razón auxiliar), entonces hay una razón completa (pero no excluyente y que puede ser derrotada por razones de mayor peso) para realizar esa acción (GONZÁLEZ LAGIER, 1994: 496).

Aplicadas las premisas anteriores a las ‘normas’ contenidas en los protocolos, al tener el carácter de *soft law*, no pueden generar genuinos deberes y, por ende, solo sugieren cursos de acción para lograr un fin determinado fin (*i.e.* realizar una investi-

<sup>42</sup> Por ello también se puede distinguir entre regla técnica y proposición anankástica porque la primera cumple una función directiva, a diferencia de las proposiciones anankásticas que tienen por fin describir qué se debe hacer para alcanzar un determinado fin (véase GONZÁLEZ LAGIER, 1994).

gación eficaz). A la vez, al estar formulados en términos exhortativos, presentan una dimensión directiva. Lo que es reconocido, según se desarrolló en el punto a) del presente apartado, por las propias disposiciones de ambos documentos internacionales. Por último, si bien no dan lugar a deberes deónticos, la jurisprudencia de la Corte IDH los vincula al cumplimiento de un deber de tal carácter.

En ese esquema a partir de una norma prescriptiva, más concretamente, de una «norma de fin» —en la clasificación de VON WRIGHT realizada en *Six Essays in Philosophical Logic*—, tal como «debes realizar una investigación eficaz para aclarar la muerte de p», pueden seguirse una serie de deberes técnicos para alcanzar ese objetivo (*v.gr.*, «si quieres realizar una investigación eficaz de la muerte de p debes realizar una autopsia a su cadáver») y una proposición anankástica (*v.gr.* «para investigar eficazmente la muerte de la muerte de p es necesario realizar una autopsia»).

Respecto de si las reglas tienen una correlativa proposición anankástica que sustente su corrección de ajuste (normas  $\rightarrow$  mundo) es relevante recordar que los protocolos tienen la pretensión de ser una «recopilación» de la *lex artis* en materia forense de investigación legal de la tortura y del homicidio.

Asimismo, dado que los protocolos fueron elaborados por expertos y, además, como se señaló, sistematizan el estado del arte en materia de investigación forense, es razonable que la Corte IDH entienda que su cumplimiento debería hacer posible el fin de realizar una investigación eficaz.

Sin embargo, las normas no podrían entenderse en su conjunto como necesarias: su aplicación al caso resultará del contexto. No todas serán de aplicación sino aquella que resulten pertinentes en función de lo que se deba probar. Lo relevante es, entonces, el cumplimiento de las reglas de actuación establecidas según cuál sea el hecho investigado (*i.e.* muerte, protocolo de Minnesota; tortura, Protocolo de Estambul) y el contexto en el que se realice la investigación. De este modo, las normas establecidas en los protocolos podrían entenderse como el conjunto de condiciones necesarias y suficientes que, según las circunstancias del caso, deberían cumplirse para realizar una investigación eficaz.

La aclaración relativa a las circunstancias del caso obedece a que no siempre van a poder colectarse todas las pruebas que los protocolos describen (por ej., si hay rastros óseos y no existen tejidos blandos, deberán trabajarse con las reglas aplicables pertinentes a la situación de hecho dada la imposibilidad fáctica apuntada). Esto, además, se encuentra aclarado en los documentos internacionales.

Por último, que los protocolos regulen normas técnicas es congruente, además, con la consecuencia que la Corte IDH sigue ante la falta de cumplimiento de esas reglas: la imposibilidad de alcanzar el fin normativo de realizar una investigación eficaz. Por ello, conlleva declaración de la responsabilidad internacional del Estado y, como forma de «reparación», la reapertura del proceso con aplicación de los estándares omitidos.

En suma, entender a esas normas como reglas técnicas es una adecuada reconstrucción del contenido de los protocolos y de la práctica de la Corte IDH. Sin embargo, como intentaré mostrar seguidamente, esto no está exento de problemas dado que ni todas las normas contenidas pueden aplicarse en el ámbito interno de los Estados ni todas las normas de los protocolos dan cuenta del estándar científico actualmente aplicable (i.e. son genuinas normas técnicas).

### 3. UNA REVISIÓN CRÍTICA

Sostener como lo hace la Corte IDH que las normas de los protocolos se aplican sin más apareja, como adelanté, una serie de problemas.

Algunas normas de los protocolos están en contradicción con normas fundamentales de derecho interno o con normas de rango legal, sea porque establecen órganos de investigación prohibidos en los ámbitos internos (*v.gr.* como es el caso establecimiento de una comisión ad hoc para la investigación prevista en el Protocolo de Estambul), porque establecen disposiciones de prueba tasada, cuando la valoración de la prueba resulta una facultad exclusiva del juzgador (*v.gr.* la clasificación de las lesiones como de tortura en el Protocolo de Estambul). Otras reglas, prevén un estándar de ciencia que actualmente, se encuentra en desuso (*v.gr.*, disposiciones relativas a la toma de las declaraciones testimoniales).

Tampoco, podría afirmarse categóricamente que la mera utilización de los protocolos sea una condición necesaria para cumplir el fin de realizar una investigación eficaz puesto que, en los términos en los que los plantea la Corte IDH, realizar una investigación eficaz implica, a su vez, el castigo a los responsables. A la vez, pierde de vista que aun usando esas normas, podría no alcanzarse aquel resultado (*i.e.* castigo) por razones basadas en infracciones a normas fundamentales durante la sustanciación del proceso interno (*v.gr.* exclusiones probatorias por aplicación de la teoría del fruto del árbol prohibido). Por ello, la aplicación de esas reglas en algunos casos podrá ser una condición necesaria, mientras que en otros resulta una condición suficiente e incluso, en algunos otros una condición contribuyente (i.e. no es una condición necesaria ni suficiente, pero ayuda a la consecución del fin).

La cuestión de la aceptación de la posición de la Corte IDH es preocupante porque existe cierta tendencia en la doctrina y jurisprudencia de Latinoamérica de aceptar sin cuestionamientos lo resuelto por el organismo, aun cuando actúe exorbitando sus competencias convencionales. Incluso, se sostiene que aquella postura importaría la primacía de las normas de derecho de internacional de derechos humanos sobre las normas de derecho interno, incluidas las normas fundamentales de derecho interno<sup>43</sup>. Llegando en algunos casos a situaciones de reapertura de procesos con lesiones a principios constitucionales fundamentales de derecho interno.

---

<sup>43</sup> Incluso, la Corte Sup. de Justicia de Argentina le llegó a reconocer el carácter de obligatorias a las recomendaciones efectuadas por la Comisión de Interamericana de Derechos Humanos en su fun-

Por ello, me parece importante plantear algunas críticas a la posición de la Corte IDH, respecto al deber que tienen los Estados miembros de seguir las normas que regulan las investigaciones de ejecuciones ilegales y/o tortura acaecidas en el ámbito interno.

### 3.1. Estándar mínimo y eficacia

La Corte IDH sostiene que los protocolos concretan los estándares mínimos de investigación. Sin embargo, la asociación entre la idea de estándar mínimo y el deber de realizar una investigación eficaz que tenga un resultado específico (esto es, el castigo), resulta problemática.

Si se parte de la base que los principios admitirían una satisfacción gradual<sup>44</sup>, si pretendo hacerlo de manera «eficaz» (*i.e.* que haga en la mayor medida probable la consecución del objetivo) debería utilizarse un estándar más exigente. No podrían apelarse a estándares mínimos, sino que debería exigirse la aplicación del mejor estándar disponible.

Los mejores estándares disponibles permitirían garantizar una mayor probabilidad de que se aplique un castigo a quien haya sido acusado de una violación grave a un derecho fundamental (*v.gr.* porque minimizarían la probabilidad de errores materiales en la investigación y por ende, de la decisión judicial que deba adoptarse) y por ello, serían los más compatibles con la búsqueda de la verdad.

### 3.2. Las normas de los protocolos no pueden reemplazar a las normas internas obligatorias

Las normas de los protocolos no pueden reemplazar a las normas obligatorias internas. Como se mostró de tanto su carácter de *soft law* como la característica de ser una norma técnica, se sigue que las reglas contenidas en los protocolos no son —ni podrían resultar— «normas» obligatorias. Como se mostró de tanto su carácter

---

ción consultiva en el precedente Caso “Espósito” (Detención ilegal y muerte de W. Bulacio, resuelto el 23/12/2004). Para un análisis de los casos remito al lector, nuevamente, a ORONESU y RODRIGUEZ, 2017. Aunque, recientemente, ha habido un giro hacia la racionalidad en la causa “Fontevicchia y otros c/ República Argentina”. Allí la Corte Suprema de Justicia argentina entendió —luego de que la Corte IDH ‘dispusiera’ dejar sin efecto una sentencia firme— que el texto de la Convención no atribuye facultades a la Corte Interamericana para ordenar la revocación de sentencias nacionales, menos aún, en el caso de que estas hayan adquirido firmeza puesto que, de otro modo, se violarían disposiciones fundamentales de Derecho interno. Sobre la falta de competencia, con el correspondiente análisis de la normativa aplicable, véase RODRÍGUEZ y ORONESU, 2017.

<sup>44</sup> En este sentido la posición de ALEXI. El autor sostiene que los principios son «mandatos de optimización». Es decir, los principios ordenan que algo sea realizado en la mayor medida de lo posible, dentro de las posibilidades jurídicas y reales existentes. Por tanto, los principios son mandatos de optimización, que se caracterizan porque pueden cumplirse en grado diferente y que la medida debida de su cumplimiento solo depende de las posibilidades reales sino jurídicas (ALEXI, 2014: 67-68).

de *soft law* como la característica de ser una norma técnica, se sigue que las reglas contenidas en los protocolos no son —ni podrían resultar— «normas» obligatorias. Por ello, las normas de los protocolos no pueden reemplazar a las normas obligatorias internas

No obstante, la función que sí pueden brindar los protocolos, dado su carácter técnico, se encuentra limitada a entenderse como normas *complementarias* de las normas internas sobre medios de prueba. Ello básicamente, frente a la ausencia generalizada de estándares forenses aplicables.

En cambio, si lo que se pretendiera la Corte IDH es la aplicación de todas sus reglas esto sería problemático porque, en algunos casos, son contradictorias con normas fundamentales de derecho interno. Y en otros casos, algunas reglas, más allá de su formulación canónica, no dan cuenta de genuinas reglas técnicas (por ej., las disposiciones relativas a las declaraciones testimoniales) por no estar vinculadas a un estándar de ciencia o resultar, incluso, contra epistémicas.

Seguidamente se mostrarán alguna de esas regulaciones problemáticas.

a) *Comisiones de Indagación «Ad Hoc»*. Una de las reglas de los protocolos que puede dificultar la aproximación a la verdad en el proceso penal es la que sugiere la investigación a través de una comisión de indagación *ad hoc*.

El PE sugiere que, cuando la imparcialidad del investigador se encuentre afectada, entonces la investigación debería ser llevada por una comisión de indagación. Esto es también sugerido por el PM, aunque en su caso no se establece ningún criterio de preferencia respecto de un tipo de proceso sobre otro, esto es de un procedimiento judicial sobre una comisión *ad hoc*<sup>45</sup>. Esa sugerencia me parece problemática dado que entra en conflicto con garantías constitucionales básicas de un Estado constitucional de Derecho.

En ese sentido, no puede perderse de vista que los órganos de investigación penal no solo se encuentran creados legalmente para garantizar la imparcialidad, sino que incluso la jurisdiccionalidad es entendida como una garantía (FERRAJOLI, 1995). Por ende, ya es difícil sostener que este tipo de comisiones pueda producir prueba *per se*, porque en definitiva habrá hecho gran parte del trabajo correspondiente a la investigación penal preparatoria fuera del lugar dónde debe producirse: el proceso penal.

Un ejemplo servirá para ilustrar la cuestión. Si se tratara de prueba definitiva e irreproducible (anticipo probatorio) su producción en el ámbito de la Comisión *Ad Hoc* puede imposibilitar su admisión en el proceso penal que deba llevarse, luego, en el ámbito interno (en este sentido, véase párr. 102 del PE). Básicamente, las pruebas

---

<sup>45</sup> Si bien en su redacción de 1991 tenía idéntica estructura al PE y básicamente, recomendaba en los mismos términos que el PE que, si no el Estado no podía garantizar que iba a realizar una investigación imparcial e independiente de los hechos, debía utilizarse el mecanismo de la comisión de indagación.

que se producirían en ese marco se realizarían *sin contradicción*<sup>46</sup> y *sin control* de ninguna autoridad que no sea la propia de la Comisión. En tal sentido ese tipo de pruebas tiene que realizarse al menos con conocimiento de los imputados de su producción y ofreciendo su control, es decir, debe, por los menos plantearse la posibilidad de contradicción<sup>47</sup>. En este sentido la utilización de una Comisión ad hoc podría impedir una correcta investigación de los hechos en el ámbito interno.

Ahora bien, si se entiende que lo que se regula no es la prueba sino actos de investigación la situación tampoco sería distinta. Como con acierto señalaba MIRANDA ESTRAMPES, los actos de investigación *no tienen valor probatorio* por lo que no pueden ser utilizados para formar la convicción fáctica del juez en el juicio, salvo aquellos supuestos excepcionales previstos legalmente (MIRANDA ESTRAMPES, 2004:11).

Las cuestiones señaladas, en mi criterio, van a dificultar, cuando no inutilizar, la incorporación de ciertas pruebas producidas en ese ámbito al proceso penal que se lleve adelante en el ámbito interno, dado que su incorporación podría vulnerar no solo reglas internas obligatorias (*v.gr.*, regulación de la prueba testimonial) sino, también, vulneraría normas fundamentales de derecho interno, dando lugar a la aplicación de la doctrina de las exclusiones probatorias.

#### b) *La regulación de las pruebas testimoniales y su valoración*

Otra regulación problemática son las previsiones relativas a la producción y valoración de la prueba testimonial.

Concretamente, el PE prevé la posibilidad de utilizar declaraciones testimoniales anónimas, así como también contienen reglas de prueba tasada, en tanto, regulan qué valor debe asignarse a la declaración del testigo de referencia. Por último, también resultan problemáticas las reglas relativas a la declaración del testigo – víctima.

Con respecto a la posibilidad de testigos anónimos y de referencia, en general, no son admitidos como medios de prueba ya que permiten el ingreso al proceso de in-

<sup>46</sup> Además, la contradicción es una de las pocas garantías a la que se le asigna valor epistemológico (FERRAJOLI, 1995; FERRER, 2010; VÁZQUEZ, 2015; UBERTIS, 2017). Esto es remarcado por, entre otros, VÁZQUEZ, quien sostiene que si partes critican mutuamente sus elementos de juicios, eso puede servir a que el juzgador pueda valorar mejor la calidad de las pruebas, aunque ello presupone un escenario en el que se discuta seriamente la calidad de estas. Es decir, solo si las partes realizan una adecuada práctica de las pruebas en contradicción entonces este (principio) podrá desplegar sus funciones epistémicas (2015: 32).

<sup>47</sup> La prueba de cargo en el proceso penal requiere ser producida dentro de este y estar sometida a contradicción, siendo este un principio fundamental de derecho interno, cuya infracción acarrearía su nulidad (cfr. FERRAJOLI, 1995; UBERTIS, 2017; ANDRÉS IBÁÑEZ, 2009, TEDH, sección V, sentencias: de 10 de Feb. De 2011; *3ACZ s.r.o. v. Republica Checa*; §39; de 25 de marzo de 1998; *BELZIUK V. POLONIA*; §37; y de 19 de diciembre, *KAMASINSKI V. AUSTRIA*, §102, entre otros). Sobre el punto se volverá *infra*.

formación de muy baja calidad, y por ello, como señala DAMÂSKA (2015) en muchos sistemas procesales, son excluidas por razones intrínsecas a la búsqueda de la verdad.

Más concretamente, los testigos de oídas no han percibido aquello sobre lo que declaran, sino sobre lo que le han contado. Es un testigo que no estuvo presente en el lugar del hecho, por lo que no puede describir su contexto y tan solo puede imaginar una perspectiva diferente a la que le relataron. Por tanto, la falsedad de su testimonio es muy difícil de descubrir (NIEVA FENOLL, 2010: 279). A la vez, su propia credibilidad puede ser puesta en duda. Como señala Nieva Fenoll, lo único que es posible valorar, en alguna medida, es que ese testigo no está mintiendo. Pero hasta eso es complejo, teniendo en cuenta que su declaración no se puede contextualizar: no se puede valorar realmente ni la coherencia del relato, ni tiene sentido su contextualización dado que no presenció el hecho (2010: 279)<sup>48</sup>.

A lo anterior cabe agregar que la inclusión de la regulación de estos testigos se da en el marco de un contexto de *insuficiencia* probatoria. Pero, esa orfandad no podría ser suplida simplemente mediante la incorporación de un testimonio de referencia, dado que por más verosímil que resulte no es probable en el sentido de que no se puede practicar prueba sobre el mismo. En esa situación, la regla de la presunción inocencia determina que ese testimonio no deba ser tenido en cuenta. Como con acierto señala NIEVA FENOLL, no se puede condenar a alguien basándose únicamente en la palabra de una persona que ni siquiera presenció los hechos. Además, de que, si le contaron los hechos, es posible que el testigo fuera amigo o familiar de la víctima, por lo que a los inconvenientes de su testimonio cabe añadir, por ejemplo, ahora su vinculación subjetiva lo que le resta objetividad (2010: 281). Por lo demás, aun contando con la declaración de la víctima, la prueba de oídas no sería independiente de esa declaración, por lo que no puede aumentar el valor inductivo de la única prueba disponible (o lo aumentaría de manera insignificante). Así, el testigo de oídas solo puede servir en el caso de que se cuenten con otras pruebas independientes y como elemento corroborador de estas. Pero su valor probatorio debería ser casi nulo, dado los graves inconvenientes señalados.

La regulación de la posibilidad de testimonios anónimos tampoco está exenta de problemas. En primer lugar, por la forma en que su declaración se encuentra regulada en el PE (2001: párr. 112). Básicamente, el documento impide su comparecencia a un eventual juicio para no poner en riesgo su seguridad. Esto impediría el ejercicio del derecho de defensa al no permitir la contradicción de esos testimonios por parte de quien es indicado como autor del hecho y por ende, impide la confrontación como garantía de confiabilidad de su testimonio<sup>49</sup>.

---

<sup>48</sup> En algunos casos excepcionales el testigo puede ser admitido dentro de un sistema de libertad probatoria y libre valoración, sin embargo, esto no disminuye los problemas epistemológicos que su inclusión genera relativos, como se señaló, a la imposibilidad de determinar si el testimonio es fiable.

<sup>49</sup> Sobre el punto, remito al lector al trabajo de ROVATTI, 2020.

Con respecto a la valoración de la declaración de la víctima, el PE establece que la ausencia de lesiones no puede entenderse en ningún caso como la inexistencia de un hecho de tortura así como también de si una víctima dice que fue torturada debe tenerse por cierto su testimonio. Esto implicaría que, si una víctima manifiesta que una persona fue su torturadora, deben tenerse por ciertos sus dichos. Esas disposiciones pierden de vista que la víctima es un interesado en el resultado del proceso y por ello, su declaración debe ser escuchada con cautela, dado que su credibilidad puede verse afectada en función de su interés (NIEVA FENOLL, 2010: 237-239).

El problema se agrava más aún si la declaración de la víctima es, además, la única prueba de cargo disponible<sup>50</sup>, porque los hechos de tortura se conocen mucho tiempo después de producidos, cuando las lesiones pudieron haber desaparecido o confundirse con otras cuya causa no es el hecho de tortura.

Ahora bien, en un contexto probatorio en que exista un testigo interesado y que no existan otros elementos de corroboración independientes de su declaración, plantea un problema de suficiencia probatoria que aún no ha sido resuelto en el ámbito interno de los Estados y que la Corte IDH resuelve mediante el uso de normas no obligatorias.

Asimismo, la utilización de una sola prueba de cargo, a la que se le asigna un determinado valor, y eventualmente de pruebas dependientes de estas para aumentar su valor inductivo, tales como los usuales informes periciales de expertos en psicólogos del testimonio y admisión de testigos de referencia, parece que no ser compatible con el sistema de libre apreciación de la prueba. A la vez que no parece muy eficiente para esclarecer los hechos, aunque sí puede serlo para imponer castigos.

c) *Reglas que presentan problemas relacionados con las condiciones de producción de la prueba pericial en los ámbitos internos*

Los protocolos establecen que las conclusiones de los informes periciales, sean del examen físico o las del examen psicológico, si bien responden a la idea de gradualidad, deben expresarse mediante conclusiones *individualizadoras*. Esto implica que su grado máximo de probabilidad es compatible con la certeza absoluta de que alguien fue objeto de tortura o de que alguien fue objeto de homicidio.

El problema es que la exigencia de tales resultados a los informes periciales pierde de vista que la ciencia no puede predicar certezas, y por ello, tampoco puede hacerlo un informe pericial. Asimismo, de la existencia de un indicio o de un conjunto de ellos no se puede seguir con certeza la ocurrencia de un determinado hecho<sup>51</sup>.

<sup>50</sup> Sobre esta cuestión, véase el excelente análisis de SANCINETTI, 2013, con amplias referencias a la jurisprudencia alemana.

<sup>51</sup> Señala GASCÓN que los expertos de estas disciplinas comunican con frecuencia sus hallazgos usando expresiones individualizadoras. De este modo las afirmaciones de que un vestigio hallado en

Como advierte GASCÓN, esa idea se basa en la tesis no comprobada de que es posible vincular un vestigio de origen desconocido a una única fuente, lo que responde a una equivocada intuición probabilística que iguala infrecuencia con unicidad. Y la unicidad no ha sido establecida en ninguna de las disciplinas forenses tradicionales (GASCÓN, 2016).

Otra de las consecuencias que sigue de la exigencia de que el informe pericial exprese conclusiones individualizadoras, es que si se siguiesen de modo estricto esas reglas, se convertiría a la prueba pericial es una especie de prueba legal tasada y por ende, no le permitiría al juez apartarse de esas conclusiones. Así, el perito se transforma en un verdadero juez ya que, además, su informe determina si un imputado cometió o no un hecho de tortura o una ejecución ilegal, lo que tornaría en innecesarias a las reglas sobre valoración de la prueba. En otros términos: «si los datos (obtenidos mediante estas pruebas) dijese directamente al juez lo que necesita saber... y el grado de probabilidad con lo que lo afirman fuese tan elevado que pudiera considerarse infalible, no tendría sentido deferir al juez la valoración de dicho enunciado, pues esta puede ser adelantada ya por el perito de manera categórica [...] lo normal, sin embargo, es que sea necesario *realizar inferencias* a partir de esos datos» (VÁZQUEZ, 2014: 67).

Por otro lado, tampoco hay que olvidar que se requieren además otras constataciones adicionales para poder afirmar la responsabilidad por el hecho.

De este modo, si bien lo relevante de los protocolos son las normas técnicas referidas a la estandarización de técnicas aplicables e informes forenses, debe realizarse una salvedad: no pueden obtenerse certezas absolutas acerca de si se cometió un hecho de tortura o una ejecución ilegal a través de un dictamen pericial. A lo sumo, el informe

---

la escena del crimen que presenta una coincidencia relevante con una fuente específica (porque ambos comparten un número suficiente de características) proviene de esa fuente. Los peritos expresan con frecuencia los resultados de las pruebas justamente en los términos en que el juez debe pronunciarse, y de este modo las pruebas aparecen como una poderosa herramienta que «cierra» la decisión judicial. Esta manera de interpretar y comunicar los resultados de las pruebas expresa lo que en el ámbito de la ciencia forense se denomina el paradigma de la individualización, que se asienta sobre la asunción teórica de que la unicidad existe. La unicidad significa que un vestigio o marca desconocida *mx* solo puede provenir de una fuente *fl*, con exclusión de todas las demás, y se refleja en la creencia de que podemos llegar a identificar plenamente un individuo o un objeto a partir de vestigios. La individualización, en palabras simples, reposa en la asunción teórica de que «la naturaleza nunca se repite». Mediante un análisis científico forense —es la idea rectora de este paradigma— se puede individualizar la marca o vestigio analizado, o sea se puede determinar con adecuada fiabilidad si esa marca o vestigio procede de una fuente específica (un individuo o un objeto), con exclusión de cualquier otra en el mundo. La mayoría de las técnicas forenses tradicionales de identificación criminal declaran en términos de individualización, asumiendo así no solo que la «unicidad» existe sino que además la individualización puede ser conseguida. Las cosas, sin embargo, no son así, pues los conocimientos científicos disponibles no permiten sostener una tesis tan fuerte. De hecho en los últimos años el paradigma de la individualización ha recibido fuertes críticas por parte de la comunidad científica. (GASCÓN, 2016: 352/353 y extensas sus citas).

pericial puede dar cuenta de que un tipo de lesión es *probablemente* compatible (en un  $x$  grado) con cierto tipo de hecho (tortura).

#### 4. A MODO DE CONCLUSIÓN

La posición de la Corte IDH de entender que las normas de los protocolos son vinculantes para los Estados miembros del sistema latinoamericano de protección de los derechos humanos porque instancian el estándar mínimo de diligencia en la investigación de graves violaciones a los derechos humanos no resulta de problemas.

En definitiva, la Corte IDH transforma en obligatorias normas, mediante la interpretación de un derecho implícito, que en el derecho internacional no pueden tener ese carácter y que no pueden tenerlo conforme a las propias normas que rigen su creación. Por otro lado, aunque pueda compartirse que las normas tienen que aplicarse (porque existe un deber técnico) no todas las normas podrían resultar aplicables.

Entiendo que la preocupación de la Corte IDH no debería estar encaminada a que se aplique un cuerpo determinado de reglas, sino lo que debería controlar es que las normas utilizadas por los Estados en su ámbito interno den cuenta del estándar de ciencia aplicable en el contexto probatorio que corresponda según el caso. Esto serviría para determinar si (concretamente) las acciones utilizadas por una Estado pueden alcanzar el objetivo de realizar una investigación eficaz.

En mi criterio, la Corte IDH impone la aplicación de normas no obligatorias sin acreditar que esas normas son el medio más idóneo para cumplir con ese deber (déontico). Esto pierde de vista, que existen, varios documentos no obligatorios que podrían ser utilizados en las investigaciones y que la Corte IDH ni siquiera analiza.

Asimismo, ordenar la reapertura de las actuaciones —dejando de lado las garantías constitucionales que pudieran lesionarse en el ámbito interno— y en consecuencia, mandar a aplicar los estándares omitidos podría arrojar el mismo resultado (i.e. la impunidad), si, por ejemplo, se trata de esclarecer hechos que tuvieron lugar hace mucho tiempo y de los que ya no queden más pruebas que colectar o la única prueba disponible es la declaración de la víctima.

En suma, he intentado brindar algunos argumentos para poder repensar lo que implica aceptar sin más la conclusión de la Corte IDH sobre que existe la obligación de aplicar los protocolos de Estambul y de Minnesota en los ámbitos internos de los Estados.

#### BIBLIOGRAFÍA

ACCATINO, D. (2016). *Prueba, verdad y justicia de transición. El caso de Cecilia Magni y Raúl Pellegrin* (Corte Suprema). *Revista de Derecho* (Valdivia), 337-361.

- (2019 (en prensa). *¿Por qué no a la impunidad? Una mirada desde las teorías comunicativas*. Política Criminal, 1-19.
- ACCATINO, D., & COLLINS, C. (2016). *Truth, evidence, truth: The Deployment of Testimony, Archives and Technical Data in Domestic Human Rights Trials*. Journal of Humans Righth Practice, 8, 81-100.
- ALCHOURRÓN, C., & BULYGIN, E. (1991). Los límites de la lógica y el razonamiento jurídico. En C. Alchourrón, & E. Bulygin, *Análisis Lógico & Derecho* (págs. 303-328). Madrid: Centro de Estudios Constitucionales.
- ANDERSON, T., SCHUM, D., & TWINING, W. (2015). *Análisis de la prueba*. Madrid: Marcial Pons.
- ANDRÉS IBAÑEZ, P. (1992). *Acerca de la motivación de los hechos en la sentencia penal*. DOXA, 257-299.
- (1994). *De nuevo sobre la motivación de los hechos. Respuesta a Manuel Atienza*. Jueces para la democracia, 87-92.
- (2009). *La prueba y convicción judicial en el proceso penal*. Buenos Aires: Hammurabi.
- ASAMOAH, O.Y. 1966. The Legal Significance of the Declarations of the General Assembly of the United Nations, The Hague, Nijhoff.
- ATIENZA, M., & RUIZ MANERO, J. (1991). *Sobre principios y reglas*. Doxa, 101-120.
- BAYÓN, C. (1991). *Razones y reglas: sobre el concepto de “razón excluyente” de Joseph Raz*. DOXA, 25-65.
- BONORINO, P. R. (1999). Lógica y prueba judicial. *Anuario de Filosofía del Derecho*, 15-24.
- BRAHM, E. (2007). *Uncovering the Truth: Examining Truth Commission Success and Impact*. International Studies Perspectives. Oxford University Press, pp. 16-35.
- CEJIL (2010). *Debida diligencia en la investigación de graves violaciones a los derechos humanos*. Buenos Aires: CEJIL.
- CORTE IDH. (2017). *Justicia Transicional - cuadernillo N° 15*. San José: OEA.
- DAMÁSKA, M. R. (2015). *El derecho probatorio a la deriva*. Madrid: Marcial Pons.
- DE LAS HERAS GARCÍA, M. A. (2005). Lex artis, onus probandi y responsabilidad médica, en Revista Jurídica de la Región de Murcia, n° 36, pp. 13-68.
- DEL TORO HUERTA, M. (2006). *El fenómeno del Soft Law y las nuevas perspectivas del derecho internacional*. Anuario Mexicano de Derecho Internacional, 513-549.
- DUQUE, M. (2010). *El papel de la medicina forense en el contexto del trabajo humanitario*. Cuadernos de medicina forense, 37-42.
- FERRAJOLI, L. (1995). *Derecho y Razón. Teoría del Garantismo Penal*. Madrid: Trotta.
- (2013). El juicio Penal. En L. Ferrajoli, *Escritos de Derecho Penal* (págs. 446-467). Buenos Aires: Hammurabi.
- FERRER BELTRÁN, J. (2002). *Derecho a la prueba y racionalidad de las decisiones judiciales en Jueces para la democracia*, N° 47, 2003, págs. 27-34.
- (2005). *Prueba y verdad en derecho*. Madrid: Marcial Pons.
- (2007). *La valoración racional de la prueba*. Madrid: Marcial Pons.
- (2013). La prueba es libertad, pero no tanto: Una teoría de la prueba cuasibenthamiana. En C. Vázquez, *Estándares de prueba y prueba científica. Ensayos de epistemología jurídica* (págs. 21-39). Madrid: Marcial Pons.
- GALAIN PALERMO, P. (2013). *Relaciones entre el “derecho a la verdad” y el proceso penal: análisis de la jurisprudencia de la Corte Interamericana de Derechos Humanos*. Revista Penal México, 65-85.
- GASCÓN ABELLÁN, M. (2010). Los hechos en el derecho. Bases argumentales de la prueba. Barcelona: Marcial Pons.
- (20102). *La prueba científica: mitos y paradigmas*. Revista de Filosofía Política y Jurídica de la Universidad de Granada, 44-60.
- GOLD, J. (1996). *Interpretation: the IMF and international law*. Springer Netherlands.
- GONZÁLEZ LAGIER, D. (1993). *Algunas cuestiones sobre las reglas técnicas*. DOXA, 473-496.
- (2003). *Hechos y argumentos (racionalidad epistemológica y prueba de los hechos en el proceso penal) (II)*. Jueces para la democracia, 35-50.
- (2005). *Questio Facti. Ensayos sobre prueba, causalidad y acción*. Lima: Pallestra.
- GUZMÁN, A., & MAYER, T. (2010). *International soft law*. Journal of Legal Analysis, 171-225.
- HART, H. (1961), *El concepto del derecho*, Buenos Aires, Abeledo-Perrot, 1998.

- HAYNER, P. (1994). *Fifteen truth commissions - 1974- 1994: a comparative study*. Human Right Quarterly, 558-655.
- (2001). *Unspeakable Truths: Confronting State Terror and Atrocity*. London: Routledge .
- IGARTÍA, J. (2007). *Prueba científica y decisión judicial (Unas anotaciones propedéuticas)*. La Ley , 6812.
- KELSEN, H. (1982). *Teoría Pura del Derecho* . Ciudad de México DF: UNAM.
- KELSEN, H., Bulygin, E., & Walter, R. (2005). *Validez y eficacia en el derecho* . Buenos Aires: Astrea.
- LAUDAN, L. (2011). *El estándar de prueba y las garantías en el proceso penal*. Buenos Aires: Hammurabi.
- (2013). *Verdad, error y proceso penal. Un ensayo sobre epistemología jurídica*. Barcelona: Marcial Pons.
- LICHTENSTEIN, C. C. (2001). *Hard Law v. Soft Law: Unnecessary Dichotomy*. In Int'l L. (Vol. 35, p. ).
- MIRANDA ESTRAMPES, M. (2004). *La prueba y su valoración en los procesos penales acusatorios latinoamericanos*, disponible on line en: [www.incipp.org.pe/archivos/publicaciones/mirandaestampres.pdf](http://www.incipp.org.pe/archivos/publicaciones/mirandaestampres.pdf)
- MAZUELOS BELLIDO A. (2004). *Soft Law: ¿Mucho ruido y pocas nueces?*, Revista Electrónica de Estudios Internacionales, N° 8. Disponible en: <http://www.reei.org/>.
- NIÉVA FENOLL, J. (2010). *La valoración de la prueba*. Madrid: Marcial Pons.
- (2013). *La duda en el proceso penal*. Madrid: Marcial Pons.
- (2013). *El principio de justicia universal: una solución deficiente para la evitación de hechos repugnantes*. Revista Española de Derecho Internacional, pp. 131-149.
- ONU. (2004). *Protocolo de Estambul. Manual de Investigación y Documentación Efectiva sobre Tortura, Castigos y Tratamientos Crueles, Inhumanos o Degradantes*. New York - Geneva: Office of the United Nations High Commissioner for Human Rights.
- (2016). *The Minnesota Protocol on the investigation of potentially unlawful death*. New York - Geneva: Office of the United Nations High Commissioner for Human Rights.
- PASTOR RIDRUEJO, J. A. (2007). *La reciente jurisprudencia del Tribunal Europeo de Derechos Humanos: temas escogidos*. Obtenido de Universitat Jaume I: <http://hdl.handle.net/10234/177702>
- PASTOR, D., & MACULAN, E. (2013). *El derecho a la verdad y su realización por medio del proceso penal*. Buenos Aires: Hammurabi.
- PASTOR, Daniel. (2006). *La deriva Neopunitivista de organismos y activistas como causa del desprestigio actual de los derechos humanos*. Obtenido de Jura Gentium: <http://www.juragentium.org/topics/latina/es/pastor.htm>
- RODRIGUEZ, J., & ORONESU, C. (2018). *Razonamiento y decisión judicial. La justificación de las decisiones judiciales en las democracias constitucionales*. Buenos Aires : Ministerio de Justicia Nación Argentina. Obtenido de Academia Edu.
- (2017). *Tres paradojas del control de convencionalidad*, artículo disponible en el siguiente link: [https://www.academia.edu/23035307/Tres\\_paradojas\\_del\\_control\\_de\\_convencionalidad](https://www.academia.edu/23035307/Tres_paradojas_del_control_de_convencionalidad)
- RODRÍGUEZ, J., & VICENTE, D. (2009). *Aplicabilidad y validez de las normas del Derecho internacional*. Doxa: Cuadernos de Filosofía del Derecho, 177-204.
- ROMERO FLORES, B. (2001). *La imputación objetiva de los delitos imprudentes*. En Anales de Derecho, Universidad de Murcia, n° 19, pp. 259-278.
- ROVATTI, P. (2020). *Testigos no disponibles y confrontación: Fundamentos epistémicos y no epistémicos*. Quaestio facti. Revista Internacional sobre Razonamiento Probatorio, n° 1, pp. 31-66.
- SALMON, J. (1999). *Les accords non formalisés ou 'solo consensu'*, Annuaire Français Droit International, vol. XLV, pp.1-28.
- SANCINETTI, M. (2013). *Testimonio único y principio de la duda*. In Dret, Nro.3, pp. 1-22.
- TARUFFO, M. (2005). *La prueba de los hechos*. Madrid: Trotta.
- (2010). *Simplemente la verdad*. Madrid: Marcial Pons.
- (2018). *Proceso y verdad en la transición* . *Derecho & Sociedad*, 295-306.
- TEITEL, R. (2003). *Transitional Justice Genealogy*. Harvard Human Rights Journal , 69-94.
- THÜRER, D. (2009). *Soft Law - Norms in the twilight between law and politics*. En D. Thüerer, *International law as progress and prospect* (págs. 159-178). Zürich: Dike.
- UBERTIS, G. (2017). *Elementos de epistemología del proceso judicial*. Trotta, Madrid.

- VÁZQUEZ, C. (2014). Sobre la cientificidad de la prueba científica en el proceso judicial. *Anuario de psicología jurídica*, 65-73.
- (2015). *De la prueba científica a la prueba pericial*. Madrid: Marcial Pons.
- ZULETA, H. (2017). *El principio de proporcionalidad. Reflexiones sobre la racionalidad de la ponderación de principios de Robert Alexy* (en prensa), 1-26.



CONJETURAS  
Y REFUTACIONES



## NATURALIZED EPISTEMOLOGY AND THE LAW OF EVIDENCE: A REPLY TO PARDO, SPELLMAN, MUFFATO, AND ENOCH

Ronald J. Allen\*

**ABSTRACT:** In «Naturalized Epistemology and the Law of Evidence Revisited», the original target article for the various refutations that I comment on here, I revisited through a slightly different lens the subject of the article that I coauthored with Brian Leiter close to twenty years ago. That article has prompted four responses from Professors Pardo, Spellman, Muffato, and Enoch. Professors Pardo and Spellman basically accept the implications of the original article and offer useful but friendly amendments. Prof. Muffato apparently does not want to dispute over my ground and so changes the subject, and in doing so offers a number of interesting points. Only the fourth, Prof. Enoch, has the same doubts about the utility of my original article as I do of the genre that gave rise to it. I can thus be quite brief in discussing professors Pardo, Spellman, and Muffato, but it will take bit more effort to lay out the limits of Prof. Enoch's analysis.

**KEYWORDS:** probability; relative plausibility; epistemology; safety; sensitivity.

**CONTENTS:** 1. PARDO.—2. SPELLMAN.—3. MUFFATO.—4. ENOCH.—BIBLIOGRAPHY

**RECOMMENDED CITATION:** RONALD J. ALLEN, 2021: «Naturalized Epistemology and the Law of Evidence: A Reply to Pardo, Spellman, Muffato, and Enoch», publicados en *Quaestio Facti* 2/2021», in *Quaestio facti*, 3: 253-272. Madrid: Marcial Pons Ediciones Jurídicas y Sociales. DOI: [http://dx.doi.org/10.33115/udg\\_bib/qf.i3.22597](http://dx.doi.org/10.33115/udg_bib/qf.i3.22597)

---

\* John Henry Wigmore Professor, Northwestern University Pritzker School of Law; President, International Association of Evidence Science; Fellow, The Forensic Science Institute, China University of Political Science and Law. I am indebted to William Lawrence for excellent research assistance and to the Julius Rosenthal Foundation Fund for supporting this research.

In «Naturalized Epistemology and the Law of Evidence Revisited», the original target article for the various refutations that I comment on here, I revisited through a slightly different lens the subject of the article that I coauthored with Brian Leiter close to twenty years ago.<sup>1</sup> At the time, we focused on the distinction between purely a priori conceptual analysis on the one hand, and empirical inquiry on the other, and the admonition that «should implies can.» I had been in the grip of this methodology long before we wrote that article, although I would not have used the language of epistemology to describe my research program, and remain in its grip today, although with a slight evolution in focus if not in meaning. In the original article, I described my research program today as involving «inquiries—analytical or empirical—into how the human mind engages with its environment and the implications of that form of engagement for western legal systems—in particular the legal systems within the United States.»<sup>2</sup> The original article was intended to do two things. First, show the fecundity of this approach by briefly describing one of its major achievements, which has been to replace probabilism as the best explanation of juridical proof in the United States<sup>3</sup> with explanationism. The second objective was to cast doubts on one common methodology used to explicate juridical proof that does not subscribe to such an approach and remains focused on a priori conceptual analysis.

These latter efforts come in two general flavors. One focuses on the interior demands of the discipline in question and is not concerned with instructing the judicial system or its scientists in what to think or do. As I made clear, my original article is not addressed to them; I would not presume to instruct people working within other disciplines how to go about their tasks. The other flavor involves purported explanation of or prescription for juridical proof—which is the manner, from beginning to end, in which a system structures and resolves legal disputes—for, I presume, real legal systems, including the American system. This flavor does presume to be instructing jurists about the object of their inquiry, and on occasion the instruction is interesting and valuable. But often it is not, and I identified three variables common to the a priori approach that seem to explain this particular slip between cup and lip. First, much of this work is focused on weird hypotheticals that have no relationship to the American legal system. Second, and relatedly, this theorizing commonly neglects critical aspects of the actual state of affairs in the object being theorized about. Third, and I must candidly say quite remarkably for many of these commentators are well versed in the intricacies of epistemology, much of the theorizing entails literally impossible epistemological demands. As I also explained, my suspicion is that these three variables are derivative of a deeper conceptual problem that often inflicts the analytic disciplines of approaching juridical proof as a static rather than dynamic system that will yield simple and direct prescriptions, solutions, analyses, whatever, and this is false. This completely misconceives legal systems (and

---

<sup>1</sup> ALLEN and LEITER, 2001.

<sup>2</sup> ALLEN, 2021: 1.

<sup>3</sup> I return to the focus on U.S. law below in my comment on Enoch's paper.

here I dare to go beyond my own) which are complex and adaptive, not static and rule bound. They are, after all, complex human institutions.

This was all done in what I referred to as an admittedly «solipsistic» fashion, by which I meant, as I think was entirely clear, that I was explaining why I find much of this work of little value in my task of trying to advance knowledge of the actual legal system that I happen to know something about—the American legal system—whatever its value from an internal perspective of the various disciplines producing this work or for other legal systems. As I say, I would not presume to comment on such foreign territory. The editors of *Quaestio Facti* solicited that paper for their «Conjectures» section, anticipating that they would also solicit various refutations, which they have done. I am very grateful to the editors for having created this opportunity, and even more so to the commentators who have spent valuable hours critiquing the original paper.

And now I get to respond. There is, however, an immediate difficulty. Two of the papers, those of professors Pardo and Spellman, basically accept the implications of the original article and offer useful but friendly amendments.<sup>4</sup> The third, Prof. Muffato, apparently does not want to dispute over my ground and so changes the subject, and in doing so offers a number of interesting points. Only the fourth, Prof. Enoch, has the same doubts about the utility of my original article as I do of the genre that gave rise to it. I can thus be quite brief in discussing professors Pardo, Spellman, and Muffato, but it will take bit more effort to lay out the limits of Prof. Enoch's analysis.

## 1. PARDO<sup>5</sup>

I am grateful to professor Pardo for taking the time to provide his thoughtful refutation. Having introduced into the literature the concept of weird hypotheticals, perhaps it is fitting that I begin a response to a refutation in the somewhat weird way of quoting the refuter's conclusions, because I agree with them entirely:

The methodological picture I have described is one way in which epistemology may contribute to the law of evidence. It is intended to be an optimistic one and an invitation to epistemologists. For anyone taking up this invitation, I think there are three important points that Allen's critique reveals. First, it is important (and sometimes difficult) to accurately describe the underlying legal details. Second, there is a complex relationship between epistemology and law. It is unlikely that useful connections will involve a simple application of some epistemic concept or issue to law—a particular concept or issue may have different features, or play different roles, in the legal context than it does in other epistemological concepts. Third, it is probably time to stop trying to «solve»

<sup>4</sup> Professor Spellman, apparently, does so somewhat begrudgingly based on my (quite friendly) efforts to give her some feedback on her developing skills as a moderator. It just goes to show no good deed goes unpunished.

<sup>5</sup> A disclosure is in order here. Professor Pardo was a student of mine at Northwestern and I have worked closely with him throughout his academic career.

issues related to the Blue Bus and Gatecrasher hypotheticals and, for those interested in contributing to our understanding of the law of evidence, to focus elsewhere.

In Pardo's always erudite and at the same time sensible fashion, he puts the discussion into a very helpful framework. His analysis cuts right to the heart of the concerns that I spent considerably more time elaborating. The first problem with the weird hypotheticals is, as he says,

that the *question* they are being used to answer is not a genuine problem for law in need of analysis or explanation. [...] Because the hypotheticals are being put in the service of answering an ill-formed question, it is not surprising that after decades of trying no one has successfully answered the question or «solved» the problem. The second problem is that, even if the question being addressed were genuine, the examples do not provide enough information to answer it. The hypotheticals are framed so as to imply that the evidence appears to be sufficient on its face to satisfy the applicable standard of proof. Implicit in this framing, however, is the assumption that the standard of proof is a probabilistic threshold, [...] which the statistic in the example surpasses. But this assumption is itself highly questionable. If, instead, standards of proof are better understood as explanatory thresholds, as Allen and I argue, then the evidence is no longer necessarily sufficient on its face to meet the standard of proof.

Exactly, and I could end this response right here. But... I will make two other points. I think that Pardo attributes to me a skepticism that I do not possess, and more importantly for purposes of this exchange, has no relationship to the original paper. He says «I remain more sanguine about the genre of scholarship that Allen critiques—in my view, epistemology may provide a greater source of understanding for legal evidence and proof than Allen's article may suggest.» My article is being misread if taken as suggesting something general about the relationship between epistemology and inquiries into the nature of juridical proof—although I will happily attribute the error to my exegetical skills. I meant to concentrate, in a laser-like fashion, on one and only one well defined issue—and that is the use of weird hypotheticals that ignore critical parts of the legal system and that make literally impossible epistemological demands—no more and no less. To be sure, one of the reasons for focusing on this limited set is that the employment of such hypotheticals is quite prevalent in the scholarship purportedly analyzing juridical proof, especially from a philosophical perspective, but at the same time that work does not uniformly employ this methodology. My intent, whether accurately conveyed or not, was to carefully limit my analysis and I was not making any broader claims. I am indebted to Pardo for bringing to my attention the need for this clarification.

The second point is that the use Pardo makes of hypotheticals actually shows even greater agreement between us than perhaps Pardo suggests. For example, in his discussion of the chemist hypothetical, he makes it less weird by adding another layer to the original problem: «Suppose that there is a small range in which the process is unreliable; sometimes the result is slightly over the true amount and sometimes it is slightly under, *but (given our best available knowledge) the former tends to happen more often than the latter*» (emphasis mine). The more layers one adds, the less weird the hypothetical becomes, and it moves in the direction of responding to my critique.

The original hypothetical posited certain facts of the matter, but it was unclear who knew those facts (another ubiquitous problem in this genre, actually). If no one knew them except some ethereal entity (like God or the people who write Gettier problems), they are completely irrelevant to the case at hand. If the facts are known in the real world, they can be presented to the fact finder and the parties can argue over their significance. Pardo has now added the last aspect, thus, as I say, moving it in the direction of my critique.

But I am still not sure what is added to our understanding of juridical proof by invoking the language of safety and sensitivity. The problem that processing evidence depends on generalizations that have complexity is not new to the legal system. The problem of false memory, for example, or stereotypical thinking, which epistemologists would call insensitive evidence, involve issues that the legal system has dealt with for hundreds of years. And so has evidence that involves a range that may or may not include the relevant variable (was the car being driven in a reckless manner, what was his blood alcohol level), which the epistemologists would call safe or unsafe depending on the case. Pardo or others may think that invoking the epistemological framework facilitates understanding the implications of such matters, and even though I do not, perhaps they are right. I have no basis nor any desire to deny that possibility. One should present one's analysis in the most effective manner possible.

His second hypothetical is not a weird hypothetical at all. It is based on a real case and presents a real problem. But the solution to the problem does not require reference to or resolution of issues of knowledge (thankfully given the record of epistemology on that score). The Commonwealth is essentially asking the appellate court to try the case on the basis of evidence not presented to the jury and that does not fit within the concept of judicial notice. Doing so would violate the Sixth Amendment right to a jury trial with a finding by a jury of guilt beyond reasonable doubt.<sup>6</sup> That is true whether whatever the fact finder does involves «knowledge» or not. It is not «knowledge» that is at stake, but a decision based on the evidence presented, which under universally accepted conventions would include the kind of data involved in this case.<sup>7</sup> Still, perhaps presenting it as he does helps clarify the issues at stake, and again perhaps he is right.

The upshot of this is that Pardo uses hypotheticals with a keen awareness of their doctrinal setting, of, in other words, their relationship to reality. Using hypotheticals in the way that he does, at least with the elaboration of the case of the chemist, is not at the core, and maybe not even at the periphery of my critique of the use of weird hypotheticals that ignore critical aspects of the legal system and that make impossible epistemological demands.

---

<sup>6</sup> The court did not elaborate the doctrinal basis of its decision, and it could have reached the same conclusion on the basis of state rather than constitutional law. The decision would be the same in either event.

<sup>7</sup> What is "evidence" is more complicated than it appears. ALLEN, 1994.

## 2. SPELLMAN

I am grateful for Professor Spellman's very interesting guided tour of the history of experimental cognitive psychology, and her situating it in the context of the present discussion. The difficulty I have in responding is that I have even less disagreement with her than with Pardo. For example, although she is in agreement that relative plausibility is currently a better explanation than probability of the nature of juridical proof, «thinking about how other people do think, and can think, (i.e., metacognition) is even a more essential foundation of the system than is relative plausibility.» I agree.<sup>8</sup> Various strands in the psychological literature were constituents in the complex set of variables that led to the abduction of the best explanation of juridical proof. Thus (and maybe the only place I actually disagree with her), it is not «ironic» at all that «all of these findings» of the work she discusses on causal attribution and the story model among other topics «are consistent with [...] relative plausibility theory.» Quite to the contrary, they were partially responsible for the emergence of the theory.

I slightly disagree that what I call weird hypotheticals are examples of or very closely analogous to the thin hypotheticals that have often been used to start research programs in cognitive psychology. They are thin in the sense of having stripped down facts, but they are not weird in the senses of being completely divorced from reality, ignoring critical variables pertinent to the object under investigation, or making impossible epistemological demands.

But even if I am wrong about that, note that Spellman flips the script so to speak. The phenomenon that I analyzed involves invoking intuitions based on these peculiar scenarios, and then engaging in heroic efforts to justify them. Spellman is describing the opposite of this process. In her cottage industry, the responses to thin hypotheticals become the occasion to dig into the reasons for the responses to try to understand the underlying thinking process rather than engaging in an effort to rationalize the initial responses. In my cottage industry, the participants do not try to understand what motivated the intuition, but instead it is taken as the gold standard that must be justified. I have no idea what explains this, and it is all very peculiar. As I noted in the original article, there is essentially no evidence in support of the operant intuition about statistical evidence within the American legal system. No matter how often this intuition is evoked in response to weird hypotheticals and reported in various literatures, the actual American legal system applies normal evidentiary standards to statistical proffers and continues to domesticate new kinds of evidence, including things that begin to look close to naked statistics like DNA evidence. If the process that Spellman discusses were in play, the conclusion of all the products of this cottage industry would be that the intuition has not borne up to scrutiny and is rejected. I do not know of a single article in this genre that reaches that conclusion.

---

<sup>8</sup> ALLEN, 2011.

I will give just one example of the different dynamics at play. In Spellman's presentation of the blue bus hypothetical, she goes through various ways in which one might try to understand why people give the answers that they do to the experimental psychologists, or as she says «I don't care much about what people think of as answers to weird hypotheticals, but I do care why they think it.» That is exactly the right question to ask. The weird hypothetical folks I focused on, by contrast, literally never ask that question but instead try to construct rationalizations that would justify the response. In my opinion, this is a perfect example of the possible beneficial effects of internalizing the lessons of naturalized epistemology.

The different tasks undertaken explains one other area where we may disagree slightly. Spellman says «Allen's claim that weird hypotheticals ignore important aspects of actual legal systems—because they do not capture all the information in trials and they tend not to represent all the values that go into a trial decision—is absolutely true. But it is mostly irrelevant.» It is irrelevant to some extent to Spellman's tasks, but as I have laid out, empirical inquiry that must begin somewhere is not what I was criticizing. I was criticizing the opposite.

### 3. MUFFATO

I am particularly grateful to Professor Muffato for his contribution. As I understand his research program, it is the most distant of all the commentators from the central objectives of mine. He is a philosopher, apparently with a particular interest in linguistics, doing philosophy; I am an empiricist trying to understand a complex, sprawling system. He, with excess of kindness, suggests that I am «a profound connoisseur of statutory and common law of evidence in the U.S.A.» and so far as I can tell, so is he of the complexities and mysteries that go under the general label «epistemology.» I suspect it is because of these different sets of knowledge that he chooses not to dispute over the central focus of the original article and my research program, which is understanding the nature of juridical proof; indeed, he apparently is convinced not only of its value but of its discoveries. No, he shifts ground to give a penetrating analysis of the conceptual framework that he believes I have adopted or promoted, and then critiques how well that framework is actually reflected in or constrains the work that has been done. The reader can immediately see where this is going, of course—in my opinion he should pursue his interests, as I do mine. Regardless of any external judgments about its compatibility with philosophical or other norms, the turn to empirical and analytical inquiries into propositions with truth value that respect the admonition that «should implies can» has been quite beneficial to reinvigorating the field of evidence. That for me is the measure that matters.

I could end this comment with the simple recognition that different disciplines, and different scholars, pursue differing objectives, but there is a bit more to say. Muffato goes beyond elaborating the implications of his own field to give advice to those pursuing one of the interesting strands in the field of evidence today that

plays an important part in the original article—the ongoing shift in conceptualizing juridical proof as probabilistic to explanatory, an example being the theory of relative plausibility. Here he unfortunately is subject to a version of the central argument in the main paper—to wit: that advice about a legal system and research into it might do well to accurately reflect the object of the inquiry.

How is this reflected in Muffato's refutation? My original paper extolled the virtues of a research agenda for the field of evidence that is an example and reflects the lessons of naturalized epistemology. It compared that methodology and one of its main products, the theory of relative plausibility, to the results of a different methodology, purely a priori conceptual work heavily informed by probability theory, that has had little success recently in advancing understanding of juridical proof. Muffato has two objections to this: first, that the evolution of the theory of relative plausibility involves at best a pale form of «naturalized epistemology,» and second, that we are not doing it very well, as we have not conducted any controlled studies of juror reasoning.

I begin with the second and more important point to demonstrate that he misunderstands what he is criticizing and misevaluates its empirical support. Thus, the criticisms that he advances miss their target. I then conclude with a few words about what comprises «naturalized epistemology.» According to Muffato, «the empirical base of [the relative plausibility theory lies] [a]lmost entirely in the [...] “story model” resulting from the researches in social and cognitive psychology conducted by Nancy Pennington and Reid Hastie.» He goes on to say that «Allen might reply that explanations in a trial context are not stories, and that holistic reasoning supplies the lacking part in RPT: but in that case he should go into details and make explicit testable hypotheses about the holism he endorses.» And, he complains, I have not provided any new evidence from controlled studies that goes beyond what is presently in the literature, and thus the only empirical support for relative plausibility is woefully inadequate. This misconceives the project.

Muffato is under the misimpression that juror (human) reasoning involving stories and their generalization, explanations, is the overwhelmingly important aspect of the relative plausibility theory, and thus its success may depend on whether well-structured empirical studies support a particular theory of juror reasoning. This is simply false. Relative plausibility arose as an abduction influenced by a cacophony of variables, of which how people reason is only one. That so far as we know humans do and must think in a certain way consistent with the theory obviously provides support for an abduction about how a sensible legal system employs that form of reasoning. If further evidence is forthcoming about that matter, it may influence the inferences drawn about the nature of juridical proof, but the fact remains that the present best explanation of juridical proof, including how people reason, is explanatory rather than probabilistic. To be sure, it would be wonderful if further evidence were forthcoming about how humans reason, but it is odd to critique a theory

that is relying on the extant evidence as one data point for failure to gather more through studies in a completely different discipline.

Perhaps recognizing the unpersuasiveness of his complaint about lack of evidence, Muffato goes on to list seven of the variables that contributed to the abduction that juridical proof is best understood as explanatory, and then comments: «It seems then that Allen and Pardo aim at explaining empirically each one of the seven objectives enumerated and their mutual connections.» And we are criticized for not having done that work through carefully controlled studies. This has it completely backwards. Those variables and others contributed to the abduction; they are not the subjects of the analysis. The objective of relative plausibility theory is not to «explain empirically» the seven (and other) variables and their almost infinite variety of interactions—and it is a good thing, too, as I doubt that could be done in the way Muffato insists it must proceed. Rather, those are data points that led to the abduction in question.

But still the strength of the evidence matters for the persuasiveness of the abduction, and according to Muffato, the only empirical evidence for relative plausibility is the psychological evidence concerning the story model. And that is simply not good enough. This is an astonishing claim both about the legal system and more generally about what it means to be evidence. So far as the legal system is concerned, evidence scholarship has explained how the variables he identifies fit into a coherent picture of juridical proof. If there is a better explanation, it has not been forthcoming. That better explanation in turn would rest on counterexamples to the various points that have been made about the actual operation of the legal system, which again would be forthcoming if they existed. And again, they have not been offered. This is not an example of random sampling that he suggests is the sole basis of empirical knowledge, but he is simply wrong that knowledge of the sort humans actually use does not come from other sources than controlled studies—like observation and induction (no matter what Hume thought).

More pointedly, all of this overlooks the massive amount of evidence that supports relative plausibility as a better explanation than probability of juridical proof (which, do not forget, is the question under examination—which is the better explanation). To identify just some of this set: probability explanations of burdens of proof lead to unresolvable paradoxes; those same explanations are internally inconsistent (a probabilistic burden of persuasion can lead to any error distributions at trial); no human can reason about real life like a Bayesian; virtually never do relative frequencies about litigated matters exist; people do, in fact, reason through a focus on the alternatives available; parties form almost without exception, and judges often make them form if there is any resistance, a theory of the case; the theories of the case are virtually always presented as alternatives to the fact finder (judge or jury); the rules of evidence nudge the trial process in the direction of competing explanations; the parties are the masters of their cases, and courts do not interfere with party presentation for the most part; the Supreme Court of the United States has instructed the lower federal

courts to bear in mind that «a syllogism is not a story»; closing arguments weave the evidence together in as compelling an explanation as possible; appellate review of fact finding is differential, and so on.

Still, the complaint might continue, most of this has not been discovered through controlled studies. That is true to some extent,<sup>9</sup> but in part because it would often be a huge waste of time. Many of the variables that have been identified are ubiquitous (and as noted above subject to counterexamples that have not been forthcoming). It would be asinine for a scholar to ask for funds to do a study of whether parties in their pleadings have to articulate plausible explanations (again in federal courts by mandate of the Supreme Court and the rules of civil procedure); the literature on advocacy and «story telling» is rich and uncontradicted; one does not have to do a random sample of cases to «know» that the standard of review of fact on appeal is deferential (one only has to know how to do legal research), and again and so on. It was not discovery of these points and their interactions that is the contribution of relative plausibility, but the weaving of them together into the best extant explanation of the nature of juridical proof.

More generally, think of the implications of the view that empirical knowledge can only come from well-structured controlled studies. This precludes geology, astronomy, much of epidemiology, paleontology, sociology, macroeconomics among other disciplines from the reach of knowledge, which to say the least is curious. None of this is to disagree with the admonition to conduct controlled studies if that is where one's interests lie, and in fact one of the advantages of the relative plausibility theory is that it produces interesting questions that can be studied and confirmed.<sup>10</sup> By comparison, the probabilistic explanation produces verifiable propositions that have been systematically disconfirmed.<sup>11</sup>

There is another ironic similarity between Muffato's critique and the original article. The original article pointed out that the work being critiqued often made impossible epistemological demands. Muffato is arguing that the work on relative plausibility has not lived up to impossible standards given the object of inquiry. But note what happens when he engages in his own factual exegesis of a sprawling phenomenon analogous in that manner to the nature of juridical proof. To support his argument about the inadequate empirical basis of the relative plausibility theory, he draws a distinction between how «social theorists and historians» on the one hand, and «the officials of a legal order» on the other go about their affairs. With respect to the latter, he asserts:

---

<sup>9</sup> But see the next footnote.

<sup>10</sup> For example, see DIAMOND et al., 2015, examining the role of alternative explanations in actual jury deliberations. This is recent empirical work of that sort that Muffato calls for providing evidence confirming the relative plausibility theory, but which he rather surprisingly ignores.

<sup>11</sup> ALLEN, 2014, 2015, 2017.

The officials of a legal order [...] do not care about all these descriptive facets (and do not praise explanations that preserve the «pluralism» of the practice) when they try to set what the law is or «says» on a specific matter. Neither do formal legal doctrine and descriptive jurisprudence/legal theory, even if they work on a different level of systematic and generality. This happens because, while operating with a (not too precise) common sense concept of law/legality, they are concerned with deriving from the law itself—not from the concept of law—the (more precise) technical criteria for its description—rectius: for the recognition of its sources and the determination of its normative contents—in order to find out the legal solution to a practical problem. Officials and formal legal doctrine assume law's criteria of self-delimitation as given: an assumption an empirical social scientist cannot accept, but must substitute with a testable theory about the effectiveness of the preferred rule of recognition.

How does Muffato know what «officials of a legal order» do or do not do, what they care or do not care about, or what they assume? And to make matters worse, which «legal order» is he talking about? Is he asserting universals applicable to all imaginable «legal orders»? How did he make such remarkable discoveries? Through controlled empirical studies with appropriate randomization as he is demanding of relative plausibility? Of course not. It is not even clear how a serious study of the sort necessary to examine such questions could be organized.

When Muffato turns to making assertions about the legal system, he reports the results of his observations over time, and thus he must think that to be a usable version or reflection of «knowledge,» certainly good enough for the human condition in these circumstances. I agree. One of the central themes of the work that goes under the umbrella of «relative plausibility» specifically or «naturalized epistemology» as it pertains to legal systems more generally is that the tool should be tailored to the task. Many questions are amenable to examination by controlled studies; others are not. That some matters resist such forms of inquiry does not mean we can know nothing about them.

Perhaps the most surprising aspect of Muffato's critique is that, while he mentions that the objective of the work on relative plausibility has been to show that explanatory considerations better explain the nature of juridical proof than does probabilism, he then proceeds to ignore the point. His critique is a series of criticisms of the empirical adequacy and conceptual purity of the efforts to date that are entirely unexplanatory of the explanatory merits of explanationism compared to probabilism. He also neglects that science is a work in progress, and in this case it is a work in progress on a complex adaptive system, which complicates matters considerably (and substantially undercuts the value of the methodology he is promoting). Perhaps his various criticisms will prove useful to guide the evolution of the field, although frankly I could provide a more compelling list of the holes in knowledge that it would be helpful to fill than those that he provides. We shall see. Providing useful criticism in this context, however, cannot neglect the comparative nature of the enterprise.

This leaves the first of Muffato's critiques, to wit: that the work leading up to the theory of relative plausibility is at best a weak and pale form of naturalized epistemol-

ogy.<sup>12</sup> He says for example: «without the support of precise laws of supervenience (or of concept- and theory-reduction), sheer reference to empirical or scientific evidence in the context of the explanation of a social practice, even if useful, is not sufficient to qualify a theory as naturalistic or empirical, in the sense of naturalized epistemology (and jurisprudence).» I have no idea what «precise laws of supervenience» means, but this is tantamount to saying that Alvin Goldman—the most important scholar in the field—is not doing naturalized epistemology. Rather obviously various scholars, like Brian Leiter and Alvin Goldman, disagree with Muffato about the extension of the field. Muffato and whomever can argue about it. For reasons I have advanced, approaching the field of evidence from the perspective that Leiter and I identified, and I slightly embellished in the original article, whatever it is called, has had a positive effect on understanding the nature of juridical proof, which for me is the objective. I most certainly have no intent to instruct epistemologists on matters internal to their own field. Moreover, I think that he knows exactly what Leiter and I were saying (and I essentially repeated), to wit: that inquiries into the field of evidence would do well to concentrate on propositions with truth value and methods that may verify them, and that any normative advice should adhere to «should implies can.» I wonder if Muffato really wishes to argue with either proposition, or whether he actually thinks it might matter that technical but debatable criticisms internal to philosophy might be directed at our characterization—or more importantly the work that has followed in its wake. I also wonder how the methodology he employs in his refutation can possibly assist in deepening our knowledge of juridical proof. Of course, that is my agenda, not his.

#### 4. ENOCH

Criticism is central to the growth of knowledge, and that is why I am particularly grateful for the time and effort Professor Enoch has expended on his refutation. Of all the refuters, he has the greatest doubts. As I do concerning his refutation. Before explaining why, I want to emphasize that I am approaching the critique of Enoch's refutation in the same spirit of my original paper, which is to say viewing it from my perspective and not his. Part of that perspective involves focusing on the payoff

---

<sup>12</sup> Muffato makes various discrete points that doth protest too much. For example, I describe my research program as generally within the purview of naturalized epistemology as it pursues propositions with truth value either analytically or empirically, to which, in what is supposed to be a criticism I believe, Muffato responds: «the engraved disjunction—«analytical or empirical»—is a bit confounding, given that naturalization and rejection of the analytic/synthetic distinction are generally seen as coextensive.» To me, the significant point is the emphasis on propositions with truth value, whatever the internal debates are within his field. On that score, much work on the nature of juridical proof has advanced through straight forward logical analysis, such as the substantial clarifying of burdens of proof. See ALLEN, 2014. I see no point to worrying about whether that process is hermitically sealed off from empirics in a fashion that might falsify Quine's argument about the analytic/synthetic distinction. QUINE, 1951.

of different forms of work for my, not his, interests. Thus, his extended inquiry into whether intuition mining can ever be of value is simply a distraction. My focus was not on intuition mining per se but instead on the role of intuition mining based on weird hypotheticals that ignore critical parts of the legal system and that make literally impossible epistemological demands, and why such work has had a minimal payoff with regard to advancing knowledge of the nature of juridical proof. I thus ignore the many discrete points that are problematic in Enoch's article<sup>13</sup> and just concentrate on the ones that might matter.

From that perspective, Enoch's argument begins on a curious note, and gets as Alice said, curiouser and curiouser. He begins by saying that he fails to recognize his aims in my original article, which focuses on the utility of a certain kind of work for advancing the understanding of the American legal system. No, his aims are much different; they are «to reach knowledge and understanding of important relevant truths.»

Put aside that his characterization means that understanding the American legal system does not involve «important relevant truths» on the order of such questions as whether some theory is «true» or «enlightening» or «promote[s] insight into the nature of statistical evidence, and indeed evidence more generally.» If he were just saying that his motivations differ from mine, I would simply agree and urge him to pursue his interests, but he goes beyond that and refers to my analysis of the American legal system as «the academic analogue of American exceptionalism.» As I said in the introduction to this article, I think it is clear why I focused on the American legal system, and that is because it is the only one that I know in depth. But regardless, as I would think Enoch would know, «American exceptionalism» is the idea that the United States is exceptional in its pursuit of freedom and liberty and is a beacon providing an ideal for other cultures to emulate. There is in my original paper not a word of comparison of any aspect of the American legal system to that of another culture or country. Unlike Enoch, I am trying to understand an embodied system, not an ethereal one.

But again, curiouser and curiouser. In Enoch's pursuit of disembodied truth in the articles that I cite, he and his coauthors have scores of citations to cases (50 or so, with some multiple citations to the same case), and every single one of them save two is to American case law. Those two exceptions are English. Perhaps that is all it takes to transform an article into a cosmopolitan pursuit of truth rather than a slightly scandalous focus on American law. When Enoch et al. articulate what motivates their analysis at least in part, they assert that «the powerful, uncompromising intuitions of pretty much all of scholars» is that there is a critical difference between

---

<sup>13</sup> Like for example, his suggestion that perhaps I am so misguided as to think that one can “dismiss ... the relevance of the probability calculus entirely when commenting on legal evidence.” Perhaps this is not just a subtle *ad hominem* but simply reflects a lack of command of the legal literature. See, e.g., ALLEN, 2017.

statistical and nonstatistical evidence and that the law is consistent with that intuition.<sup>14</sup> Those «intuitions» for decades have come from scholarship focused on American and, to a lesser extent, English case law—just as his citations demonstrate. Indeed, what would it mean to have intuitions about the use of evidence in «the law» unless the phrase «the law» were referring to something real—like American law for example?

This might appear as little better than the «academic analogue» to name calling, but to the contrary there is a serious point lurking here. Maybe it does not matter where one's examples come from if one is investigating timeless truth, which Enoch informs that he is. But the truths that he is reaching for are not disembodied truths concerning statistical or other forms of evidence; the issues that he focuses on—sensitivity, safety, incentive effects, deterrence rationales for procedural and evidentiary matrices—are well known in the literatures of philosophy, economics, among others, and yes, even law. His efforts do not advance understanding of these in disembodied forms but instead prescribe how courts should use evidence in a certain way and attempt to justify the reported but erroneous intuitions about how they would use it.

But, what courts? All the courts in the world? Does he think the intuitions he relates would apply to those structuring or working within Kadhi courts? And how do those in Kenya differ from those in Zanzibar? How about the courts in Germany, Azerbaijan, China or Tanganyika? Sharia courts? Rabbinic courts? Are the scholars whose intuitions he is mining up on the evolution and present conditions of the world's varied systems of justice so that we can count on their intuitions to be informed and universal? Or is this just the paradigmatic example of what I am trying to avoid in my work and advising against in the work of others if they want to advance the understanding of real legal systems—sitting in a well-appointed office in the Western tradition thinking about «evidence» or maybe «evidence law» and dreaming things up?

I suspect so. I have worked on reform of the law of evidence and procedure in multiple countries and on multiple continents entailing widely divergent histories and economic, social and political contexts, and not to put too fine a point on the matter, it verges on the ridiculous to analyze juridical proof in a disembodied fashion. The critical question facing legal systems across the world is not some undiscovered truth about statistical or any other kind of evidence; rather, the critical questions are how to structure dispute resolution and use different kinds of evidence given the historical, economic, social, and political context of very much embodied legal systems.

Enoch thinks to the contrary. He thinks that examining an idealization such as naked statistical evidence may, like the mathematics of frictionless surfaces, yield insights into the real world. Further, he says:

---

<sup>14</sup> ENOCH et al., 2012

think about the suspiciousness of relying on statistical evidence in the real world; now assume all these «friction»-like factors away, so that you remain with an unrealistic clean case of statistical evidence. Does it feel like you've assumed away the problem? Of course not. These other factors are much more like friction to Newtonian mechanics, and much less like air resistance to aerodynamics.

Maybe you have not assumed away the problem, but you have ended the argument. The argument is about, and only about, naked statistical evidence. Once you assume away naked statistics and have non-naked evidence, no one—not even the individuals whose intuitions he is relying on—have any doubts about the approach of American courts—normal standards of relevancy, materiality and probative value apply; there is nothing unique about statistical evidence. I realize that what American courts actually do may not matter because knowledge of such things does not rise to the level of disembodied, timeless truths for which Enoch and his colleagues are striving. But if real court systems are not behaving consistently with the intuitions at the base of Enoch's argument, it is a complete mystery what the point of the intuitions might be.

There is, however, an ambiguity in Enoch's presentation. At times, Enoch refers to «unrealistic» examples of naked statistics, yet at other times he seems to think that the unrealistic case exists in the real world. For example, he criticizes my and Pardo's realistic examples of statistical evidence to demonstrate the erroneousness of the reputed intuitions in play because he claims that they mix up different types of statistical evidence and evidentiary contexts. He is focusing, he says, on «the phenomenon sometimes called base-rate evidence, sometimes market-share evidence, or sometimes naked statistical evidence...» as though these were synonyms. Yet both market share and base-rate evidence are potentially admissible evidence (and thus not «unrealistic») in every jurisdiction in the United States; thus, precisely what is being distinguished from what and why is somewhat opaque.<sup>15</sup>

Thankfully, Enoch's views on the ontology of naked statistical evidence does not matter, because the unrealistic case litigated on the basis purely of naked statistics does not exist in the real world. It is literally not possible to have and only have such evidence in an American case. At a very minimum, and perhaps unknown to Enoch, there is one and only one universal law of (American) evidence, even though it is nowhere written down, and that is that everything needs to be shown to be what it purports to be. Evidence, naked or otherwise, does not walk into court on its own power, hop up onto the witness stand and offer itself up for analysis. Someone purportedly with pertinent knowledge has to present it, and thus the idealization of a case decided solely on naked statistics never in fact materializes. Consequently, and critically, for Enoch's argument to be analogous to frictionless surfaces and to

---

<sup>15</sup> However, he may not be keeping evidentiary niceties straight, admissibility, sufficiency, and probative value being different concepts, and the question of the use of any particular statistical proffer could vary over these legal principles. For more on failing to keep distinct different evidentiary matters, see n. 19, *infra*.

have any meaningful content, it would need to extend to the realistic condition of «friction,» which here means non-naked evidence, and show how the idealization shines light onto the real world. That demonstration is entirely lacking in his and his colleagues' work.<sup>16</sup> Nowhere of which I am aware, at any rate, is any insight offered from an examination of the frictionless naked statistics to the world of friction inflicted non-naked evidence, and thus the entire body of work fails to live up to Enoch's own articulated standards.

Enoch's argument about the incentive effects involved with statistical proffers fares no better than his more abstract attempt to justify rootless intuition mining as a method of obtaining disembodied knowledge. Ironically, given that my primary point was the lack of utility to analyzing weird and unrealistic hypotheticals that make impossible epistemological demands, Enoch adds to the genre in responding to my demonstration that he and his coauthors' employment of incentive effects does not justify the intuitions in question. Their arguments made the critical mistakes of failing adequately to explore crucial distinctions, for example between bus drivers and bus companies, and offering wildly implausible assertions concerning likely human responses to differing conditions.<sup>17</sup> Consider his replay of the gate-crashing scenario:

But if you know that you may be convicted on the basis of statistical evidence alone, you know that your own action has little effect on your legal fate, or anyway, less effect than it has in a legal regime that is more suspicious of statistical evidence: After all, if so many people crash the gates, then the statistical evidence will be available against you even if you happen to buy a ticket. And if so, whatever incentive we may have hoped the criminal law supplied to buy the ticket rather than crash the gates is seriously eroded by the willingness to rely on naked statistical evidence. Notice how modest our claim here is: We're not saying that accepting statistical evidence always, necessarily, eliminates the relevant incentives. What we are saying is that the evidence law regime—and in particular, the attitude towards statistical evidence—may affect the relevant incentives, and at least often, in a counter-productive way.

---

<sup>16</sup> Remember that the intuitions he is mining include that courts would treat naked statistics as though there is a critical difference between naked statistics and something else. If naked statistics do not exist, what possible meaning can be given to an intuition about how courts would treat the nonexistent thing? It is like saying intuitions about the size and speed of unicorns may render insights into thoroughbreds. This is quite unlike friction that may range from a coefficient of 0.0 to 1.0. Indeed, rather than relying on people's untutored intuitions, he might have consulted the views of people who have actually studied the American court system and who fairly uniformly announce that this «intuition» is inconsistent with reality. For a discussion and citations, see PARDO, 2019: 262 n. 128.

<sup>17</sup> Tongue in cheek, I referred to my analysis of some of these issues as my own form of a priori analysis. I see that the joke was not understood. Making predictions as to how humans would react based on decade upon decade of observations of humans is not a priori. It is a form of clinical knowledge. I thought the point to be obvious. If Enoch and the other scholars he relies on have substantial experience with «courts» that lead to the pertinent intuitions (inferences, really), he should disclose how that can be done without observing actual courts—like those in the United States for example. It seems rather mysterious to me. Moreover, he ignores that the difficulty is not with conceptual analysis per se but with purely conceptual analysis. See ALLEN AND LEITER, 2001.

I will put aside the internal inconsistency in arguing that this is a «modest claim» even though the appropriate incentives of the criminal law will be «seriously eroded» «at least often [...] in a counter-productive way» and just stick to the facts. It is difficult to imagine a less perspicacious assertion than the claim about the American legal system that if statistical evidence is admitted, «you know that your own action has little effect on your legal fate.» A person will «know» this and have her incentives eroded even slightly, let alone seriously, only if the person is massively ignorant of the American legal system (massive stupidity by contrast is required for his blue bus arguments to go through, which I will turn to next). Buying a ticket always creates admissible evidence, from the receipt to a record of the form of payment (credit or debit card for example) to a friend who attended with the person to—I would have thought obviously—the admissible firsthand testimony of the ticket buyer. As one is walking toward the rodeo gate, reflecting on the legal system’s treatment of statistical evidence, the operative incentive is to create admissible evidence so that you will not be harassed with legal proceedings concerning whether you crashed the gate. Like remember what you did.

No judge in the land would let a case survive peremptory motions that involves only naked statistics of the gatecrasher variety and any plausible contrary evidence including firsthand knowledge of the person in question, for both epistemological and policy reasons, and not because there otherwise would be incentive crushing consequences that need to be suppressed in order to secure compliance with the law.<sup>18</sup> This is just a case of lousy evidence caused by the plaintiff offset by a plausible explanation.<sup>19</sup> Nothing more complicated than that would be needed to explain the phenomenon, were it ever to arise. When the proffered evidence is not ridiculous,<sup>20</sup> the courts treat statistical evidence more or less like any other form—as my list of ex-

---

<sup>18</sup> Enoch might respond that the gatecrasher hypothetical includes that NO other source of evidence is admissible, including the firsthand knowledge of the defendant. That is why it is a weird hypothetical that yields no insights into the real world. To the extent that there is an interesting incentive argument here, it has to do with parties not creating ridiculous phenomena like rodeos without gates or fences and then trying to take advantage of the public. There is no doubt that the government function of evidence law often involves the effort to create appropriate incentives, but that is not support for the argument Enoch is making about naked statistical evidence. On incentives and the governmental functions of the law of evidence, see ALLEN, 2015. Note that DNA is not a counterexample to any of this. DNA identifies a person on occasion when there is no other evidence, and the party offering the evidence is not responsible for that state of affairs, which among other things is missing in the gatecrasher hypothetical.

<sup>19</sup> I am not sure that evidentiary technicalities are being kept straight, so I will simply note that relevance, admissibility and sufficiency are different matters. For another example of Enoch and his colleagues jumbling up evidence law, see PARDO, 2018: 62. Although Enoch and I reach similar results in the gatecrasher case, the routes are considerably different. He is saying that there is something in our intuitions about the case that are informative about a class of evidence—naked statistics presumably. I am saying that the hypothetical just involves lousy evidence in a context that argues for incentives to rodeo operators to organize their affairs differently—in other words, normal approaches to probative value, the production of evidence, and policy concerns generate what would be the outcome. And of course, it is absurd to think that there would be no other evidence, as I pointed out in the original article.

<sup>20</sup> Nonstatistical proffered evidence can be ridiculous, too, just to be clear.

amples in the original article demonstrates. That in turn explains why to my knowledge no American court has ever excluded evidence or entered a preemptory motion on any ground remotely like Enoch is suggesting is a fundamental disembodied «truth» about statistical evidence. Enoch's explication of the gatecrasher case simply reinforces the inutility of studying weird hypotheticals. There is literally no evidence supporting his intuitions other than implausible assertions about how humans in the real world (in the United States) would react.

Enoch's argument in his refutation concerning the other aspect of the incentive argument that I critiqued involving blue buses is equally curious, although for different reasons. In essence, I made two points: First, that bus drivers had to be carefully sorted out from bus companies, and second, although the analysis would differ over the two categories in neither case would one get the socially undesirable incentive effects upon which Enoch et al. based their entire argument, unless the bus companies involved were massively stupid. As for bus drivers, it is just silly to think that statistics concerning the accident rate of bus companies will have any effect on their behavior (not based on purely a priori analysis but instead the observations of human beings over a lifetime), given the other interests of the drivers at stake.

Enoch now apparently agrees with this, and thus I presume that he rejects his earlier argument that bus drivers would be noticeably disincentivized from avoiding accidents, but he goes on to say, as though he were criticizing my argument generally:

This, however, is just not how incentive-explanations work, partly because this is not how incentives work. Sure, one way in which agents—people, organizations, etc.—respond to incentives is by explicitly deliberating about the incentives' significance for them and their interests. But it is an extremely impoverished picture of action and its responsiveness to incentives (and other reasons) to think of this as the only way in which incentives may work. Incentives serve to structure choice situations, and they infiltrate actions in other ways as well.

Reading this passage, one would think that I had neglected the bus companies, but rather than critiquing my original argument, this passage adopts it. As I pointed out, one has to take into account the adaptive nature of the phenomenon, and as again one observes over a lifetime in the economy, businesses adapt to new conditions (they are usually not massively stupid). And here the adaptation would be simple and cheap.

That is not what Enoch et al. originally thought. Inadvisedly relying on what must be one of the worst misapplications of microeconomics in the legal literature, they thought:

In addition to these precautionary distortions, the admissibility of statistical evidence may also impair market competition: Each of the bus companies will be incentivized to hold less than fifty-one percent of the market share so as not to suffer from the evidentiary disadvantage that a larger market share imposes. Moreover, the company holding the larger market share will absorb higher liability costs, which may lead to a decrease in its market share and to a possible exit from the market.<sup>21</sup>

---

<sup>21</sup> ENOCH AND FISHER, 2015: 583, n. 83, relying on an argument made by Richard Posner.

Rather than relying on a misguided application of microeconomics to the law, Enoch et al. would have been better advised to adopt one of the basic principles of microeconomic theory that having parties internalize costs contributes to efficient outcomes. As I pointed out in the original paper, the result of this form of liability would surely be (observational data again) lower social costs due to the reduction of accidents and the correct allocation of their costs.<sup>22</sup> Were the law to instead suppress statistical or other evidence in order to avoid what Enoch et al. feared, the law would indeed be misguided. Which probably explains why it does not do so. At least not in the United States.<sup>23</sup>

In any event, all the angst over «naked statistical evidence» has yielded literally nothing in decades<sup>24</sup> on non-naked evidence, which is the only form of evidence that exists in the United States.<sup>25</sup> It has not yielded anything of juridical import on naked statistical evidence because either the intuition is wrong that the courts would reject it were it actually to exist (DNA cold hits being the best recent close example)<sup>26</sup> or because such evidence does not exist (DNA evidence is produced by securing evidence and testing, either of which can introduce the possibility of error, and has to be authenticated at trial).

Although I am pleased to have this opportunity to continue the dialogue with Enoch and his colleagues, I must say that I am disappointed that he did not respond to the deeper points about the nature of evidence and the differences between statistical and nonstatistical evidence. As I pointed out, the distinguishing feature of statistical and nonstatistical evidence seems to be primarily how much one knows about the evidence in question. Realistic statistical evidence<sup>27</sup> for the most part carries its

---

<sup>22</sup> For example, a cheap adaptation (cheaper than leaving the market surely) that would both reduce accidents and assist in allocating costs correctly would be the installation of cameras on buses.

<sup>23</sup> There are points at which Enoch's argument becomes incomprehensible. After critiquing my explanation of how incentives would actually play out in the real world, he says:

Does Allen really want to suggest—on the basis of a priori reasoning, no less!—that nothing in evidence law supplies any incentive at all for primary behavior? Indeed, if his objection to our incentive story is to be taken seriously, he must reject anything remotely resembling law-and-economics explanations of pretty much everything, and vast parts of the social sciences.

Having already demonstrated that Enoch et al. misapplied microeconomics and that the incentive structure would differ (not disappear) from what he and his colleagues asserted, I will let this pass without further comment. As I will his statement, because I literally have no idea what it is referring to, «if Allen is sincere about his attempts to implement in legal theory the lessons of naturalized epistemology, he cannot consistently reject all attempts to incorporate philosophical insight into studying the law.»

<sup>24</sup> When these arguments started in the juridical context fifty years or so ago, it was a different matter.

<sup>25</sup> I am putting aside a different literature on Bayesian approaches to evidence that have nothing to do with the present discussion.

<sup>26</sup> For what might be the most recent example, see <https://www.cnn.com/2020/11/19/us/norcal-rapist-roy-charles-waller-convicted-trnd/index.html>.

<sup>27</sup> I say «realistic» statistical evidence because it is impossible to decipher the kind of assertions typical in the weird hypothetical literature, like «The witness identified the bus as blue and is 70% reliable.»

ambiguity on its surface, whereas nonstatistical evidence for the most part does not. In that sense, statistical evidence often is better than nonstatistical evidence, as its limitations are more obvious. Thus, the arguments about preferring one form of evidence to the other often seem to have it backwards in preferring the more to the less ambiguous evidence. I would love to see individuals with the talents of Enoch and his colleagues tackle that issue, including pointing out my mistakes, rather than being distracted by weird hypotheticals that ignore critical aspects of real legal systems and make impossible epistemological demands.

## BIBLIOGRAPHY

- ALLEN, R.J., 1994: «Factual ambiguity and a theory of evidence», in *Northwestern Law Review*, 88(2): 604-640.
- ALLEN, R.J., 2011: «Rationality and the Taming of Complexity», in *Alabama Law Review*, 62(5): 1047-1068 (2011).
- ALLEN, R.J., 2014: «Burdens of Proof», in *Law, Probability and Risk* 13(3-4): 195-219.
- ALLEN, R.J., 2015: «A note to my philosophical friends about expertise and legal systems», *Humana. Mente Journal Of Philosophical Studies*, 8(28):79-97.
- ALLEN, R.J., 2017: «The nature of juridical proof: probability as a tool in plausible reasoning», in *International Journal of Evidence and Proof*, 21: 133-142.
- ALLEN, R.J., 2021: «Naturalized epistemology and the law of evidence revisited», in *Quaestio facti. Revista Internacional sobre Razonamiento Probatorio*, 2: 1-32.
- ALLEN, R.J. and LEITER, B., 2001: «Naturalized epistemology and the law of evidence», in *Virginia Law Review*, 87(8): 1491-1550.
- DIAMOND, SHARI S., MURPHY, BETH AND ROSE, MARY R., 2015: «The “Kettleful of Law” in Real Jury Deliberations: Successes, Failures, and Next Steps», in *Northwestern University Law Review*: 106(4) 1537-1608.
- ENOCH, D., SPECTRE, L AND FISHER, L., 2012: «Statistical evidence, sensitivity, and the legal value of knowledge», in *Philosophy & Public Affairs*, 40(3): 197-224.
- ENOCH, D. AND FISHER, T., 2015: «Sense and sensitivity: Epistemic and instrumental approaches to statistical evidence», in *Stanford Law Review*, 67(3): 557-611.
- PARDO, MICHAEL S., 2018: «Safety vs. sensitivity: possible worlds and the law of evidence», in *Legal Theory*, 24 (1): 50–75.
- PARDO, MICHAEL S., 2019: «The paradoxes of legal proof: A Critical Guide», in *Boston University Law Review*, 99(1): 233-290.
- QUINE, WILLARD V.O., 1951: «Two dogmas of empiricism», in *The Philosophical Review* 60 (1): 20-43.

INFORME DEL PCAST:  
CIENCIA FORENSE  
EN LOS TRIBUNALES PENALES



## PRESENTACIÓN DE LA TRADUCCIÓN AL CASTELLANO DEL INFORME DEL PCAST SOBRE LA CIENCIA FORENSE EN LOS TRIBUNALES PENALES

Carmen Vázquez  
Universitat de Girona

En el ámbito de la prueba pericial es fundamental el análisis de la experiencia estadounidense, dados los desarrollos que han caracterizado ese sistema prácticamente desde 1923, con la resolución del caso *Frye*. No obstante, dicha experiencia muchas veces ha sido reducida al estudio del famoso caso *Daubert* y sus cuatro criterios establecidos por la Corte Suprema en 1993 para decidir si una prueba pericial era fiable y, con ello, si era admisible: que un método o técnica pueda ser (y haya sido) sometido a test o pruebas, que haya sido sometido a revisión por pares y publicado, que se indique su rango de error conocido o posible y se muestre que goza de una amplia aceptación en la comunidad experta de referencia<sup>1</sup>.

Esos cuatro criterios han sido traducidos en muy diferentes idiomas para ser utilizados por la jurisprudencia<sup>2</sup>, la dogmática procesal<sup>3</sup> o incluso por el propio

---

<sup>1</sup> Un análisis de los diversos criterios que conforman los casos *Frye* y *Daubert*, puede verse en HAACK (2020) y en VÁZQUEZ (2015).

<sup>2</sup> Es el caso, por ejemplo, de México, donde la Suprema Corte de Justicia de la Nación los introdujo en la contradicción de tesis 154/2005-PS, diciendo que: «Para que un órgano jurisdiccional pueda apoyarse válidamente en una opinión de algún experto en una rama de la ciencia, es necesario que esa opinión tenga las siguientes características:

1. Que la evidencia científica sea relevante para el caso concreto de estudio, es decir, que a través de la misma pueda efectivamente conocerse la verdad de los hechos sujetos a prueba, y

Nota 3 en página siguiente

legislador<sup>4</sup> con el objetivo de establecer criterios de admisibilidad o, bien, para sugerirlos como criterios de valoración de este tipo de elementos de juicio. Sin embargo, la experiencia estadounidense va muchísimo más allá de la mera identificación de criterios para orientar las decisiones judiciales, pues sobre todo en los últimos años se han puesto en marcha esfuerzos institucionales diversos para obtener información relevante sobre la fiabilidad de métodos, técnicas, etc., *más allá de un proceso judicial concreto*. Dos rasgos destacan en estos esfuerzos:

*i)* Que muestran un interés en conocer la fiabilidad de los métodos, técnicas, teorías, etc., que usan los expertos (que luego fungen como peritos). Que dicha información es, en parte, independiente de un proceso judicial.

Aunque parezca una cuestión banal afirmar que estamos interesados en conocer la fiabilidad del conocimiento que emplea un experto cuando funge como perito en un proceso, en realidad no lo es y, de hecho, nuestros sistemas jurídicos han actuado como si no lo estuviéramos. ¿Por qué? Básicamente porque han prestado atención fundamentalmente a quién es el experto en turno para decidir si le creen o no<sup>5</sup> o

---

2. Que la evidencia científica sea fidedigna, esto es, que se haya arribado a ella a través del método científico, para lo cual se requiere, generalmente, que la teoría o técnica científica de que se trate:

- a. Haya sido sujeta a pruebas empíricas, o sea, que la misma haya sido sujeta a pruebas de refutabilidad;
- b. Haya sido sujeta a la opinión, revisión y aceptación de la comunidad científica;
- c. Se conozca su margen de error potencial, y
- d. Existan estándares que controlen su aplicación”.

<sup>3</sup> En España, por ejemplo, NIEVA FENOLL (2010: 294 ss.) sugiere que el dictamen debe seguir parámetros científicos de calidad, entre los cuales estaría: «que las técnicas y teorías científicas utilizadas para obtener datos y conclusiones han sido aplicadas previamente, son relevantes y están generalmente aceptadas por la comunidad científica internacional»; «que las técnicas utilizadas se han aplicado según los estándares y normas de calidad de vigentes»; o «que el dictamen contenga información sobre el posible grado o nivel de error y también el nivel o gradación de variabilidad e incertidumbre de los datos obtenidos por la citada técnica o teoría científica».

<sup>4</sup> En Colombia, el legislador en el Código procesal penal, Ley 906 de 2004, en el artículo 422, estableció: «Admisibilidad de publicaciones científicas y de prueba novel. Para que una opinión pericial referida a aspectos noveles del conocimiento sea admisible en el juicio, se exigirá como requisito que la base científica o técnica satisfaga al menos uno de los siguientes criterios:

1. Que la teoría o técnica subyacente haya sido o pueda llegar a ser verificada.
2. Que la teoría o técnica subyacente haya sido publicada y haya recibido la crítica de la comunidad académica.
3. Que se haya acreditado el nivel de confiabilidad de la técnica científica utilizada en la base de la opinión pericial.
4. Que goce de aceptabilidad en la comunidad académica».

<sup>5</sup> Por supuesto, hay que atender, por ejemplo, a las credenciales de los expertos para asegurarse de que tienen el conocimiento relevante para analizar el caso y que, entonces, puedan ser nombrados como peritos. No obstante, la práctica de preocuparse por la titulación de un experto solo a efectos de atribuirle valor probatorio es ampliamente cuestionable, no solo porque la titulación no nos brinda gran información sobre los fundamentos que tiene un perito para afirmar lo que afirma, sino incluso porque la etapa en la que deberíamos atender a ese criterio es en la admisión. No debe admitirse a un sujeto que

han empleado criterios más bien formales relacionados con quién nombra al perito<sup>6</sup> o, en el peor de los casos, han atendido exclusivamente a ciertos comportamientos procesales del perito<sup>7</sup>. Ninguno de esos aspectos nos informa sobre la fiabilidad del conocimiento experto empleado y, por ello, solo podríamos decir que nos hemos interesado en la *credibilidad* que genera alguien que se presenta como experto o como un perito imparcial o que hace un peritaje que parece completo y coherente, etc. En este contexto, obviamente, credibilidad y fiabilidad son términos muy distintos: el primero es ampliamente subjetivo, dependiente de las creencias de cada uno sobre un sujeto; el segundo, en cambio, depende del mundo, de cómo de hecho funciona un método o una técnica con independencia de lo que cada uno de nosotros cree sobre cómo lo hace. Nótese que no solo se trata de parámetros más o menos objetivos, sino de qué es aquello que se está analizando, es decir, el sujeto o el conocimiento, método, técnica, etc. que emplea ese sujeto. Claramente la balanza de nuestros sistemas se ha inclinado vertiginosamente hacia el sujeto, descuidando los fundamentos que tiene para afirmar lo que afirma<sup>8</sup>.

Vale la pena introducir un matiz en la distinción entre el sujeto y el conocimiento que emplea ese sujeto, pues hay métodos o técnicas que dependen en buena medida del criterio o juicio de una persona y, por ello, ciertas características suyas cobran suma relevancia para poder valorar mínimamente la calidad de sus afirmaciones. Un ejemplo de método que depende fuertemente de juicios humanos, de la percepción e interpretación del experto y que es empleado en algunas pericias antropológico-sociales es la llamada «observación participante» que, como su nombre indica, a grandes rasgos consiste en que el experto observa activamente a grupos o comunidades en su vida cotidiana para luego describir algún aspecto de ellas, por ejemplo, sobre sus costumbres en determinada actividad. Esto significa que el investigador, sus ojos, son en sí mismos el instrumento principal en este tipo de pruebas, pues su percepción, sus decisiones sobre qué observar, sus interpretaciones filtran todo el estudio de la

---

no es experto en la materia relevante para el caso, no podemos perder tiempo y esfuerzos en un proceso judicial en pruebas que son claramente irrelevantes.

<sup>6</sup> Sigue siendo común que se entienda que un experto seleccionado y pagado por una de las partes es necesariamente parcial y que no amerita valor probatorio; en cambio, por el contrario, que un perito oficial en sí mismo amerita mayor valor probatorio porque no ha sido seleccionado y pagado por las partes. No obstante, un perito de parte podría realizar su trabajo de manera totalmente correcta, usar métodos fiables, hacer un razonamiento impecable y, por ello, ameritar mayor valor probatorio que un perito oficial que podría usar métodos no tan fiables, cometer errores o incluso sufrir de ciertas parcialidades cognitivas. Precisamente por todo ello, la calidad de una prueba pericial no necesariamente depende de qué tipo de perito la haga.

<sup>7</sup> ABEL LLUCH (2017: 237 ss.) brinda como ejemplos «la ratificación y matización del dictamen en la vista o en el juicio» o «el dominio de la oralidad por el perito (silencios, dudas, comunicación no verbal, etc.)», etc.

<sup>8</sup> Esta situación es un problema compartido por las pruebas periciales y las pruebas testificales. En ambas, a efectos de su valoración, la atención se ha centrado en valorar quién es el sujeto que está afirmando X y no en las afirmaciones que hace.

comunidad que analiza. Así pues, podríamos decir que este método es más subjetivo que objetivo, en el sentido en que depende muy fuertemente de un sujeto.

No obstante, hay un buen conjunto de pruebas periciales con mucha mayor independencia del criterio del sujeto que las realiza y, en ese sentido, más objetivas. Entre esas encontramos las pruebas periciales que se basan en la comparación de ciertas características de personas u objetos que permitirían hacer una identificación relevante en un proceso judicial: por ejemplo, de un sujeto a través de su ADN o de un arma en función de determinados patrones. La validez y la fiabilidad de este tipo de pruebas constituye el objeto de análisis del informe del *President's Council of Advisors on Science and Technology* (PCAST) de Estados Unidos, realizado en septiembre de 2016, y titulado «Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature – Comparisons Methods», cuya traducción se publica aquí.

Y, precisamente respecto de la mencionada distinción entre métodos más o menos dependientes del criterio de un sujeto, el PCAST nos brinda un párrafo que no tiene desperdicio alguno:

Queremos resaltar, finalmente, que ni la experiencia, ni el buen juicio, ni las buenas prácticas profesionales (tales como programas de certificación y acreditación, protocolos estandarizados, pruebas de aptitud técnica y códigos éticos) pueden sustituir las pruebas genuinas sobre la validez de sus fundamentos y su fiabilidad. La frecuencia con la que un patrón o un conjunto de características particulares se observa en diferentes muestras, que es un elemento esencial para inferir las conclusiones, no es un asunto de «juicio». Es un hecho empírico para el cual solo es relevante la evidencia empírica. De forma similar, la expresión de confianza utilizada por un analista, basada en su experiencia profesional, o las expresiones sobre un consenso entre analistas respecto a la precisión a la que llegan en su campo de trabajo, no pueden sustituir a las tasas de error estimadas a partir de estudios relevantes. Para los métodos forenses de comparación de características, el establecimiento de la validez de sus fundamentos a partir de evidencia empírica es entonces una condición *sine qua non*. Nada puede sustituirla.

El punto clave para entender por qué deberíamos estar interesados en conocer la validez y la fiabilidad de los métodos empleados para la identificación de objetos y personas que analiza el PCAST (y otros tantos empleados en los tribunales) es que la validez y la fiabilidad de un método es un hecho del mundo y que para evaluarlas solo nos sirve la evidencia empírica. Si se trata de un hecho que es susceptible de ser comprobado empíricamente, entonces se requiere hacer los experimentos necesarios para mostrar que los métodos o técnicas en comento tienen la capacidad de funcionar exitosamente. Quizá para el lector no familiarizado con la experimentación empírica, la siguiente narración de Luis MOCHÁN (2010: 34), un físico de la Universidad Nacional Autónoma de México, le brinde cierta luz:

Imagina que te lastimas un dedo jugando a voleibol. Un sanador te coloca un anillo imantado y una semana después tu dedo ha sanado y el dolor ha desaparecido. ¿Qué podrías concluir de ello? Que el magnetismo es terapéutico.

No. El anillo magnético podría no haber jugado ningún papel. Todos los días hay gente que se lastima y sana sin ninguna ayuda. Para saber si el campo magnético sirve necesitamos repetir la experiencia muchas veces y hacer estadística con los resultados.

Yo no estaría dispuesto a lastimarme repetidas veces para hacer estadística.

Claro, pero un médico tiene acceso a muchos pacientes, y hay muchos médicos. Imagina que se consigue un grupo numeroso de pacientes con dedos lastimados y que la mitad de ellos se les coloca imanes y a la otra mitad no. A la semana han sanado un número determinado, que llamaremos P, de miembros del primer grupo y S miembros del segundo grupo. ¿Qué demostraría que P fuese mayor que S?

Ahora sí, que los anillos magnéticos curan.

No, todavía no. El primer grupo pudo haber sanado más rápido por sentir que su dolencia fue atendida y por su confianza en el tratamiento, mientras que la mejoría entre los pacientes del segundo grupo pudo haberse retrasado por la decepción de no haber recibido cuidado alguno. Quizás los miembros del primer grupo hubieran sanado igual con cualquier otro tratamiento o incluso con un tratamiento falso. Este es el conocido efecto placebo. Sería importante demostrar que el tratamiento es mejor que la administración de meros placebos. Para que el experimento sirva, se le tendría que poner un anillo similar a todos los pacientes, de los cuales solo la mitad estuvieran magnetizados y la otra mitad no. El paciente no debería saber qué clase de anillo le tocó. En este sentido, el experimento debe ser ciego.

Si P es mayor que S, ahora sí sabríamos que el magnetismo es benéfico.

Aún no del todo. Podría ser que el médico se mostrará más alegre y comunicara su confianza a los pacientes del primer grupo. Además, al registrar los resultados podrían manipularlos e ignorar pequeñas dolencias en los miembros del primer grupo y pequeñas mejorías en los del segundo grupo.

¡Eso sería deshonesto!

No necesariamente. El médico podría favorecer los resultados del primer grupo inconscientemente, motivado por el deseo de que el tratamiento funcione. Para evitar esto, requeriríamos que el médico que vaya a distribuir los anillos y realizar las observaciones no sepa cuáles son magnéticos y cuáles no. La prueba debe ser doble ciego.

Pero si nadie sabe nada sobre los anillos, ¿cómo podremos averiguar si los anillos magnéticos funcionaron?

Claro, alguien debe saber algo de los anillos. Un participante en el experimento podría inscribir un número al azar en cada anillo y anotar qué números corresponden a qué tipo de anillo, pero sin informar al médico ni a los pacientes sino hasta después de haberse registrado los resultados.

Ahora sí, ¿si P es mayor que S habremos demostrado las propiedades curativas de los imanes?

Ya casi. Necesitaríamos que la diferencia sea suficientemente grande para ser estadísticamente significativa.

Para ser ¿qué?

Diferencias pequeñas podrían deberse al azar: por ejemplo, si tiras bolados honestos deberías obtener alrededor de 50 soles. No sería sorprendente que salieran 55, ¡pero si obtuvieses más de 80, yo podría sacar ciertas conclusiones estadísticamente significativas sobre tu poca honorabilidad!

Este tipo de experimentación científica, en dónde se pone en acción un cierto método o técnica (en el ejemplo, los anillos imantados) en una cantidad suficiente de varios casos en las mismas condiciones (*v. gr.*, dedos lastimados) para lograr un objetivo (en el relato, sanar el dedo), de ninguna manera puede ser sustituida por la mera experiencia de un sujeto o por acuerdos de grupos de expertos, etc. ¿Por qué la experiencia de un sujeto no contaría para ello? Porque se corren distintos riesgos importantes: como que el criterio de corrección de lo que dice un experto termine siendo el mismo experto o que el sujeto sufra de sesgos que le impidan ver sus errores o las pruebas contrarias a sus creencias o que no sea capaz de analizar un número suficiente de casos, entre muchos otros. Precisamente para evitar esos riesgos que

invalidarían el experimento, es que este tipo de investigación se debe llevar de forma controlada, es decir, estableciendo medidas que garanticen, en la medida de lo posible, que los resultados se deben exclusivamente al funcionamiento correcto de un método o técnica y no a otras cuestiones.

En todo lo anterior, la idea de «replicabilidad» es clave. Si un método o técnica empleado por un experto no puede ser puesto en marcha de igual manera por otros expertos para tratar el tipo de casos en que supuestamente funciona, entonces no hay manera de comprobar si efectivamente lo hace o no. Esto supone que la experimentación es sobre *un* método estandarizado, por ejemplo, que permite identificar las huellas dactilares, y no sobre distintos métodos que podrían llevar a identificar huellas dactilares; aunque, obviamente, se podrían someter a experimentaciones distintas esos diversos métodos. En todo caso, la idea es preguntarse si este método *X realmente* funciona o no para llegar a conocer algo sobre el mundo. Esa idea intuitiva sobre si “funciona” o no, al trasladarse a términos técnicos, corresponde a la «validez» y «fiabilidad» de un método o técnica. En términos muy llanos, como dijo la Corte Suprema estadounidense, un método es válido si mide lo que pretende medir y, en cambio, su grado de fiabilidad nos informará sobre cuán bien mide eso que efectivamente mide. Esto significa que primero habría que mostrar la validez y luego la fiabilidad; siendo la primera una cuestión de todo o nada, es válido o inválido, mientras que la segunda, por el contrario, es gradual, más o menos fiable. Si un método no es válido, poco sentido tiene entonces preguntarnos por su fiabilidad.

Pues bien, el PCAST en este informe, «Ciencia forense en los tribunales penales: asegurando la validez científica de los métodos forenses basados en la comparación de características», nos ofrece un conjunto de datos empíricos fundamentales sobre la validez y la fiabilidad de varios métodos forenses: análisis de ADN, marcas de mordeduras, huellas dactilares latentes, armas de fuego, huellas de calzado y cabellos. No lo hace mediante la realización de experimentos como los arriba mencionados, sino haciendo «meta-análisis» de los experimentos publicados. Esto es, analiza un conjunto de estudios publicados en los que se da cuenta de los resultados que arrojaron diversos experimentos que tuvieron como objetivo mostrar la validez y fiabilidad de los métodos o técnicas que se estudian.

Los meta-análisis, como su nombre indica, son análisis de los análisis, esto es, integran de forma estructurada y sistemática los resultados obtenidos en diferentes estudios empíricos. Para ello se recopila de forma exhaustiva un conjunto de estudios que se considera que fueron hechos en las condiciones adecuadas, ofreciendo entonces un estudio de más casos que cualquiera de los tomados en cuenta en los estudios individuales.

Y aquí es donde el PCAST se encuentra con una gran sorpresa: para la mayoría de los métodos o técnicas que analiza, no hay un número importante de estudios empíricos publicados que hayan pretendido demostrar que efectivamente funcionan y cómo lo hacen. Ello es en sí mismo un gran problema para el uso de esos métodos o técnicas en el ámbito jurídico, pues: ¿cómo es que seguimos usando métodos o

técnicas para decidir casos que impactan en los derechos y libertades de nuestros ciudadanos cuando ni siquiera sabemos si funcionan? Y esta pregunta, crucial en el ámbito jurídico, primero debería hacerse desde el ámbito de los expertos, en este caso, de las ciencias forenses.

Una de las explicaciones que se han dado es que, en ambos contextos, el jurídico<sup>9</sup> y el forense<sup>10</sup>, existe una peligrosa *incultura* de la investigación empírica y todo lo que esta supone. Por ello, los propios forenses no se cuestionan si está demostrado científicamente que los métodos o técnicas que emplean de hecho funcionan, simplemente actúan como si lo estuviese. Y peor aún, incluso siguen afirmando que sus métodos o técnicas son infalibles, lo que es imposible desde el punto de vista empírico y científico. No hay ningún método o técnica 100% fiable y es precisamente por ello que la investigación empírica debe mostrar cuáles son las condiciones en que mejor funciona. Se trata, entonces, de una labor colectiva, en la que las diferentes piezas del engranaje científico deberían operar para mostrar que determinado instrumento puede ser usado en un proceso judicial. En cambio, en muchas ocasiones, parece que todas esas piezas solo *suponen* que funciona y, con esa mera creencia, se termina privando de la libertad o restringiendo los derechos de las personas. Es un sinsentido y una grave irresponsabilidad colectiva.

Esa situación de ignorancia debe cambiar, no podemos seguir tomando decisiones públicas que se apoyan, aunque sea parcialmente, en métodos y técnicas que ni los propios expertos (o supuestos expertos) saben si funcionan o cuán bien lo hacen. En esta línea, la traducción al castellano de este informe del PCAST puede jugar un rol importante, no solo brindándonos información relevante sobre la validez y fiabilidad de los métodos que analiza, sino incluso formándonos en el tipo de información que los sistemas de justicia deberían exigir a los expertos. Hago referencia a los expertos, y no a los peritos, porque la investigación que se requiere para obtener la información relevante supone el trabajo de comunidades de expertos, no de una persona o ni siquiera de alguien que funge como perito en los procesos judiciales. Esto nos recuerda la importancia de las comunidades de expertos a las que aluden los criterios establecidos en los casos *Frye* y *Daubert*, pero no por su mera existencia o porque podamos consultarles si aceptan o no un método o técnica, sino porque son

---

<sup>9</sup> EDMOND (2020a y 2020b), por ejemplo, al tratar el caso de las huellas dactilares latentes, nos muestra cómo la jurisprudencia y las prácticas de distintos países del *common law* han ignorado las diversas investigaciones empíricas que se han desarrollado sobre la fiabilidad de estos métodos de identificación, muchas veces argumentando que las herramientas jurídicas son suficientes para mostrar si son o no fiables.

<sup>10</sup> MNOOKIN, *et al.* (2011: 726) defienden explícita y contundentemente que las ciencias forenses en general «actualmente no poseen -y, sin lugar a duda, deberían desarrollar- una base científica bien establecida. Esto sólo puede lograrse mediante el desarrollo de una cultura de investigación que impregne todo el campo de la ciencia forense. Una cultura de investigación [...] debe basarse en los valores del empirismo, la transparencia y el compromiso con un análisis crítico permanente. Las disciplinas de las ciencias forenses deben acrecentar sustancialmente su compromiso con las pruebas de la investigación empírica como base de sus conclusiones».

quienes pueden demostrar adecuadamente si aquellos funcionan y cuán bien funcionan. Por ello, para mejorar nuestras prácticas con las pruebas periciales y, con ello, nuestros sistemas de justicia, no basta trasladar los criterios *Daubert*, sino entender las implicaciones más profundas que tiene su establecimiento: generar conocimiento actualizado sobre la validez y fiabilidad de los métodos y técnicas que emplean los peritos.

Ahora bien, antes mencioné que la información sobre el funcionamiento de los métodos, técnicas, etc., es independiente de *un* proceso judicial. Ello quiere decir, por un lado, que esa información debe ser el resultado de investigaciones que no deberían estar relacionadas con un caso específico y, por el otro, que un si se atribuye validez y fiabilidad a un método o técnica, ello nada dice sobre la corrección de su *aplicación* a un caso concreto. Claro, un método válido y tan fiable como la identificación de ADN en casos de paternidad, podría ser aplicado de manera inadecuada en un caso concreto y llevarnos a una identificación errónea por causas diversas: contaminación en la muestra o en el laboratorio<sup>11</sup>, incapacidad del analista que llevó a cabo la comparación, etc. Es precisamente en este punto donde tiene importancia analizar *quién* realiza la labor pericial, si tiene una adecuada formación e incluso experiencia; pero, como ya se mencionó antes, ello no puede substituir a los datos empíricos sobre la validez y fiabilidad de los métodos o técnicas empleados. Por eso, el lector de este informe erraría si asumiera que la razón por la que en él se omite la cuestión de quién es el experto o cómo hace su labor en un proceso judicial radica en que estos extremos no son relevantes: lo son, sin duda, pero en un nivel distinto y mucho más limitado al que habitualmente se le atribuye en nuestros sistemas.

Este informe del PCAST es claramente ejemplo de un esfuerzo institucional por indagar sobre la validez y la fiabilidad de determinados métodos, técnicas, etc., que utilizan los peritos. Fue precedido por otro, conocido como NAS Report, realizado por la National Academy of Science de Estados Unidos, titulado «Strengthening Forensic Science in the United States: A Path Forward» (2009)<sup>12</sup>, donde se pusieron sobre la mesa un conjunto importante de serias debilidades de las ciencias forenses en Estados Unidos. Ambos informes fueron realizados en la administración de Barack Obama, que con ello mostró gran sensibilidad sobre el tema, quizá debido a un par de acontecimientos que impulsaron fuertemente su presencia en la agenda pública: escándalos por malas prácticas en varios laboratorios forenses oficiales y la identificación de un amplio conjunto de condenas erróneas basadas en pruebas periciales poco

---

<sup>11</sup> Evidentemente, no basta con que un área de conocimiento desarrolle métodos y técnicas válidos y fiables, los Estados deben invertir en tener todas las condiciones adecuadas para que esos métodos y técnicas puedan ser aplicados de manera fiable a los casos que se deben resolver. Así, por ejemplo, de poco sirve a un sistema de justicia que las pruebas de ADN realizadas en ciertas condiciones puedan alcanzar una fiabilidad del 99.999%, si ese Estado no tiene de hecho esas condiciones.

<sup>12</sup> Puede consultarse en: <https://www.ojp.gov/pdffiles1/nij/grants/228091.pdf>

sólidas<sup>13</sup>. Es de resaltar que en lugar de esconder el problema o dejarlo en el olvido mediante el paso del tiempo, se transparentaron e hicieron públicos varios problemas reales de las pruebas periciales provenientes de las ciencias forenses. Esa transparencia tuvo como efecto que varias de las ciencias forenses que salieron mal informadas en el 2009 mejoraran mucho mediante el trabajo arduo y conjunto de las comunidades expertas, entre ellas, la identificación de huellas dactilares. Este es claramente un rasgo de la ciencia, identificar los errores, aprender de ellos e intentar resolverlos.

Muchos de nuestros sistemas han dado el paso de copiar los criterios *Daubert*. Sin embargo, no se han hecho políticas públicas dirigidas a detectar las debilidades que pudieran tener los métodos y técnicas que emplean los peritos en un proceso judicial, si se están empleando solo aquellos que son válidos y, en todo caso, cuál es su nivel de fiabilidad. La protección de los derechos subjetivos en un Estado de Derecho exige evitar los errores en las decisiones judiciales. Para hacerlo, no basta con elaborar buenas leyes de procedimiento: es necesario también fomentar la investigación sobre la validez y la fiabilidad de los métodos y técnicas que se usan por parte de las ciencias forenses. Una vez detectadas sus debilidades, es imprescindible, por una parte, promover la mejora de esas disciplinas y, por otra, informar de esas debilidades a los operadores jurídicos para que no tomen decisiones basadas en pruebas que carecen de la validez o fiabilidad requeridas o que no las sobrevaloren.

Por eso, estimo de la mayor importancia que el contenido de este informe del PCAST sea conocido por todos los operadores de nuestros sistemas de justicia. Ojalá pronto veamos un estudio así en nuestras latitudes. Mientras tanto, la esperanza es que la traducción de este informe sirva para concienciar sobre esa necesidad, la de que nuestros sistemas de justicia usen solo métodos y técnicas que hayan mostrado funcionar adecuadamente.

## BIBLIOGRAFÍA

- ABEL LLUCH, X. (2017): «Criterios orientadores de la valoración de la prueba pericial», en PICÓ I JUNOY, J.: *Peritaje y prueba pericial*, Barcelona: Bosch Editor.
- DUCE, M. (2020): «Prácticas probatorias y riesgos de condenas erróneas: una visión empírica», en FERRER BELTRÁN, J. y VÁZQUEZ, C., *El razonamiento probatorio en el proceso judicial. Un encuentro entre diferentes tradiciones*, Madrid – Barcelona: Marcial Pons.
- EDMOND, G. (2020a): «Cuando el derecho es poco fiable. Respuestas jurídicas a la prueba de huellas dactilares latente. Parte I», en *Quaestio Facti. Revista Internacional sobre razonamiento probatorio*, no. 1.
- (2020b): «Cuando el derecho es poco fiable. Respuestas jurídicas a la prueba de huellas dactilares latente. Parte II», en Ferrer Beltrán, J. y Vázquez, C., *El razonamiento probatorio en el proceso judicial. Un encuentro entre diferentes tradiciones*, Madrid – Barcelona: Marcial Pons.
- HAACK, S. (2020): *Filosofía del derecho y de la prueba*, Madrid – Barcelona: Marcial Pons.

<sup>13</sup> Sobre el rol de las pruebas periciales en las decisiones judiciales erróneas, fundamentalmente en las condenas erradas, véase DUCE (2020).

- MNOOKIN, J.; COLE, S.; DROR, I.; FISHER, B.; HOUCK, M.; INMAN, K.; KAYE, D.; KOEHLER, J.; LANGENBURG, G.; RISINGER, M.; RUDIN, N.; SIEGEL, J.; y STONEY, D., (2011): «The Need for a Research Culture in the Forensic Sciences», en *UCLA Law Review*, no. 58.
- MOCHÁN, L., (2010): «Magia, Ciencia, Salud y Seguridad Nacional», *Academia de Ciencias de Morelos A.C.*, lunes 6 de septiembre de 2010.
- NIEVA FENOLL, J. (2010): *La valoración de la prueba pericial*, Madrid – Barcelona: Marcial Pons.
- VÁZQUEZ, C. (2015): *De la prueba científica a la prueba pericial*, Madrid – Barcelona: Marcial Pons.

**INFORME AL PRESIDENTE.**

**CIENCIA FORENSE  
EN LOS TRIBUNALES PENALES:  
ASEGURANDO LA VALIDEZ CIENTÍFICA  
DE LOS MÉTODOS FORENSES BASADOS  
EN COMPARACIÓN DE CARACTERÍSTICAS**

**OFICINA EJECUTIVA DEL PRESIDENTE  
CONSEJO DE ASESORES DEL PRESIDENTE  
EN CIENCIA Y TECNOLOGÍA**

**SEPTIEMBRE DE 2016**

**TRADUCCIÓN DE JOSÉ JUAN LUCENA MOLINA**

**SUMARIO:** I. SOBRE EL CONSEJO DE ASESORES DEL PRESIDENTE EN CIENCIA Y TECNOLOGÍA.—II. EL CONSEJO DE ASESORES DEL PRESIDENTE EN CIENCIA Y TECNOLOGÍA.—III. GRUPO DE TRABAJO DEL PCAST.—IV. ASESORES SENIOR.—V. CARTA AL PRESIDENTE OBAMA.—VI. RESUMEN EJECUTIVO: 1. Trabajos previos sobre validez científica de disciplinas de ciencia forense. 2. El contexto jurídico. 3. Criterios científicos para la validez y la fiabilidad de los métodos forenses de comparación de características. 4. Evaluación de la validez científica de siete métodos de comparación de características. 5. Observaciones finales de las siete evaluaciones.—VII. INFORME: 1. Introducción. 2. Trabajo previo sobre validación de métodos de ciencia forense. 3. El papel de la validez científica en los tribunales. 4. Criterios científicos para la validez y fiabilidad de los métodos forenses de comparación de características. 5. Evaluación de la validez científica de siete métodos de comparación de características. 6. Acciones para asegurar la validez científica en la ciencia forense: recomendaciones al NIST y a la OSTP. 7. Acciones para garantizar la validez científica en la ciencia forense: recomendaciones para el Laboratorio del FBI. 8. Acciones para garantizar la validez científica en la ciencia forense: recomendaciones al Fiscal General. 9. Acciones para garantizar la validez científica en la ciencia forense: recomendaciones al poder judicial. 10. Hallazgos científicos.—APÉNDICE A: CUESTIONES ESTADÍSTICAS: 1. Sensibilidad y tasa de falsos positivos; 2. Intervalos de confianza; 3. Calculando resultados para pruebas concluyentes; 4. Análisis bayesiano.—APÉNDICE B: EXPERTOS ADICIONALES QUE REALIZAN APORTACIONES.—BIBLIOGRAFÍA: 1. Informes; 2. Libros/Artículos.—FRECUENTES SIGLAS EN INGLÉS TRADUCIDAS AL ESPAÑOL.

## I. SOBRE EL CONSEJO DE ASESORES DEL PRESIDENTE EN CIENCIA Y TECNOLOGÍA

El Consejo de Asesores del Presidente en Ciencia y Tecnología (PCAST, por sus siglas en inglés) es un grupo consultivo de prestigiosos científicos e ingenieros de la nación, elegidos por el Presidente para mejorar el asesoramiento científico y tecnológico disponible a su persona tanto desde la Casa Blanca como desde los Departamentos del Gabinete y de otras Agencias federales. PCAST es consultado acerca de, y frecuentemente realiza recomendaciones sobre, una amplia variedad de asuntos en los que la comprensión de aspectos relacionados con la ciencia, la tecnología y la innovación poseen potencial influencia en las decisiones de carácter político que conciernen al Presidente.

Para más información sobre PCAST, consúltese [www.whitehouse.gov/ostp/pcast](http://www.whitehouse.gov/ostp/pcast).

## II. EL CONSEJO DE ASESORES DEL PRESIDENTE EN CIENCIA Y TECNOLOGÍA

### COPRESIDENTES

John P. Holdren: *Asistente del Presidente para la Ciencia y la Tecnología. Director, Oficina de Política en Ciencia y Tecnología*

Eric S. Lander: *Presidente. Instituto Broad de Harvard y del MIT*

## VICEPRESIDENTES

William Press: *Cátedra Raymer en Ciencia Computacional y Biología Integrativa. Universidad de Texas en Austin*

Maxine Savitz: *Honeywell (ret.)*

## MIEMBROS

Wanda M. Austin: *Presidente y CEO. Corporación Aerospace*

Rosina Bierbaum: *Catedrática, Facultad de Recursos Naturales y Medioambiente, Universidad de Michigan. Cátedra Roy F. Westin en Economía Natural, Facultad de Política Pública, Universidad de Maryland*

Christine Cassel: *Decano de Planificación. Facultad de Medicina Kaiser Permanente*

Christopher Chyba: *Catedrático, Ciencias Astrofísicas y Asuntos Internacionales. Universidad de Princeton*

S. James Gates Jr.: *Cátedra John S. Toll de Física. Director, Centro para la Teoría de Cuerdas y Partículas. Universidad de Maryland, College Park*

Mark Gorenberg; *Miembro Directivo. Zeta Venture Asociados*

Susan L. Graham: *Catedrática Distinguida Emérita en Ingeniería Eléctrica y Ciencias de la Computación. Universidad de California, Berkeley*

Michael McQuade: *Vicepresidente senior para la Ciencia y la Tecnología. Corporación United Technologies*

Chad Mirkin: *Cátedra de Química George B. Rathmann. Director, Instituto Internacional para la Nanotecnología. Northwestern University*

Mario Molina: *Catedrático Distinguido, Química y Bioquímica. Universidad de California, San Diego. Catedrático, Centro para las Ciencias Atmosféricas. Institución Scripps de Oceanografía*

Craig Mundie: *Presidente. Mundie Asociados*

Ed Penhoet: *Director. Alta Asociados. Catedrático Emérito, Bioquímica y Salud Pública. Universidad de California, Berkeley*

Barbara Schaal: *Decana de la Facultad de Humanidades y Ciencias. Catedrática Distinguida Mary-Dell Chilton de Biología. Universidad de San Luis en Washington*

Eric Schmidt: *Director Ejecutivo. Alphabet, Inc.*

Daniel Schrag: *Cátedra Sturgis Hooper de Geología. Catedrático, Ciencia e Ingeniería Ambiental. Director, Centro de la Universidad Harvard para el Medioambiente. Universidad de Harvard*

## STAFF

Ashley Predith; *Director Ejecutivo*

Diana E. Pankevich; *Becaria de Políticas de Ciencia y Tecnología de la AAAS*

Jennifer J. Michael; *Especialista de Apoyo al Programa*

## III. GRUPO DE TRABAJO DEL PCAST

Los miembros del Grupo de Trabajo participaron en la preparación de este informe. Todos los miembros del PCAST lo revisaron y aprobaron.

## GRUPO DE TRABAJO

Eric S. Lander (Director del Grupo de Trabajo): *Presidente. Instituto Broad de Harvard y el MIT*

S. James Gates Jr.: *Cátedra John S. Toll de Física. Director, Centro para la Teoría de Cuerdas y Partículas. Universidad de Maryland, College Park*

Susan L. Graham: *Catedrática Distinguida Emérita en Ingeniería Eléctrica y Ciencias de la Computación. Universidad de California, Berkeley*

Michael McQuade; *Vicepresidente senior para la Ciencia y la Tecnología. Corporación United Technologies*

William Press: *Cátedra Raymer en Ciencia Computacional y Biología Integrativa. Universidad de Texas en Austin*

Daniel Schrag: *Cátedra Sturgis Hooper de Geología. Catedrático, Ciencia e Ingeniería Ambiental. Director, Centro de la Universidad Harvard para el Medioambiente. Universidad de Harvard*

## STAFF

Diana E. Pankevich: *Becaria de Políticas de Ciencia y Tecnología de la AAAS*

Kristen Zarrelli: *Consultor, Política Pública & Proyectos Especiales. Instituto Broad de Harvard y del MIT*

## EDITORIA

Tania Simoncelli: *Consejera senior del Director. Instituto Broad de Harvard y del MI*

## IV. ASESORES SENIOR

El PCAST consultó a un grupo de expertos jurídicos para que proporcionaran orientación sobre aspectos prácticos relacionados con la interacción entre la ciencia

y el derecho. El PCAST también buscó orientación y la aportación de dos estadísticos con experiencia en este ámbito. Se dio a los consejeros senior la oportunidad de revisar los primeros borradores para asegurar la exactitud fáctica. El PCAST expresa su gratitud a los que aquí se listan. Su consentimiento para participar con el PCAST en puntos específicos no implica que respalden los puntos de vista expresados en este informe. La responsabilidad de las opiniones, hallazgos y recomendaciones en este informe, así como cualquier error de hecho o de interpretación, descansa únicamente sobre el PCAST.

#### COPRESIDENTES DE ASESORES SENIOR

Honorable Harry T. Edwards: *Juez. Corte de Apelaciones de los Estados Unidos. Distrito del Circuito de Columbia*

Jennifer L. Mnookin: *Decana, Cátedra de Derecho David G. Price y Dallas P. Price. Universidad de California, Facultad de Derecho de Los Ángeles*

#### ASESORES SENIOR

Honorable James E. Boasberg: *Juez de Distrito. Corte de Distrito de los Estados Unidos. Distrito de Columbia*

Honorable Andre M. Davis: *Magistrado Emérito. Corte de Apelaciones de los Estados Unidos. Cuarto Circuito*

David L. Faigman: *Rector interino y Decano. Facultad de Derecho Hastings de la Universidad de California*

Stephen Fienberg: *Catedrático de Estadística y Ciencias Sociales de la Universidad Maurice Falk (Emérito). Universidad Carnegie Mellon*

Honorable Pamela Harris: *Magistrada. Corte de Apelaciones de los Estados Unidos. Cuarto Circuito*

Karen Kafadar: *Rector interino y Decano. Facultad de Derecho Hastings de la Universidad de California*

Honorable Alex Kozinszi: *Magistrado Emérito. Corte de Apelaciones de los Estados Unidos. Noveno Distrito*

Honorable Cornelia T.L. Pillard: *Magistrada. Corte de Apelaciones de los Estados Unidos. Distrito del Circuito de Columbia*

Honorable Charles Fried: *Catedrático de Derecho. Facultad de Derecho de Harvard. Universidad de Harvard*

Honorable Nancy Gertner: *Profesora Titular de Derecho. Facultad de Derecho de Harvard. Universidad de Harvard*

Honorable Jed S. Rakoff: *Juez de Distrito. Corte de Distrito de los Estados Unidos. Distrito Sur de Nueva York*

Honorable Patti B. Saris: *Magistrada Presidente. Corte de Distrito de los Estados Unidos. Distrito de Massachusetts*

## V. CARTA AL PRESIDENTE OBAMA

OFICINA EJECUTIVA DEL PRESIDENTE  
CONSEJO DE ASESORES DEL PRESIDENTE  
EN CIENCIA Y TECNOLOGÍA  
WASHINGTON, D.C. 20502

Presidente Barack Obama  
La Casa Blanca  
Washington, DC 20502

Estimado Sr. Presidente:

Nos complace enviarle este informe del PCAST sobre *Ciencia forense en los Tribunales Penales de Justicia: Asegurando la validez científica de los métodos basados en comparaciones de características*. El estudio que condujo a este informe fue una respuesta a una pregunta suya al PCAST, en 2015, acerca de si hay pasos adicionales desde la vertiente científica que pudieran asegurar la validez de la prueba forense en el sistema jurídico nacional, más allá de los llevados a cabo por la Administración tras el informe extremadamente crítico del Consejo de Investigación Nacional del año 2009 sobre el estado de las ciencias forenses.

El PCAST concluye que hay dos importantes carencias: (1) la necesidad de clarificar los estándares científicos sobre la validez y fiabilidad de los métodos forenses y (2) la necesidad de evaluar métodos forenses específicos para determinar si han sido científicamente declarados como válidos y fiables. Nuestro estudio tiene como objetivo ayudar a resolver estas carencias para una serie de métodos forenses basados en «comparación de características» —específicamente, métodos de comparación de muestras de ADN, mordeduras, impresiones de huellas latentes, señales balísticas, huellas de calzado y cabello—.

Nuestro estudio, que incluye una extensa revisión de la literatura científica, tiene en cuenta también contribuciones de investigadores forenses del FBI (Oficina Federal de Investigación) y del NIST (Instituto Nacional de Estándares y Tecnología), así como de muchos otros científicos y analistas forenses, jueces, fiscales, abogados defensores, investigadores académicos, defensores de la reforma de la justicia penal y representantes de las Agencias federales. Los resultados y recomendaciones transmitidos en este informe son, desde luego, únicamente del PCAST.

Nuestro informe revisa estudios previos relacionados con la práctica forense y las acciones federales puestas recientemente en marcha para fortalecer la ciencia forense,

discute el papel de la validez científica dentro del sistema jurídico, explica los criterios que permiten juzgar la validez científica de los métodos forenses de comparación de características y aplica estos criterios a los métodos de comparación anteriormente mencionados.

Basándonos en nuestros resultados sobre la «validez de los fundamentos» de los métodos indicados, así como sobre la «validez de su aplicación» en la práctica ante los Tribunales, ofrecemos recomendaciones sobre las acciones que pudieran llevar a cabo el NIST, la Oficina sobre Política de Ciencia y Tecnología, y el Laboratorio del FBI, para fortalecer los fundamentos científicos de las disciplinas forenses, así como las acciones que pudieran realizar el Fiscal General y el Poder Judicial para promover el uso más riguroso de esas disciplinas en los Tribunales.

Atentamente,

John P. Holdren y Eric S. Lander, copresidentes.

## VI. RESUMEN EJECUTIVO

«La ciencia forense» se ha definido como la aplicación de prácticas científicas o técnicas al reconocimiento, recogida, análisis e interpretación de evidencia relacionada con el derecho penal, civil o de cuestiones reglamentarias. Los acontecimientos de las últimas dos décadas —incluyendo: (i) exoneraciones de acusados que habían sido erróneamente condenados basados, en parte, en la prueba científico-forense, (ii) una serie de estudios sobre la base científica de las disciplinas forenses, (iii) revisiones de testimonios de analistas\* basados en indicios forenses, y (iv) escándalos en laboratorios de criminalística estatales— han llamado cada vez más la atención sobre la validez y la fiabilidad de algunas importantes formas de prueba forense y el testimonio fundamentado en ellas<sup>1</sup>.

---

\* N. del T.: una de las primeras dificultades en la traducción está relacionada con lo que en el ámbito forense profesional español se entiende por «experto». Como ocurre en otros muchos países, quienes prestan testimonio experto de carácter forense en procesos penales como peritos pueden estar vinculados o no a laboratorios oficiales. En países como España, en los que rige un sistema procesal-penal continental, los «expertos» suelen ser peritos oficiales. Sin embargo, la formación académica de esos «expertos» es muy dispar y depende de la institución a la que pertenecen. Respecto a los cinco grandes laboratorios forenses españoles pertenecientes a la red ENFSI (European Network of Forensic Science Institute), puede consultarse el Informe del Ministerio del Interior Español (2018). Debido a esa disparidad formativa, se ha preferido utilizar el término «analista» en la traducción, también utilizado en el ámbito cultural y jurídico anglosajón. Ese término intenta independizar al perito de su titulación académica, suponiéndole siempre capacitado para realizar las pericias de las que se responsabiliza oficialmente. En el ámbito cultural y jurídico mencionado se distingue entre *practitioners, experts, academics*, etc., de forma que esos términos se diferencian entre sí, fundamentalmente, por la preparación académica de los que ejercen la pericia ante los tribunales. Por todo ello, usaré pues el término «analista».

<sup>1</sup> Las citas de la bibliografía que sustentan aspectos mencionados en el Resumen Ejecutivo se encuentran en el cuerpo principal del informe.

Un estudio plurianual del *National Research Council* (Consejo de Investigación Nacional)<sup>2</sup> solicitado por el Congreso sobre este asunto, *Strengthening Forensic Science in the United States: A Path Forward* (Fortaleciendo la ciencia forense en los Estados Unidos: un paso hacia delante), y entregado en el año 2009, fue particularmente crítico en señalar la debilidad en la fundamentación científica de una serie de disciplinas utilizadas de manera habitual en el sistema judicial penal. Ese informe condujo a una amplia discusión sobre el camino a seguir dentro y fuera del gobierno federal y finalmente llevó al establecimiento de dos grupos: la *National Commission on Forensic Science* (Comisión Nacional sobre Ciencia Forense), integrada en el *Department of Justice* (Departamento de Justicia), y la *Organization for Scientific Area Committees for Forensic Science* (Organización de Comités de Áreas Científicas para la Ciencia Forense), del *National Institute of Standards and Technology* (Instituto Nacional de Estándares y Tecnología).

Tras la pregunta del Presidente Obama al *President's Council of Advisors on Science and Technology* (Consejo de Asesores del Presidente en Ciencia y Tecnología) (PCAST), en el año 2015, sobre la posibilidad de dar pasos adicionales que pudieran ser útiles desde el punto de vista científico para fortalecer las disciplinas forenses y así asegurar la validez de la prueba científica utilizada en el sistema jurídico nacional, el PCAST concluye que hay dos carencias importantes: (1) la necesidad de clarificar los estándares científicos sobre la validez y la fiabilidad de los métodos forenses y (2) la necesidad de evaluar métodos forenses específicos para determinar si han sido científicamente declarados como válidos y fiables.

Este informe tiene como objetivo ayudar a resolver estas carencias para el caso de métodos forenses basados en la «comparación de características» —es decir, métodos que intentan determinar si una muestra indiciaria (*e.g.* de la escena del crimen) está o no asociada con una muestra de una potencial «fuente» (*e.g.* de un sospechoso), fundamentándose en la presencia de patrones, impresiones u otras características que sean similares en la muestra y en la fuente—. Ejemplos de tales métodos son el análisis de ADN, cabellos, huellas latentes, armas y municiones disparadas, marcas de herramientas y marcas de mordeduras, huellas de calzado y de neumáticos y escritura manuscrita.

En el transcurso de su estudio, el PCAST ha recopilado y revisado más de 2000 artículos de diversas fuentes —incluyendo bibliografía preparada por el Subcomité sobre Ciencia Forense del *National Science and Technology Council* (Consejo de Ciencia y Tecnología Nacional) y por los Grupos de Trabajo relevantes organizados por el Instituto Nacional de Estándares y Tecnología (NIST); la información proporcionada por la comunidad interesada en la ciencia forense en respuesta a preguntas realizadas por PCAST; y las búsquedas en literatura propia del PCAST—.

---

<sup>2</sup> El National Research Council (Consejo de Investigación Nacional) es el órgano director de los estudios de las *National Academies of Science, Engineering, and Medicine* (Academias Nacionales de Ciencias, Ingeniería y Medicina).

Para informarse bien sobre problemas reales en la interacción entre la ciencia y el derecho, el PCAST ha consultado a un grupo de consejeros *senior* compuesto por nueve jueces federales (que ejercen o han ejercido como tales), un anterior procurador general de los Estados Unidos, un anterior juez de una Corte Suprema estatal, dos decanos de facultades de derecho y dos distinguidos estadísticos especializados en esta materia. Se obtuvo información adicional del Laboratorio del *Federal Bureau of Investigation* (Oficina Federal de Investigación) (FBI) y de científicos del NIST, así como de otros muchos científicos y prácticos forenses, jueces, fiscales, abogados defensores, académicos, defensores de la reforma de la justicia penal y representantes de agencias federales. La disposición de estos grupos e individuos para colaborar con el PCAST no implica que aprueben las opiniones expresadas en el informe. Las conclusiones y recomendaciones que se transmiten en el informe son de exclusiva responsabilidad del PCAST.

El informe resultante —resumido aquí sin las extensas elaboraciones técnicas y densas citas del texto principal que sigue a continuación— comienza con una revisión de estudios previos relacionados con la práctica forense y de las acciones federales que actualmente están en marcha para fortalecer la ciencia forense; reflexiona sobre el papel de la validez científica dentro del sistema jurídico; explica los criterios necesarios para valorar la validez científica de los métodos forenses basados en comparación de características; aplica estos criterios a seis de tales métodos en detalle y revisa una evaluación de un séptimo método que fue realizada por otros; finalmente, ofrece recomendaciones sobre acciones federales que pudieran adoptarse para fortalecer la ciencia forense y promover su más riguroso uso en los tribunales.

Creemos que las conclusiones y recomendaciones serán útiles tanto para el poder judicial como para aquellos que trabajan para fortalecer la ciencia forense.

## 1. Trabajos previos sobre validez científica de disciplinas de ciencia forense

Irónicamente, fue la aparición y proceso de madurez de una *nueva* ciencia forense, el análisis de ADN, en los años 90, el que condujo al cuestionamiento serio de la validez de la mayoría de las disciplinas forenses tradicionales. Cuando se introdujo por primera vez la prueba de ADN en los tribunales, a finales de los 80, se consideró inicialmente como una técnica infalible; pero los métodos utilizados en los primeros casos acabaron calificándose como poco fiables: los laboratorios carecían de procedimientos validados y consistentemente aplicados para definir los patrones de ADN de las muestras, para establecer cuándo dos patrones eran coincidentes dentro de una determinada tolerancia y para determinar la probabilidad de que tales coincidencias pudieran producirse por casualidad entre la población. Cuando a consecuencia de una decisión judicial sobre un caso acaecido en Nueva York, la prueba de ADN fue declarada inadmisibile en el año 1989, los científicos involucrados en el análisis

de ADN, tanto en aplicaciones forenses como no forenses, aunaron esfuerzos para promover el desarrollo de principios y métodos fiables, lo que ha permitido que el análisis de ADN de muestras procedentes de un único individuo se califique como el «estándar de oro» de la ciencia forense tanto para la investigación como para el enjuiciamiento.

Una vez que el ADN llegó a ser una metodología fiable, el poder de la tecnología —que incluye su capacidad para analizar pequeñas muestras y distinguir entre individuos— hizo posible no sólo identificar y condenar a los verdaderos culpables, sino también contrarrestar falsas acusaciones antes del procesamiento y reexaminar condenas del pasado. Informes del *National Institute of Justice* (Instituto Nacional de Justicia) y otros institutos afirman que los tests de ADN durante el curso de las investigaciones han descartado a decenas de miles de sospechosos y que la revisión de condenas de casos cerrados mediante tests de ADN ha conducido a la exoneración de hasta 342 acusados. Informes independientes sobre esos casos han revelado que muchos se produjeron, en parte, por deficientes testimonios expertos de científicos forenses que comunicaron incorrectamente a los miembros del jurado que el hecho de encontrar similares características en un par de muestras (cabellos, muestras balísticas, marcas de mordedura, huellas de neumáticos o de calzado, u otros vestigios) tomadas, una del sospechoso y otra de la escena del crimen, implicaba al acusado en la comisión del crimen con un alto grado de certeza.

Los interrogantes que el análisis de ADN había suscitado sobre la validez científica de las disciplinas forenses tradicionales y el testimonio basado en ellas condujeron, naturalmente, a un incremento de los esfuerzos para comprobar empíricamente los métodos que tales disciplinas utilizaban. Los estudios relevantes que se produjeron son los siguientes:

— Un reexamen del FBI en 2002 de las comparaciones microscópicas de cabello que científicos de la agencia habían realizado en casos penales, en el que las pruebas de ADN revelaron que el 11% de las muestras de cabello que se encontraron con coincidencia microscópica en realidad provenían de individuos diferentes;

— Un informe realizado por el *National Research Council* (Consejo de Investigación Nacional) en 2004, a petición del FBI, sobre la prueba balística basada en la naturaleza del plomo, reveló que no existía ni suficiente investigación ni datos que permitieran concluir que hubiera una conexión definitiva entre dos proyectiles basándose en la similitud de la composición del plomo que contenían;

— Un informe del año 2005 realizado por un comité internacional, a solicitud del FBI, para revisar el uso de una prueba de una huella latente en el caso de un ataque terrorista con explosivos en España, constató que el «sesgo de confirmación» —es decir, la inclinación a confirmar una sospecha basada en otros fundamentos— contribuyó a una falsa identificación y a una detención incorrecta; y

— Estudios realizados en los años 2009 y 2010 sobre la prueba de mordedura, constataron que los actuales procedimientos para comparar marcas de mordedura no

eran capaces de excluir o incluir de modo fiable a un sospechoso como causante potencial de la mordedura.

Más allá de esta clase de deficiencias con respecto a «métodos fiables» en disciplinas forenses basadas en comparación de características, existen informes que concluyen que los testimonios expertos han sobrevalorado frecuentemente el valor probatorio de sus pruebas, yendo más allá de lo que la ciencia relevante pudiera justificar. A veces, los analistas han testificado que sus conclusiones tienen «una certeza del 100%»; o tienen una tasa de error igual a «cero», «esencialmente cero», o «insignificante». Como muchos estudios han resaltado —incluyendo el muy apreciado informe del Consejo de Investigación Nacional del año 2009—, sin embargo, tales expresiones no son científicamente defendibles: ninguno de los ensayos de los laboratorios y de los análisis de comparación de características tiene una tasa de error igual a cero.

Comenzando en el año 2012, el Departamento de Justicia (DOJ) y el FBI llevaron a cabo un estudio sin precedentes de las pruebas periciales en más de 3000 casos penales con análisis microscópico de cabellos. Los resultados iniciales, comunicados en el año 2015, mostraron que los analistas del FBI habían ofrecido un testimonio científicamente inválido en más del 95% de los casos en los que se utilizó para inculpar al acusado en el juicio oral. En marzo de 2016, el Departamento de Justicia anunció su intención de expandir a otros métodos de la ciencia forense la revisión de los testimonios forenses realizados por el laboratorio del FBI en casos penales cerrados. Esta revisión ayudará a valorar hasta qué punto se han dado similares testimonios sobreestimados en otras disciplinas forenses.

El informe del Consejo de Investigación Nacional del año 2009 es hasta la fecha la revisión más completa de las ciencias forenses en los Estados Unidos. El informe dejó claro que algunos tipos de problemas, irregularidades y errores judiciales, no pueden atribuirse simplemente a un puñado de analistas malintencionados o a laboratorios con bajo rendimiento, sino que son sistemáticos y generalizados —resultado de factores que tienen un alto nivel de fragmentación (incluyendo exigencias de entrenamiento y de formación, así como recursos y capacidades de los laboratorios dispares y frecuentemente inadecuados), carencia de estandarización de las disciplinas, insuficiente educación e investigación de alta calidad y escasez de estudios revisados por pares que establezcan la base científica y la validez de muchos métodos forenses rutinariamente utilizados—.

El informe del año 2009 resaltó que las deficiencias en las ciencias forenses eran especialmente prevalentes en las disciplinas de comparación de características, muchas de las cuales, decía el informe, carecían de sistemas bien definidos para determinar las tasas de error y en las que no se habían realizado estudios para establecer la unicidad, o la rareza o habitualidad de la frecuencia relativa de las marcas o características particulares examinadas. Además, las pruebas para evaluar la aptitud de los analistas, en los lugares en los que han sido realizadas, mostraron ejemplos de bajo rendimiento de analistas específicos. En resumen, el informe concluía que «muchas pruebas forenses —incluyendo, por ejemplo, las identificaciones por mordedura, por análisis de armas

de fuego y de marcas de herramientas— son introducidas en los procesos penales sin una significativa validación científica, una determinación de las tasas de error o una comprobación de su fiabilidad para explicar los límites de la disciplina.»

## 2. El contexto jurídico

Históricamente, la ciencia forense ha sido primordialmente utilizada en dos fases del procedimiento penal: (1) en la *investigación*, que busca identificar al presunto delincuente y (2) en la *acusación*, que busca probar la culpabilidad del acusado más allá de toda duda razonable. En los últimos años, la ciencia forense —particularmente el análisis de ADN— ha sido ampliamente utilizada para desafiar condenas pasadas.

Resulta importante ser conscientes de que las fases de investigación y procesamiento exigen diferentes estándares para el uso de la ciencia forense y de otras herramientas de investigación. En la investigación, las perspectivas y la información pueden proceder tanto de una ciencia bien establecida como de enfoques exploratorios. En la fase de procesamiento, la ciencia forense debe satisfacer un estándar más alto. Concretamente, las *Federal Rules of Evidence* (Reglas Federales de la Prueba) (Regla 702[c, d]) exigen para la admisibilidad del testimonio experto que esté fundamentado, entre otras cosas, sobre «principios y métodos fiables» que hayan sido «aplicados con fiabilidad» a los hechos del caso. Y la Corte Suprema ha establecido que los jueces deben determinar «si el razonamiento o la metodología subyacente al testimonio son científicamente válidos».

Aquí es donde los estándares jurídicos y los científicos se cruzan. Las decisiones de los jueces sobre la admisibilidad de la prueba científica descansan únicamente sobre estándares *jurídicos*; son un área exclusiva de los tribunales y el PCAST no opina sobre ellos. Pero esas decisiones requieren realizar determinaciones sobre la validez científica. Es facultad de la comunidad científica proporcionar una guía sobre los estándares científicos respecto a la validez científica y es sobre esos estándares *científicos* sobre los que se centra aquí el PCAST.

Distinguimos dos tipos de validez científica: la validez de los fundamentos y la validez en la aplicación.

(1) La *validez de los fundamentos* de un método científico-forense requiere que se haya demostrado, basándose en estudios empíricos, que es *repetible, reproducible y preciso* en los niveles que se hayan medido, y que sean apropiados para la aplicación prevista. La validez de los fundamentos, entonces, significa que un método puede, en principio, considerarse fiable. Pretendemos que el concepto *científico* se corresponda con el requisito jurídico previsto en la Regla 702(c), que alude a «principios y métodos fiables».

(2) La *validez en la aplicación* significa que el método ha sido fiablemente aplicado en *la práctica*. Pretendemos que el *concepto científico* pensamos se corresponda

con el requisito jurídico previsto en la Regla 702(d), que exige que un analista «haya aplicado fiablemente los principios y métodos a los hechos del caso».

### 3. Criterios científicos para la validez y la fiabilidad de los métodos forenses de comparación de características

La sección 4 del informe principal proporciona una detallada descripción de los criterios científicos para establecer la validez de los fundamentos y la fiabilidad de los métodos forenses de comparación de características, incluyendo tanto los métodos objetivos como los subjetivos<sup>3</sup>.

Los métodos subjetivos requieren un escrutinio particularmente cuidadoso porque su fuerte dependencia de los juicios humanos conlleva que sean especialmente vulnerables a los errores humanos, a inconsistencias entre analistas y a sesgos cognitivos. En las disciplinas de comparación de características forenses, los sesgos cognitivos incluyen los fenómenos que, en ciertos entornos, los seres humanos podemos tender a fijarnos de forma natural en las similitudes entre las muestras y descartar sus diferencias, así como también podemos ser influenciados por información extraña al área de conocimiento y por presiones externas sobre el caso.

Las ideas esenciales sobre la validez de los fundamentos son las siguientes:

(1) La validez de los fundamentos requiere que un método haya estado sujeto a estudios *empíricos* realizados por múltiples grupos, en condiciones apropiadas para el uso pretendido. Los estudios deben (a) demostrar que el método es repetible y reproducible; y (b) proporcionar estimaciones válidas sobre la precisión del método (es decir, cuántas veces el método alcanza una conclusión incorrecta) que indiquen que es apropiado para la aplicación pretendida.

(2) Para métodos objetivos, la validez de los fundamentos del método puede establecerse estudiando las medidas de precisión, reproducibilidad y consistencia de cada una de sus etapas individuales.

(3) Para métodos subjetivos de comparación de características, dado que las etapas individuales no están objetivamente especificadas, el método debe evaluarse como si fuera una «caja negra» para la mente del analista. Las evaluaciones de validez y fiabilidad han de basarse en «estudios de caja negra», en los que muchos analistas

---

<sup>3</sup> Los métodos de comparación de características pueden clasificarse en métodos objetivos y subjetivos. Por métodos de comparación de características objetivos entendemos aquellos métodos consistentes en procedimientos definidos con suficiente detalle en cuanto a su estandarización y cuantificación de forma que puedan ser realizados por un sistema automatizado o por expertos humanos que no emitan juicios o apenas lo hagan. Por métodos subjetivos entendemos aquellos métodos que incluyen procedimientos esenciales que conllevan un significativo juicio humano —por ejemplo, sobre las características que deben seleccionarse dentro de un patrón, o cómo determinar si las características comparadas son lo suficientemente similares como para que pueda decirse que conforman una probable coincidencia—.

toman decisiones sobre una gran variedad de pruebas independientes (generalmente con muestras «dubitadas» y una o más muestras «indubitadas») y en los que se determinan las tasas de error.

(4) Si no existen estimaciones apropiadas de precisión, la aseveración de un analista consistente en que dos muestras son similares —o, incluso, indistinguibles— carece de significado científico: no tiene valor probatorio y, en cambio, sí tiene un considerable potencial para causar un impacto perjudicial.

Una vez que un método ha sido considerado válido en sus fundamentos como resultado de estudios empíricos apropiados, las afirmaciones acerca de la precisión del método y el valor probatorio de las identificaciones propuestas, para que sean válidas, deben fundamentarse en tales estudios empíricos. *Las declaraciones que afirmen o impliquen mayor certeza que la demostrada por la evidencia empírica son científicamente inválidas.* Los analistas forenses deben, por consiguiente, informar sobre los hallazgos de una identificación propuesta con claridad y con las restricciones debidas, explicando en cada caso que el hecho de que dos muestras satisfagan los criterios de un método para proponer un cotejo positivo (*match*) no significa que las muestras procedan de la misma fuente. Por ejemplo, si la tasa de falsos positivos de un método es de 1 en 50 comparaciones, los analistas no deben dar a entender que el método es capaz de producir resultados con una precisión mayor.

Para que pueda decirse que los criterios científicos de validez son aplicables, han de superarse dos pruebas:

(1) El analista debe haber demostrado que es *capaz* de aplicar de manera fiable el método y debe *realmente* haberlo hecho así. Demostrar que un analista es *capaz* de aplicar fiablemente el método es crucial —especialmente en métodos subjetivos, en los que el juicio humano desempeña un papel central—. Desde el punto de vista científico, la habilidad de aplicar fiablemente un método puede demostrarse únicamente mediante pruebas empíricas que midan cuántas veces el analista ofrece respuestas correctas. La determinación de si *realmente* un analista ha aplicado fiablemente el método, requiere que el procedimiento realmente utilizado en el caso, los resultados obtenidos y las notas de laboratorio, estén disponibles para que pueda llevarse a cabo una revisión científica por terceros.

(2) Las aseveraciones del analista sobre el valor probatorio de las identificaciones propuestas deben ser científicamente válidas. El analista debe informar sobre la tasa de falsos positivos y la sensibilidad del método, establecidas en los estudios sobre la validez de sus fundamentos y debe demostrar que las muestras utilizadas en los estudios sobre los fundamentos sean relevantes para los hechos del caso. Cuando sea aplicable, el analista debe informar sobre el valor probatorio del cotejo positivo (*match*) observado basándose en las características específicas observadas en el caso. Y el analista no debe realizar afirmaciones o implicaciones que vayan más allá de la evidencia empírica y de la aplicación de principios estadísticos válidos al tipo de prueba en juego.

Queremos resaltar, finalmente, que ni la experiencia, ni el buen juicio, ni las buenas prácticas profesionales (tales como programas de certificación y acreditación, protocolos estandarizados, pruebas de aptitud técnica y códigos éticos) pueden sustituir las pruebas genuinas con validez de los fundamentos y su fiabilidad. La frecuencia con la que un patrón o un conjunto de características particulares se observa en diferentes muestras, que es un elemento esencial para inferir las conclusiones, no es un asunto de «juicio». Es un hecho empírico para el cual solo es relevante la evidencia empírica. De forma similar, la expresión de *confianza* utilizada por un analista, basada en su experiencia profesional, o las expresiones sobre un *consenso* entre analistas respecto a la precisión a la que llegan en su campo de trabajo, no pueden sustituir a las tasas de error estimadas a partir de estudios relevantes. Para los métodos forenses de comparación de características, el establecimiento de la validez de sus fundamentos a partir de evidencia empírica es entonces una condición *sine qua non*. Nada puede sustituirla.

#### 4. Evaluación de la validez científica de siete métodos de comparación de características

Para este estudio, el PCAST aplicó los criterios explicados más arriba a seis métodos de comparación de características forenses: (1) análisis de ADN de muestras procedentes de una única fuente y con una mezcla simple, (2) análisis de ADN de muestras con mezclas complejas, (3) marcas de mordedura, (4) huellas latentes, (5) identificación de armas de fuego y (6) análisis de huellas de calzado. Para cada método, la sección 5 del informe principal proporciona una breve explicación de la metodología, analiza estudios e información básicos, proporciona una evaluación sobre la validez científica y ofrece sugerencias de mejora para el futuro. Con respecto a un séptimo método de comparación de características —el análisis de cabellos— no disponemos de una evaluación completa de la validez científica, sino de una revisión de material de soporte recientemente entregado por el Departamento de Justicia para la realización de comentarios. Este resumen ejecutivo proporciona únicamente un breve resumen de algunos hallazgos clave concernientes a estos siete métodos.

##### 4.1. *Análisis de ADN de muestras procedentes de una única fuente y con una mezcla simple*

La amplia mayoría de los análisis de ADN tienen que ver actualmente con muestras procedentes de un único individuo o de una mezcla simple de dos individuos (como el contenido en el llamado «kit de violación»<sup>\*</sup>). El análisis de ADN en tales

---

\* N. del T.: un *rape kit* es un paquete que contiene los instrumentos médicos necesarios para recoger y preservar evidencias físicas cuando se alega una violación sexual.

casos es un método objetivo en el que los protocolos de laboratorio están definidos con precisión y la interpretación conlleva muy poco o nada de juicio humano.

Para evaluar la validez de los fundamentos de un método objetivo, podemos examinar la fiabilidad de cada uno de los pasos individuales en lugar de tener que confiar en estudios de caja negra. En el caso del análisis de ADN de muestras procedentes de una única fuente o de mezclas simples, ha podido demostrarse que cada uno de esos pasos es «repetible, reproducible y preciso» con niveles que han sido medidos y que son «apropiados para la aplicación pretendida» (para citar el requisito de validez de los fundamentos expresado más arriba), y donde la probabilidad de que pueda encontrarse una coincidencia (*match*) por casualidad en la población ha sido directamente estimada a partir de bases de datos genéticas apropiadas y es extremadamente baja.

Con respecto a la validez en la aplicación de los análisis de ADN, al igual que sucede con todos los análisis forenses, no es infalible en la práctica. Puede haber errores y de hecho ocurren. Aunque la probabilidad de que dos muestras procedentes de fuentes distintas tengan el mismo perfil de ADN es pequeña, la posibilidad de un error humano es mucho mayor. Tales errores pueden derivarse de mezcla de muestras, contaminación, interpretación incorrecta y errores en la presentación de los informes.

Para minimizar el error humano, el FBI requiere, como una condición para participar en el *National DNA Index System* (Sistema nacional de base de datos de ADN) que los laboratorios sigan los *FBI's Quality Assurance Standards* (estándares de garantía de calidad del FBI). Estos requieren que el analista cumpla una serie de controles para comprobar una posible contaminación y asegurarse de que la prueba PCR transcurra apropiadamente. Los estándares también requieren evaluaciones semestrales respecto la aptitud de todos los analistas que realicen pruebas de ADN en casos penales. Hemos advertido, no obstante, la necesidad de mejorar las evaluaciones de aptitud.

#### 4.2. *Análisis de ADN de muestras con mezclas complejas*

Algunas investigaciones conllevan realizar análisis de ADN con mezclas complejas de muestras biológicas procedentes de múltiples individuos desconocidos y en proporciones desconocidas (Tales muestras surgen, por ejemplo, de manchas de sangre mezcladas y, cada vez más, de múltiples individuos que tocan una superficie). La diferencia fundamental entre el análisis de ADN de muestras con mezclas complejas y el análisis de ADN de una única fuente y mezcla simple no estriba en el proceso del laboratorio sino en la interpretación del perfil de ADN resultante.

El análisis de ADN de mezclas complejas es intrínsecamente difícil. Tales muestras presentan un perfil de ADN que es la superposición de múltiples perfiles de

ADN individuales. Por muchas razones, la interpretación de un perfil mezcla es diferente y más complicada que la de un perfil simple. Con frecuencia es imposible decir con certeza qué variantes genéticas están presentes en la mezcla o cuántos individuos diferentes contribuyen a ella y mucho menos inferir con precisión el perfil de ADN de cada una de las personas.

Las preguntas que un analista debe hacerse, entonces, son: «¿podría estar el perfil de ADN de un sospechoso presente *dentro* del perfil de la mezcla? Y ¿cuál es la probabilidad de que eso pueda ocurrir por azar?». Dado que muchos perfiles diferentes de ADN pueden encajar dentro de algunos perfiles mezcla, la probabilidad de que un sospechoso «no pueda ser excluido» como posible contribuyente a un perfil mezcla puede ser *mucho más alta* (en algunos casos, millones de veces más alta) que las probabilidades encontradas para perfiles de ADN procedentes de una única fuente.

Las aproximaciones iniciales que se hicieron para la interpretación de mezclas complejas se apoyaron en juicios subjetivos de analistas y cálculos simplificados. Este enfoque es problemático porque las elecciones subjetivas hechas por los analistas pueden afectar drásticamente la respuesta y al valor probatorio estimado al incorporar un significativo riesgo tanto de error analítico como de sesgo de confirmación. El PCAST concluye que el análisis subjetivo de las mezclas de ADN complejas no ha sido establecido sobre la validez de sus fundamentos y no es una metodología fiable.

Dados los problemas encontrados en la interpretación subjetiva de mezclas complejas de ADN, una serie de grupos iniciaron esfuerzos para desarrollar programas informáticos que aplican varios algoritmos para interpretar mezclas complejas de manera objetiva. Estos programas representan claramente una mejora importante con respecto a la interpretación meramente subjetiva. Sin embargo, aún requieren escrutinio científico para determinar (1) si los métodos son científicamente válidos, incluyendo la delimitación de los márgenes de fiabilidad (es decir, las circunstancias en las cuales pueden ofrecer resultados no fiables) y (2) si el software implementa correctamente los métodos.

El PCAST encuentra que, hasta la fecha, los estudios han establecido la validez de los fundamentos de algunos métodos objetivos bajo determinadas circunstancias (específicamente, en una mezcla de tres personas en la que el contribuyente menor constituya, al menos, el 20% del ADN intacto en la mezcla), pero se considera que aún se necesita más evidencia sustancial para establecer la validez de los fundamentos en entornos más amplios.

#### 4.3 *Análisis de marcas de mordeduras*

Los análisis de marcas de mordeduras normalmente conllevan examinar marcas ocasionadas sobre una víctima o sobre un objeto en la escena del crimen y compararlas con impresiones dentales tomadas de un sospechoso. La comparación de las marcas de mordeduras se basa en las premisas de que (1) las características dentales,

particularmente la disposición de los dientes delanteros, difieren sustancialmente entre las personas y (2) la piel (o alguna otra superficie marcada en la escena del crimen) puede recoger fiablemente esas características. El análisis de las marcas de mordedura comienza con el analista decidiendo si una herida es una marca causada por dientes humanos. Si ello es así, el analista hace fotografías o saca impresiones de la mordedura cuestionada y de la dentadura del sospechoso, compara la mordedura con la marca de la dentadura, y determina si la marca de la dentadura (1) no puede excluirse de haber producido la mordedura, (2) puede excluirse de haberla producido o (3) es inconclusa.

El análisis de marcas de mordeduras es un método subjetivo. Los actuales protocolos no proporcionan estándares bien definidos sobre la identificación de características o el grado de similitud que debe hallarse para sostener una conclusión fiable sobre si la marca pudo o no haber sido producida por la dentadura en cuestión. Estas preguntas quedan sometidas al juicio del analista.

Como se resaltó más arriba, la validez de los fundamentos de un método subjetivo solo puede ser establecida a través de estudios, múltiples y apropiadamente diseñados, de caja negra. Pocos estudios — y ningún estudio apropiado de caja negra— se han realizado para analizar la capacidad de los analistas para identificar con precisión la fuente de una marca de mordedura. En esos estudios, las tasas de falsos positivos observadas son muy altas —normalmente por encima del 10% o mucho más—. Además, algunos de estos estudios emplearon diseños de conjunto cerrado inapropiados que probablemente *subestiman* la verdadera tasa de falsos positivos. En efecto, la evidencia científica disponible sugiere fuertemente que los analistas no sólo no pueden identificar la fuente de la marca de la mordedura con razonable precisión, sino que incluso ni siquiera pueden estar consistentemente de acuerdo sobre si la herida *es* una mordedura humana. Por estas razones, el PCAST encuentra que el análisis de marcas de mordeduras está lejos de cumplir los estándares científicos para la validez de sus fundamentos.

Observamos que algunos analistas forenses han expresado su preocupación por el hecho de que la exclusión de las marcas de mordedura por parte de los tribunales pudiera obstaculizar los esfuerzos para condenar a los acusados en algunos casos. Si así fuera, la solución correcta, desde una perspectiva científica, no sería la de admitir testimonio experto basado en métodos inválidos y no fiables sino, por el contrario, intentar desarrollar métodos científicamente válidos. Pero el PCAST considera que las perspectivas de desarrollo de análisis de marcas de mordedura para que sean métodos científicamente válidos son bajas. Desaconsejamos que se dediquen recursos significativos a tales esfuerzos.

#### 4.4. *Análisis de huellas latentes*

El análisis de huellas latentes normalmente conlleva comparar (1) una «impresión latente» (una impresión completa o parcial de las crestas papilares de un sujeto desconocido) que ha sido revelada u observada sobre una superficie con (2) una o más «impresiones conocidas» (impresiones dactilares deliberadamente recogidas bajo una configuración controlada de sujetos conocidos; también referida como «reseña decadactilar»), para valorar si las dos pueden haber sido originadas por la misma fuente. Puede conllevar también la comparación de impresiones latentes entre sí. Al analista se le puede pedir que (1) compare una impresión latente con las impresiones decadactilares de un sospechoso conocido que ha sido identificado por otros medios («sospechoso identificado») o (2) que busque en una gran base de datos de impresiones decadactilares para identificar a un sospechoso («búsqueda en base de datos»).

El análisis de impresiones latentes fue propuesto para su uso en la identificación criminal en el siglo XIX y ha sido utilizado durante más de una centuria. El método fue considerado durante mucho tiempo como infalible, a pesar de la falta de estudios empíricos apropiados para valorar su tasa de error. Como respuesta a las críticas sobre este particular vertidas por el informe del Consejo de Investigación Nacional en el año 2009, aquellos que trabajaban en el campo de los análisis de las impresiones latentes reconocieron la necesidad de realizar estudios empíricos para valorar la fundamentación de la validez y medir la fiabilidad, a partir de ahí ha habido progresos en esos menesteres. El mayor crédito se lo lleva el laboratorio del FBI, que guió el camino para realizar estudios de caja negra para valorar la validez y estimar la fiabilidad, así como los llamados estudios de «caja blanca» para comprender los factores que afectan a las decisiones de los analistas. El PCAST aplaude los esfuerzos del laboratorio del FBI. Existen también esfuerzos primigenios para comenzar a mover el campo desde un método puramente subjetivo hacia un método objetivo —aunque todavía falta un considerable camino para alcanzar esta importante meta—.

El PCAST encuentra que el análisis de impresiones latentes es una metodología subjetiva fundamentadamente válida —aunque con una tasa de falsos positivos sustancial y probablemente más alta que la esperada por muchos jurados debido a que tradicionalmente se han considerado infalibles los análisis de impresiones de huellas dactilares—. La tasa de falsos positivos pudiera llegar a ser tan alta como de 1 error cada 306 casos de acuerdo con el estudio del FBI y de 1 error cada 18 casos según un estudio realizado por otro laboratorio criminalístico<sup>4</sup>. A la hora de informar sobre

---

<sup>4</sup> El informe principal trata sobre los cálculos apropiados de las tasas de error, incluyendo los mejores estimadores (que son 1 de cada 604 y 1 de cada 24, respectivamente, para los dos casos citados) y los límites de confianza (los mencionados más arriba). También trata cuestiones con estudios específicos, incluyendo problemas con estudios que pueden conducir a diferencias en las tasas (como en los dos estudios citados).

resultados de exámenes de impresiones latentes, es importante establecer las tasas de falsos positivos basadas en estudios de validación apropiadamente diseñados.

Con respecto a la validez en la aplicación, se detectan, sin embargo, una serie de cuestiones abiertas, principalmente:

(1) *Sesgo de confirmación*. El trabajo de investigación de científicos del FBI ha demostrado que los analistas alteran con frecuencia las características que marcan inicialmente sobre una impresión latente cuando se compara con una huella conocida aparentemente coincidente. Tal razonamiento circular introduce un serio riesgo de sesgo de confirmación. Debe requerirse a los analistas que completen y documenten su análisis de huella dactilar latente *antes* de que observen cualquier impresión dactilar conocida y deberían documentar, separadamente, cualquier dato adicional utilizado durante sus comparaciones y evaluaciones.

(2) *Sesgo contextual*. El trabajo de académicos ha demostrado que los juicios de los analistas pueden verse influenciados por información irrelevante sobre los hechos del caso. Deberían realizarse esfuerzos para asegurar que los analistas no estén expuestos a información que pueda potencialmente producir sesgos.

(3) *Pruebas de aptitud*. Las pruebas de aptitud son esenciales para asegurar la capacidad y rendimiento del analista para emitir juicios precisos. Como se dijo en otro lugar de este informe, la prueba de aptitud necesita ser mejorada de forma que sea más rigurosa, incorporándola sistemáticamente en el flujo de trabajo de casos y divulgando pruebas para su evaluación por parte de la comunidad científica.

La validez científica en la aplicación requiere, por tanto, que un analista: (1) haya realizado pruebas de aptitud relevantes para probar su precisión y que informe de los resultados de la prueba de aptitud; (2) revele si documentó las características encontradas en la impresión latente por escrito antes de compararla con la impresión conocida; (3) proporcione un análisis escrito que explique la selección y la comparación de las características; (4) revele si, cuando realizó el examen, era consciente de algún hecho del caso que pudiera influirle en la conclusión; y (5) verifique que la impresión latente del caso en cuestión es similar, en calidad, dentro del rango de impresiones latentes consideradas en los estudios de fundamentación.

En lo referente al camino a seguir, se necesitan realizar esfuerzos continuos para mejorar el estado del arte de los análisis de impresiones latentes —y esos esfuerzos producirán claros beneficios en el sistema de justicia penal—. Una dirección posible es la de continuar mejorando el análisis de las impresiones latentes en cuanto método subjetivo. Hay necesidad de realizar estudios empíricos adicionales para estimar las tasas de error de los análisis de impresiones latentes de acuerdo con la variabilidad de su calidad y su integridad, utilizando mediciones bien definidas.

Una segunda —y más importante— dirección es la de convertir los análisis de impresiones latentes de un método subjetivo en un método objetivo. En la pasada década se han realizado extraordinarios avances en el análisis automatizado de las imágenes basado en *machine learning* y en otros enfoques —conduciendo a

mejoras notables en tareas tales como el reconocimiento facial y la interpretación de imágenes médicas—. Este progreso sostiene prometedoramente la posibilidad de que el análisis de impresiones latentes pueda hacerse completamente automático en un futuro próximo. Ha habido ya pasos en esta dirección tanto en el mundo académico como en la industria.

El recurso más importante para impulsar el desarrollo de métodos objetivos sería la creación de enormes bases de datos que contengan impresiones dactilares conocidas, cada una con muchas impresiones latentes «simuladas» de diferentes cualidades y grado de integridad, que podrían facilitarse a investigadores científicos capacitados en el ámbito académico y en la industria. Las impresiones latentes simuladas podrían crearse «modelando» («*morphing*») las impresiones conocidas, basándose en transformaciones derivadas de recolecciones de pares de impresión latente real – impresión rodada.

#### 4.5. *Análisis de armas de fuego*

En el análisis de armas de fuego, los analistas intentan determinar si la munición está o no asociada a un arma de fuego específica basándose en «marcas de herramientas» producidas por las armas de fuego en la munición. La disciplina se basa en la idea de que las marcas de herramientas producidas por diferentes armas de fuego varían sustancialmente (debido a las variaciones en la fabricación y el uso) como para permitir que los componentes de cartuchos disparados se identifiquen con armas de fuego particulares. Por ejemplo, los analistas pueden comparar los casquillos «cuestionados» de un arma, recuperados de la escena del crimen, con disparos de prueba de un arma sospechosa. El examen comienza con una evaluación de las características de clase de los proyectiles y casquillos, que son características permanentes y predefinidas anteriores a la fabricación. Si estas características de clase son diferentes, corresponde una conclusión de eliminación. Si las características de la clase son similares, el examen procede a identificar y comparar características individuales, como las marcas que surgen durante el disparo de un arma en particular.

Los analistas de armas de fuego han declarado durante mucho tiempo que su disciplina tiene una precisión casi perfecta; sin embargo, el estudio del Consejo de Investigación Nacional de 2009 sobre todas las disciplinas forenses concluyó acerca de los análisis de armas de fuego que «no se han realizado estudios suficientes para comprender la fiabilidad y reproducibilidad de los métodos», es decir, que la validez de los fundamentos de la disciplina no se ha establecido.

Nuestra extensa revisión de la literatura pertinente antes de 2009 es coherente con la conclusión del Consejo de Investigación Nacional. Encontramos que muchos de estos primeros estudios fueron inapropiadamente diseñados para evaluar la validez de sus fundamentos y estimar la fiabilidad. De hecho, hay evidencia interna entre

los propios estudios que indican que muchos estudios anteriores subestimaron al menos 100 veces la tasa de falsos positivos.

Identificamos un avance notable desde 2009: la finalización del primer estudio de caja negra apropiadamente diseñado sobre armas de fuego. El trabajo fue encargado y financiado por el *Defense Department's Forensic Science Center* (Centro de Ciencias Forenses del Departamento de Defensa) y fue llevado a cabo por un laboratorio de pruebas independiente (el Laboratorio Ames, un laboratorio nacional del *Department of Energy* (Departamento de Energía) afiliado a la Universidad Estatal de Iowa). La tasa de falsos positivos se estimó en 1 de cada 66, con un intervalo de confianza que indicaba que la tasa podría llegar hasta 1 de cada 46. Aunque el estudio está disponible como informe al gobierno federal, no se ha publicado en una revista científica.

Los criterios científicos para la validez de los fundamentos exigen que haya más de un estudio de este tipo para demostrar la reproducibilidad y que estos estudios deberían ser idealmente publicados en la literatura científica revisada por pares. Conforme a ello, las pruebas actuales aún no cumplen con los criterios científicos para la validez de los fundamentos.

Si el análisis de armas de fuego debe considerarse admisible sobre la base de las pruebas actuales es una decisión que corresponde a los tribunales. Si el análisis de armas de fuego *fuese* permitido en los tribunales, debería entenderse que los criterios científicos de validez en su aplicación exigen informar claramente de las tasas de error percibidas en el único estudio de caja negra apropiadamente diseñado. En la actualidad, no están justificadas científicamente las afirmaciones que asuman mayor precisión.

La validez en la aplicación también requeriría, desde un punto de vista científico, que un analista que testifique sobre el análisis de armas de fuego (1) se haya sometido a rigurosas pruebas de aptitud en un gran número de tests para medir su precisión y divulgue los resultados de las pruebas de aptitud y, además, (2) revele si al realizar el examen era consciente de cualquier otro hecho del caso que pudiera influir en la conclusión.

En cuanto al camino a seguir, en el análisis de armas de fuego al igual que con el análisis de huellas dactilares latentes, hay dos direcciones disponibles para fortalecer los fundamentos científicos de las disciplinas. La primera consiste en mejorar el análisis de armas de fuego como un método subjetivo, que requeriría estudios adicionales de caja negra para evaluar la validez científica y la fiabilidad, así como pruebas de aptitud más rigurosas a los analistas, utilizando problemas que sean apropiadamente desafiantes y que estos sean divulgados públicamente una vez hechas las pruebas.

La segunda dirección, al igual que con el análisis de impresiones latentes, es convertir el análisis de armas de fuego de un método subjetivo a un método objetivo. Esto implicaría desarrollar y probar algoritmos de análisis de imágenes para comparar la similitud de las marcas de herramientas en los proyectiles. Ha habido pasos alentadores hacia este objetivo. El mismo enorme progreso, llevado a cabo en la

última década en el análisis de imágenes, que aporta razones para esperar un logro temprano de un análisis de impresiones latentes totalmente automatizado, mueve al optimismo para que el análisis de armas de fuego totalmente automatizado pueda ser posible en un futuro próximo. Sin embargo, los esfuerzos en esta dirección se ven obstaculizados actualmente por la falta de acceso a bases de datos realmente grandes y complejas que puedan utilizarse para continuar el desarrollo de estos métodos y validar las propuestas iniciales.

El NIST, en coordinación con el Laboratorio del FBI, debe desempeñar un papel de liderazgo en la propulsión de la transformación necesaria mediante la creación y difusión de grandes conjuntos de datos apropiados. Estas agencias también deben proporcionar subvenciones y contratos para apoyar el trabajo y los procesos sistemáticos para evaluar los métodos. En particular, creemos que las competencias «premiadas», que estén basadas en grandes colecciones de imágenes públicamente disponibles podrían atraer un interés significativo del ámbito académico y de la industria.

#### 4.6. *Análisis de huellas de calzado*

El análisis de huellas de calzado es un proceso que ordinariamente conlleva comparar un objeto conocido, como un zapato, con una impresión completa o parcial hallada en la escena del crimen, para valorar si es probable que el objeto pudiera ser el origen de la impresión. El proceso se realiza secuencialmente, comenzando por una comparación de «características de clase» (tales como diseño, tamaño físico y desgaste general) y luego con una «identificación de características» o «características adquiridas aleatoriamente» (tales como marcas en el calzado causadas por cortes, incisiones y boquetes producidos en su uso).

El PCAST no ha abordado la cuestión de si los analistas pueden fiablemente determinar las características de clase —por ejemplo, si una impresión de calzado particular fue producida por un zapato de talla 12 de una específica manufactura—. Aunque es importante que se realicen estudios para estimar la fiabilidad de los análisis de huellas de calzado dirigidos a determinar las características de clase, el PCAST no quiso centrar su atención en este aspecto del examen de huellas de calzado porque la determinación de las características de clase, estimar la frecuencia de los tipos de calzado con una característica de clase particular o (a efectos de la tarea de los jurados) comprender la naturaleza de las propiedades en cuestión no es, *inherentemente*, un problema de medición desafiante.

En su lugar, el PCAST se centró en la fiabilidad de las conclusiones respecto a la probabilidad de que una impresión proceda de un calzado *específico*. Se trata de un problema mucho más difícil porque requiere saber con qué precisión los analistas pueden identificar propiedades específicas compartidas entre un calzado y una impresión, cuánto fallan en identificar propiedades que los distinguirían y qué valor probatorio puede otorgarse a una particular «característica aleatoriamente adquirida».

El PCAST encuentra que no hay estudios de caja negra apropiados que soporten la validez de los fundamentos de los análisis de huellas de calzado consistentes en asociar impresiones con un calzado concreto mediante marcas identificativas específicas. Esas asociaciones no están fundamentadas en pruebas significativas o estimaciones de su precisión, por lo que no son científicamente válidas.

#### 4.7. *Análisis de cabellos*

El análisis forense de cabellos es un proceso por el que los analistas comparan características microscópicas del cabello para determinar si una persona en particular puede ser el origen del cabello cuestionado. Mientras el PCAST completaba su informe, el Departamento de Justicia (DOJ) publicó, para que se hicieran comentarios, una propuesta de directrices sobre testimonios de exámenes de cabello, que incluía un documento de respaldo concerniente a la validez y fiabilidad de la disciplina. Aunque el PCAST no ha realizado una evaluación en profundidad de la disciplina de análisis de cabello como lo ha hecho en otras disciplinas basadas en comparación de características aquí tratadas, llevó a cabo una revisión del documento emanado del DOJ para arrojar más luz sobre los estándares para llevar a cabo una evaluación científica de una disciplina forense basada en comparación de características.

El documento establece que «la comparación microscópica de cabellos ha demostrado ser una metodología científica válida y fiable», mientras que subraya que «las comparaciones microscópicas de cabellos no pueden, por sí solas, conducir a una identificación personal y es crucial que esta limitación se transmita tanto en el informe escrito como en la comparecencia oral». Sin embargo, en apoyo a su conclusión de que el examen de cabellos es válido y fiable el documento sólo analiza un puñado de estudios sobre la comparación de cabello humano que abarcan los años 70 y 80. Los documentos de apoyo referidos fallan al no considerar que posteriores estudios encontraron defectos sustanciales en la metodología y en los resultados de los documentos clave. La propia revisión de los citados artículos realizada por el PCAST encuentra que esos estudios no establecen la validez de los fundamentos y fiabilidad de los análisis de cabellos.

El documento de apoyo del DOJ también cita un estudio del FBI realizado en el año 2002 que utilizó ADN mitocondrial para reexaminar 170 muestras de casos previos en los que el FBI había realizado análisis microscópicos de cabellos. Pero la conclusión principal de este estudio *no* respalda la conclusión de que el análisis de cabellos sea «una metodología científicamente válida y fiable». Los autores del FBI realmente encontraron que en 9 de 80 casos (11%) el Laboratorio del FBI había dictaminado que los cabellos eran microscópicamente indistinguibles, mientras que los análisis de ADN mostraron que los cabellos procedían de *diferentes* individuos.

Estas deficiencias ilustran tanto la dificultad de esas evaluaciones científicas como la razón por la que es mejor ellas sean llevadas a cabo por una agencia de carácter

científico que no esté involucrada en la aplicación de la ciencia forense dentro del sistema jurídico. Ellas también subrayan por qué es importante que la información *cuantitativa* sobre la fiabilidad de los métodos (por ejemplo, la frecuencia de falsas asociaciones en los análisis de cabellos) sea claramente establecida en el testimonio experto.

## 5. Observaciones finales de las siete evaluaciones

Aunque hemos realizado evaluaciones detalladas solo de seis métodos específicos —y una revisión de una evaluación realizada por otros de un séptimo— nuestro enfoque podría aplicarse para valorar tanto la validez de los fundamentos como la validez en la aplicación de cualquier método forense basado en la comparación de características, incluyendo tanto las disciplinas forenses tradicionales como también métodos aún pendientes de desarrollo (tales como el análisis microbiano o los patrones de navegación por Internet).

Hacemos notar, finalmente, que la evaluación de la validez científica ha de basarse, necesariamente, en la evidencia científica disponible en un momento dado. Algunos métodos que no han demostrado poseer validez en sus fundamentos podrían, al final, ser declarados fiables, aunque se requieran modificaciones significativas para que esos métodos alcancen esa meta. Algunos métodos puede que no la consigan, como puede ser el caso del análisis de la composición del plomo de los proyectiles y muy probablemente es el caso de las marcas de mordeduras. Otros pueden acabar siendo subsumidos por métodos distintos y más fiables, tal como el análisis de ADN ha reemplazado a otros métodos en algunos casos.

### 5.1. Recomendaciones al NIST y a la OSTP

#### 1.<sup>a</sup> Recomendación. Valoración de la validez de los fundamentos.

Es importante que las evaluaciones científicas de la validez de los fundamentos se realicen de forma continua para poder valorar la validez de los fundamentos de tecnologías de comparación de características forenses actuales y recientemente desarrolladas. Para asegurar que los juicios científicos no sean sesgados y sean independientes, tales evaluaciones deben estar dirigidas por un organismo científico que no tenga interés en el resultado.

(A) El Instituto Nacional de Estándares y Tecnología (NIST) debe realizar esas evaluaciones y debe emitir un informe público anual evaluando la validez de los fundamentos de los métodos clave de comparación de características.

(i) Las evaluaciones deben (a) valorar si cada método revisado ha sido adecuadamente definido, si ha sido adecuadamente establecida la validez de sus fundamentos

y si el nivel estimado de precisión se basa en evidencia empírica; (b) basarse en estudios publicados en la literatura científica de laboratorios y organismos de Estados Unidos y de otros países, así como cualquier trabajo dirigido por la propia plantilla y concesionarios del NIST; (c) como mínimo, elaborar evaluaciones sobre la línea del presente informe, actualizándolos según corresponda; (d) llevarse a cabo bajo los auspicios del NIST, con asesoramiento adicional realizado por analistas ajenos a la ciencia forense cuando se considere necesario.

(ii) El NIST debe establecer un comité asesor de científicos experimentales y estadísticos ajenos a la comunidad de ciencias forenses que proporcione asesoramiento sobre las evaluaciones y debe asegurarse de que sean rigurosos e independientes. Los miembros del comité consultivo deben seleccionarse conjuntamente por el NIST y la Oficina de Políticas Científicas y Tecnológicas.

(iii) El NIST debe priorizar aquellos métodos forenses de comparación de características que más necesiten de una evaluación, incluyendo los actualmente en uso y los que estén en desarrollo en etapas avanzadas, partiendo de las aportaciones del Departamento de Justicia y de la comunidad científica.

(iv) Cuando el NIST evalúe que un método ha sido establecido como válido en sus fundamentos, debe (a) aportar estimaciones apropiadas de tasas de error basadas en estudios sobre los fundamentos y (b) identificar cualquier problema relevante para la validez de su aplicación.

(v) Cuando el NIST evalúe que un método no ha sido establecido como válido en sus fundamentos, debe sugerir qué medidas, si las hubiere, podrían adoptarse para establecer la validez del método.

(vi) NIST no debe tener responsabilidades regulatorias con respecto a la ciencia forense.

(vii) NIST debe alentar a una o más revistas científicas líderes ajenas a la comunidad forense a que desarrolle mecanismos para promover la rigurosa revisión por pares y la publicación de artículos relacionados con la validez de los fundamentos de los métodos forenses de comparación de características.

(B) El Presidente debe solicitar y el Congreso debe proporcionar un aumento de financiamiento al NIST de (a) 4 millones de dólares para sustentar las actividades de evaluación anteriormente descritas y (b) de 10 millones de dólares para sustentar el aumento de las actividades de investigación en ciencia forense, incluyendo las mezclas complejas de ADN, las huellas dactilares latentes, el reconocimiento de voz y del hablante y la biometría facial y del iris.

2.<sup>a</sup> Recomendación. Desarrollo de métodos objetivos para el análisis de ADN de muestras de mezclas complejas, análisis de huellas dactilares latentes y análisis de armas de fuego.

El Instituto Nacional de Estándares y Tecnología (NIST) debe desempeñar el papel de líder en la transformación de tres importantes métodos de comparación de características que son actualmente subjetivos —análisis de huellas dactilares latentes, análisis de armas de fuego y, en algunas circunstancias, análisis de ADN de mezclas complejas— en métodos objetivos.

(A) El NIST debe coordinar esos esfuerzos con el Laboratorio de la Oficina Federal de Investigaciones (FBI), el Centro de Ciencias Forenses de la Defensa, el Instituto Nacional de Justicia y otras agencias relevantes.

(B) Esos esfuerzos deben incluir (i) la creación y difusión de grandes bases de datos y materiales de pruebas (tales como mezclas de ADN complejas) que permitan el desarrollo y la prueba de los métodos, tanto por el sector privado como por investigadores académicos, (ii) subvenciones y contratos, y (iii) procesos de patrocinio, tales como competiciones premiadas, para evaluar métodos.

3.<sup>a</sup> Recomendación. Mejora del proceso de organización para los Comités de Áreas Científicas.

(A) El Instituto Nacional de Estándares y Tecnología (NIST) debería mejorar la organización para los Comités de Áreas Científicas (OSAC), que fue establecida para desarrollar y promulgar estándares y directrices que permitan las mejores prácticas en la comunidad de ciencia forense.

(i) El NIST debe establecer un Comité de Recursos de Metrología, compuesto por metrologos, estadísticos y otros científicos ajenos a la comunidad de ciencia forense. Un representante del Comité de Recursos de Metrología debe prestar servicio en cada uno de los Comités de Área Científica (SACs) para proporcionar orientación sobre la aplicación de los principios estadísticos y de medición de los estándares documentados en desarrollo\*.

(ii) El Comité de Recursos de Metrología, en su conjunto, debe revisar y aprobar o desaprobado públicamente todos los estándares propuestos por los Comités de Áreas Científicas antes de que sean transmitidos a la Junta Directiva de Estándares de Ciencia Forense.

(B) El NIST debe asegurar que el contenido de los estándares y directrices registrados por la OSAC estén disponibles libremente para cualquier parte que los desee

---

\* N. del T.: los estándares documentados son los generados por los organismos de normalización de procedimientos técnicos.

en relación con un caso jurídico, o para evaluación o investigación, incluso alineándose con las políticas relacionadas con la disponibilidad razonable de estándares en la Oficina de Gestión y Circular presupuestaria A-119, en la Participación Federal en el Desarrollo y Uso de Estándares de Consenso Voluntario y en las Actividades de Evaluación de la Conformidad y la Oficina del Registro Federal, Manual IBR (incorporación por referencia).

#### 4.<sup>a</sup> Recomendación. Estrategia de I+D para la ciencia forense.

(A) La Oficina de Política de la Ciencia y de la Tecnología (OSTP) debe coordinar la creación de una estrategia nacional de investigación y desarrollo de la ciencia forense. La estrategia debe abordar los planes y las necesidades de financiación para:

(i) una gran expansión y fortalecimiento de la comunidad de investigación académica que trabaja en las ciencias forenses, incluyendo un aumento sustancial de la financiación tanto para la investigación como para la formación;

(ii) estudios de validez de los fundamentos de los métodos forenses de comparación de características;

(iii) mejora de los métodos forenses actuales, incluyendo la transformación de métodos subjetivos en métodos objetivos, y el desarrollo de nuevos métodos forenses;

(iv) el desarrollo de bases de datos de características forenses, con las adecuadas protecciones de la privacidad, que puedan utilizarse para la investigación;

(v) salvar la brecha entre los investigadores científicos y los profesionales forenses;

(vi) supervisión y revisión periódica de la investigación en ciencias forenses.

(B) Al preparar la estrategia, la OSTP debe solicitar la opinión de las agencias federales apropiadas, incluyendo especialmente el Departamento de Justicia, el Departamento de Defensa, la Fundación Nacional de la Ciencia y el Instituto Nacional de Estándares y Tecnología; a profesionales de la ciencia forense de ámbito federal y estatal; a investigadores científicos forenses y no forenses; y a otras partes interesadas.

#### 5.2 Recomendación al Laboratorio del FBI

##### 5.<sup>a</sup> Recomendación. Agenda para la ampliación de las ciencias forenses en el Laboratorio del FBI.

(A) Programas de investigación. El Laboratorio de la Oficina Federal de Investigación (FBI) debería emprender un vigoroso programa de investigación para mejorar la ciencia forense, sobre la base de su reciente e importante trabajo en análisis de huellas dactilares latentes. El programa debe incluir:

(i) la realización de estudios sobre la fiabilidad de los métodos de comparación de características, junto a terceros independientes sin interés en el resultado;

(ii) el desarrollo de nuevos enfoques que mejoren la fiabilidad de los métodos de comparación de características;

(iii) la ampliación de los programas de colaboración con científicos externos; y

(iv) la garantía de que los científicos externos tengan acceso adecuado a los conjuntos de datos y colecciones de muestras para que puedan realizar estudios independientes.

(B) Estudios de caja negra. Basándose en su experiencia en la investigación científica forense, el Laboratorio del FBI debería ayudar en el diseño y ejecución de estudios adicionales de caja negra para métodos subjetivos, incluido los análisis de huellas dactilares latentes y de armas de fuego. Esos estudios deben ser realizados por o en conjunto con terceros independientes sin interés en el resultado.

(C) Desarrollo de métodos objetivos. El Laboratorio del FBI debe trabajar con el Instituto Nacional de Estándares y Tecnología para transformar tres importantes métodos de comparación de características que actualmente son subjetivos —análisis de huellas dactilares latentes, análisis de armas de fuego y, en algunas circunstancias, análisis de ADN de mezclas complejas—, en métodos objetivos. Estos esfuerzos deben incluir (i) la creación y difusión de grandes bases de datos que sustenten el desarrollo y las pruebas de métodos, tanto por parte de las empresas como de investigadores académicos, (ii) becas y apoyo contractual, y (iii) patrocinio de competiciones premiadas para evaluar los métodos.

(D) Pruebas de aptitud. El Laboratorio del FBI debe promover un mayor rigor en las pruebas de aptitud (i) dentro de los próximos cuatro años, instituyendo pruebas de aptitud ciegas de rutina dentro del flujo de casos en su propio laboratorio, (ii) ayudando a otros laboratorios federales, estatales y locales a hacerlo igualmente, y (iii) fomentar el acceso rutinario y la evaluación utilizadas en las pruebas comerciales de aptitud.

(E) Análisis de huellas dactilares latentes. El Laboratorio del FBI debe promover vigorosamente la adopción, por todos los laboratorios que realizan análisis de huellas dactilares latentes, de las normas que requieren un proceso de «Análisis, Comparación y Evaluación lineales» —en el que los analistas deben completar y documentar sus análisis de huellas dactilares latentes antes de mirar a una impresión dactilar conocida y deben, separadamente, documentar cualquier dato adicional utilizado en la evaluación y en la comparación—.

(F) Transparencia en materia de calidad en casos. El Laboratorio del FBI, así como otros laboratorios forenses federales, debe informar y regular públicamente sobre problemas de calidad en el análisis de casos (de forma similar a las prácticas empleadas por el Instituto Forense Holandés, descritas en la sección 5), como un medio para mejorar la calidad y promover la transparencia.

(G) Presupuesto. El Presidente debe solicitar y el Congreso debe proporcionar un incremento de la partida presupuestaria del FBI que restaure el presupuesto del Laboratorio del FBI para las actividades de investigación en ciencias forenses partiendo de su nivel actual con 30 millones de dólares adicionales, y debe evaluar la necesidad de aumentar la financiación para otras actividades de investigación en ciencias forenses en el Departamento de Justicia.

### 5.3. *Recomendación al Fiscal General*

6.<sup>a</sup> Recomendación. Uso de los métodos de comparación de características en procesos judiciales federales.

(A) El Fiscal General debe ordenar a los fiscales que comparezcan en nombre del Departamento de Justicia (DOJ) se aseguren de que los testimonios ante los tribunales de los analistas sobre métodos de comparación de características cumplan los estándares científicos de validez científica.

Si bien las investigaciones previas al juicio pueden basarse en una gama más amplia de métodos, el testimonio de los analistas en el debate sobre métodos forenses de comparación de características en casos penales —que pueden ser muy influenciados y que ha conducido a muchas condenas injustas— debe cumplir un estándar más alto. Particularmente, los fiscales que comparezcan en nombre del Departamento de Justicia se deben asegurar de que:

(i) Los métodos forenses de comparación de características sobre los que se basa el testimonio hayan sido establecidos como válidos en sus fundamentos, como lo demuestran los estudios empíricos apropiados y la consistencia con las evaluaciones realizadas por el Instituto Nacional de Estándares y Tecnología (NIST), cuando estén disponibles; y

(ii) El testimonio sea científicamente válido, con las declaraciones del analista sobre la precisión de los métodos y el valor probatorio de las identificaciones propuestas constreñidas por la evidencia empírica que les sirve de apoyo, sin implicar un mayor grado de certeza.

(B) El Departamento de Justicia debe llevar a cabo una revisión inicial, con la asistencia del NIST, de los métodos de comparación de características subjetivos utilizados por el DOJ, con el fin de identificar qué métodos (más allá de los revisados en este informe) carecen de estudios apropiados de caja negra necesarios para evaluar la validez de los fundamentos. Como consecuencia de que se presume que tales métodos subjetivos no son válidos en sus fundamentos, el Departamento de Justicia debe evaluar si es apropiado que se presenten ante los Tribunales conclusiones basadas en tales métodos.

(C) Cuando existan métodos relevantes que aún no se hayan establecido como válidos en sus fundamentos, el DOJ debe fomentar y proporcionar apoyo para que se

realicen estudios apropiados de caja negra para evaluar la validez de los fundamentos y medir la fiabilidad. El diseño y la ejecución de estos estudios debe llevarse a cabo por o en unión de terceros independientes sin participación alguna en el resultado.

#### 7.<sup>a</sup> Recomendación. Directrices del Departamento de Justicia sobre el testimonio experto.

(A) El Fiscal General debe revisar y publicar nuevamente para recibir comentarios públicos, la propuesta de «Lenguaje Uniforme para Testimonios e Informes» y documentos de apoyo con el fin de alinearlos con los estándares científicos de validez científica.

(B) El Fiscal General debe emitir instrucciones que indiquen que:

(i) Cuando existan estudios empíricos y/o modelos estadísticos que arrojen luz sobre la precisión de un método de comparación de características, el analista debe proporcionar información cuantitativa sobre las tasas de error, de acuerdo con las directrices que establecerán el DOJ y el Instituto Nacional de Estándares y Tecnología, basada en el asesoramiento de la comunidad científica.

(ii) Cuando no existan estudios empíricos y/o modelos estadísticos apropiados para proporcionar información significativa sobre la precisión de un método forense de comparación de características, los fiscales del Departamento de Justicia y los analistas no deben ofrecer testimonio basado en ese método. Si es necesario dar testimonio en relación con el método, deben reconocer claramente ante los tribunales la falta de tales pruebas.

(iii) En el testimonio, los analistas deben siempre afirmar, claramente, que pueden ocurrir errores y que, de hecho, ocurren, debido tanto a la similitud entre las características como a errores humanos en el laboratorio.

#### 5.4. Recomendación al Poder Judicial

#### 8.<sup>a</sup> Recomendación. La validez científica como fundamento del testimonio experto.

(A) Al decidir sobre la admisibilidad de testimonios de analistas, los jueces federales deben tener en cuenta los criterios científicos apropiados para evaluar la validez científica, que incluyen:

(i) *La validez de los fundamentos*, con respecto al requisito de la Regla 702(c) de que el testimonio sea el producto de principios y métodos fiables; y

(ii) *La validez en la aplicación*, con respecto al requisito de la Regla 702(d) de que el analista haya aplicado fiablemente los principios y métodos a los hechos del caso.

Estos criterios científicos se describen en el Hallazgo 1.

(B) Los jueces federales, cuando permitan que un analista testifique sobre un método de comparación de características válido en sus fundamentos, deben asegurarse de que el testimonio sobre la precisión del método y el valor probatorio de las identificaciones propuestas sea científicamente válido en el sentido de que esté limitado a lo que la evidencia empírica sustente. Las declaraciones que sugieran o impliquen una mayor certeza no son científicamente válidas y no deben permitirse. En particular, los tribunales nunca deben permitir aseveraciones científicamente indefendibles como: tasa de error «cero», «muy pequeña», «esencialmente cero», «despreciable», «mínima» o «microscópica»; «certeza del 100%» o prueba «hasta un grado razonable de certeza científica»; identificación «hasta la exclusión de todas las demás fuentes posibles»; o una probabilidad de error tan remota como para que sea una «imposibilidad práctica».

(C) Para ayudar a los jueces, la Conferencia Judicial de los Estados Unidos, a través de su Comité Consultivo Permanente sobre las Reglas Federales de la Prueba, debe preparar, con el asesoramiento de la comunidad científica, un manual de buenas prácticas y un documento del Comité Consultivo, proporcionando orientación a los jueces federales sobre la admisibilidad bajo la Regla 702 de testimonios de analistas basados en métodos forenses de comparación de características.

(D) Para ayudar a los jueces, el Centro Judicial Federal debe desarrollar programas sobre criterios científicos para la validez científica de los métodos forenses de comparación de características.

## VII. INFORME

### 1. Introducción

La «ciencia forense» ha sido definida como la aplicación de prácticas científicas y técnicas al reconocimiento, recogida, análisis e interpretación de la evidencia en asuntos jurídicos penales o civiles o de otros ámbitos normativos<sup>5</sup>. La ciencia forense abarca un amplio rango de disciplinas, cada una con su propio conjunto de tecnologías y prácticas. El Instituto Nacional de Justicia (NIJ) divide esas disciplinas en doce categorías: toxicología general, balística y trazas instrumentales; documentos cuestionados; evidencia de traza (tales como análisis de cabellos y fibras); sustancias controladas; exploración biológica/serológica (incluyendo el análisis de ADN); análisis de incendios y de sus escombros; evidencia impresa (huellas de calzado, neumáticos, etc.); evidencia de patrones de sangre; inspección ocular; investigación médico-legal de la causa de la muerte; y la evidencia digital<sup>6</sup>. En los próximos años, la ciencia

---

<sup>5</sup> Definición de «ciencia forense» proporcionada por la Comisión Nacional de Ciencia Forense en su documento estratégico, «Defining forensic science and related terms». Publicado el 30 de mayo de 2015. [www.justice.gov/ncfs/file/786571/download](http://www.justice.gov/ncfs/file/786571/download).

<sup>6</sup> Véase: National Institute of Justice, 2006. [www.ojp.usdoj.gov/nij/pubs-sum/213420.htm](http://www.ojp.usdoj.gov/nij/pubs-sum/213420.htm).

y la tecnología probablemente nos ofrezcan poderosas herramientas en el dominio forense —quizá la capacidad de comparar poblaciones de bacterias en el intestino o patrones de búsqueda en Internet—.

Históricamente, la ciencia forense ha sido utilizada primordialmente en las dos fases del proceso de justicia penal: (1) *investigación*, que busca identificar al probable perpetrador de un delito, y (2) *procesamiento*, que busca probar la culpabilidad de un acusado más allá de toda duda razonable (En años recientes, la ciencia forense —particularmente el análisis de ADN— ha sido también ampliamente utilizada para cuestionar condenas del pasado). Es importante resaltar que las fases de investigación y de procesamiento conllevan diferentes estándares en el uso de la ciencia forense, así como otras herramientas de investigación. En las investigaciones, las perspectivas e informaciones pueden proceder de ciencia bien establecida o de aproximaciones exploratorias<sup>7</sup>. En la fase de procesamiento, la ciencia forense debe satisfacer un estándar más alto. Específicamente, las reglas de la prueba federales requieren que el testimonio experto esté basado, entre otras cosas, sobre «principios y métodos fiables» que hayan sido «fiablemente aplicados» a los hechos del caso<sup>8</sup>. Y la Corte Suprema ha establecido que los jueces deben determinar «si el razonamiento o metodología subyacente al testimonio experto es científicamente válido»<sup>9</sup>.

Aquí es donde los estándares jurídicos y científicos se interseccionan. Las decisiones de los jueces sobre la admisibilidad de la prueba científica descansan únicamente en estándares *jurídicos*; ellas son exclusiva potestad de los tribunales. Pero la cuestión preponderante a analizar para tomar esa decisión judicial es la validez científica<sup>10</sup>. Y pertenece a la competencia de la comunidad científica proporcionar directrices sobre los estándares científicos de validez científica<sup>11</sup>.

Resulta oportuna una mirada al lado científico de esta intersección porque ha quedado cada vez más claro en estos años que la falta de rigor en la valoración de la validez científica de la prueba forense no es un problema hipotético sino una real y significativa debilidad del sistema judicial. Como se relata en la sección 2, las revisiones que han hecho organismos competentes de los fundamentos científicos de las disciplinas forenses y el uso en los juzgados de la prueba basada en esas disciplinas han revelado una preocupante frecuencia en el uso de pruebas forenses que no superan una prueba objetiva de validez científica.

---

<sup>7</sup> Aunque los métodos de investigación no necesitan cumplir los estándares de fiabilidad requeridos bajo las Reglas de la Evidencia Federal, deben estar basados en principios y prácticas científicas sanas para evitar falsas acusaciones.

<sup>8</sup> Regla de la Evidencia Federal 702.

<sup>9</sup> *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579 (1993) en 592.

<sup>10</sup> *Daubert*, en 594.

<sup>11</sup> En este informe, PCAST aborda únicamente los estándares *científicos* para la validez y fiabilidad científicas. No ofrecemos opinión sobre los criterios *jurídicos*.

La revisión más completa hecha hasta la fecha fue realizada por un comité del Consejo de Investigación Nacional (NRC) codirigido por el Juez Harry Edwards del Tribunal de Apelación para el Distrito del Circuito de Columbia y Constantine Gatsonis, director del Centro para las Ciencias Estadísticas de la Universidad de Brown. Con el mandato del Congreso en un proyecto de ley convertido en ley a finales del año 2005, el estudio empezó en otoño de 2006 y el comité emitió su informe en febrero de 2009<sup>12</sup>.

El informe del NRC (2009) describió un perturbador patrón de deficiencias común a muchos de los métodos rutinariamente utilizados en el sistema de justicia penal, de forma particularmente importante la falta de rigor y de estudios apropiados que estableciesen su validez científica, concluyendo que «muchas pruebas forenses—incluyendo, por ejemplo, identificaciones por marcas de mordeduras, balística y de herramientas— se introducen en los procesos penales sin ninguna validación científica significativa, sin determinación de tasas de error o sin una comprobación de su fiabilidad que explique los límites de la disciplina<sup>13</sup>.

En el año 2013, después de una prolongada discusión de los resultados y recomendaciones del informe NRC, dentro y fuera del gobierno federal, el Departamento de Justicia (DOJ) —en colaboración con el Instituto Nacional de Estándares y Tecnología (NIST)— estableció la Comisión Nacional de la Ciencia Forense (NCFCS) como un cuerpo consultivo federal con el encargo de proporcionar directrices en ciencia forense y recomendaciones políticas a la Fiscalía General; codirigido por el Fiscal General Adjunto y el Director del NIST, los 32 miembros del NCFCS, incluidos ocho científicos del ámbito académico y otros cinco científicos doctorados; los demás miembros incluyen jueces, abogados y analistas forenses. Para fortalecer la ciencia forense de forma más general, en el año 2014 el NIST estableció la Organización para los Comités de Áreas Científicas para la Ciencia Forense (OSAC) con el fin de «coordinar el desarrollo de los estándares y directrices... para mejorar la calidad y consistencia del trabajo en la comunidad de la ciencia forense»<sup>14</sup>.

En septiembre de 2015, el Presidente Obama pidió a su Consejo de Asesores en Ciencia y Tecnología (PCAST) que explorara, a la luz del trabajo hecho por NCSF y OSAC, qué esfuerzos adicionales pudieran contribuir a fortalecer las disciplinas de la ciencia forense y asegurar la fiabilidad científica de la prueba forense utilizada en el sistema jurídico de la Nación. Después de revisar las actividades en marcha y la literatura jurídica y científica relevante—incluyendo, particularmente, las valoraciones científicas y jurídicas del informe del NRC (2009)— el PCAST concluyó que hay dos importantes lagunas: (1) la necesidad de clarificar el significado científico de los «principios y métodos fiables» y de la «validez científica» en el contexto de ciertas

---

<sup>12</sup> National Research Council, 2009.

<sup>13</sup> *Ibid.*, 107-8.

<sup>14</sup> Véase: [www.nist.gov/forensics/organization-scientific-area-committees-forensic-science](http://www.nist.gov/forensics/organization-scientific-area-committees-forensic-science)

disciplinas forenses y (2) la necesidad de evaluar métodos forenses específicos para determinar si han sido científicamente establecidos como válidos y fiables.

Dentro de la amplia gama de disciplinas forenses, elegimos reducir nuestra atención hacia técnicas que referimos aquí como métodos forenses de «comparación de características» (véase el cuadro 1)<sup>15</sup>. La limitación del análisis a estos métodos, fue una motivación para hacer nuestra tarea manejable dentro de los límites de tiempo y recursos disponibles, no obstante elegimos esta particular clase de métodos porque: (1) son comúnmente utilizados en casos penales; (2) han atraído un alto grado de preocupación con respecto a su validez (*e.g.*, el informe del NRC [2009]); y (3) todos pertenecen a la misma amplia disciplina científica, la *metrología*, que es la «ciencia de la medición y de su aplicación», en este caso midiendo y comparando características<sup>16</sup>.

**CUADRO 1.**  
**MÉTODOS FORENSES DE COMPARACIÓN DE CARACTERÍSTICAS**

El PCAST utiliza el término «métodos de comparación de características» para referirse a la amplia variedad de métodos que se dirigen a determinar si una muestra que constituye una evidencia (por ejemplo, recogida en la escena del crimen) está o no asociada con una muestra de su origen potencial (por ejemplo, del sospechoso) basada en la presencia de patrones, impresiones, propiedades, o características similares en la muestra y en la fuente. Como ejemplos señalamos los análisis de ADN, cabellos, huellas dactilares latentes, de armas y munición utilizada, herramientas y sus marcas, de huellas de calzado y de neumáticos, mordeduras y escritura a mano.

El PCAST comenzó este estudio formando un grupo de trabajo con seis de sus miembros recogiendo información para su consideración<sup>17</sup>. Para instruirse a sí mismo sobre aspectos prácticos relacionados con la interacción entre ciencia y derecho,

<sup>15</sup> Vale la pena señalar que hay cuestiones relacionadas con la validez científica de otros tipos de pruebas forenses que se escapan del ámbito de este informe, pero que requieren urgente atención —incluyendo principalmente la ciencia que estudia los incendios y el traumatismo craneoencefálico abusivo, comúnmente referido como «Síndrome del bebé sacudido»—. Además, un área de gran importancia no abordada por este informe es la de los métodos científicos para la evaluación de la causalidad —por ejemplo, si la exposición a determinada sustancia fue la causa que probablemente ocasionó un daño a un individuo—.

<sup>16</sup> *International Vocabulary of Metrology – Basic and General Concepts and Associated Terms*, 2012.

<sup>17</sup> Dos de los miembros han estado implicados en la ciencia forense. El codirector del PCAST, Eric Lander, ha trabajado en varios roles científicos (como testigo experto en *People v. Castro* 545 N.Y.S. 2d 985 (Corte Suprema, 1989), un caso relevante y novel sobre la calidad del análisis de ADN que se expone en la página 31; testigo del Tribunal en *U.S. v. Yee*, 134 F.R.D. 161 en 1991; miembro del panel del NRC sobre el análisis forense de ADN en 1992; coautor científico con un científico forense del Laboratorio del FBI en 1994; y miembro de la Junta de Directores del Proyecto Inocencia desde el año 2004 hasta el presente). Ninguno de estos roles ha sido remunerado. El miembro del PCAST S. James Gates Jr. ha sido miembro desde su inicio de la Comisión Nacional sobre Ciencia Forense.

el PCAST consultó a un grupo de asesores con experiencia (listados al comienzo del documento), compuesto por nueve jueces federales, actuales o anteriores, un ex Procurador General de los Estados Unidos y un juez de la Corte Suprema Estatal, dos decanos de facultades de derecho y dos estadísticos con experiencia en esta materia. El PCAST buscó información de un grupo adicional de expertos y partes interesadas, que incluyó científicos y analistas forenses, jueces, fiscales, abogados defensores, defensores de la reforma de la justicia penal, estadísticos, académicos y representantes de agencias federales (véase Apéndice B). La información se recabó a través de múltiples reuniones personales y conferencias telefónicas, incluyendo una sesión en una reunión del PCAST el 15 de enero de 2016. El PCAST también llevó a cabo el inusual paso de iniciar una convocatoria abierta en línea para ampliar las aportaciones, en particular de la comunidad de analistas en ciencia forense; se recibieron más de 70 respuestas<sup>18</sup>.

El PCAST también compartió un borrador de este informe con el NIST y el DOJ, que proporcionaron comentarios detallados y útiles que fueron cuidadosamente considerados en la revisión de este informe.

El PCAST expresa su gratitud a todos los que han compartido sus opiniones. Su voluntad de colaborar con el PCAST no implica su apoyo a los puntos de vista expresados en este informe. La responsabilidad de las opiniones, resultados y recomendaciones expresados en este informe, así como como cualquier error de hecho o de interpretación, sólo es imputable al PCAST.

El resto de nuestro informe está organizado de la siguiente manera:

— La sección 2 proporciona una breve visión general de los resultados de otros estudios relacionados con la práctica forense y el testimonio basado en ella y revisa también las acciones federales recientemente llevadas a cabo para fortalecer la ciencia forense.

— La sección 3 revisa brevemente el papel de la validez científica dentro del sistema jurídico. Describe la importante distinción entre estándares jurídicos y estándares científicos.

— La sección 4 después describe los estándares científicos que estarían implicados cuando se habla de «principios y métodos fiables» y «validez científica» en la medida en que se aplican a métodos de comparación de características, también ofrece criterios claros para que puedan ser fácilmente aplicados por los tribunales.

— La sección 5 ilustra la aplicación de los criterios indicados para usarlos en la evaluación de la validez científica de seis importantes métodos de comparación de características: análisis de ADN de muestras de una única fuente y mezcla simple, análisis de ADN con mezclas complejas, análisis de mordeduras, análisis de impre-

---

<sup>18</sup> Véase: [www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast\\_forensics\\_request\\_for\\_information.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensics_request_for_information.pdf).

siones latentes, análisis balísticos y análisis de huellas de calzado. También analizamos una evaluación llevada a cabo por otros respecto de un séptimo método, el análisis de cabellos.

— En las Secciones 6 a 9 ofrecemos recomendaciones basadas en los resultados de las Secciones 4 y 5, concernientes a acciones federales que pudieran ponerse en marcha para fortalecer la ciencia forense y promover su más riguroso uso en las salas de audiencia.

## 2. Trabajo previo sobre validación de métodos de ciencia forense

Los desarrollos de las dos últimas décadas —incluyendo la exoneración de acusados que habían sido erróneamente condenados debido en parte a la evidencia en ciencia forense, a una serie de estudios sobre los fundamentos científicos de las disciplinas forenses, revisiones del testimonio de experto basado en hallazgos forenses y a escándalos en laboratorios criminalísticos estatales— han suscitado una atención creciente sobre la cuestión de la validez y fiabilidad de algunos métodos de prueba forense y el testimonio basados en ellos (Sobre definiciones de términos clave como validez y fiabilidad científica, véase Cuadro 2 en la página 57).

En esta sección revisamos brevemente esta historia para dar nuestra valoración del estado actual de los métodos de la ciencia forense y su validez, así como el camino a seguir<sup>19</sup>.

### 2.1. Prueba de ADN y condenas erróneas

Irónicamente, fue la aparición y maduración de una nueva ciencia forense, el análisis de ADN, la primera en cuestionar seriamente la validez de muchas de las disciplinas forenses tradicionales. Cuando los acusados que fueron condenados con la ayuda de prueba forense de esas disciplinas tradicionales comenzaron a ser exonerados sobre la base de persuasivas comparaciones de ADN, se puso en marcha una investigación más profunda sobre su validez científica. La manera en la que esto llegó a suceder nos proporciona un contexto útil para nuestra actual investigación.

Cuando la prueba de ADN se introdujo por primera vez en los tribunales, comenzando a finales de los 80, fue inicialmente considerada como infalible. Pero los métodos utilizados en los primeros casos resultaron poco fiables: los laboratorios de pruebas carecían de procedimientos validados y aplicados consistentemente para definir los patrones de ADN a partir de las muestras, para declarar si dos patrones

---

<sup>19</sup> Para realizar este resumen, nos apoyamos particularmente en el informe del Consejo de Investigación Nacional (National Research Council, 2009) y en el informe de las Academias Nacionales de Ciencias, Ingeniería y Medicina (National Academies of Sciences, Engineering, and Medicine, 2015).

coincidían dentro de una cierta tolerancia y para determinar la probabilidad de que tales coincidencias pudieran ocurrir por mero azar en la población<sup>20</sup>.

Cuando la prueba de ADN se consideró inadmisibile en *People v. Castro*, un caso acaecido en Nueva York en 1989, los científicos —incluyendo los de la Academia Nacional de Ciencias y de la Oficina Federal de Investigación (FBI)— convinieron en promover conjuntamente el desarrollo de principios y métodos fiables, haciendo posible que el análisis de ADN de muestras procedentes de una única fuente haya llegado a ser el «estándar de oro» de la ciencia forense tanto para la investigación como para el procesamiento<sup>21</sup>.

Tanto el reconocimiento inicial de serios problemas como el desarrollo subsiguiente de procedimientos fiables se vieron favorecidos por la existencia de una robusta comunidad de biólogos moleculares que utilizaban los análisis de ADN en aplicaciones no forenses, tales como las ciencias biomédicas y las agrícolas. También contaron con la ayuda de jueces que reconocieron que este poderoso método forense solo debía ser admitido en los tribunales una vez que fuera establecida apropiadamente su fiabilidad.

Una vez que el análisis de ADN llegó a ser un método fiable, el poder de la tecnología —incluyendo su capacidad para analizar pequeñas muestras y distinguir entre individuos— hizo posible no solo la identificación y condena de los verdaderos autores de los delitos sino también el descarte de sospechosos falsamente acusados antes del procesamiento y el reexamen de una serie de condenas del pasado. Algunos informes del Instituto Nacional de Justicia (NIJ)<sup>22</sup> y de otros organismos, han resaltado que los tests de ADN en el curso de las investigaciones han descartado a decenas de miles de sospechosos. El proceso de reexamen de casos pasados basado en el ADN, además, ha conducido hasta la fecha a la exoneración de 342 condados, incluyendo 20 que habían sido sentenciados a pena de muerte, y a la identificación de 147 verdaderos culpables<sup>23</sup>.

---

<sup>20</sup> Véanse: LANDER, 1989: 501-5; LANDER y BUDOWLE, 1994: 735-8; KAYE, 1993: 101-72; ROBERTS, 1991: 1721-1723; THOMPSON y FORD, 1990: 38-43; NEUFELD y COLMAN, 1991: 46-53.

<sup>21</sup> *People v. Castro* 545 N.Y.S.2d 985 (Corte Suprema 1989). Un caso en el que un conserje fue acusado del asesinato de una mujer en el Bronx fue uno de los primeros casos penales en los que se aplicó el análisis de ADN en los Estados Unidos. El Tribunal estuvo un total de quince semanas en audiencia previa al juicio debatiendo la admisibilidad de la prueba de ADN. Al término de la audiencia, los analistas independientes, tanto de la defensa como de la acusación, acordaron unánimemente que la prueba de ADN presentada no era científicamente fiable —y el Juez declaró la prueba inadmisibile—. Véase: LANDER, 1989: 501-505. Estos sucesos condujeron finalmente a dos informes del NRC sobre análisis de ADN forense, en 1992 y en 1996, y a la fundación del Proyecto Inocencia ([www.innocenceproject.org](http://www.innocenceproject.org)).

<sup>22</sup> Los tests de ADN han excluido entre un 20-25% de los sospechosos iniciales en casos de agresiones sexuales. U.S. Department of Justice, Oficina de Programas de Justicia, National Institute of Justice, 1996.

<sup>23</sup> Innocence Project, «DNA Exonerations in the United States». Véase: [www.innocenceproject.org/dna-exonerations-in-the-united-states](http://www.innocenceproject.org/dna-exonerations-in-the-united-states).

Revisiones independientes de estos casos han revelado que muchos se basaron, en parte, en testimonios periciales defectuosos de científicos forenses que habían dicho a los miembros del Jurado que las características similares en un par de muestras, una tomada del sospechoso y otra encontrada en la escena del crimen (por ejemplo, cabellos, proyectiles, mordeduras, bandas de rodadura de neumáticos y huellas de calzado, u otros artículos), implicaban al acusado en el delito con un alto grado de certeza<sup>24</sup>. De acuerdo con las revisiones, estos errores no fueron simplemente un asunto consistente en una serie de analistas testificando conclusiones que se revelaron incorrectas; sino que, por el contrario, reflejaban un problema sistémico —el testimonio se basó en métodos e incluyó afirmaciones sobre su precisión que estaban encubiertas por una supuesta respetabilidad científica, pero realmente nunca habían estado sujetos a un escrutinio científico significativo<sup>25</sup>—.

## 2.2. *Estudios de métodos de ciencia forense específicos y prácticas de laboratorio*

Las preguntas que el análisis de ADN había suscitado sobre la validez científica de las tradicionales disciplinas forenses y el testimonio basado en ellas condujeron, naturalmente, a crecientes esfuerzos para comprobar empíricamente la fiabilidad de los métodos que empleaban esas disciplinas. Análogamente, se realizaron indagaciones dirigidas a examinar las prácticas de recogida, almacenamiento y análisis de la evidencia forense en los laboratorios de criminalística de todo el país. El laboratorio del FBI, ampliamente considerado como uno de los mejores del país, jugó un importante papel en las investigaciones recién mencionadas, reevaluando sus propias prácticas y las de los otros. En lo que sigue, resumimos algunos de los resultados clave de los estudios de los métodos y las prácticas que se produjeron en el caso de las disciplinas de «comparaciones» que constituyen el centro de atención de este informe.

### 2.2.1. Examen del plomo de los proyectiles

Desde los años 60 hasta el año 2005, el FBI utilizó el análisis de la composición del plomo de los proyectiles como una herramienta forense para identificar su origen. Sin embargo, un informe de NRC encargado por el FBI y emitido en el año 2004 puso en entredicho la validez de los fundamentos de las identificaciones basadas en esa disciplina. La técnica conllevaba comparar la cantidad de los distintos elementos de los proyectiles encontrados en la escena del crimen con la de proyectiles no utilizados, con el fin de determinar si los proyectiles provenían de la misma caja de

---

<sup>24</sup> Por ejemplo, véase: GROSS Y SHAFFER (2012) disponible en: [www.law.umich.edu/special/exoneration/Documents/exonerations\\_us\\_1989\\_2012\\_full\\_report.pdf](http://www.law.umich.edu/special/exoneration/Documents/exonerations_us_1989_2012_full_report.pdf). Véase también: SAKS, Y KOEHLER, 2005: 892-5.

<sup>25</sup> GARRET Y NEUFELD, 2009: 1-97; National Research Council, 2009, 42-43.

munición. El informe NRC 2004 dictaminó que no había base científica para realizar tal determinación<sup>26</sup>. Mientras que el método para determinar las concentraciones de los diferentes elementos dentro de un proyectil fue considerado fiable, el informe dictaminó que no había suficiente investigación y datos para sostener una conexión, basada en la similitud de composición, entre un proyectil particular y un lote determinado de munición, lo que comúnmente es lo relevante en un caso penal<sup>27</sup>. En el año 2005, el FBI anunció que abandonaba la práctica del examen del plomo de los proyectiles, señalando que mientras «apoyaba firmemente el fundamento científico del análisis del plomo de los proyectiles», la fabricación y la distribución de los proyectiles era demasiado variable como para hacer fiable el cotejo positivo de una comparación<sup>28</sup>.

### 2.2.2. Huellas dactilares latentes

En el año 2005, un comité internacional establecido por el FBI dictó un informe sobre defectos en las prácticas del FBI en identificación dactilar que había conducido a una relevante falsa identificación. Basado casi enteramente en una huella latente recuperada en el atentado con bombas en el sistema de trenes de cercanías de Madrid en el año 2004, el FBI detuvo erróneamente a un estadounidense en Portland (Oregón) y lo mantuvo durante dos semanas como *testigo material*<sup>29</sup>. Un analista del FBI concluyó que las huellas coincidían con «un 100% de certeza», aunque las autoridades españolas no pudieron confirmar el cotejo positivo<sup>30</sup>. El comité de revisión concluyó que la falsa identificación del FBI había ocurrido, principalmente,

---

<sup>26</sup> National Research Council, 2004. El examen del plomo de los proyectiles, también conocido como Análisis Composicional del Plomo de los Proyectiles (CABL), conlleva la comparación de la composición elemental de los proyectiles encontrados en la escena del crimen con la de los cartuchos no utilizados en posesión del sospechoso. Esta técnica asume que (1) la fuente fundida utilizada para producir un mismo «lote» de proyectiles tiene la misma composición por todas partes, (2) que no es posible que dos fuentes fundidas distintas tengan la misma composición, y (3) que los proyectiles con distintas composiciones no se han mezclado durante el proceso de fabricación o de envío. Sin embargo, en la práctica esto no es así. El informe del NRC (2004) dictaminó que los volúmenes composicionalmente indistinguibles de plomo pueden producir desde pequeños números de lotes de proyectiles —del orden de 12.000— a lotes más grandes —con más de 35 millones—. El informe también dictaminó que no hay seguridad de que distintos volúmenes de plomo producidos en distinto tiempo y lugar sean indistinguibles. Ni los científicos ni los fabricantes de proyectiles son capaces de declarar de forma contundente el significado de una asociación entre proyectiles en el curso de un examen de plomo de proyectil. Lo más que se puede decir es que los proyectiles que son indistinguibles por CABL podrían haber procedido de la misma fuente.

<sup>27</sup> FAIGMAN, CHENG, MNOOKIN, MURPHY, SANDER y SLOBOGIN Eds., 2016.

<sup>28</sup> Oficina Federal de Investigación. *FBI Laboratory Announces Discontinuation of Bullet Lead Examinations*. (1 de septiembre de 2005, nota de prensa) [www.fbi.gov/news/pressrel/press-releases/fbi-laboratory-announces-discontinuation-of-bullet-lead-examinations](http://www.fbi.gov/news/pressrel/press-releases/fbi-laboratory-announces-discontinuation-of-bullet-lead-examinations) (última consulta el 6 de mayo de 2016).

<sup>29</sup> STACEY, 2005.

<sup>30</sup> Application for Material Witness Order and Warrant Regarding Witness: Brandon Bieri Mayfield, In re Federal Grand Jury Proceedings 03-01, 337 F. Supp. 2d 1218 (D. Or. 2004) (No. 04-MC-9071).

como resultado de un «sesgo de confirmación»<sup>31</sup>. De forma similar, un informe de la Oficina del Inspector General del DOJ subrayaba que fue «el razonamiento inverso», desde la huella conocida a la imagen de la huella latente, lo que condujo a centrar la atención de forma exagerada en las aparentes similitudes y a una inadecuada falta de atención a las diferencias entre las imágenes<sup>32</sup>.

### 2.2.3. Análisis de cabellos

En el año 2002, los científicos del FBI utilizaron la secuenciación del ADN mitocondrial para reexaminar 170 comparaciones de cabellos microscópicas que los científicos de la agencia habían realizado en casos penales. El análisis de ADN demostró que en un 11% de los casos en los que los analistas del FBI habían dictaminado un cotejo positivo entre las muestras de cabellos comparadas microscópicamente el test de ADN de las muestras reveló que realmente procedían de individuos diferentes<sup>33</sup>. Estas falsas asociaciones pueden no haber sido fallos del analista a la hora de realizar los análisis; en lugar de ello, las características coincidentes pudieron haberse producido por casualidad. El estudio demostró que el poder de la comparación microscópica de cabellos entre muestras procedentes de diferentes fuentes era mucho menor que el previamente asumido. (Por ejemplo, anteriores estudios sugerían que la tasa de falsos positivos de los análisis de cabellos entraba en un rango de 1 en 40.000)<sup>34</sup>.

### 2.2.4. Mordeduras

Un estudio llevado a cabo en 2010 sobre mordeduras realizadas por personas claramente identificadas en el experimento halló que la deformación de la piel distorsiona las mordeduras tan sustancial y variablemente que los procedimientos actuales para comparar mordeduras son incapaces de excluir o implicar fiablemente a un sospechoso como potencial mordedor. («Los datos derivados demostraron carencias de correlación y reproducibilidad, es decir, la misma dentición no podía crear una impresión medible que fuera consistente en todos los parámetros, sea cuales fueren las circunstancias del test»)<sup>35</sup>. Un estudio reciente de la *American Board of Foren-*

---

<sup>31</sup> Específicamente, la semejanza entre las dos impresiones, combinada con la presión inherente de trabajar en casos de alto perfil, influyó en el juicio inicial del analista: características ambiguas fueron interpretadas como puntos semejantes y las diferencias entre las impresiones fueron dejadas de lado. Un segundo analista, no protegido frente a las conclusiones del primero, simplemente confirmó sus resultados. Véase: STACEY, 2005.

<sup>32</sup> U.S. Department of Justice, Office of the Inspector General, 2006. [oig.justice.special/s0601/final.pdf](https://oig.justice.special/s0601/final.pdf).

<sup>33</sup> HOUCK y BUDOWLE, 2002: 964-967.

<sup>34</sup> GAUDETTE y KEEPING, 1975: 599-606. Este estudio fue citado recientemente por el DOJ para sostener la afirmación de que el análisis de cabellos es una metodología científica válida y fiable. [www.justice.gov/dag/file/877741/download](http://www.justice.gov/dag/file/877741/download). El tema del análisis de cabellos se trata en la sección 5.

<sup>35</sup> BUSH, COOPER y DORION, 2010: 976-83. Véase también: BUSH, M.A., MILLER, BUSH, P.J., ET AL., 2009: 167-76.

*sic Odontology* (Junta Directiva Norteamericana de Odontología Forense) también mostraba una inquietante falta de consistencia en la forma en que los odontólogos forenses analizaban las marcas de mordeduras, que incluían incluso decisiones sobre si había prueba suficiente para determinar si una mordedura fotografiada era una mordedura humana<sup>36</sup>. En febrero de 2016, tras seis meses de investigación, la *Texas Forensic Science Commission* (Comisión de Ciencia Forense de Texas) recomendó una moratoria en el uso de las identificaciones de mordeduras en juicios penales, concluyendo que la validez de la técnica no había sido científicamente establecida<sup>37</sup>.

Estos ejemplos ilustran cómo diversos métodos de comparación de características que han tenido extendido uso no han sido sometidos, sin embargo, a tests significativos sobre su validez científica o para medir su fiabilidad.

### 2.3. Testimonio sobre la prueba forense

Al revisarse transcripciones de juicios orales, se ha comprobado que los testimonios periciales han sobreestimado frecuentemente el valor probatorio de sus evidencias, yendo mucho más allá de lo que la ciencia relevante puede justificar. Por ejemplo, algunos analistas han testificado:

— Que sus conclusiones son «100% ciertas»; tienen tasas de error igual a «cero», «esencialmente cero», «vanamente apreciable», «despreciable», «mínima» o «microscópica»; o tienen una probabilidad de error tan remota que es «prácticamente imposible»<sup>38</sup>. Como muchas revisiones han señalado, sin embargo, tales expresiones no son científicamente defendibles. Ningún test de laboratorio o análisis de comparación de características tiene error cero, incluso si un analista recibiera la máxima puntuación en la ejecución de un test particular consistente en un número limitado de muestras<sup>39</sup>. Incluso los tests altamente automatizados no tienen tasa de error igual a cero<sup>40 41</sup>.

— Que pueden «individualizar» la evidencia —por ejemplo, utilizando marcas sobre un proyectil para atribuirlo a un arma específica «hasta la exclusión de cual-

<sup>36</sup> BALKO, 2015. [www.washingtonpost.com/news/the-watch/wp/2015/04/08/a-bite-mark-matching-advocacy-group-just-conducted-a-study-that-dicredits-bite-mark-evidence/](http://www.washingtonpost.com/news/the-watch/wp/2015/04/08/a-bite-mark-matching-advocacy-group-just-conducted-a-study-that-dicredits-bite-mark-evidence/); FREEMAN y PRETTY, 2015. (los datos están disponibles por parte de los autores previa solicitud).

<sup>37</sup> Texas Forensic Science Commission, 2016. [www.fsc.texas.gov/sites/default/files/FinalBite-MarkReport.pdf](http://www.fsc.texas.gov/sites/default/files/FinalBite-MarkReport.pdf).

<sup>38</sup> THOMPSON, TARONI y AITKEN, 2003: 1-8. THOMPSON, 2013; COLE, 2005: 985-1078; y KOEHLER, J.J. «Forensics or fauxrensic? Ascertainning accuracy in the forensic sciences» [papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2773255](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2773255) (última consulta el 28 de junio de 2016).

<sup>39</sup> COLE, 2005: 985-1078; y KOEHLER, «Forensics or fauxrensic? Ascertainning accuracy in the forensic sciences» [papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2773255](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2773255) (última consulta el 28 de junio de 2016).

<sup>40</sup> THOMPSON, TARONI y AITKEN, 2003: 1-8.

<sup>41</sup> Los resultados de falsos positivos pueden proceder de dos fuentes: (1) similitud casual entre dos características y (2) fallos humanos/técnicos. Véase discusión en la sección 4 páginas 60-62.

quier otra arma de fuego en el mundo»— es una aseveración que no es sostenible por la ciencia relevante<sup>42</sup>.

— Que un resultado sea verdadero «hasta un razonable grado de certeza científica». Esta frase no tiene un significado generalmente aceptado en la ciencia y está abierta a amplias y diversas interpretaciones por parte de diferentes científicos<sup>43</sup>. Además, la expresión puede tomarse como si implicara certeza.

### 2.3.1. Revisión del DOJ del testimonio sobre el análisis de cabellos

En el año 2012, el DOJ y el FBI anunciaron que iniciarían una revisión formal del testimonio forense en 3000 casos penales relacionados con el análisis microscópico de cabellos. Los resultados iniciales de esta revisión sin precedentes, realizados en consulta con el *Innocence Project* (Proyecto Inocencia) y la *National Association of Criminal Defense Lawyers* (Asociación Nacional de Abogados Defensores en materia penal), mostraron que los analistas del FBI habían proporcionado testimonio científicamente inválido en el 95 % de los casos en los que esa prueba fue utilizada para inculpar al acusado en el juicio. Estos problemas fueron sistemáticos: 26 de 28 analistas de cabellos del FBI que testificaron en 328 casos proporcionaron testimonio científicamente inválido<sup>44,45</sup>.

La importancia de la revisión del análisis de cabellos realizada por el FBI fue puesta de relieve por la decisión de enero de 2016 del juez de la Corte Superior de Massachusetts, Robert Kane, de anular la condena de George Perrot basada en parte en el reconocimiento por parte del FBI de los errores en el análisis de cabellos<sup>46</sup>.

<sup>42</sup> Véanse: National Research Council, 2008.; SAKS y KOEHLER, 2008: 199-218.

<sup>43</sup> National Commission of Forensic Science, «Recommendations to the Attorney General Regarding Use of the Term “Reasonable Scientific Certainty”», 2016, disponible en: [www.justice.gov/ncfs/file/839726/download](http://www.justice.gov/ncfs/file/839726/download). El NCFS dictamina que «las conclusiones de las disciplinas forenses a menudo se testifican como apoyadas hasta un razonable grado de certeza científica» o «hasta un cierto grado de certeza [de la disciplina]». Estos términos no tienen significado científico y pueden confundir a los investigadores de los hechos sobre el nivel de objetividad envuelto en el análisis, su fiabilidad científica y limitaciones, así como sobre la capacidad del análisis para alcanzar una conclusión».

<sup>44</sup> Federal Bureau of Investigation. *FBI Testimony on Microscopic Hair Analysis Contained Errors in at Least 90 Percent of Cases in Ongoing Review*, (20 de abril de 2015, nota de prensa). [www.fbi.gov/news/pressrel/press-releases/fbi-testimony-on-microscopic-hair-analysis-contained-errors-in-at-least-90-percent-of-cases-in-ongoing-review](http://www.fbi.gov/news/pressrel/press-releases/fbi-testimony-on-microscopic-hair-analysis-contained-errors-in-at-least-90-percent-of-cases-in-ongoing-review).

<sup>45</sup> Las expresiones erróneas se clasificaron en tres categorías, en las que el analista: (1) declaró o insinuó que el cabello encontrado pudiera estar asociado con un específico individuo hasta la exclusión de todos los demás; (2) asignó hasta la asociación positiva una valoración estadística o una probabilidad de que el cabello probatorio procedía de una fuente en particular; o (3) citó el número de casos realizados en el laboratorio y el número de comparaciones positivas exitosas para sustentar una conclusión de que el cabello encontrado pertenecía a un específico individuo. REIMER, «The hair microscopy review project: An historic breakthrough for law enforcement and a daunting challenge for the defense bar». *The Champion*, (Julio 2013): 16. [www.nacdl.org/champion.aspx?id=29488](http://www.nacdl.org/champion.aspx?id=29488).

<sup>46</sup> *Commonwealth v. Perrot*, No. 85-5415, 2016 WL 380123 (Mass. Super. Man. 26, 2016).

### 2.3.2. Revisión expandida del DOJ

En marzo de 2016, el DOJ anunció su intención de expandir su revisión del testimonio pericial por el laboratorio del FBI en casos penales cerrados a métodos de ciencia forense adicionales. La revisión proporcionará oportunidades de valorar hasta qué punto hayan podido ocurrir similares testimonios sobrevalorados en otras disciplinas<sup>47</sup>. El DOJ planea establecer una estructura para auditar muestras de testimonios procedentes de Unidades del FBI que emplean tipos de pruebas adicionales basadas en características como el rastro de las impresiones que las armas dejan sobre los proyectiles, huellas de calzado, fibras, suelo y otras evidencias en la escena del crimen.

### 2.4. *Sesgos cognitivos*

Además de las cuestiones anteriormente descritas, los científicos han estudiado un problema más sutil pero igualmente importante que afecta a la fiabilidad de las conclusiones en muchos campos, incluyendo la ciencia forense: los sesgos cognitivos. Los sesgos cognitivos son los modos en que las percepciones y juicios humanos pueden alterarse por factores distintos a los relevantes para la decisión que se ha de tomar en un momento dado. Incluye el «sesgo de contexto» por el que los individuos se ven influenciados por información de contexto irrelevante; el «sesgo de confirmación» por el que los individuos interpretan información, o buscan nueva evidencia, de modo tal que se adecúe a las creencias o asunciones preexistentes; y «evitar la disonancia cognitiva» por la que los individuos son reacios a aceptar nueva información que sea inconsistente con sus primeras conclusiones. La comunidad científica biomédica, por ejemplo, hace todo lo posible para minimizar los sesgos cognitivos mediante el empleo de protocolos estrictos, tales como pruebas de doble ciego en ensayos clínicos.

Ha habido estudios que han demostrado que los sesgos cognitivos pueden ser un serio problema en la ciencia forense. Por ejemplo, un estudio realizado por Itiel Dror y sus colegas demostró que el juicio de los analistas en huellas dactilares latentes puede estar influenciado por el conocimiento de las decisiones de otros analistas forenses (una forma de sesgo de confirmación)<sup>48</sup>. Estos estudios se explican con mayor detalle en la Sección 5.4. Estudios similares han replicado estos resultados en otros campos forenses, incluyendo la interpretación de mezclas de perfiles de ADN, análisis microscópico de cabellos y en la investigación de incendios<sup>49, 50</sup>.

<sup>47</sup> Véase: [www.justice.gov/dag/file/870671/download](http://www.justice.gov/dag/file/870671/download).

<sup>48</sup> DROR, CHARLTON Y PERON, 2006: 74-8.

<sup>49</sup> Véase, por ejemplo: DROR Y HAMPIKIAN, 2011: 204-8; MILLER, 1987: 157; y BIEBER, 2014, disponible en [onlinelibrary.wiley.com/doi/10.1002/9780470061589.fsa1119/abstract](http://onlinelibrary.wiley.com/doi/10.1002/9780470061589.fsa1119/abstract).

<sup>50</sup> Véase, generalmente, DROR, 2016: 121-127.

Se han propuesto diversas estrategias para mitigar los sesgos cognitivos en los laboratorios forenses, incluyendo la gestión del flujo de información en un laboratorio criminalístico para minimizar la exposición de los analistas forenses a información contextual irrelevante (tales como confesiones o identificaciones de testigos oculares) y asegurarse de que los analistas trabajan de modo lineal, documentando sus hallazgos sobre las evidencias procedentes de la escena del crimen *antes* de realizar comparaciones con muestras procedentes de un sospechoso<sup>51</sup>.

## 2.5. Estado de la ciencia forense

El estudio del NRC (2009) concluyó que muchas de las dificultades de la ciencia forense posiblemente se deriven de la realidad histórica de que muchos métodos fueron ideados como heurística básica para ayudar en las investigaciones penales y no se fundamentaron en prácticas de validación de la investigación científica<sup>52</sup>. Aunque muchos laboratorios forenses requieren ahora que los analistas en ciencia forense recién contratados tengan un grado en ciencias, muchos analistas en laboratorios forenses no tienen grados avanzados en una disciplina científica<sup>53</sup>. Además, hasta el año 2015 no existían programas de doctorado específicos en ciencia forense en los Estados Unidos (aunque tales programas existían en Europa)<sup>54</sup>. Ha habido una financiación muy limitada para la investigación en ciencia forense, especialmente para estudiar la validez o fiabilidad de estas disciplinas. Resulta bastante reducido el número de revistas de ciencia forense con revisión por pares que sean especializadas en los campos de comparación de características.

Como señalaron el estudio del NRC 2009 y otros, fundamentalmente las ciencias forenses no han alcanzado todavía una «cultura investigadora» bien desarrollada<sup>55</sup>. Es importante destacar que una cultura investigadora incluye los principios: (1) debe presumirse que los métodos no son fiables hasta que se haya establecido la validez de los fundamentos basada en evidencia empírica y (2) incluso entonces, el cuestionamiento científico y la revisión de los métodos debe continuar de manera continua.

---

<sup>51</sup> KASSIN, DROR Y KAKUCKA, 2013: 42-52. Véase también: KRANE, FORD, S., GILDER, ET AL., 2008: 1006-1007.

<sup>52</sup> National Research Council, 2009: 128.

<sup>53</sup> National Research Council, 2009: 223-230. Véase también: CONNEY, 2010: 223-58. («Las áreas en las que no hubo consenso incluían requerimientos de grado (hubo casi una proporción de 50/50 entre agencias que requerían una titulación de 4 años o más frente a las que requerían menos de 4 años o ninguna titulación»).

<sup>54</sup> National Research Council, 2009: 223. Mientras existen diversos programas de doctorado en justicia penal, psicología forense, antropología forense o programas en química o en disciplinas relacionadas que ofrecen una concentración en la ciencia forense, solo el *Sam Houston State University College of Criminal Justice* (Instituto Universitario Estatal de Justicia Criminal Sam Houston) ofrece un programa doctoral en ciencia forense. Véase: [www.shsu.edu/programs/doctorate-of-philosophy-in-forensic-science](http://www.shsu.edu/programs/doctorate-of-philosophy-in-forensic-science).

<sup>55</sup> MNOOKIN, COLE, DROR, ET AL., 2011: 754-8.

En particular, algunos analistas forenses abrazan la noción de que una prolongada «experiencia» en resolver casos puede sustituir a los estudios empíricos de validez científica<sup>56</sup>. El ejercicio de la pericia no es una investigación científicamente válida y solo la experiencia no puede establecer la validez científica. En particular, no es posible que podamos estimar tasas de error a partir trabajo de casos porque ordinariamente no disponemos de conocimiento independiente sobre «fundamento de verdad» o «la respuesta correcta»<sup>57</sup>.

Más allá de la cuestión fundamental de la validez científica, históricamente, la mayoría de las disciplinas de comparación de características no prestaron suficiente atención a la importancia de blindar a los analistas frente a información potencialmente sesgada; al desarrollo de medidas objetivas de evaluación e interpretación; a prestar una cuidadosa atención a las tasas de error y su medición; y a desarrollar evaluaciones objetivas del significado de una asociación entre una muestra y su fuente potencial<sup>58</sup>.

El informe del NRC (2009) estimuló a algunos miembros de la comunidad científica forense a reconocer esos defectos. Algunos científicos forenses han abrazado la necesidad de colocar la ciencia forense sobre un fundamento científico sólido y han iniciado esfuerzos para lograrlo<sup>59</sup>.

## 2.6. *Estado de la práctica forense*

Las investigaciones de la práctica forense han desenterrado igualmente problemas procedentes de la falta de una fuerte «cultura sobre la calidad». Específicamente, docenas de investigaciones de laboratorios criminalísticos —primordialmente de nivel estatal o local— han revelado fallos repetidos en el manejo y procesado de las evidencias, así como interpretación incorrecta de los resultados de los análisis forenses<sup>60</sup>.

<sup>56</sup> Véase Sección 4.7.

<sup>57</sup> Véase Sección 4.7.

<sup>58</sup> National Research Council, 2009: 8, 124, 184-185, 188-191. Véase también: KOPPL y KRANE, 2016.

<sup>59</sup> Véase Sección 4.8.

<sup>60</sup> Unos cuantos ejemplos de esas investigaciones incluyen:

(1) una investigación independiente durante 2 años en el laboratorio de criminalística del Departamento de Policía de Houston que conllevó la revisión de 3500 casos (Informe Final del Investigador Independiente para el Laboratorio de Criminalística del Departamento de Policía de Houston, preparado por Michael R. Bromwich, 13 de junio de 2007, [www.hpdlabinvestigation.org/reports/070613report.pdf](http://www.hpdlabinvestigation.org/reports/070613report.pdf));

(2) la investigación y cierre de la Unidad Balística del Laboratorio de Criminalística de la Policía de Detroit tras el descubrimiento de pruebas contaminadas y fallo en el debido mantenimiento de los equipos de pruebas (véase BUNKLEY, «Detroit police lab is closed after audit finds serious errors in many cases». *New York Times*, 25 de septiembre de 2008: [www.nytimes.com/2008/09/26/us/26detroit.html?\\_r=0](http://www.nytimes.com/2008/09/26/us/26detroit.html?_r=0));

Varios comentaristas han resaltado una cuestión fundamental que puede subyacer a estos serios problemas: el hecho de que casi todos los laboratorios de criminalística están estrechamente vinculados a la acusación en los casos penales. Esta estructura socava la mayor objetividad que típicamente puede encontrarse en los laboratorios que realizan pruebas en otros campos y crea situaciones en las que el personal puede cometer errores debido a sutiles sesgos cognitivos o a una presión palmaria<sup>61</sup>.

El informe del NRC (2009) recomendó que todos los laboratorios forenses públicos y sus instalaciones fueran desvinculados del control administrativo de las agencias encargadas de hacer cumplir las leyes o de las fiscalías<sup>62</sup>. Por ejemplo, Houston —después de disolver su laboratorio criminalístico dos veces en tres años— siguió esta recomendación y, a pesar de una importante oposición política, consiguió el éxito en transformar el laboratorio en un centro de ciencia forense independiente<sup>63</sup>.

## 2.7. Informe del Consejo de Investigación Nacional

El informe del NRC (2009), *Fortaleciendo la Ciencia Forense en los Estados Unidos: un Paso Hacia Adelante*, ha sido la revisión más completa de las ciencias forenses en los Estados Unidos hasta la fecha. El informe dejó claro que los tipos de problemas, irregularidades y errores judiciales que resume no podían atribuirse simplemente a

---

(3) una investigación realizada en 2010 en el laboratorio de criminalística de la Oficina de Investigación del Estado de Carolina del Norte que descubrió que los agentes constantemente retenían pruebas exculpatorias o distorsionaban las pruebas en más de 230 casos en un periodo de 16 años (véase SWECKER y WOLF, «An Independent Review of the SBI Forensic Laboratory»: [images.bimedia.net/documents/SBI+Report.pdf](https://images.bimedia.net/documents/SBI+Report.pdf)); y

(4) una revisión en el año 2013 de la utilización de la prueba de ADN, realizada por la Oficina del Médico forense jefe del Hospital de la Ciudad de Nueva York, en más de 230 casos de violación (véase Estado de Nueva York, Oficina del Inspector General. Diciembre de 2013, [www.ig.ny.gov/sites/default/files/pdfs/OCMEFinalReport.pdf](http://www.ig.ny.gov/sites/default/files/pdfs/OCMEFinalReport.pdf)). Un análisis estimó que, al menos, en 50 laboratorios principales, se emitieron informes falsos, se destruyeron pruebas, fallaron en tests de aptitud, tergiversaron resultados en el testimonio o manipularon drogas entre los años 2005 y 2011. 28 de esos laboratorios estaban acreditados a nivel nacional. Memorando de Marvin Schechter a la Comisión de Ciencia Forense del Estado de Nueva York (25 de marzo de 2011): 243-4. Véase: [www.americanbar.org/content/dam/aba/administrative/legal\\_aid\\_indigent\\_defendants/ls\\_sclaid\\_def\\_train\\_memo\\_schechter.authcheckdam.pdf](http://www.americanbar.org/content/dam/aba/administrative/legal_aid_indigent_defendants/ls_sclaid_def_train_memo_schechter.authcheckdam.pdf).

<sup>61</sup> El informe del National Research Council, 2009: 24-25 dice: «La mejor ciencia se lleva a cabo en un entorno científico y no en un organismo encargado de hacer cumplir la ley. Como consecuencia de que los científicos forenses se conducen en su trabajo por la necesidad de responder a una pregunta particular relacionada con las cuestiones de un caso específico, a veces se enfrentan a la presión de sacrificar su propia metodología en aras de la conveniencia». Véase también: GIANNELLI, 2010: 247-66 y THOMPSON, 2015.

<sup>62</sup> National Research Council, 2009: 24.

<sup>63</sup> El Centro de Ciencia Forense de Houston abrió sus puertas en abril de 2014, reemplazando al Laboratorio de Criminalística del Departamento de Policía de Houston. El Centro opera como una «corporación gubernamental local» con sus propios directores, funcionarios y empleados. La estructura fue diseñada a propósito para aislar al Centro de la influencia de policías, fiscales, oficiales o grupos con algún interés especial. Véase: THOMPSON, 2015: 214.

un conjunto de analistas deshonestos o a laboratorios con bajo rendimiento. En su lugar, el informe descubrió que los problemas que asolan a la comunidad de la ciencia forense son sistémicos y generalizados —resultado de factores entre los que se incluyen un alto grado de fragmentación (incluyendo requerimientos de instrucción y formación, recursos y capacidades de los laboratorios dispares e inadecuados); una falta de estandarización de las disciplinas e insuficiente investigación y formación de alto nivel; y escasez de estudios revisados por pares que establezcan las bases científicas y la validez de muchos métodos forenses rutinariamente utilizados—.

Fueron especialmente prevalentes en las ciencias forenses las deficiencias en las disciplinas de comparación de características. El informe 2009 descubrió que muchas de esas disciplinas carecían de sistemas bien definidos para determinar las tasas de error y que no se habían realizado estudios para establecer la unicidad o rareza relativa o lo que resulta común en las marcas o características particulares examinadas. Además, las pruebas de aptitud, cuando se habían realizado, mostraban casos de bajo rendimiento por parte de determinados analistas. En resumen, el informe concluye que se introduce en los juicios penales «abundante prueba forense sin una seria validación científica, sin determinación de tasas de error, o sin pruebas de fiabilidad que expliquen los límites de la disciplina —incluyéndose, por ejemplo, las identificaciones por mordeduras, balísticas y de trazas instrumentales—»<sup>64</sup>.

El informe del NRC (2009) descubrió que los problemas que asolan las ciencias forenses eran tan graves que solo podían ser abordados mediante «un compromiso nacional para revisar la actual estructura que sostiene la comunidad de la ciencia forense en este país»<sup>65</sup>. Subyacente a las 13 recomendaciones básicas del informe yacía una llamada al liderazgo al más alto nivel de los gobiernos federales y estatales y a la promoción y adopción de una agenda a largo plazo para impulsar el desarrollo de la ciencia forense desde su actual situación de debilidad.

El informe del NRC (2009) requirió la realización de estudios para comprobar si una serie de métodos forenses eran válidos en sus fundamentos, incluyendo la realización de tests empíricos sobre la precisión de los resultados. También solicitó la creación de una nueva agencia federal independiente que proporcionara la necesaria supervisión al sistema de ciencia forense; la estandarización de la terminología utilizada en los dictámenes y en el testimonio rendido sobre los resultados de las ciencias forenses; la desvinculación de los laboratorios públicos forenses del control administrativo de las agencias encargadas de hacer cumplir las leyes; la implementación de requisitos de certificación de analistas y programas de acreditación de laboratorios obligatorios; la investigación sobre el sesgo del observador y las fuentes de error humano en los exámenes forenses; el desarrollo de herramientas de medición avanzada, validación, fiabilidad y pruebas de aptitud en la ciencia forense; y el fortalecimiento y desarrollo de grados y programas de formación e instrucción continua.

---

<sup>64</sup> National Research Council, 2009: 107-108.

<sup>65</sup> National Research Council, 2009.

## 2.8. *Progresos recientes*

En respuesta al informe del NRC (2009), la Administración Obama inició una serie de reformas dirigidas a fortalecer las ciencias forenses, comenzando con la creación en 2009 de un Subcomité en Ciencia Forense dentro del Comité de Ciencia del Consejo Nacional de Ciencia y Tecnología que recibió el encargo de considerar cómo conseguir del mejor modo posible los objetivos del informe NRC. Debajo se describen con algún detalle las actividades resultantes.

### 2.8.1. Comisión Nacional de la Ciencia Forense

En el año 2013, el DOJ y el NIST, con el apoyo de la Casa Blanca, firmaron un Memorando de Entendimiento que perfilaba un marco de cooperación y colaboración entre las dos agencias juntando esfuerzos para fortalecer la ciencia forense.

En el año 2013, el DOJ estableció una Comisión Nacional de Ciencia Forense (NCFS), que es un comité consultivo federal para informar al Fiscal General, codirigida por el Fiscal General Adjunto y el Director del NIST; entre los 32 miembros del NCFS se incluyen siete científicos del ámbito académico y otros cinco científicos doctorados; los demás miembros incluyen jueces, abogados y analistas forenses. La Comisión está encargada de proporcionar las recomendaciones políticas al Fiscal General<sup>66</sup>. La NCFS emite recomendaciones formales al Fiscal General, así como «documentos de opiniones» que reflejan el punto de vista de la mayoría de dos tercios del NCFS y que no requieren acciones específicas por parte del Fiscal General. Hasta la fecha, el NCFS ha emitido diez recomendaciones, entre otras cosas, sobre la acreditación de los laboratorios forenses y a la certificación de los analistas forenses, el avance de la interoperatividad de sistemas de información de huellas dactilares impresas, el desarrollo de protocolos de análisis de causa raíz para los proveedores de servicios forenses y la mejora de las comunicaciones entre los médicos examinadores y las oficinas forenses<sup>67</sup>. Hasta la fecha, el Fiscal General ha adoptado formalmente el primer conjunto de recomendaciones sobre acreditación<sup>68</sup> y ha ordenado al Departamento que comience a dar pasos dirigidos hacia alguna de las otras recomendaciones puestas de manifiesto hasta el presente<sup>69</sup>.

---

<sup>66</sup> Véase: [www.justice.gov/ncfs](http://www.justice.gov/ncfs).

<sup>67</sup> Para una lista completa de documentos aprobados por el NCFS, véase [www.justice.gov/ncfs/work-products-adopted-commission](http://www.justice.gov/ncfs/work-products-adopted-commission).

<sup>68</sup> Departamento de Justicia. «El Departamento de Justicia anuncia nuevas políticas de acreditación para hacer progresar la ciencia forense» (7 de diciembre de 2015, nota de prensa). [www.justice.gov/opa/pr/justice-department-announces-new-accreditation-policies-advance-forensic-science](http://www.justice.gov/opa/pr/justice-department-announces-new-accreditation-policies-advance-forensic-science).

<sup>69</sup> Memorando del Fiscal General a los jefes de departamento con respecto a las recomendaciones de la Comisión Nacional de Ciencia Forense, 17 de marzo de 2016. [www.justice.gov/ncfs/file/841861/download](http://www.justice.gov/ncfs/file/841861/download).

En el año 2014, el NIST estableció la Organización de Comités de Áreas Científicas (OSAC), un cuerpo colaborador de más de 600 miembros voluntarios en gran parte provenientes de la comunidad de la ciencia forense<sup>70</sup>. La OSAC se estableció para fomentar el desarrollo de estándares y directrices voluntarias para la consideración de la comunidad de analistas forenses<sup>71</sup>. Su estructura consta de seis Comités de Área Científica (SACs) y 25 Subcomités que trabajan para desarrollar estándares, directrices y códigos deontológicos para cada una de las disciplinas y metodologías forenses<sup>72</sup>. Tres Comités de Recursos Globales proporcionan orientación sobre cuestiones jurídicas, factores humanos y de garantía de la calidad. Todos los documentos desarrollados por las SACs se aprueban por la Junta Directiva de Estándares de Ciencia Forense (FSSB), que es un componente de la estructura de OSAC, para su incorporación al listado del Registro de OSAC de Estándares Aprobados. La OSAC no es un comité consultivo federal.

### 2.8.2. Fondos federales para la investigación

El gobierno federal ha dado también pasos para abordar un factor que contribuye a los problemas que se constatan en la ciencia forense —la falta de una comunidad de investigación científica robusta y rigurosa en muchas disciplinas de la ciencia forense—. Aunque existen múltiples razones que explican la ausencia de tal comunidad científica, una de ellas es que, a diferencia de la mayoría de las disciplinas científicas, ha habido muy pocos fondos que atraigan y sostengan a un grupo de científicos de excelencia centrados en la *investigación básica* en la ciencia forense.

La Fundación Nacional de la Ciencia (NSF) ha iniciado recientemente esfuerzos para ayudar a reparar esta deficiencia fundacional de la ciencia forense. En el año 2013, la NSF resaltó su interés en esta área y animó a los investigadores a presentar propuestas de investigación dirigidas a las cuestiones fundamentales que puedan hacer avanzar el conocimiento y la formación en las ciencias forenses<sup>73</sup>. Como resultado de un proceso de interagencias conducido por la OSTP y la NSF, en colaboración con el Instituto Nacional de Justicia (NIJ), se invitó a la presentación de propuestas para la creación de nuevos centros de investigación multidisciplinar para su financiación con los presupuestos de 2014<sup>74</sup>. Sobre la base de nuestra revisión de

---

<sup>70</sup> Entre los miembros se incluyen analistas en ciencia forense y otros analistas que representan a agencias federales, estatales y locales, así como al mundo académico y a la industria.

<sup>71</sup> Para más información, véase: [www.nist.gov/forensics/osac.cfm](http://www.nist.gov/forensics/osac.cfm).

<sup>72</sup> Los seis Comités del Área Científica de OSAC son: Biología/ADN; Química/Análisis Instrumental; Escena del Crimen/Investigación Médico-forense, Digital/Multimedia; Física/Interpretación de patrones ([www.nist.gov/forensics/upload/OSAC-Block-Org-Chart-3-17-2015.pdf](http://www.nist.gov/forensics/upload/OSAC-Block-Org-Chart-3-17-2015.pdf)).

<sup>73</sup> Véase: Dear Colleague Letter: Forensic Science – Opportunity for Breakthroughs in Fundamental and Basic Research and Education. [www.nsf.gov/pubs/2013/nsf13120/nsf13120.jsp](http://www.nsf.gov/pubs/2013/nsf13120/nsf13120.jsp).

<sup>74</sup> Los centros que NSF está proponiendo crear son centros de cooperación en investigación industria/universidad (I/UCRCs). I/UCRCs son colaborativos por diseño y pueden ser eficaces para ayudar a tender puentes entre las brechas culturales y científicas de los académicos que trabajan en áreas rele-

los resúmenes de subvenciones, PCAST estima que la NSF compromete un total de 4,5 millones de dólares anuales para fomentar proyectos de investigación extracurriculares sobre ciencia forense fundacional.

El NIST ha dado pasos también dirigidos a este tema creando un nuevo Centro de Excelencia de Ciencia Forense que se denomina Centro para la Estadística y Aplicaciones en Pruebas Forenses (CSAFE), que centrará sus esfuerzos investigadores en mejorar la fundamentación estadística de los análisis de las huellas latentes, balísticas, neumáticos, escritura manuscrita, patrones de manchas de sangre, trazas instrumentales y análisis de comparación de patrones, así como en análisis de sistemas de computación e información, dispositivos móviles, tráfico en la red, redes sociales y en evidencia digital GPS<sup>75</sup>. El CSAFE está financiado bajo un acuerdo de cooperación con la Universidad del Estado de Iowa, para establecer un centro en asociación con investigadores de la Universidad Carnegie Mellon, la Universidad de Virginia y la Universidad de California, Irvine; con una dotación total de veinte millones de dólares en cinco años. El PCAST estima que el NIST compromete un total de cinco millones de dólares al año para financiar proyectos de investigación básica extracurricular sobre ciencia forense, asignando, aproximadamente, 4 millones de dólares al CSAFE y 1 millón aproximadamente a otros proyectos.

El NIJ no tiene presupuesto específicamente asignado para la investigación en ciencia forense. Para que pueda apoyar actividades de investigación, el NIJ debe extraer de su presupuesto base financiación para esas actividades utilizando los programas de asistencia de la Oficina de Programas de Justicia para investigación y estadísticas o los programas de reducción de atrasos en informes de ADN<sup>76</sup>. La mayoría de la ayuda a la investigación está dirigida a la investigación aplicada. Aunque es difícil clasificar los proyectos de investigación del NIJ, estimamos que destina un total de aproximadamente cuatro millones de dólares al año para fomentar proyectos de investigación básica extramuros sobre ciencia forense<sup>77</sup>.

---

vantes de la ciencia para el ámbito forense y los analistas forenses. [www.nsf.gov/pubs/2014/nsf14066/nsf14066.pdf](http://www.nsf.gov/pubs/2014/nsf14066/nsf14066.pdf).

<sup>75</sup> National Institute of Standards and Technology, 2015. [www.nist.gov/forensics/center-excellence-forensic052615.cfm](http://www.nist.gov/forensics/center-excellence-forensic052615.cfm).

<sup>76</sup> National Academies of Sciences, Engineering, and Medicine, 2015. De acuerdo con el informe «La financiación otorgada por el Congreso para fomentar los programas de investigación del NIJ disminuyeron desde inicios hasta mediados de la década del 2000 y permanecen insuficientes, especialmente a la luz de los crecientes retos a los que se enfrenta la comunidad científica forense... Con financiación de partida limitada, el NIJ la financia y desarrolla con los créditos de los programas de reducción de atrasos en informes de ADN y otros programas asistenciales. Estos fondos para salir del paso están apoyando esencialmente la cartera actual de las ciencias forenses del NIJ, pero hay presiones para reducir la cantidad asignada a la investigación para estos programas. En los tres últimos años, la financiación que asiste a estos programas ha disminuido; por tanto, la financiación disponible para la investigación también se ha reducido».

<sup>77</sup> U.S. Department of Justice, National Institute of Justice, 2016.

Incluso con los incrementos recientes, es probable que la financiación extramuros total para la investigación básica en ciencia forense entre la NSF, el NIST y el NIJ esté en el rango de solo 13,5 millones de dólares anuales. El informe del NRC (2009) dijo que:

La investigación en ciencia forense no está [en general] bien sustentada... En comparación a otras áreas de la ciencia, las disciplinas de la ciencia forense tienen oportunidades extremadamente limitadas para financiar la investigación. Aunque el FBI y el NIJ han sostenido algunas investigaciones en las disciplinas de la ciencia forense, el nivel de soporte ha sido muy inferior con respecto a lo necesario para que la comunidad de ciencia forense establezca lazos fuertes con una amplia base de universidades dirigidas hacia la investigación y la comunidad de investigadores a nivel nacional. Además, la financiación de la investigación en la academia es limitada..., lo que puede inhibir la búsqueda de soluciones a cuestiones científicas fundamentales, esenciales para establecer la base de la ciencia forense. Finalmente, la comunidad de investigadores, en un sentido más amplio, no está generalmente involucrada en llevar a cabo investigación relevante para desarrollar las disciplinas de la ciencia forense<sup>78</sup>.

Un informe del NRC del año 2015, *Support for Forensic Science Research: Improving the Scientific Role of the National Institute of Justice* (Apoyo a la Investigación en Ciencia Forense: Mejorando el Rol Científico del Instituto Nacional de Justicia), encontró que la situación de la financiación para la investigación en ciencia forense no había mejorado mucho desde el informe del NRC (2009)<sup>79</sup>.

Además, el Centro de Ciencias Forenses de la Defensa ha comenzado recientemente a financiar investigación externa que abarca las disciplinas de la ciencia forense como parte de su misión de proporcionar capacidades especializadas de investigación forense y biométrica y apoyo al Departamento de Defensa. Rediseñado como DFSC en el año 2013, el Centro fue primeramente el Laboratorio de Investigación Criminalística del ejército estadounidense, originalmente encargado de apoyar las investigaciones penales dentro del ámbito militar, pero adicionalmente encargado en el año 2007 de proporcionar «duraderas capacidades forenses expedicionarias», en respuesta en parte a las necesidades de investigación y enjuiciamiento por ataques con explosivos en Irak y Afganistán. Mientras que la mayor parte del soporte de DFSC ha sido tradicionalmente la investigación en análisis de ADN y en bioquímica, el Centro ha destinado recientemente recursos hacia proyectos que abordan carencias básicas críticas en otras disciplinas, incluyendo los análisis de balística y de huellas dactilares latentes.

De forma destacada, el DFSC ha ayudado a estimular la investigación en la comunidad de la ciencia forense. Fruto de las conversaciones entre DFSC y la Sociedad estadounidense de directores de laboratorios de criminalística (ASCLD), esta última organizó una reunión en el año 2011 para identificar las prioridades de investigación para la comunidad de la ciencia forense. El DFSC acordó financiar dos estudios fundamentales para abordar las necesidades de investigación de mayor prioridad identi-

<sup>78</sup> National Research Council, 2009: 78.

<sup>79</sup> National Academies of Sciences, Engineering, and Medicine, 2015: 15.

ficadas por el Comité de Investigación Forense de ASCLD: el primer estudio independiente de análisis de «caja negra» sobre análisis de armas de fuego y un estudio de interpretación de mezclas de perfiles de ADN (véase Sección 5). En el año Fiscal 2015, DFSC asignó aproximadamente 9,2 millones de dólares a la investigación externa en ciencia forense. El 75% del crédito de DFSC sustentó proyectos relacionados con el ADN y bioquímica; el 9% sobre la prueba digital; el 8% sobre pruebas de comparación de patrones distintos al ADN; y el 8% a la química<sup>80</sup>. Como en el caso de NIJ, no existe línea de créditos dedicados a la investigación en ciencia forense en los presupuestos del DFSC; el DFSC, en su lugar, debe solicitar créditos de múltiples fuentes dentro del Departamento de Defensa para sostener esta investigación.

### 2.8.3. Una carencia crítica: la validez científica

La Administración ha dado pasos importantes y muy necesarios creando mecanismos para discutir la política pública, desarrollar mejores prácticas entre los analistas de métodos específicos y sostener la investigación científica. Al mismo tiempo, el trabajo a fecha de hoy no ha abordado la demanda del informe del NRC (2009) a examinar la validez científica fundamental y la fiabilidad de muchos métodos forenses utilizados cada día en los tribunales. El resto de nuestro informe se centra en ese tema.

## 3. El papel de la validez científica en los tribunales

El tema central de este informe es la validez científica de las pruebas de ciencia forense —más específicamente, la prueba de métodos científicos de comparación de características (por ejemplo, en muestras de ADN, huellas latentes, marcas balísticas y otros vestigios)—. La fiabilidad de los métodos para interpretar las pruebas es una consideración fundamental en toda ciencia. En consecuencia, cada especialidad científica dispone de un dominio específico de comprensión de lo que implica la validez científica de los métodos bien desarrollado.

El concepto de validez científica también juega un importante papel en el sistema jurídico. Particularmente, como se dijo en la sección 1, las *Federal Rules of Evidence* (FRE) (Reglas Federales de la Prueba) exigen que el testimonio pericial sobre ciencia forense sea resultado de «principios y métodos fiables» y de que estos hayan sido «fiablemente aplicados ... a los hechos del caso».

Este informe explica los criterios científicos para la validez científica en el caso de métodos forenses de comparación de características, tanto para su uso dentro del sistema jurídico como en el ámbito de los que trabajan para fortalecer los fundamen-

---

<sup>80</sup> Centro de Ciencia Forense de la Defensa, oficina del jefe científico, informe anual dlla sección de investigación, 5 de enero de 2016.

tos científicos de esas disciplinas. Antes de profundizar en esa explicación científica, proporcionamos en esta sección un resumen muy breve, dirigido principalmente a científicos y a lectores no especializados, sobre cuestiones jurídicas y términos relevantes en el ámbito estadounidense, así como sobre la naturaleza de esta intersección entre el derecho y la ciencia.

### 3.1. Evolución de los criterios de admisibilidad

A lo largo del siglo xx, el enfoque para decidir la admisibilidad de las pruebas científicas evolucionó en respuesta a los avances de la ciencia. En 1923, en el caso *Frye v. United States*,<sup>81</sup> el Tribunal de Apelación del Distrito de Columbia consideró la admisibilidad del testimonio concerniente a los resultados de un supuesto «detector de mentiras», un test basado en la presión sanguínea sistólica que fue precursor del polígrafo. Después de describir el dispositivo y su operación, el Tribunal rechazó el testimonio, estableciendo:

Si bien los Tribunales recorren un largo camino para admitir el testimonio pericial deducido de un principio o descubrimiento científico bien reconocido, aquello de lo que parte la deducción debe estar lo suficientemente establecido para haber ganado la aceptación general en el área del conocimiento específica a la que pertenece<sup>82</sup>.

El Tribunal estimó que el test sistólico «no había aún conseguido tal nivel de establecimiento y reconocimiento científico entre las autoridades fisiológicas y psicológicas» y, por consiguiente, fue declarado inadmisibile.

Más de medio siglo después, las *Federal Rules of Evidence* (FRE) fueron promulgadas en 1975 para regular los litigios civiles y penales en los tribunales federales. La Regla 702, en su forma original, estableció que:

Si el conocimiento científico, técnico o cualquier otro conocimiento especializado ayudará al Juez a comprender la prueba o a determinar el hecho en disputa, un testigo cualificado por su conocimiento, habilidad, experiencia, entrenamiento o educación puede testificar en forma de una opinión o de otro modo<sup>83</sup>.

Hubo un considerable debate entre los abogados litigantes, los jueces y los académicos sobre si la regla adoptaba el criterio *Frye* o establecía un nuevo criterio<sup>84</sup>. En 1993, la Corte Suprema de los Estados Unidos trató de resolver estas cuestiones en su histórico fallo del caso *Daubert v. Merrell Dow Pharmaceuticals*. Interpretando la Regla 702, la Corte Suprema mantuvo que las FRE reemplazaban a *Frye* como criterios de admisibilidad de la prueba pericial en los tribunales federales. El Tribunal rechazó

<sup>81</sup> *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923).

<sup>82</sup> *Ibid.*, 1014.

<sup>83</sup> Act of January 2, 1975, Pub. Law No. 93-595, 88 Stat. 1926 (1975). Véase: [federalevidence.com/pdf/FRE\\_Amendments/1975\\_Orig\\_Enact/1975-Pub.L.\\_93-595\\_FRE.pdf](http://federalevidence.com/pdf/FRE_Amendments/1975_Orig_Enact/1975-Pub.L._93-595_FRE.pdf)

<sup>84</sup> Véanse: GIANNELLI, 1980; McCABE, 1996: 455-82; PAGE, TAYLOR y BLENKIN, 2011: 913-7.

«la aceptación general» como el criterio de admisibilidad y en su lugar mantuvo que la admisibilidad del testimonio científico dependía de su fiabilidad científica.

Mientras que *Frye* establecía a los jueces deferir al juicio de la comunidad científica relevante, *Daubert* asignó a los jueces el papel de «guardianes» encargados de asegurar que el testimonio pericial «tuviera un fundamento fiable»<sup>85</sup>.

El Tribunal dictaminó que para su admisión «el juez debe determinar ... si el razonamiento o la metodología subyacente al testimonio es científicamente válida»<sup>86</sup>. Identificó cinco factores que un juez debe, entre otros, considerar ordinariamente a la hora de valorar la validez de la metodología subyacente. Estos factores son: (1) si la teoría o la técnica puede comprobarse (y ha sido comprobada); (2) si la teoría o la técnica ha estado sujeta a revisión por pares y publicación; (3) la tasa de error conocida o potencial de la técnica científica específica; (4) la existencia y mantenimiento de estándares que controlen la operación de la técnica; y (5) el grado de aceptación de la técnica en la comunidad científica relevante.

El Tribunal Supremo en el caso *Daubert* también señaló que los jueces para admitir los ofrecimientos de testimonio científico pericial debían considerar otras reglas aplicables, incluyendo la:

— Regla 403, que permite la exclusión de una prueba relevante «si su valor probatorio es sustancialmente superado por el peligro de un prejuicio injusto, de una confusión en los asuntos a tratar, o lleve a engaño al jurado» (señalando que la prueba pericial puede ser «tan contundente como engañosa a causa de la dificultad en valorarla»); y

— Regla 706, que permite al tribunal, a su propia discreción, solicitar la asistencia de un analista de su elección<sup>87</sup>.

El Congreso modificó la Regla 702 en el año 2000 para hacerla más precisa, y realizó ulteriores cambios de estilo en el 2011. En su actual versión, la regla 702 impone cuatro requisitos:

Un testigo cualificado como experto por su conocimiento, habilidad, experiencia, entrenamiento o educación, puede testificar en forma de opinión o de cualquier otro modo si:

- (a) el conocimiento científico, técnico, o cualquier otro especializado que tenga el experto ayuda al juez a comprender las pruebas o a determinar los hechos en cuestión;
- (b) el testimonio se fundamenta en suficientes hechos o datos;
- (c) el testimonio es resultado de principios y métodos fiables;
- (d) el analista ha aplicado fiablemente los principios y métodos a los hechos del caso.

El llamado «*Advisory Committee*»<sup>\*</sup> de las FRE realizó una Nota a la Regla 702 especificando también una serie de factores de fiabilidad que completaba los cinco

<sup>85</sup> *Daubert*, en 587.

<sup>86</sup> *Daubert*, en 580. Véase también, Nota 9 («En un caso en el que haya una prueba científica, la fiabilidad de la prueba estará fundamentada en la *validez científica*». [énfasis en el original]).

<sup>87</sup> *Daubert*, en 595, citando a Weinstein, 138 F.R.D., en 632.

factores enumerados en *Daubert*. Entre esos factores se encuentra «si el área experta invocada por el perito es conocida por alcanzar resultados fiables»<sup>88, 89</sup>.

Muchos estados han adoptado reglas de prueba que siguen en sus aspectos clave a las reglas federales. Esas reglas son actualmente el derecho en aproximadamente la mitad de los estados, mientras que otros continúan rigiéndose por el criterio *Frye* o variaciones de este<sup>90</sup>.

### 3.2. Validez de los fundamentos y validez en su aplicación

Como se describió en *Daubert*, el sistema jurídico prevé un diálogo importante entre el derecho y la ciencia:

El análisis judicial previsto en la Regla 702 es, lo enfatizamos, flexible. Su tema principal es la validez científica —y, por tanto, la relevancia y fiabilidad probatoria— de los principios que subyacen a la propuesta ofrecida<sup>91</sup>.

Entonces, tanto consideraciones de carácter jurídico como científico juegan roles importantes:

---

\* N. del T.: la Corte Suprema de los Estados Unidos estableció por primera vez un comité asesor de normas en junio de 1935 para ayudar a redactar las Reglas Federales de Procedimiento Civil, que entraron en vigor en 1938. Los Comités Asesores sobre el Reglamento de Apelación, Quiebra, Civil, Procedimiento Penal y las Reglas sobre la prueba realizan un estudio continuo de las normas que les competen y recomiendan cambios en la denominada *Judicial Conference*, a través de un Comité Permanente de Práctica y Procedimiento. Cfr. Committee Membership Selection | United States Courts (uscourts.gov)

<sup>88</sup> Véase: Nota del *Advisory Committee* a la FRE 702 (2000). Los siguientes factores pueden ser relevantes bajo la Regla 702: si la investigación subyacente se realizó independientemente del litigio; si el experto, a partir de una premisa aceptada, infirió injustificadamente una conclusión infundada; si el experto ha dado cuenta adecuadamente de explicaciones alternativas obvias; si el experto fue tan cuidadoso como lo hubiera sido en su trabajo profesional, fuera de su actividad pagada en un litigio; y si el área de especialización a la que el experto considera pertenecer es conocida por alcanzar resultados fiables [énfasis añadido].

<sup>89</sup> Esta nota ha sido señalada como un apoyo a los esfuerzos para desafiar campos enteros de la ciencia forense, incluidas las huellas dactilares y las comparaciones de cabellos. Véase: GIANELLI, 2003: 1096.

<sup>90</sup> Incluso bajo *Frye*, lo que piensan los científicos sobre el significado de la fiabilidad es relevante. *Frye* exige que una técnica o método científico «tenga aceptación general» en la comunidad científica relevante para ser admisible. Como una cuestión científica, la comunidad científica relevante para asesorar sobre la fiabilidad de las ciencias de comparación de características incluye a metrólogos (incluidos estadísticos), así como otros científicos de la física o de la vida desde disciplinas sobre las que se fundamentan los métodos específicos. Es importante señalar que la comunidad no está constreñida a los científicos forenses que practican el método específico. Por ejemplo, el tribunal del caso *Frye* evaluó si el detector de mentiras ofrecido había conseguido un «reconocimiento científico estable entre las autoridades médicas y psicológicas» y no entre analistas en detectores de mentiras. *Frye v. United States*, 293 F. 1013 (D. C. Cir. 1923).

<sup>91</sup> *Daubert*, p. 594.

(1) La admisibilidad del testimonio pericial depende de un criterio gradual como, entre otras cosas, de si cumple ciertos requisitos *jurídicos* previstos en la Regla 702. Estas decisiones sobre admisibilidad son de competencia exclusiva de los tribunales.

(2) Ahora bien, como se advirtió más arriba, el tema general de análisis de los jueces bajo la regla 702 es «la validez científica.» Está en las competencias de la comunidad científica proporcionar guías sobre los estándares *científicos* para la validez científica.

El PCAST no opina aquí sobre los criterios jurídicos, sino que busca solo clarificar los estándares científicos subyacentes. Para lograr una completa claridad sobre lo que nos proponemos, hemos adoptado términos específicos para referirnos a los estándares científicos para dos tipos clave de validez científica, que queremos que se correspondan, como estándares científicos, con los estándares jurídicos previstos en la Regla 702 (c, d):

(1) por «validez de los fundamentos» queremos decir que el *estándar* científico se corresponde con el criterio jurídico de que la prueba se base en «principios y métodos fiables», y

(2) por «validez en la aplicación» queremos decir que el *estándar* científico se corresponde con el criterio jurídico de que el experto o analista «aplique fiablemente los principios y métodos».

En el siguiente apartado discurriremos los estándares científicos de estos conceptos. Cerramos esta sección advirtiendo que la respuesta a la cuestión de la validez científica de las disciplinas forenses es importante no solo para los tribunales sino también porque establece estándares de calidad que se extienden a todas ellas, afectando a la práctica y definiendo la investigación necesaria.

#### 4. Criterios científicos para la validez y fiabilidad de los métodos forenses de comparación de características

— En este informe, el PCAST ha elegido centrarse en la validez y fiabilidad de un área específica dentro de la ciencia forense: los métodos forenses de comparación de características. Hemos hecho eso porque es posible e importante hacerlo para esta específica clase de métodos.

— Es *posible* porque la comparación de características es una actividad científica común y la ciencia posee estándares claros para determinar si tales métodos son fiables. En particular, los métodos de comparación de características pertenecen a la disciplina de la metrología —la ciencia de la medida y su aplicación—<sup>92,93</sup>.

<sup>92</sup> International Vocabulary of Metrology – Basic and General Concepts and Associated Terms, 2012.

<sup>93</sup> Que los métodos de comparación de características pertenecen a la disciplina de la metrología es claro partiendo del hecho de que el NIST —cuya misión es asistir a la Nación mediante «ciencia de me-

— Es *importante* porque en la pasada década se ha evidenciado que una deficiente comparación de características forense ha conducido a numerosos errores judiciales<sup>94</sup>. También se ha revelado que los problemas no son debidos simplemente a una pobre ejecución de unos pocos analistas sino más bien al hecho de que la fiabilidad de muchos métodos de comparación de características forenses nunca ha sido significativamente evaluada<sup>95</sup>.

Comparado con muchos tipos de testimonio pericial, el basado en métodos de comparación de características posee exclusivos peligros de llevar al error a los miembros de un jurado por dos razones:

— La gran mayoría de los miembros de un jurado no tienen una habilidad independiente para interpretar el valor probatorio de los resultados basados en la detección, comparación y frecuencia de la evidencia científica. Si dos mitades de una

---

dición avanzada, estándares y tecnología», y que es el laboratorio metrológico líder mundial— es la sede del Gobierno Federal que se ocupa de los esfuerzos de investigación en ciencia forense. Los programas del NIST incluyen investigación interna, financiación externa de la investigación externa, conferencias y preparación de los materiales y estándares de referencia. Véanse: [www.nist.gov/public\\_affairs/mission.cfm](http://www.nist.gov/public_affairs/mission.cfm) y [www.nist.gov/forensics/index.cfm](http://www.nist.gov/forensics/index.cfm). Los métodos de comparación de características conllevan determinar si dos conjuntos de características concuerdan dentro de una cierta tolerancia en la medida.

<sup>94</sup> La reexaminación mediante ADN de casos cerrados de informes ha conducido a la exoneración de 342 condenados hasta la fecha, incluyendo 20 que habían sido sentenciados a muerte, y a la identificación de 147 verdaderos culpables. Véase: Innocence Project, «DNA Exonerations in the United States», [www.innocenceproject.org/dna-exonerations-in-the-united-states](http://www.innocenceproject.org/dna-exonerations-in-the-united-states). La revisión de estos casos ha revelado que, aproximadamente la mitad, se deben en parte al testimonio de analistas, basados en métodos que no habían sido sometidos a un escrutinio científico significativo o que incluían conclusiones científicas afirmadas con niveles de seguridad inválidos. Véanse: GROSS Y SHAFFER, 2012. Disponible en: [www.law.umich.edu/special/exoneration/Documents/exonerations\\_us\\_1989\\_2012\\_full\\_report.pdf](http://www.law.umich.edu/special/exoneration/Documents/exonerations_us_1989_2012_full_report.pdf); GARRETT Y NEUFELD, 2009: 1-97; National Research Council, 2009, 42-43. La naturaleza de los asuntos se ilustra en los documentos citados mediante ejemplos específicos: Levon Brooks y Kennedy Brewer, cada uno condenado por asesinato por asesinato de dos niños, separadamente, en los años 90, casi enteramente sobre la base de un testimonio experto de análisis de marcas de mordedura, que conllevó que pasaran más de 13 años en prisión antes de que el análisis de ADN identificara al auténtico culpable, que confesó ambos crímenes; Santae Tribble, condenado por asesinato después de que el analista del FBI testificara que los cabellos encontrados en una máscara de media implicaba a Tribble en el crimen al coincidir (con su cabello) «en todas las características microscópicas», pasó más de 20 años en prisión antes de que el análisis de ADN revelara que ninguno de los 13 cabellos examinados pertenecían a Tribble y que uno procedía de un perro; Jimmy Ray Bromgard de Montana pasó 15 años de prisión por violación antes de que el análisis de ADN mostrase que los cabellos recogidos de la cama de la víctima y considerados coincidentes con los de Bromgard no podían ser suyos; Stephan Cowans, condenado por asesinato por disparar a un oficial de policía de Boston después de que dos analistas en huellas dactilares testificaran que la impresión del pulgar dejada por el autor del crimen era «única e idéntica» (a la de Bromgard), pasó más de 5 años en prisión antes de que el análisis de ADN sobre múltiples muestras le excluyera como autor del crimen; y Steven Barnes, del norte del estado de New York, pasó 20 años en prisión por un delito de violación y asesinato que no cometió después de que un analista en criminalística testificara que una superposición fotográfica entre la tela de los jeans de la víctima y una impresión del camión de Barnes mostraba patrones que eran «similares» y que los cabellos recogidos del camión eran similares a los cabellos de la víctima.

<sup>95</sup> Véase: Sección 5.

nota de rescate que fueran coincidentes se encontrasen en la escena del crimen y en el domicilio del acusado, los miembros del jurado podrían confiar en su propia experiencia para valorar la probabilidad de que dos trozos rotos coincidan si no procediesen de un solo original. Si un testigo describiera a un criminal como «alto y de cabello tupido», los miembros del jurado podrían realizar un juicio razonable sobre cuánta gente se ajusta a esa descripción. Pero, si un testigo experto dijera que en dos muestras de ADN el tercer exón del gen *DYNC1H1* tiene precisamente 174 nucleótidos de longitud, la mayoría de los miembros del jurado no tendrían manera de saber si deberían estar impresionados por la coincidencia; dependerían completamente de las declaraciones revestidas con el manto de la ciencia hechas por los analistas (No deberían sentirse impresionados por la aseveración precedente: en el marcador de la cadena de ADN citada, más del 99.9 por ciento de la gente tiene un fragmento del tamaño indicado)<sup>96</sup>.

— El potencial impacto perjudicial es inusualmente alto porque los miembros del jurado suelen sobreestimar el valor probatorio de un cotejo positivo entre muestras. De hecho, el mismo DOJ históricamente sobreestimó el valor probatorio de cotejos positivos en su prolongada posición, reconocida ahora como inapropiada, sobre que los cotejos de huellas latentes eran «infalibles»<sup>97</sup>. De forma similar, un jefe de la unidad de huellas dactilares del FBI testificó que el FBI tenía «una tasa de error de 1 por cada 11 millones de casos»<sup>98</sup>. En un experimento online, los investigadores preguntaron a miembros del jurado ficticios que estimaran la frecuencia de que un científico forense experimentado y cualificado pudiera erróneamente concluir que dos muestras de tipos específicos provinieran de la misma persona cuando realmente lo eran de personas distintas. Los miembros del jurado ficticios creyeron que tales errores era probable que ocurrieran en 1 de cada 5.5 millones de comparaciones de huellas dactilares; 1 de cada millón de comparaciones de marcas de mordedura; y 1 de cada 100.000 comparaciones de escritura manuscrita<sup>99</sup>. Mientras que se desconocen las tasas de error de la mayoría de esas técnicas, todos los indicios apuntan a que las tasas de error son de mayores órdenes de magnitud. Por ejemplo, los propios estudios del FBI sobre análisis de huellas latentes indican tasas de error en el rango de 1 en varios cientos de cotejos<sup>100</sup> (Como consecuencia de que el término «match» o coincidencia es probable que implique un valor probatorio inapropiadamente alto, debiera utilizarse un término más neutral para expresar la creencia del analista de que las dos muestras proceden de la misma fuente. Sugerimos el término «*identificación propuesta*» para transmitir adecuadamente la conclusión del analista y, a la vez, la posibilidad de que quizá sea errónea. Usaremos este término a lo largo de este informe).

<sup>96</sup> Véase: Base de datos ExAC: <http://exac.broadinstitute.org/gene/ENSG0000019710>.

<sup>97</sup> Véase: [www.justice.gov/olp/file/861906/download](http://www.justice.gov/olp/file/861906/download).

<sup>98</sup> *U.S. v. Baines* 573 F.3d 979 (2009) en 984.

<sup>99</sup> KOEHLER, 2016. Disponible en [papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2817443](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2817443)

<sup>100</sup> Véase: Sección 5.4.

Este apartado recoge las conclusiones del PCAST sobre los criterios científicos de validez científica. Las conclusiones se basan en el principio fundamental del «método científico» —aplicable a lo largo de toda la ciencia— respecto a que el conocimiento científico válido *sólo* puede obtenerse a través de la comprobación *empírica* de las proposiciones específicas<sup>101</sup>. Las conclusiones del PCAST de esta sección pueden resumirse brevemente como sigue:

La validez y fiabilidad científicas requieren que un método haya sido sometido a comprobación empírica, bajo condiciones apropiadas a su pretendido uso, que proporcione estimaciones válidas de la frecuencia con que el método llega a conclusiones incorrectas. Para los métodos de comparación subjetiva de características se requieren estudios de caja negra apropiadamente diseñados, en los que muchos analistas tomen decisiones sobre muchas pruebas independientes (típicamente incluyendo muestras «cuestionadas» y una o más muestras «conocidas») y en los que las tasas de error se determinen. Sin las apropiadas estimaciones de acierto en las comparaciones, una proposición del analista que diga que dos muestras comparadas son similares —o, incluso, indistinguibles— carece de significado científico: carece de valor probatorio y tiene potencial considerable de causar un impacto perjudicial. Nada —ni personal formado y con experiencia, ni prácticas profesionales— pueden sustituir a una demostración empírica adecuada sobre el nivel de acierto.

La sección se organiza como sigue:

— La primera sección describe la distinción entre dos tipos fundamentalmente diferentes de métodos de comparación de características: objetivos y subjetivos.

— Las cinco secciones siguientes tratan sobre los criterios científicos para los dos tipos de validez científica: la validez de los fundamentos y la validez en la aplicación.

— Las dos secciones finales reflejan puntos de vista mantenidos en la comunidad forense.

#### 4.1. *Métodos de comparación de características: métodos objetivos y subjetivos*

Un método de comparación de características es un procedimiento mediante el cual un analista busca determinar si un vestigio probatorio (por ejemplo, de la escena del crimen) está o no asociado con una muestra de origen conocido (por ejemplo, del sospechoso)<sup>102</sup> basándose en características similares.

Los métodos de comparación de características pueden clasificarse en objetivos o subjetivos. Por métodos de comparación de características objetivos entendemos

<sup>101</sup> Por ejemplo, el diccionario online de Oxford define *scientific method* como «un método o procedimiento que ha caracterizado a las ciencias naturales desde el siglo xvii, consistente en la observación, medida y experimentación sistemáticas, y en la formulación, comprobación y modificación de hipótesis». «Scientific Method» *Oxford Dictionaries Online*. Oxford University Press (última consulta el 19 de agosto de 2016).

<sup>102</sup> Una «muestra de origen conocido» se refiere a un individuo u objeto específico (por ejemplo, un neumático o una pistola).

aquellos consistentes en procedimientos definidos con un detalle suficientemente estandarizado y cuantificable que pueden ejecutarse bien mediante un sistema automático, bien mediante analistas que emiten escasos o ningún juicio. Por métodos subjetivos entendemos métodos que incluyen procedimientos clave que conllevan juicios humanos significativos —por ejemplo, sobre qué características han de seleccionarse o cómo determinamos si las características son suficientemente similares para considerarlas para proponer una identificación—.

Los métodos objetivos son, en general, preferibles a los subjetivos. Los análisis que dependen de juicios humanos (en lugar de una medida cuantitativa de similitud) son obviamente más susceptibles al error humano, los sesgos y la variabilidad de ejecución entre los analistas<sup>103</sup>. Por contra, los métodos objetivos y cuantificados tienden a producir una mayor precisión, repetibilidad y fiabilidad, incluyendo una reducción en la variación de los resultados entre los analistas. Los métodos subjetivos pueden evolucionar hacia métodos objetivos o ser reemplazados por ellos<sup>104</sup>.

#### 4.2. *Validez de los fundamentos: requerimiento de estudios empíricos*

Para que un método metrológico sea científicamente válido y fiable, debe mostrar se el procedimiento en el que consiste, fundamentado en estudios empíricos, ser *repetible*, *reproducibile* y *preciso*, en los niveles que ha sido medido, y ser apropiado para la aplicación requerida<sup>105, 106</sup>.

El método no necesita ser perfecto, pero es claramente *esencial* que su precisión se haya medido mediante comprobaciones empíricas apropiadas y sea suficientemente alta para ser adecuada a la aplicación. Sin una apropiada estimación de su precisión, un método metrológico es inútil —porque no hay forma de interpretar sus resultados—. La importancia de conocer la precisión del método fue resaltada en el informe del NRC (2009) sobre la ciencia forense y por un informe NRC del año 2010 sobre tecnologías biométricas<sup>107</sup>.

<sup>103</sup> DROR, 2016: 121-127.

<sup>104</sup> Por ejemplo, antes del desarrollo de la prueba de alcoholemia, los tribunales tenían que confiar exclusivamente en el testimonio de los oficiales de policía y de otros que, a su vez, se basaban en indicios conductuales de embriaguez y en la presencia de alcohol en la respiración. El desarrollo de tests químicos objetivos condujo a un cambio de estándares subjetivos por objetivos.

<sup>105</sup> National Physical Laboratory, 2010. Disponible en: [www.npl.co.uk/upload/pdf/NPL-Beginners-Guide-to-Measurement.pdf](http://www.npl.co.uk/upload/pdf/NPL-Beginners-Guide-to-Measurement.pdf); PAVESE, 2009.

<sup>106</sup> Los métodos de comparación de características que dan resultados erróneos con demasiada frecuencia tienen, por definición, bajo valor probatorio. Como se dijo más arriba, el impacto perjudicial es probable que de este modo supere al valor probatorio.

<sup>107</sup> «El reconocimiento biométrico es una tarea inherentemente probabilística ... En consecuencia, incluso aunque la tecnología y el sistema incrustado en ella estén funcionando como se diseñaron, hay inevitablemente incertidumbre y existe riesgo de error», National Research Council, 2010: viii-ix.

**CUADRO 2.  
DEFINICIÓN DE TÉRMINOS CLAVE**

Por «repetible», entendemos que, con probabilidad conocida, un perito obtiene el mismo resultado cuando analiza muestras procedentes de las mismas fuentes.

Por «reproducibile» entendemos que, con probabilidad conocida, diferentes peritos obtienen el mismo resultado cuando analizan las mismas muestras.

Por «preciso», entendemos que, con probabilidades conocidas, un perito obtiene resultados correctos tanto para (1) muestras procedentes de la misma fuente (verdaderos positivos) como para (2) muestras procedentes de distintas fuentes (verdaderos negativos).

Por «fiabilidad», entendemos repetibilidad, reproducibilidad y precisión <sup>108</sup>.

Por «científicamente válido», entendemos que un método ha demostrado, basado en estudios empíricos, ser fiable con niveles de repetibilidad, reproducibilidad y precisión apropiados a la aplicación pretendida.

Por un «estudio empírico» entendemos una prueba en la que se usa un método para analizar un gran número de conjuntos independientes de muestras, similares en aspectos relevantes a las encontradas en los casos en los que hemos de trabajar, para estimar la repetibilidad, reproducibilidad y precisión del método.

Por «estudio de caja negra», entendemos un estudio empírico que valora un método subjetivo mediante examinadores que analizan muestras y opinan sobre el origen o las similitudes entre las muestras.

Para cumplir los criterios científicos de validez en los fundamentos se requieren dos elementos:

(1) un procedimiento reproducible y consistente para (a) identificar características en las muestras que son aportadas como pruebas; (b) comparar las características de dos muestras; y (c) determinar, basándose en la similitud entre las características de dos muestras, si las muestras deben considerarse aptas para una propuesta de identificación («regla de coincidencia»).

(2) medidas empíricas, resultantes de múltiples estudios independientes, de (a) la tasa de falsos positivos del método —es decir, la probabilidad de que se declare una propuesta de identificación entre muestras que realmente proceden de *diferentes* fuentes— y (b) la sensibilidad del método —es decir, la probabilidad de que declare una propuesta de identificación entre muestras que realmente proceden de la *misma* fuente—.

Trataremos sobre estas cuestiones en su momento.

<sup>108</sup> Queremos resaltar que «fiabilidad» tiene también un significado próximo a «consistencia», usado en estadística —es decir, hasta qué punto un método produce el mismo resultado, con independencia de si el resultado es preciso—. Ese no es el sentido en que «fiabilidad» es utilizado en este informe, o en la ley.

#### 4.2.1. Procedimientos reproducibles y consistentes

Para que un método sea objetivo, deben definirse *cada uno* de los tres pasos (identificación de características, comparación de características y reglas de coincidencia) de forma precisa, reproducible y consistente. Los analistas forenses deben identificar las características relevantes de la misma manera y alcanzar el mismo resultado. Deben comparar las características de la misma forma cuantitativa. Para plantear una propuesta de identificación deben calcular si las características de una muestra ofrecida como prueba y las de una muestra procedente de un sospechoso se encuentran dentro de una pre-especificada tolerancia de medición (regla de coincidencia)<sup>109</sup>. De un método objetivo podemos establecer la validez del fundamento de cada uno de los pasos individuales midiendo su precisión, reproducibilidad y consistencia.

Para los métodos subjetivos, los procedimientos han de ser cuidadosamente definidos —pues conllevan un sustancial juicio humano—. Por ejemplo, diferentes analistas pueden reconocer o centrarse en diferentes características, pueden otorgar diferente importancia a las mismas características y pueden tener diferentes criterios a la hora de hacer propuestas de identificación. Como consecuencia de que los procedimientos de identificación de características, las reglas de coincidencia y las determinaciones frecuenciales de las características no están objetivamente especificados, el procedimiento en su conjunto ha de tratarse como una especie de «caja negra» dentro de la cabeza del analista.

Los métodos subjetivos requieren cuidadoso escrutinio, normalmente más, pues su fuerte dependencia de los juicios humanos lleva consigo que sean especialmente vulnerables al error humano, a la inconsistencia entre analistas y a los sesgos cognitivos. En las disciplinas forenses de comparación de características, los sesgos cognitivos incluyen el fenómeno de que, en ciertos entornos, los seres humanos (1) pueden tender de modo natural a centrarse sobre las similitudes entre las muestras y a descartar las diferencias y (2) pueden también ser influenciados por información no pertinente y presiones externas sobre el caso<sup>110</sup> (Estos últimos temas se ilustran con la equivocada identificación de una huella latente por el FBI relacionada con los atentados con bombas en los trenes de Madrid, que se aborda en la página 34).

Puesto que la caja negra en la cabeza del analista no puede examinarse directamente con respecto a su fundamentación en la ciencia, la validez de los fundamentos

---

<sup>109</sup> Si se declara que una fuente *no* comparte las mismas características, el test la excluye. La regla de coincidencia debe elegirse cuidadosamente. Si se elige una «regla de coincidencia» demasiado estricta, muestras que realmente proceden de la misma fuente se declararán como no coincidentes (falsos negativos). Si es demasiado laxa, entonces el método no tendrá un alto poder discriminatorio porque la probabilidad de coincidencia aleatoria será demasiado alta (falsos positivos).

<sup>110</sup> Véanse, por ejemplo: BORODITSKY, 2007: 118- 128; HASSIN, 2001: 728-731; MEDIN, GOLDSTONE y GENTNER, 1993: 254-278; TVERSKY, 1977: 327-352; KIM, NOVEMSKY Y DHAR, 2012: 225-229; LARKEY y MARKMAN, 2005: 1061-1076; MEDIN, GOLDSTONE y MARKMAN, 1995: 1-19; GOLDSTONE, 1994: 125-157; NOSOFSKY, 1986: 39-57.

de los métodos subjetivos *solo* puede establecerse a través de estudios empíricos sobre las acciones del analista para determinar si pueden proporcionar respuestas precisas; a tales estudios se les llama «estudios de caja negra» (Cuadro 2). En los estudios de caja negra, a muchos analistas se les presentan muchos problemas de comparación independientes —normalmente, incluyendo muestras «cuestionadas» y una o más muestras «conocidas»— y se les pregunta que declaren si las muestras cuestionadas proceden de la misma fuente que la de las muestras conocidas<sup>111</sup>. Los investigadores, posteriormente, determinan cuántas veces los analistas llegan a conclusiones erróneas.

Como ejemplo sobresaliente, el FBI condujo recientemente un estudio de caja negra en análisis de huellas latentes, implicando a 169 analistas que examinaron 744 pares de huellas dactilares, y publicó los resultados del estudio en una revista científica líder<sup>112</sup>. (Algunos científicos forenses han advertido que prestar demasiada atención a aspectos subjetivos de los métodos forenses —tales como estudios de sesgos cognitivos y cajas negras— puede distraer del fin de mejorar el conocimiento sobre las características objetivas de la evidencia forense y el desarrollo de métodos verdaderamente objetivos<sup>113</sup>. Otros han resaltado que eso no es actualmente un problema, porque los esfuerzos y los medios económicos para afrontar los retos asociados a los métodos forenses subjetivos son muy limitados<sup>114</sup>).

#### 4.2.2. Medidas empíricas de precisión

Es necesario que tengamos medidas empíricas apropiadas de la tasa de falsos positivos y de la sensibilidad del método. Como se explicó en el Apéndice A, resulta necesario conocer esas dos medidas para estimar el valor probatorio de un método. La tasa de falsos positivos es la probabilidad de que el método declare una propuesta de identificación entre muestras que realmente procedan de fuentes *diferentes*. Por ejemplo, una tasa de falsos positivos del 5% significa que dos muestras de fuentes *diferentes* serán (debido a las limitaciones del método) incorrectamente declaradas como procedentes de la misma fuente un 5% de las veces. (La cantidad igual a 1 menos la tasa de falsos positivos —95% en el ejemplo— es referida como la especificidad).

La sensibilidad del método es la probabilidad de que el método declare una propuesta de identificación entre muestras que realmente procedan de una *misma* fuente. Por ejemplo, una sensibilidad del 90% significa que dos muestras procedentes de una misma fuente se declararán como tales el 90% de las veces y declaradas como procedentes de fuentes diferentes el 10% de las veces. (La última cantidad se conoce como tasa de falsos negativos).

---

<sup>111</sup> Las respuestas pueden expresarse en términos de «cotejo positivo / cotejo negativo / inconclusión» o «identificación / exclusión / inconclusión».

<sup>112</sup> ULERY, HICKLIN, BUSCAGLIA y ROBERTS, 2011: 7733-7738.

<sup>113</sup> CHAMPOD, 2014: 107-109.

<sup>114</sup> RISINGER, THOMPSON, JAMIESON, ET AL., 2014: 508-509.

La tasa de falsos positivos es especialmente importante porque los resultados de falsos positivos pueden conducir directamente a condenas erróneas<sup>115</sup>. En algunas circunstancias es posible estimar una tasa de falsos positivos relacionada con características específicas de las pruebas del caso (Por ejemplo, la probabilidad de coincidencia aleatoria calculada en los análisis de ADN depende en parte del genotipo específico observado en una muestra de la prueba. La tasa de falsos positivos de los análisis de huellas latentes depende de la calidad de la huella latente). Para otros métodos de comparación de características, solo es posible tener una estimación media de la tasa de falsos positivos en las muestras.

Para métodos objetivos, la tasa de falsos positivos se compone de dos fuentes distinguibles —cotejos positivos por azar (en los que muestras de diferentes fuentes tienen, sin embargo, *características* que caen dentro de las tolerancias de las reglas de coincidencia objetivas)— y fallos humano-técnicos (en los que las muestras presentan características fuera de las reglas de coincidencia, pero en los que existió, sin embargo, una declaración de propuesta de identificación debido a un fallo humano o técnico).

Para métodos objetivos en los que la probabilidad de coincidencia aleatoria sea muy baja (como sucede en los análisis de ADN), la tasa de falsos positivos aplicada a un caso real estará dominada por la tasa de fallos humano-técnicos —que puede ser cientos de veces mayor—. En cuanto a los métodos subjetivos, los tipos de errores descritos —coincidencia aleatoria y fallos humano-técnicos— ocurren igualmente, pero sin una «regla de coincidencia» objetiva, no podrán distinguirse esas dos fuentes de error. Para el establecimiento de la validez de los fundamentos es esencial realizar estudios de caja negra que midan empíricamente la tasa de error global de muchos analistas. (Véase el Cuadro 3 sobre el término «error»).

### CUADRO 3. LOS SIGNIFICADOS DE «ERROR»

El término «error» tiene diferentes significados en las ciencias y en el derecho, que pueden conducir a confusión. En sentido jurídico, el término «error» a menudo implica una falta —por ejemplo, que una persona haya cometido un error cuando pudo haber sido evitado si él o ella hubiera seguido un procedimiento adecuado o que una máquina haya dado un resultado erróneo que pudo haber sido evitado si hubiera sido apropiadamente calibrada—. En ciencias, el término «error» también incluye la situación en la que el propio procedimiento, siendo propiamente aplicado, no da la respuesta correcta debido a una circunstancia imprevista.

Cuando aplicamos un método forense de comparación de características con el fin de valorar si dos muestras procedieron o no de la misma fuente, los cotejos positivos por

<sup>115</sup> Véase nota 94. Bajo algunas circunstancias, los resultados de falsos negativos pueden contribuir a también a condenas erróneas.

azar y los fallos humano-técnicos se consideran, desde el punto de vista estadístico, como «errores» porque pueden inducir a conclusiones incorrectas.

Los estudios que se diseñan para estimar la tasa de falsos positivos y la sensibilidad de un método son necesariamente realizados utilizando solo un limitado número de muestras. Como consecuencia, no pueden proporcionar valores «exactos» para esas cantidades (y no deben sostener que lo hacen), sino solo «intervalos de confianza», cuyos límites reflejan, respectivamente, el rango de valores que es razonablemente compatible con los resultados. Cuando se informa de un falso positivo a un jurado, es científicamente importante establecer que por encima del límite de confianza superior unilateral del 95% se refleja el hecho de que la tasa de falsos positivos real pudiera ser razonablemente tan alta como ese valor. (Para más información, véase el Apéndice)<sup>116</sup>.

Con frecuencia los estudios categorizan los resultados como concluyentes (por ejemplo, identificación o exclusión) o no concluyentes (sin una determinación realizada)<sup>117</sup>. Cuando se informa de la tasa de falsos positivos a un jurado, es científicamente importante calcular la tasa fundamentándose en la proporción en los dictámenes concluyentes, en lugar de en la proporción en todos los dictámenes. Esto es apropiado porque la evidencia utilizada contra un acusado estará basada generalmente en dictámenes concluyentes en lugar de dictámenes inconclusos. Para ilustrar este punto, consideremos un caso extremo en el que un método se ha testeado 1000 veces y se han obtenido 990 resultados inconclusos, 10 falsos positivos y ningún resultado correcto. Sería engañoso decir que la tasa de falsos positivos fue del 1% (10/1000 dictámenes). En su lugar, deberíamos informar que en el 100% de los dictámenes concluyentes se obtuvieron falsos positivos (10/10 dictámenes).

Mientras que los estudios científicos exploratorios pueden ser de muchas formas, los estudios de validación científica —dirigidos a valorar la validez y la fiabilidad de un método metrológico para una aplicación de comparación de características forense particular— deben satisfacer una serie de criterios, los cuales se describen en el Cuadro 4.

<sup>116</sup> El límite superior de confianza incorpora propiamente la precisión de la estimación basada en el tamaño muestral. Por ejemplo, si un estudio no encontró errores en 100 pruebas, sería engañoso informar al jurado que la tasa de error es cero. De hecho, si las pruebas son independientes, el límite superior de confianza del 95% de la verdadera tasa de error es el 3%. De acuerdo con eso, al jurado se le debe decir que la tasa de error pudiera ser tan alta como el 3% (es decir, 1 por cada 33). La verdadera tasa pudiera ser más alta, pero con bastante poca probabilidad (menor al 5%). Si el estudio fuera más pequeño, el límite superior de confianza del 95% será más alto. Para un estudio sin errores en 10 pruebas, el límite superior de confianza del 95% es el 26% —es decir, la tasa de falsos positivos real pudiera ser aproximadamente de 1 por cada 4— (véase Apéndice A).

<sup>117</sup> Véase: Sección 5.

**CUADRO 4.**  
**CRITERIOS CLAVE PARA ESTUDIOS DE VALIDACIÓN**  
**QUE ESTABLEZCAN LA VALIDEZ DE LOS FUNDAMENTOS**

Los estudios de validación científica —dirigidos a valorar la validez y la fiabilidad de un método metrológico para una particular aplicación de comparación de características forense— han de satisfacer una serie de criterios.

(1) Los estudios deben tener en cuenta un número suficientemente grande de examinadores y han de basarse en conjuntos de muestras *conocidas* y *representativas* suficientemente *grandes* de poblaciones *relevantes* que reflejen el rango de características o combinaciones de características que ocurran en la aplicación. En particular, los conjuntos de muestras deben ser:

(a) representativos de la calidad de las muestras de pruebas vistas en casos reales. (Por ejemplo, si un método va a ser utilizado sobre huellas dactilares latentes parciales y distorsionadas, debemos determinar la *probabilidad de coincidencia aleatoria* —es decir, la probabilidad de que la coincidencia ocurra por casualidad— para huellas dactilares latentes parciales y distorsionadas; la probabilidad de coincidencia aleatoria para huellas dactilares completas escaneadas, o incluso de huellas latentes de muy alta calidad, no sería relevante.

(b) elegidos de poblaciones relevantes a casos reales. Por ejemplo, para características de muestras biológicas, la tasa de falsos positivos ha de determinarse sobre la totalidad de la población estadounidense y para los principales grupos étnicos, como se hace con los análisis de ADN.

(c) suficientemente grande para proporcionar estimaciones apropiadas de las tasas de error.

(2) Los estudios empíricos deben conducirse de forma que ni el examinador ni aquellos con los que el examinador interactúa tengan información alguna sobre la respuesta correcta.

(3) El diseño del estudio y el marco de referencia del análisis se deben especificar de antemano. En los estudios de validación es inapropiado modificar el protocolo a posteriori basándose en los resultados<sup>118</sup>.

(4) Los estudios empíricos deben ser dirigidos o supervisados por individuos u organizaciones que no tengan beneficio en la entrega de los estudios<sup>119</sup>.

<sup>118</sup> La situación análoga en medicina es un test clínico para asegurarnos de la seguridad y eficacia de un fármaco para una aplicación particular. En el diseño de los tests clínicos, la FDA requiere que los criterios para los análisis sean previamente especificados y las notas que se añaden a posteriori a los análisis comprometen la validez del estudio. Véase: FDA Guidance, 2016. Disponible en: [www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm446729.pdf](http://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm446729.pdf); ALOSH, FRITSCH, HUQUE, ET AL., 2015: 286-303; FDA Guidance, 2013. Disponible en: [www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm373750.htm](http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm373750.htm); FDA Guidance for Industry, 1998. Disponible en: [www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm073137.pdf](http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm073137.pdf); Pocock, 1983.

<sup>119</sup> Cuando se inician ensayos clínicos, quien patrocina la prueba (una farmacéutica, una empresa que fabrica dispositivos o una empresa biotecnológica o, en algunos casos, una institución académica)

(5) Los datos, el software y los resultados de los estudios de validación deben estar disponibles para permitir que otros científicos revisen las conclusiones.

(6) Para asegurarse que las conclusiones son reproducibles y robustas, debe haber múltiples estudios de grupos distintos que lleguen a conclusiones similares.

Una medida empírica de las tasas de error no es solo una característica deseable, es esencial para determinar si un método es válido en sus fundamentos. En la ciencia, un procedimiento de prueba —como el test de embarazo o la comprobación de si el agua pudiera estar contaminada— no se considera válido hasta que no se haya medido su fiabilidad. Por ejemplo, necesitamos conocer la frecuencia con la que el test de embarazo indica que hay un embarazo cuando no es cierto y viceversa. Los mismos principios científicos se aplican, en no menor grado, a las pruebas forenses, que pueden contribuir a que un acusado pierda su vida o su libertad.

Es particularmente importante no inferir las tasas de error desde la casuística forense, sino que debe determinarse basándose en muestras de las que se conozca la respuesta correcta. Por ejemplo, el exjefe de la Unidad de Huellas Dactilares del FBI testificó que el FBI tenía «una tasa de error de 1 cada 11 millones de casos» basándose en el hecho de que la agencia era conocida por haber cometido solo un error en los últimos 11 años, en los que habían realizado 11 millones de identificaciones<sup>120</sup>. La falacia es obvia: el analista simplemente *asumió sin pruebas* que habían salido a la luz todos los errores en casos reales.

¿Por qué es esencial conocer la tasa de falsos positivos y la sensibilidad de un método? Porque sin la apropiada medida empírica de la precisión de un método, el hecho de que dos muestras, en un caso particular, muestren similares características, *no tiene valor probatorio* —y, como se dijo anteriormente, puede tener un impacto perjudicial considerable, porque los miembros del jurado adjudicarán un significado equivocado a la observación—<sup>121</sup>.

Desde una perspectiva científica, la absoluta necesidad de datos empíricos es elegantemente expresada en una analogía del juez de distrito norteamericano John Potteren en su decisión del caso *U.S. v. Yee (1991)*, un caso temprano sobre el uso del análisis de ADN:

---

financia e inicia el estudio, pero la prueba es dirigida por individuos independientes del patrocinador (frecuentemente, médicos académicos), para asegurar la fiabilidad de los datos generados en el estudio y minimizar la posibilidad de que haya sesgos. Véanse, por ejemplo, 21 C.F.R. § 312.3 y 21 C.F.R. § 54.4(a).

<sup>120</sup> *U.S. v. Baines* 573 F.3d 979 (2009) en 984.

<sup>121</sup> Bajo la regla federal de la prueba (Fed. R. Evid.) n° 403, la prueba debe ser excluida si «su valor probatorio es sustancialmente sobrevalorado por el temor de un perjuicio injusto».

Sin la valoración de la probabilidad, el jurado no sabe qué hacer ante el hecho de que dos patrones comparados coincidan: el jurado no sabe si los patrones son tan comunes como las imágenes con dos ojos o tan únicos como la Mona Lisa<sup>122,123</sup>.

#### 4.3. *Validez de los fundamentos: requerimiento para el testimonio científicamente válido*

Debe considerarse obvio —pero merece que se enfatice— que una vez que un método ha sido establecido como válido en sus fundamentos por estar basado en estudios empíricos apropiados, la forma de asegurar que las afirmaciones sobre la precisión de un método y el valor probatorio de las identificaciones propuestas sean válidas es que estén basadas en tales estudios empíricos. *Los enunciados que aseveren o impliquen mayor certeza que la demostrada mediante la evidencia empírica, son científicamente inválidos*. Los analistas forenses deben, por tanto, informar de sus hallazgos sobre una propuesta de identificación con claridad y mesura, explicando en cada caso que el hecho de que dos muestras satisfagan los criterios de un método para una propuesta de identificación no implica necesariamente que las muestra procedan de la misma fuente. Si la tasa de falsos positivos de un método se cifra en 1 cada 50, los analistas no pueden deducir que el método sea capaz de producir resultados con un mayor grado de precisión.

Es preocupante que los peritos en ocasiones vayan más allá de la evidencia empírica sobre la frecuencia de las características —incluso hasta el extremo de sostener o deducir que una muestra procedió de una específica fuente con casi certeza absoluta o con ella, a pesar de carecer de base para tales opiniones—<sup>124</sup>. Desde el punto de vista de la validez científica, a los analistas no se les debe permitir establecer ante los tribunales que puedan emitir conclusiones con certeza o con cercanía a la certeza (tales como tasas de error igual a «cero», «imperceptiblemente pequeña», «esencialmente cero», «despreciable», «mínima» o «microscópica»; «certeza del 100%» o «un razonable grado de certeza científica»; o una identificación «hasta la exclusión de todas las fuentes»<sup>125</sup>.

---

<sup>122</sup> *U.S. v. Yee*, 134 F.R.D. 161 (N.D. Ohio 1991).

<sup>123</sup> Algunos tribunales han dictaminado que no hay peligro en admitir pruebas basadas en comparación de características sobre la base de que los jurados pueden ver las características con sus propios ojos y decidir por sí mismos si comparten características. *U.S. v. Yee* muestra por qué ese razonamiento es falaz: los jurados no tienen forma de conocer con qué frecuencia dos muestras diferentes comparten sus características, y con qué nivel de especificidad.

<sup>124</sup> Como se dijo con anterioridad, la larga historia de afirmaciones exageradas sobre la precisión de los métodos forenses incluye el propio enunciado inicial de DOJ de que los análisis de huellas latentes son «infalibles», que DOJ ha juzgado después como inapropiado. [www.justice.gov/olp/file/861906/download](http://www.justice.gov/olp/file/861906/download).

<sup>125</sup> COLE, 2004. Véase también: National Research Council, 2009: 87, 104, y 143.

Que esos testimonios sean inapropiados es certeramente advertido en una analogía que la juez del Tribunal de Apelación del Distrito de Columbia Catharine Easterly emite en su opinión concurrente en el caso *Williams v. United States*, un caso en el que un analista testificó que las marcas sobre ciertos proyectiles eran únicas y procedentes de una pistola recogida en el apartamento del acusado:

Tal y como están las cosas actualmente, una declaración de certeza con respecto a la coincidencia de patrones de marcas de herramientas tiene el mismo valor probatorio que la visión de un vidente: no refleja nada más que la fe individual sin fundamento en lo que cree que es verdad. Esa no es la clase de prueba en la que podamos confiar en buena conciencia, particularmente en casos penales, donde exigimos que esté probado—prueba real—más allá de la duda razonable, precisamente porque lo que se juega vale mucho<sup>126</sup>.

En ciencias, la afirmación de que un método metrológico es más preciso que lo que ha sido demostrado empíricamente se considera con razón como una mera especulación, no una conclusión válida que merezca credibilidad.

#### 4.4. *Ni la experiencia ni las prácticas profesionales pueden sustituir la validez de los fundamentos*

En algunas situaciones, un científico puede ser científicamente capaz de emitir juicios basándose principalmente en su «experiencia» y en su «juicio personal». Basándose en su experiencia, un cirujano podría estar cualificado para ofrecer una opinión sobre si otro doctor actuó apropiadamente o no en la sala de operaciones o un psiquiatra podría estar científicamente cualificado para ofrecer una opinión sobre si un acusado está mentalmente bien para defenderse.

Por contraste, ni la «experiencia» ni el «juicio personal» pueden utilizarse para establecer la validez científica y fiabilidad de un método metrológico, como lo es un método de comparación de características forense. La frecuencia con la que un patrón particular o un conjunto de características puede observarse en diferentes muestras, que es un elemento esencial para emitir conclusiones, no es un asunto de «juicio personal». Es una cuestión empírica para la que solo es relevante información empírica. Más aún, la «experiencia» de un analista forense a partir de muchos casos reales no es informativa —porque las «respuestas correctas» no se conocen ordinariamente en los casos reales y por eso los analistas no conocen con precisión con qué frecuencia declaran cotejos positivos erróneos y no es fácil que perfeccionen su precisión aprendiendo de sus errores mientras hacen informes periciales—.

Es importante destacar que las buenas prácticas profesionales —tales como la existencia de sociedades profesionales, programas de certificación, programas de acreditación, artículos revisados por pares, protocolos estandarizados, tests de apti-

---

<sup>126</sup> *Williams v. United States*, DC Court of Appeals, dictaminado el 16 de enero de 2016 (voto concurrente de Easterly).

tud y códigos éticos— no pueden ser sustitutos de una prueba real sobre la validez y fiabilidad científicas<sup>127</sup>.

Análogamente, la manifestación de *confianza* que haga un analista basado en su experiencia profesional o afirmaciones sobre el *consenso* entre quienes practican una técnica respecto la fiabilidad de sus procedimientos no sustituye a las estimaciones de las tasas de error a partir de estudios relevantes. Para que un método sea *fiable*, se requiere una prueba empírica de su validación, como se mencionó anteriormente.

Finalmente, lo que se ha dicho subraya que la validez científica de un método debe valorarse dentro de un marco científico más amplio del que aquel forma parte (por ejemplo, la ciencia de la medida en el caso de métodos de comparación de características). El hecho de que los analistas de marcas de mordedura defiendan la validez de los cotejos de marcas de mordedura significa poco.

#### 4.5. Validez en la aplicación: elementos clave

La validez de los fundamentos significa que un método, *en principio*, puede ser fiable. Validez en la aplicación significa que el método ha sido fiablemente aplicado *en la práctica*. Se trata del concepto científico que se corresponde con el requerimiento jurídico previsto en la Regla 702(d): que un experto «haya aplicado fiablemente los principios y métodos a los hechos del caso».

Desde un punto de vista científico, existen ciertos criterios esenciales para establecer que un analista forense ha aplicado fiablemente un método a los hechos de un caso. Esos elementos se describen en el Cuadro 5.

#### CUADRO 5. CRITERIOS CLAVE PARA LA VALIDEZ EN LA APLICACIÓN

El analista forense debe haber demostrado que es capaz de aplicar fiablemente el método y debe realmente haberlo hecho.

La demostración de que un analista es *capaz* de aplicar fiablemente el método es crucial —especialmente para métodos subjetivos, en los que los juicios humanos juegan un papel central. Desde el punto de vista científico, la capacidad para aplicar fiablemente un método puede demostrarse solo mediante comprobación empírica que mida con qué frecuencia el analista consigue el resultado correcto. (Los tests de aptitud se tratan más extensamente en las páginas 68-71). La determinación de si un analista ha *realmente* aplicado fiablemente el método requiere que el procedimiento que ha sido de hecho utilizado en el caso, los resultados obtenidos y las notas de laboratorio estén disponibles para una revisión científica llevada a cabo por terceros.

Las aserciones sobre la probabilidad de que las características observadas ocurran por casualidad han de ser científicamente válidas.

<sup>127</sup> Por ejemplo, tanto las disciplinas científicas como las pseudocientíficas emplean tales prácticas.

El analista forense debe informar de la tasa global de falsos positivos y de la sensibilidad del método establecida en los estudios de validez de los fundamentos y debe demostrar que las muestras utilizadas en los estudios sobre su fundamentación son relevantes para los hechos del caso<sup>128</sup>.

Cuando sea aplicable, el analista debe informar la probabilidad de coincidencia aleatoria basada en las características específicas observadas en el caso.

Un analista no debe hacer aseveraciones o realizar implicaciones que vayan más allá de la evidencia empírica y de la aplicación de los principios estadísticos válidos para esa prueba.

#### 4.6. Validez en la aplicación: pruebas de aptitud

Incluso aunque un método sea válido en sus fundamentos, hay muchas razones que explican por qué los analistas no siempre consiguen resultados correctos<sup>129</sup>. Como se dijo anteriormente, el único modo de establecer científicamente que un analista es capaz de aplicar un método válido en sus fundamentos es a través de pruebas empíricas apropiadas que midan la frecuencia con la que el analista consigue resultados correctos. A esas pruebas empíricas se les conoce frecuentemente como «pruebas de aptitud». Queremos advertir que el término «prueba de aptitud» es a veces utilizado para referirse a otros tipos de pruebas —tales como (1) tests que determinan si un analista de laboratorio sigue los pasos establecidos en un protocolo, sin que se *valore* la precisión de sus conclusiones, y (2) práctica de ejercicios que ayuden a los analistas de laboratorio a mejorar sus habilidades poniendo de manifiesto sus errores, sin que reflejen con precisión las circunstancias de los casos reales.

---

<sup>128</sup> Por ejemplo, para el análisis de ADN, se conoce que la frecuencia de las variantes genéticas varía entre los grupos étnicos; de este modo, es importante que la recogida de muestras refleje los grupos étnicos relevantes para el caso en cuestión. Para huellas latentes, el riesgo de declarar falsamente una identificación es más alto cuando poseen escasa calidad; así, para ser relevantes, los conjuntos de muestras utilizados para estimar la precisión deben basarse en huellas latentes impresas comparables en calidad y completitud al caso en cuestión.

<sup>129</sup> J.J. Koehler ha enumerado una serie de problemas que podrían producirse, en principio: las características pueden medirse de forma deficiente; las muestras pueden intercambiarse, etiquetarse o codificarse mal, alterarse o contaminarse; el equipamiento puede descalibrarse; fallos técnicos y fallos en general pueden ocurrir sin previo aviso y sin que claramente se manifiesten; y los resultados pueden leerse equivocadamente, malinterpretarse, registrarse deficientemente, mal etiquetarse, mezclarse, descolocarse o descartarse.

KOEHLER, «Forensics or fauxrensic? Ascertaining accuracy in the forensic sciences», [papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2773255](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2773255) (última consulta el 28 de junio de 2016).

En este informe, utilizamos el término prueba de aptitud para referirnos a pruebas empíricas en curso que «evalúen la capacidad y la ejecución del trabajo de los analistas»<sup>130, 131, 132</sup>.

Los tests de aptitud deben realizarse en condiciones que sean representativas de casos reales y sobre muestras de las que se conozca la verdadera respuesta, y que sean representativas del entero rango de tipos de muestras y calidad que es probable encontrar en un caso real en la aplicación pretendida. (Por ejemplo, el hecho de que un analista supere una prueba de aptitud que contenga un análisis de muestras simples y procedentes de una única fuente de ADN, no demuestra que sea capaz de hacer análisis de ADN de mezclas complejas del tipo que se encuentra en casos reales; véanse pp. 91-100).

Para garantizar su integridad, los tests de aptitud deben ser supervisados por una tercera parte desinteresada que carezca de incentivo institucional o económico que sesgue su realización. Resaltamos que los llamados *Testing Services* han declarado que la comunidad forense prefiere que estas pruebas no sean demasiado difíciles<sup>133</sup>.

Como se advirtió previamente, la tasa de falsos positivos consiste tanto en la tasa de coincidencias aleatorias como en fallos técnicos o humanos. Para algunas tecnologías (como el ADN), los últimos pueden ser cientos de veces mayores que los primeros.

La prueba de aptitud es especialmente importante en los métodos subjetivos: como consecuencia de que el procedimiento no está basado únicamente en criterios objetivos, sino que se apoya en el juicio humano, es inherentemente vulnerable al error y a la variabilidad inter-analista. Cada analista debe pasar esos tests, dado que en estudios empíricos se han constatado considerables diferencias en cuanto a la precisión entre los analistas<sup>134, 135</sup>.

---

<sup>130</sup> ASCLD/LAB Supplemental Requirements for Accreditation of Forensic Testing Laboratories. [des.wa.gov/SiteCollectionDocuments/About/1063/RFP/Add7\\_Item4ASCLD.pdf](https://des.wa.gov/SiteCollectionDocuments/About/1063/RFP/Add7_Item4ASCLD.pdf).

<sup>131</sup> Queremos resaltar que las pruebas de aptitud no están dirigidas a estimar las tasas de error inherentes a un método; esas tasas deben valorarse mediante estudios de validez de los fundamentos.

<sup>132</sup> El término en inglés «*proficiency testing*» debe también distinguirse de «*competency testing*», que es «la evaluación del conocimiento y capacidad de una persona antes de realizar informes de forma independiente en casos forenses reales». [des.wa.gov/SiteCollectionDocuments/About/1063/RFP/Add7\\_Item4ASCLD.pdf](https://des.wa.gov/SiteCollectionDocuments/About/1063/RFP/Add7_Item4ASCLD.pdf).

<sup>133</sup> Christopher Czyryca, presidente de *Collaborative Testing Services, Inc.*, la empresa líder en Estados Unidos de tests de aptitud, ha declarado públicamente que «la comunidad prefiere los tests sencillos». Reunión de agosto de 2015 de la *National Commission on Forensic Science*, una presentación en el *Accreditation and Proficiency Testing Subcommittee* (Subcomité de Acreditación y Pruebas de Aptitud). [www.justice.gov/ncfs/file/761061/download](http://www.justice.gov/ncfs/file/761061/download).

<sup>134</sup> Por ejemplo, un estudio sobre decisiones de cotejos de huellas latentes en 2011 observó que los analistas frecuentemente diferían sobre si las huellas eran adecuadas para alcanzar una conclusión. ULERY, HICKLIN, BUSCAGLIA, ET AL., 2011: 7733-7738.

<sup>135</sup> No es suficiente hacer las pruebas de aptitud con voluntarios en un laboratorio, porque los analistas que mejor realizan su tarea son más proclives a participar. KOEHLER, «Forensics or fauxrensic?»

Los problemas contenidos en los test utilizados en las pruebas de aptitud deben hacerse públicos una vez que la prueba haya terminado, lo que permitiría a los científicos valorar si el test ha sido apropiado y adecuado para el propósito perseguido.

Por último, las pruebas de aptitud deben *idealmente* conducirse a ciegas —es decir, con muestras insertadas dentro del flujo de trabajo de forma que los analistas desconozcan que están siendo sometidos a una prueba— (Por ejemplo, la *Transportation Security Administration* (Administración de Seguridad en el Transporte) lleva a cabo pruebas blindadas de envío de armas y explosivos dentro de maletas en los puntos de control para ver la frecuencia con que los controladores los detectan). Es un hecho comprobado en muchos campos de trabajo (incluyendo el de los análisis de huellas latentes) que cuando los individuos son conscientes de que están siendo inspeccionados, realizan sus tareas de forma distinta que cuando los hacen en el transcurso de un día de trabajo ordinario (conocido como «efecto Hawthorne») <sup>136, 137</sup>.

Mientras que un test de aptitud ciego es ideal, hay desacuerdo en la comunidad forense sobre su viabilidad en todos los entornos. Por una parte, los laboratorios varían considerablemente con respecto al tipo de casos que reciben, cómo se gestiona y procesa la prueba y qué información se proporciona a un analista sobre la prueba o el caso en cuestión. En consecuencia, los tests de aptitud ciegos interlaboratorios pueden ser difíciles de diseñar y dirigir a gran escala <sup>138</sup>. Por otra parte, los tests de aptitud ciegos se han utilizado en análisis de ADN <sup>139</sup> y algunos selectos laboratorios han comenzado a implementar este tipo de pruebas internamente, como parte de sus programas de aseguramiento de la calidad <sup>140</sup>. Resaltamos que los tests de aptitud ciegos son mucho más fáciles de adoptar en laboratorios que hayan adoptado «procedimientos de gestión de contextos» para reducir sesgos contextuales <sup>141</sup>.

El PCAST cree que los tests de aptitud ciegos de los analistas forenses deben fomentarse vigorosamente, con la esperanza de que tengan amplio uso, al menos en los laboratorios grandes, en los próximos cinco años. Sin embargo, el PCAST cree que

---

Ascertaining accuracy in the forensic sciences.» [papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2773255](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2773255) (última consulta el 28 de junio de 2016).

<sup>136</sup> Sobre el efecto Hawthorne, véanse, por ejemplo: BRACHT Y GLASS, 1968: 437-474; WEECH, Y GOLDHOR, 1982: 305-324; BOUCHET, GUILLEMIN Y BRAINCON, 1996: 15-20; MANGIONE-SMITH, ELLIOTT, McDONALD, ET AL., 2002: 1603-1623; MUJIS, 2006: 53-74; McCARNEY, WARNER, ILIFFE, ET AL., 2007.

<sup>137</sup> Para demostraciones de que los analistas forenses cambian su comportamiento cuando conocen que están siendo inspeccionados de una determinada forma, véase LANGENBURG, 2009.

<sup>138</sup> Algunos de los retos asociados con tests ciegos de aptitud interlaboratorio podrían abordarse si los laboratorios forenses avanzaran hacia un sistema donde el conocimiento de un caso por parte de los analistas fuera limitado al dominio de la información relevante.

<sup>139</sup> Véase: PETERSON, LIN, HO, ET AL., 2003: 32-40.

<sup>140</sup> Por ejemplo, el *Houston Forensic Science Center* ha implementado pruebas ciegas de aptitud rutinarias para sus balísticos y la unidad de análisis químico y está planeando llevar a cabo pruebas similares para sus analistas en ADN y en huellas latentes.

<sup>141</sup> Para más detalle, véase [www.justice.gov/ncfs/file/888586/download](http://www.justice.gov/ncfs/file/888586/download).

no es aún realista exigir esos tests de aptitud ciegos porque los procedimientos aún no han sido diseñados y evaluados.

Mientras no se utilicen otros tests de aptitud, más allá de los no ciegos, para garantizar la validez en la aplicación, resulta científicamente importante informar de esta limitación, incluyendo a los jurados —porque como se dijo más arriba, los tests de aptitud no ciegos tienden a sobrevalorar la precisión porque los analistas conocían que estaban siendo inspeccionados—.

#### 4.7. *Puntos de vista no empíricos en la comunidad forense*

Mientras que la validez científica de los métodos metrológicos requiere demostración empírica de la precisión, históricamente han existido esfuerzos en la comunidad forense para justificar aproximaciones no empíricas. Esto requiere atención particular porque tales puntos de vista son en ocasiones erróneamente codificados en políticas o en prácticas. Estos modos heterodoxos de ver las cosas repetidamente involucran cuatro temas que revisamos a continuación.

##### 4.7.1. Teorías» de identificación

Un argumento común es que las prácticas forenses deben considerarse válidas porque se asientan sobre «teorías» científicas similares a las leyes fundamentales de la física, que deben ser aceptadas porque han sido comprobadas y no «falsadas»<sup>142</sup>.

Un ejemplo es la «Teoría de la Identificación en su relación con las marcas de herramientas», emitida en 2011 por la Asociación de analistas de Marcas de Armas y Herramientas<sup>143, 144</sup>. Afirma rotundamente:

1. La teoría de la identificación, en lo que se refiere a la comparación de las marcas de herramientas, permite opiniones de procedencia que deben hacerse cuando la superficie única de dos marcas de herramientas se halla en «concordancia suficiente».

2. Esta «concordancia suficiente» se relaciona con la duplicación significativa de marcas de herramientas aleatorias, como lo demuestra la correspondencia de un patrón o una combinación de patrones de contornos de superficies. La significación está determinada por el examen comparativo de dos o más conjuntos de patrones de contorno de superficies, compuesto por picos individuales, crestas y surcos. Específicamente, la altura o profundidad relativa, la anchura, la curvatura y la relación espacial entre los picos, crestas y surcos individuales dentro de un conjunto de contornos

<sup>142</sup> Véanse: [www.swggun.org/index.php?option=com\\_content&view=article&id=66:the-foundations-of-firearm-and-toolmarkidentification&catid=13:other&Itemid=43](http://www.swggun.org/index.php?option=com_content&view=article&id=66:the-foundations-of-firearm-and-toolmarkidentification&catid=13:other&Itemid=43) y [www.justice.gov/ncfs/file/888586/download](http://www.justice.gov/ncfs/file/888586/download).

<sup>143</sup> ASSOCIATION OF FIREARM AND TOOL MARK EXAMINERS, 2011: 287.

<sup>144</sup> El análisis de armas de fuego es tratado en detalle en la sección 5.

de superficie se definen y comparan con las correspondientes características del segundo conjunto de contornos de superficie. La concordancia es significativa cuando la hallada entre las características individuales supera la mejor concordancia demostrada entre marcas de herramientas que se sabe que han sido producidas por diferentes herramientas y es consistente con la concordancia demostrada por marcas de herramientas que se sabe que han sido producidas por la misma herramienta. La declaración de «concordancia suficiente» entre dos marcas de herramientas significa que la concordancia entre las características individuales es de una cantidad y calidad tal que la probabilidad de que otra herramienta pudiera haber hecho la marca es tan remota como para considerarla una imposibilidad práctica.

3. Actualmente, la interpretación de identificación/individualización es de naturaleza subjetiva, fundamentada en principios científicos y basada en la formación y experiencia del analista.

La declaración claramente no es una teoría científica, definida por la Academia Nacional de Ciencias como «una comprehensiva explicación de algún aspecto de la naturaleza sustentada por un vasto cuerpo de evidencia»<sup>145</sup>. En su lugar, es una declaración de que los analistas que aplican aproximaciones subjetivas pueden individualizar el origen de una marca de herramienta con precisión. Sin embargo, no es una «teoría» lo que se necesita. Lo que se necesita son pruebas empíricas que permitan ver cuán bien funciona el método.

Más importante aún, el método mencionado es circular. Declara que un analista puede establecer que dos marcas de herramientas tienen un «origen común» cuando sus características están en «concordancia suficiente». Entonces, define «concordancia suficiente» como el que se da cuando el analista considera la «imposibilidad práctica» de que las marcas de herramienta tengan distintos orígenes. (En respuesta a la objeción del PCAST sobre esta circularidad, el laboratorio del FBI replicó que: «la imposibilidad práctica es la certeza que existe cuando hay suficiente concordancia en la cantidad y calidad de características individuales»<sup>146</sup>. Esta respuesta no resolvió la circularidad).

#### 4.7.2. Apoyándose en «entrenamiento y experiencia» en vez de en la demostración empírica de la precisión

Muchos analistas creen honestamente que son capaces de hacer juicios precisos sobre identificación basándose en su capacitación y experiencia. Esta noción es explícita en la Teoría de la Identificación de AFTE, que resalta que la interpretación es subjetiva por naturaleza, «basada en la capacitación y experiencia del analista». Análogamente, el libro de texto líder en el análisis de huellas de calzado afirma,

<sup>145</sup> Véase: [www.nas.edu/evolution/TheoryOrFact.html](http://www.nas.edu/evolution/TheoryOrFact.html)

<sup>146</sup> Comunicación del Laboratorio del FBI a PCAST (6 de junio de 2016).

Las identificaciones positivas pueden hacerse con tan solo una característica identificativa aleatoria, pero solo si esa característica es confirmable; tiene suficiente definición, claridad y propiedades; está en la misma ubicación y orientación en la suela del zapato; y *en opinión del analista experimentado, no se presentaría en otro zapato*<sup>147</sup> [énfasis añadido].

En efecto, dice el libro, la identificación positiva depende de que el analista diga que hay una identificación *positiva*.

La «experiencia» es un fundamento inadecuado para emitir dictámenes sobre si dos conjuntos de propiedades pudieran haber sido producidos por (o encontrados en) fuentes diferentes. Incluso si los analistas pudieran recordar con suficiente detalle todos los patrones o conjunto de características que hubieran visto no tendrían modo de saber con precisión en qué casos dos patrones procedieron de fuentes diferentes, porque las respuestas correctas rara vez se conocen en casos reales.

La falacia de confiar en la «experiencia» fue evidente en el testimonio de un jefe de la Unidad de Huellas Dactilares del FBI (explicado anteriormente) al declarar que el FBI tenía «una tasa de error de 1 cada 11 millones de casos», basándose en el hecho de que la agencia era únicamente consciente de haber cometido un único error<sup>148</sup>. Por contraste, recientes estudios empíricos realizados por el laboratorio del FBI (explicados en la sección 5) indican tasas de error de aproximadamente uno en varios cientos.

La «capacitación» es incluso un fundamento más débil. El mero hecho de que un individuo haya sido capacitado en un método no significa que el método, en sí mismo, sea científicamente válido, ni que el individuo sea capaz de producir respuestas fiables cuando aplique el método.

#### 4.7.3. Centrarse en la «unicidad» en lugar de hacerlo en la «precisión»

Muchas disciplinas forenses de comparación de características se basan en la premisa de que diversos conjuntos de características (por ejemplo, impresiones dactilares, marcas de herramienta sobre proyectiles, la dentición humana, etc.) son «únicas»<sup>149</sup>.

<sup>147</sup> BODZIAK, 2000.

<sup>148</sup> *U.S. v. Baines* 573 F.3d 979 (2009) en 984

<sup>149</sup> Para huellas dactilares, véanse, por ejemplo, WERTHEIM, 2002: 669 («La ley de unicidad biológica afirma que no puede ocurrir en la naturaleza una replicación exacta (la naturaleza nunca se repite a sí misma) y, por consiguiente, nunca una entidad biológica será exactamente igual que otra») y BUDOWLE, BUSCAGLIA y PERLMAN 2006. («El uso de las comparaciones de crestas papilares de la piel como medio de identificación se fundamenta en que asume que el patrón de crestas papilares de la piel es único y permanente»). Para armas de fuego, véanse, por ejemplo, RIVA y CHRISTOPE, 2014: 637 («La capacidad para identificar un arma de fuego como fuente de un casquillo o proyectil de origen desconocido se fundamenta en dos premisas que constituyen el fundamento científico de la disciplina. La primera asume la unicidad de las impresiones dejadas por las armas») y SWGGUN Admis Kit (ARK): Foundational Overview of Firearm/Toolmark Identification. disponible en: [afte.org/resources/swggun-ark](http://afte.org/resources/swggun-ark) («La base para la identificación en la Identificación de Marcas de Herramientas está fundamentada

La literatura en ciencia forense contiene muchos estudios sobre la «unicidad» que hacen todo lo posible para establecer la exactitud de esa premisa<sup>150</sup>. Por ejemplo, un artículo de 2012 que estudió 39 zapatillas para correr de la marca Adidas Supernova Classic (talla 12), calzadas por un único corredor durante 8 años, corriendo un mismo tiempo diariamente y sobre los mismos tipos de superficie<sup>151</sup>. Después de aplicar crema de zapatos negra en las suelas de los zapatos, el autor pidió al corredor que produjera cuidadosamente marcas de pisadas en hojas de papel legal sobre un suelo de madera. El autor demostró que era posible distinguir pequeñas diferencias identificativas entre las marcas de pisada producidas por los diferentes pares de zapatillas.

No obstante, los estudios de unicidad olvidan la cuestión fundamental. El tema no es si los *objetos* o las *propiedades* difieren; seguramente lo hacen si miramos con suficiente nivel de detalle. El tema es en qué medida y bajo qué circunstancias aplicando un método metrológico *los analistas* pueden *detectar* fiablemente diferencias significativas en las propiedades para identificar fiablemente si comparten un origen común. Los estudios de unicidad que se centran en las propiedades de las características en sí mismas no pueden nunca, por tanto, establecer si un *método* particular para medir y comparar propiedades es válido en sus fundamentos. Solo lo pueden hacer los estudios empíricos.

Más aún, no es *necesario* que las propiedades sean únicas para ser útiles porque reducen las fuentes posibles de una característica. Más bien, es esencial que haya evidencia empírica sobre la frecuencia con que un método atribuye incorrectamente la fuente de una característica.

#### 4.7.4. Disociando conclusiones sobre identificación desde estimaciones sobre precisión

Por último, algunos mantienen la opinión de que cuando la aplicación de un método científico conduce a una conclusión de propuesta de asociación o identificación, es *innecesario* informar al tribunal sobre la fiabilidad del método<sup>152</sup>. A modo

---

en el principio de unicidad . . . por el cual, todos los objetos son únicos por sí mismos, y, de este modo, pueden ser diferenciados unos de otros»). Para marcas de mordedura, véanse, por ejemplo, KIESER, BERNAL, WADDELL, ET AL., 2007: 671-677 («Hay dos postulados que subyacen en todos los análisis de marcas de mordedura: primero, que las características de la dentadura anterior actuante en una mordedura son únicas y, segundo, que esta unicidad es registrada con precisión en el material mordido.») y PRETTY, 2011 («El análisis de marcas de mordedura está basado en dos postulados: (a) las características dentales de los dientes anteriores actuantes en una mordedura son únicas entre individuos y (b) esta unicidad aseverada es transferida y grabada en la lesión.»).

<sup>150</sup> Algunos autores han criticado los intentos de afirmar la proposición de unicidad basada en observaciones, resaltando que se apoyan en el puro razonamiento inductivo, un método para la investigación científica que «cayó en desgracia durante la época de Sir Francis Bacon en el siglo XVI» PAGE, TAYLOR y BLENKIN, 2011: 12-18.

<sup>151</sup> WILSON, 2012: 194-204.

<sup>152</sup> Véase: [www.justice.gov/olp/file/861936/download](http://www.justice.gov/olp/file/861936/download).

de fundamento racional, se ha argumentado en ocasiones que no es posible medir las tasas de error perfectamente o que es imposible conocer la tasa de error en el *específico* caso que llevamos entre manos.

Esta noción es contraria al principio fundamental de validez científica en metrología —concretamente, la afirmación de que dos objetos tras ser comparados presenten las mismas propiedades (longitud, peso o patrón dactiloscópico) carece de sentido si no existe información cuantitativa sobre la fiabilidad del proceso de comparación—.

Es una práctica estándar en medicina estudiar e informar sobre tasas de error —tanto para establecer en principio la fiabilidad de un método como para valorar su implementación en la práctica—. Nadie argumenta que medir o informar sobre tasas de error clínicas sea inapropiado porque quizá no reflejen perfectamente la situación de un paciente *específico*. Si la transparencia sobre las tasas de error es apropiada para el acierto en la detección del tipo de sangre antes de una transfusión, es también apropiada para el acierto en el cotejo positivo entre muestras forenses —donde los errores pueden tener consecuencias similares a las amenazas contra la vida—.

Volveremos sobre este tema en la sección 8, donde veremos que la guía sobre testimonio experto recientemente propuesta por el DOJ está basada, en parte, sobre este modo de pensar científicamente inapropiado.

#### 4.8. *Puntos de vista empíricos en la comunidad forense*

Aunque algunos de los integrantes de la comunidad forense mantengan modos de pensar como el descrito en la anterior sección, un creciente segmento de la comunidad de ciencia forense ha respondido al informe NRC del año 2009 con un mayor reconocimiento sobre la necesidad de realizar estudios empíricos y con esfuerzos iniciales para llevarlos a cabo. Como ejemplos podemos citar algunos estudios de investigación publicados que han sido realizados por investigadores forenses, evaluaciones de necesidades de investigación por Grupos de Trabajo Científicos y Comités de la OSAC y declaraciones del NCFCS.

A continuación, resaltamos algunos ejemplos de artículos recientes realizados por científicos forenses: Investigadores de la Academia Nacional de Ciencias y de otros lugares (por ejemplo, SAKS y KOELER, 2005; SPINNEY, 2010) han argumentado que hay una urgente necesidad de desarrollar medidas objetivas de precisión en la identificación mediante huellas dactilares. Aquí presentamos esos datos<sup>153</sup>.

— La prueba de impresión de marcas de herramientas, por ejemplo, se ha presentado ante los tribunales con éxito durante décadas, pero su examen ha carecido de prueba científica, estadística, que de forma independiente corroboraría conclusiones basadas en características morfológicas (2-7). En nuestro estudio, aplicaremos méto-

<sup>153</sup> TANGEN, THOMPSON y MCCARTHY, 2011: 995-997.

dos de reconocimiento de patrones estadístico (por ejemplo, «machine learning») a los análisis de impresiones de marcas de herramienta<sup>154</sup>.

— El informe NAS llamó a incrementar las investigaciones en el área de las marcas de mordedura para demostrar que existe un grado de valor probatorio y para una posible restricción del uso de los análisis a la exclusión de individuos. Esta llamada ha de ser escuchada si la prueba de marcas de mordedura quiere ser defendida en adelante como disciplina<sup>155</sup>.

— El National Research Council of the National Academies (Consejo de Investigación Nacional de las Academias Nacionales) y las comunidades jurídica y de ciencias forenses han convocado a investigar con el objetivo de medir la precisión y fiabilidad de las decisiones de los analistas de huellas latentes, un problema complejo que constituye un reto que necesita un análisis sistemático. Nuestra investigación está dirigida al desarrollo de aproximaciones empíricas para estudiar este problema<sup>156</sup>.

— Pensamos que este informe debe impulsar a la comunidad jurídica a requerir que el emergente campo de la neuroimagen forense, incluyendo el detector de mentiras fundamentado en tecnología FMRI, tenga la apropiada fundamentación científica antes de que sea admitido por los tribunales<sup>157</sup>.

— Es preferible una solución empírica que trata al sistema [referido a las huellas acústicas de las voces] como una caja negra y sus salidas como valores puntuales<sup>158</sup>.

Análogamente, la OSAC y otros grupos han reconocido grietas críticas en investigación respecto a la evidencia que sostiene a diversas disciplinas de ciencia forense y ha comenzado a desarrollar planes para cerrar alguna de esas grietas. Vale la pena resaltar algunos ejemplos:

— Aunque se han desarrollado algunos estudios de validación de programas de análisis de armas de fuego y marcas de herramienta, la mayoría se han hecho con conjuntos de datos relativamente pequeños. Si se hiciera un gran estudio bien diseñado y tuviera suficiente participación, anticipamos que podrían aprenderse lecciones similares para las disciplinas de armas de fuego y marcas de herramienta<sup>159</sup>.

— No conocemos estudio alguno que valore la capacidad de las disciplinas de armas de fuego y marcas de herramientas para categorizar la evidencia correcta/con-

---

<sup>154</sup> PETRACO, SHENKIN, SPEIR, ET AL., 2012: 900-911.

<sup>155</sup> PRETTY Y SWEET, 2010: 38-44.

<sup>156</sup> ULERY, HICKLIN, BUSCAGLIA, ET AL., 2011: 7733-7738.

<sup>157</sup> LANGLEBEN Y MORIARTY, 2013: 222-234.

<sup>158</sup> MORRISON, ZHANG Y ROSE, 2011: 59-65.

<sup>159</sup> OSAC Research Needs Assessment Form (Formulario de evaluación de necesidades de investigación de la OSAC), «Study to Assess the Accuracy and Reliability of Firearm and Toolmark.» Emitido en octubre de 2015 (Aprobado en enero de 2016). Disponible en: [www.nist.gov/forensics/osac/upload/FATM-Research-NeedsAssessment\\_Blackbox.pdf](http://www.nist.gov/forensics/osac/upload/FATM-Research-NeedsAssessment_Blackbox.pdf).

sistentemente mediante características de clase, identificar marcas de subclase y descartar vestigios utilizando características individuales<sup>160</sup>.

— Actualmente no existe una valoración fiable de la fuerza discriminatoria de los tipos específicos de puntos característicos de las crestas de fricción<sup>161</sup>.

— Hasta la fecha, hay pocos datos científicos que cuantifiquen el riesgo global de cotejos negativos dudosos en las bases de datos AFIS. Es difícil crear estándares relativos a la suficiencia para realizar un examen o para realizar búsquedas en AFIS sin este tipo de investigación<sup>162</sup>.

— Se precisa investigación para averiguar si el desenmascaramiento secuencial reduce los efectos negativos de sesgos durante el análisis de huellas latentes<sup>163</sup>.

— El IAI ha buscado, durante muchos años, apoyo para realizar investigaciones que validasen muchos de los análisis comparativos conducidos por sus miembros como analistas forenses. Aunque existe una gran cantidad de evidencia empírica que apoya a esos análisis, no se ha hecho una validación independiente<sup>164</sup>.

La Comisión Nacional de Ciencia Forense ha reconocido análogamente la necesidad de realizar una rigurosa evaluación empírica de los métodos forenses en un documento que compila diversas opiniones al respecto aprobado por la comisión:

Todas las metodologías de la ciencia forense deben ser evaluadas por un cuerpo científico independiente que defina sus capacidades y limitaciones para responder con precisión y fiabilidad a una específica y claramente definida cuestión forense<sup>165</sup>.

El PCAST aplaude este creciente esfuerzo por centrarse en la evidencia empírica. Resaltamos que será necesaria una mayor inversión en investigación para alcanzar estos objetivos críticos (véase Sección 6).

---

<sup>160</sup> OSAC Research Needs Assessment Form (Formulario de evaluación de necesidades de investigación de la OSAC). «Assessment of Examiners' Toolmark Categorization Accuracy.» Emitido en octubre de 2015 (Aprobado en enero de 2016). Disponible en: [www.nist.gov/forensics/osac/upload/FATM-Research-NeedsAssessment\\_Class-and-individual-marks.pdf](http://www.nist.gov/forensics/osac/upload/FATM-Research-NeedsAssessment_Class-and-individual-marks.pdf).

<sup>161</sup> OSAC Research Needs Assessment Form (Formulario de evaluación de necesidades de investigación de la OSAC). «Assessing the Sufficiency and Strength of Friction Ridge Features.» Emitido en octubre de 2015. Disponible en: [www.nist.gov/forensics/osac/upload/FRS-Research-Need-Assessment-of-Features.pdf](http://www.nist.gov/forensics/osac/upload/FRS-Research-Need-Assessment-of-Features.pdf).

<sup>162</sup> OSAC Research Needs Assessment Form (Formulario de evaluación de necesidades de investigación de la OSAC). «Close Non-Match Assessment», Emitido en octubre de 2015. Disponible en: [www.nist.gov/forensics/osac/upload/FRS-Research-Need-Close-Non-Match-Assessment.pdf](http://www.nist.gov/forensics/osac/upload/FRS-Research-Need-Close-Non-Match-Assessment.pdf).

<sup>163</sup> OSAC Research Needs Assessment Form (Formulario de evaluación de necesidades de investigación de la OSAC). «ACE-V Bias.» Emitido en octubre de 2015. Disponible en: [www.nist.gov/forensics/osac/upload/FRS-Research-Need-ACE-V-Bias.pdf](http://www.nist.gov/forensics/osac/upload/FRS-Research-Need-ACE-V-Bias.pdf).

<sup>164</sup> International Association for Identification (Asociación Internacional para la Identificación). Letter to Patrick J. Leahy, Chairman, Senate Committee on the Judiciary, March 18, 2009. Disponible en: [www.theiai.org/current\\_affairs/nas\\_response\\_leahy\\_20090318.pdf](http://www.theiai.org/current_affairs/nas_response_leahy_20090318.pdf).

<sup>165</sup> National Commission on Forensic Science, (Comisión Nacional sobre Ciencia Forense): «Views of the Commission Technical Merit Evaluation of Forensic Science Methods and Practices». Disponible en: [www.justice.gov/ncfs/file/881796/download](http://www.justice.gov/ncfs/file/881796/download).

#### 4.9. *Resumen de hallazgos científicos*

##### **HALLAZGO 1: CRITERIOS CIENTÍFICOS PARA LA VALIDEZ CIENTÍFICA DE UN MÉTODO DE COMPARACIÓN DE CARACTERÍSTICAS FORENSE**

(1) *Validez de los fundamentos.* Para establecer la validez de los fundamentos de un método de comparación de características, se requieren los siguientes elementos:

(a) un procedimiento reproducible y consistente para (i) la identificación de las propiedades en muestras que son aportadas como pruebas; (ii) la comparación de las propiedades entre dos muestras; (iii) la determinación, basándose en la similitud de las propiedades entre dos conjuntos de características, sobre si se debe declarar que las muestras proceden probablemente del mismo origen (regla de coincidencia”); y

(b) estimaciones empíricas, tras estudios apropiadamente diseñados a partir de múltiples grupos, que establezcan: (i) la tasa de falsos positivos del método —es decir, la probabilidad de una propuesta de identificación entre muestras que realmente proceden de diferentes fuentes— y (ii) la sensibilidad del método —es decir, la probabilidad de una propuesta de identificación entre muestras que realmente proceden de la misma fuente—.

Como se describió en el Cuadro 4, la validación científica debe satisfacer una serie de criterios: (a) debe basarse en conjuntos suficientemente grandes de muestras conocidas y representativas de poblaciones relevantes; (b) debe conducirse de tal modo que los analistas no tengan información sobre las respuestas correctas; el diseño del estudio y el plan de análisis deben especificarse por adelantado y no modificarse posteriormente después de alcanzar los resultados; (d) el estudio debe ser conducido o supervisado por individuos u organizaciones sin beneficio por los resultados; (e) los datos, el software y los resultados deben estar disponibles para permitir que otros científicos puedan revisar las conclusiones; y (f) para asegurar que los resultados sean robustos y reproducibles, debe haber diversos estudios independientes realizados por grupos separados que alcance resultados similares.

Una vez que un método ha sido establecido como válido en sus fundamentos mediante adecuados estudios empíricos, las afirmaciones sobre la precisión y el valor probatorio de las propuestas de identificación, para que sean válidas, deben basarse sobre tales estudios empíricos.

Para métodos objetivos, la validez de los fundamentos puede establecerse demostrando la fiabilidad de cada uno de los pasos individuales (identificación de características, comparación de propiedades, regla de coincidencia, probabilidad de coincidencia falsa y sensibilidad).

Para métodos subjetivos, la validez de los fundamentos *solo* puede establecerse mediante estudios de caja negra que midan la frecuencia en la que muchos examinadores alcancen conclusiones precisas en muchos problemas de comparación de características que contengan muestras representativas del uso pretendido. En ausencia de tales estudios, un método subjetivo de comparación de características no puede considerarse científicamente válido.

La validez de los fundamentos es un *sine qua non* y solo puede demostrarse mediante estudios empíricos. Es muy importante resaltar que las buenas prácticas profesionales —tales como la existencia de asociaciones profesionales, programas de certificación, programas de acreditación, artículos revisados por pares, protocolos estandarizados, pruebas de aptitud y códigos éticos— no pueden sustituir a la evidencia empírica de la validez y fiabilidad científica.

(2) *Validez en la aplicación.* Una vez que un método de comparación de características forense ha sido considerado como válido en sus fundamentos, es necesario establecer su validez en la aplicación en un caso dado.

Como se describió en el Cuadro 5, la validez en la aplicación requiere que: (a) el examinador forense debe haber demostrado que es *capaz* de aplicar el método fiablemente, lo que se hace sometiéndose a tests de aptitud apropiados (véase Sección 4.6) y debe *de hecho* haberlo llevado a cabo, como se ha demostrado en el procedimiento *de hecho* utilizado en el caso, los resultados obtenidos y las notas de laboratorio, que deben estar disponibles para que otros puedan revisarlas científicamente; y (b) las aserciones sobre el valor probatorio de las propuestas de identificación deben ser científicamente válidas —incluyendo que los examinadores deben informar sobre la tasa completa de falsos positivos y la sensibilidad del método establecidas en los estudios de validez de los fundamentos—; se demuestre que las muestras utilizadas en los estudios de los fundamentos son relevantes para los hechos del caso; y cuando sea aplicable, que se demuestre el valor probatorio de la coincidencia observada a partir de las características específicas observadas en el caso; además de no realizar afirmaciones o implicaciones que vayan más allá de la evidencia empírica.

## 5. Evaluación de la validez científica de siete métodos de comparación de características

En la sección precedente, describimos los criterios científicos que un método de comparación de características debe cumplir para ser considerado científicamente válido y fiable, y subrayamos la necesidad de contar con evidencia empírica sobre su precisión y fiabilidad.

En esta sección, ilustraremos el significado de esos criterios aplicándolos a seis métodos de comparación de características forense específicos: (5.1) análisis de ADN de un único origen y de muestras con mezclas simples; (5.2) análisis de ADN de muestras con mezclas complejas; (5.3) marcas de mordedura; (5.4) huellas dactilares latentes; (5.5) identificación de armas de fuego y (5.6) análisis de calzado<sup>166</sup>. Con

<sup>166</sup> La American Association for the Advancement of Science (AAAS) (Asociación estadounidense para el avance de la ciencia) está llevando a cabo un análisis de las bases científicas subyacentes de las herramientas y métodos forenses utilizados actualmente en el sistema de justicia penal. A fecha 1 de septiembre de 2016, todavía no han emitido informe alguno. Véase: [www.aaas.org/page/forensic-science-assessments-quality-and-gap-analysis](http://www.aaas.org/page/forensic-science-assessments-quality-and-gap-analysis).

respecto al séptimo método de comparación de características, el análisis de cabellos, no realizamos una evaluación completa, pero revisamos una evaluación reciente del DOJ.

En adelante, se evaluará si esos métodos han sido considerados como válidos en sus fundamentos y fiables y, si así fuera el caso, qué estimación de precisión debería acompañar al testimonio pericial concerniente a una propuesta de identificación que está fundamentada en estudios científicos actuales. También discutiremos brevemente algunos asuntos relacionados con la validez en la aplicación.

El PCAST compiló una lista de 2019 artículos procedentes de diversas fuentes —incluyendo bibliografías preparadas por el Subcomité sobre Ciencia Forense del Consejo Nacional de Ciencia y Tecnología, los Grupos de Trabajo Científicos relevantes (predecesores del actual OSAC)<sup>167</sup> y los comités de OSAC relevantes; documentos que se presentaron en respuesta a la solicitud del PCAST para obtener información de los miembros interesados de la comunidad de la ciencia forense; y nuestras propias búsquedas de literatura especializada<sup>168</sup>—. Miembros y personal del PCAST identificaron y revisaron esos artículos relevantes para establecer la validez científica. Después de alcanzar un conjunto de conclusiones iniciales, se recibió información del laboratorio del FBI y de científicos individuales del NIST, así como de otros analistas —incluyendo la petición de identificación de artículos adicionales que sustentaran la validez científica y que quizá habíamos omitido—.

En cada uno de los métodos proporcionaremos una breve visión general de la metodología, explicaremos sus antecedentes y estudios y revisamos la evidencia sobre la validez científica.

Como se trató en la sección 4, los métodos objetivos tienen procedimientos bien definidos para (1) identificar las características en las muestras, (2) medir las propiedades, (3) determinar si las propiedades de dos muestras coinciden dentro de una medida establecida de tolerancia (regla de coincidencia) y (4) estimar la probabilidad de que dos muestras de diferentes fuentes coincidan (probabilidad de coincidencia falsa). Es posible examinar la validez y fiabilidad de cada uno de estos pasos diferentes. De los seis métodos considerados en esta sección, solo los dos primeros métodos (que conllevan análisis de ADN) emplean métodos objetivos. Los restantes cuatro métodos son subjetivos.

En los métodos subjetivos los procedimientos no están definidos con precisión, por lo que se requiere sustancialmente del juicio humano experto. Los analistas pueden centrarse en ciertas propiedades mientras ignoran otras, pueden compararlas de distintos modos y pueden tener distintos estándares para declarar una propuesta de identificación entre muestras. Como se describió en la sección 4, la única forma

---

<sup>167</sup> Véase: [www.nist.gov/forensics/workgroups.cfm](http://www.nist.gov/forensics/workgroups.cfm).

<sup>168</sup> Véase: [www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast\\_forensics\\_references.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensics_references.pdf)

de establecer la validez de los fundamentos es a través de múltiples e independientes estudios de «caja negra» que midan la frecuencia en que los analistas alcanzan conclusiones precisas en muchos problemas de comparación de características que contengan muestras representativas para el uso pretendido. En ausencia de tales estudios, un método de comparación de características no puede considerarse como científicamente válido.

El PCAST encontró pocos estudios de caja negra apropiadamente diseñados para valorar la validez científica de métodos subjetivos. Dos notables excepciones, tratadas en esta sección, fueron un estudio sobre huellas dactilares latentes dirigido por el Laboratorio del FBI y un estudio sobre identificación de armas de fuego patrocinado por el Ministerio de Defensa y dirigido por el Laboratorio AMES del Departamento de Energía.

Nos hemos pronunciado sobre si las pruebas de aptitud, dirigidas por organizaciones comerciales en algunas disciplinas, pudieran ser útiles para establecer la validez de los fundamentos. A la fecha, concluimos que no por diversas razones. En primer lugar, las pruebas de aptitud no están dirigidas a establecer la validez de los fundamentos. En segundo lugar, los problemas de los tests o el conjunto de tests utilizados en pruebas de aptitud comerciales ordinariamente no se hacen públicos —haciendo imposible determinar si los tests valoran apropiadamente el método en el rango de aplicaciones en el que es utilizado—. La publicación y la revisión crítica de los métodos y datos es un componente esencial en el establecimiento de la validez científica. En tercer lugar, la empresa dominante en el mercado, *Collaborative Testing Services, Inc. (CTS)*, afirma explícitamente que sus pruebas de aptitud no son apropiadas para estimar las tasas de error de una disciplina porque (a) los resultados de los tests, que están abiertos a cualquiera, puede que no reflejen las habilidades de los analistas forenses y (b) «los resultados de los que se informa no reflejan respuestas “correctas” o “incorrectas”, sino más bien respuestas que están de acuerdo o no con las conclusiones consensuadas entre la población participante<sup>169</sup>». En cuarto lugar, los tests para métodos de comparación de características forenses constan normalmente solo de uno o dos problemas al año. Y, en quinto lugar, «la comunidad favorece los tests fáciles», pues tests que sean demasiado difíciles podría poner en peligro la repetición del negocio para un proveedor comercial<sup>170</sup>.

<sup>169</sup> Véase: [www.ctsforensics.com/assets/news/CTSErrorRateStatement.pdf](http://www.ctsforensics.com/assets/news/CTSErrorRateStatement.pdf)

<sup>170</sup> El PCAST agradece al Presidente de Collaborative Testing Services, Inc. (CTS), Christopher Czyryca, las útiles conversaciones con él mantenidas sobre los tests de aptitud. Czyryca explicó que (1) CTS define que hay consenso cuando se alcanza acuerdo, al menos, entre el 80% entre quienes responden y (2) las pruebas de aptitud para huellas latentes solo ocasionalmente incluyen un problema en el cual la impresión cuestionada *no coincide con alguna* de las posibles respuestas. Czyryca resaltó que la comunidad forense no es favorable a tests más retadores —y que las empresas de tests temen que podrían ver mermado su negocio si sus tests fueron vistos como demasiado difíciles—. Un ejemplo de test «retador» es el importantísimo escenario en el que ninguna de las muestras cuestionadas coincide con algunas de las muestras conocidas: a causa de que los analistas pueden esperar encontrar alguna coincidencia, tales escenarios proporcionan una oportunidad para valorar con qué frecuencia los analistas

Las observaciones y conclusiones del PCAST que a continuación se recogen son en gran medida coherentes con las conclusiones de informes anteriores de la NRC<sup>171</sup>.

### 5.1. *Análisis de ADN de muestras procedentes de una única fuente y de mezclas simples*

El análisis de ADN de una muestra procedente de una única fuente o de una mezcla simple incluye excelentes ejemplos de métodos objetivos en los que la validez de sus fundamentos ha sido apropiadamente establecida<sup>172</sup>.

#### 5.1.1. Metodología

El análisis de ADN implica comparar perfiles genéticos de diferentes muestras para ver si una muestra conocida pudo haber sido el origen de una muestra cuestionada (habiendo sido aportada como prueba).

Para generar un perfil genético, se *extrae* el ADN mediante un proceso químico de una muestra que contiene material biológico, como sangre, semen, cabello o células de la piel. Después, se *amplifica* un conjunto predeterminado de segmentos de ADN («marcadores») que contienen pequeñas secuencias repetidas<sup>173</sup>, utilizando la *Polymerase Chain Reaction* (PCR) (Reacción en Cadena de la Polimerasa): un proceso enzimático que replica un segmento de ADN dirigido una y otra vez a producir millones de copias. Después de la amplificación, las longitudes de los fragmentos de ADN resultantes se miden utilizando una técnica llamada electroforesis capilar, basada en el hecho de que los fragmentos más largos se mueven más lentamente que los fragmentos más cortos a través de una solución de polímero. Los datos originales recogidos de este proceso se analizan mediante un software que produce una imagen gráfica (un electroferograma) y una lista de números (el perfil de ADN) que se corresponden con los tamaños de cada uno de los fragmentos («mediante la comparación con tamaños de moléculas conocidos y estandarizados»).

Como práctica actual, el método utiliza 13 marcadores específicos y el proceso de amplificación se diseña para que los fragmentos de ADN correspondientes a los diferentes marcadores ocupen diferentes rangos de tamaño —facilitando reconocer

---

declaran coincidencias que son falsos positivos. (Véase también la presentación a la Comisión Nacional de la Ciencia Forense por el Presidente de CTS, el Sr. Czyryca, resaltando que «La comunidad favorece los tests fáciles». [www.justice.gov/ncfs/file/761061/download](http://www.justice.gov/ncfs/file/761061/download).)

<sup>171</sup> National Research Council, 2009; National Research Council, 2008.

<sup>172</sup> El análisis forense de ADN pertenece a dos disciplinas anteriores de las que depende —metrología y genética molecular— y se ha beneficiado de la extensa aplicación de la tecnología de ADN en investigación biomédica y en aplicaciones médicas.

<sup>173</sup> Las repeticiones, denominadas repeticiones cortas en tandem (STRs), consisten en copias consecutivas repetidas de segmentos de 2 a 6 pares de bases.

qué fragmentos vienen de cada marcador<sup>174</sup>—. En cada marcador, cada ser humano lleva dos variantes (denominadas ‘alelos’), uno heredado de la madre y otro del padre—que pueden ser de la misma o diferente longitud—<sup>175</sup>.

### 5.1.2. Análisis de muestras procedentes de una única fuente

El análisis de ADN de una muestra procedente de un único individuo es un método objetivo. Además de que los protocolos del laboratorio están definidos con precisión, la interpretación también conlleva poco o ningún juicio humano.

Un analista puede valorar si una muestra procedió de una única fuente basándose en si el perfil de ADN contiene, ordinariamente y para cada marcador, exactamente un fragmento de cada cromosoma que contiene el marcador, lo que produce una o dos longitudes de fragmentos distintos de cada marcador<sup>176</sup>. El perfil de ADN puede compararse luego con el perfil de ADN de un sospechoso conocido. Puede también ser introducido en el *FBI's National DNA Index System (NDIS)* (Sistema Indexado de Perfiles de ADN Nacional del FBI) y buscado en una base de datos de perfiles de ADN de condenados (y detenidos, en más de la mitad de los estados) o de crímenes aún no resueltos.

Se declara que dos perfiles de ADN coinciden si las listas de alelos son iguales<sup>177</sup>. La probabilidad de que dos perfiles de ADN de *diferentes* fuentes tengan el mismo perfil (la probabilidad de coincidencia aleatoria) se calcula después fundamentándose

---

<sup>174</sup> El actual kit utilizado por el FBI (Identifiler Plus) tiene 16 marcadores en total: 15 marcadores STR y el marcador «Amelogenina». Más adelante, durante este año, se implementará un kit que tiene 24 marcadores.

<sup>175</sup> El FBI anunció en 2015 que planea expandir los marcadores que conforman su núcleo añadiendo los 7 que normalmente más se utilizan en las bases de datos de otros países. (Los datos poblacionales han sido publicados para el conjunto expandido, incluyendo frecuencias en 11 poblaciones étnicas [www.fbi.gov/about-us/lab/biometric-analysis/codis/expanded-fbi-str-2015-final-6-16-15.pdf](http://www.fbi.gov/about-us/lab/biometric-analysis/codis/expanded-fbi-str-2015-final-6-16-15.pdf).) A principios de 2017, se requerirán esos marcadores para introducir y buscar perfiles de ADN en el sistema nacional. Se espera que los datos expandidos en cada perfil provean de mayor capacidad de potencial discriminante para la identificación, especialmente en muestras coincidentes, aunque con solo perfiles parciales de ADN, investigaciones de personas desaparecidas y en casos de aplicación de leyes internacionales y contra el terrorismo.

<sup>176</sup> El analista revisa el electroferograma para determinar si cada uno de los picos es un verdadero alelo o un artefacto (por ejemplo, ruido de fondo en forma de pequeños picos antes de un pico alélico, agujas y otros fenómenos) y para determinar si más de un individuo pudiera haber contribuido al perfil. En casos raros, un individuo puede tener más de dos fragmentos en un marcador debido a una extraña variación del número de copias en el genoma humano.

<sup>177</sup> Cuando solo es posible conseguir un perfil parcial en la muestra que constituye la prueba (por ejemplo, en casos con limitadas cantidades de ADN, degradación de la muestra o la presencia de inhibidores de PCR), el analista puede también informar de una «inclusión» si el perfil parcial es *consistente* con el perfil de ADN obtenido de una muestra de referencia. El analista puede también informar de una inclusión cuando los resultados de ADN de una muestra de referencia estén presentes en una mezcla. Estos casos generalmente requieren significativamente mayor análisis e interpretación humanos que las muestras de una única fuente.

en las frecuencias medidas empíricamente de cada alelo y en principios establecidos de la genética poblacional (véase p. 92)<sup>178</sup>.

### 5.1.3. Análisis de mezclas simples

Muchos casos de agresiones sexuales conllevan mezclas de ADN de dos individuos, donde uno de ellos (por ejemplo, la víctima) es conocido. El análisis de ADN de esas mezclas simples es también relativamente sencillo. Se han utilizado métodos durante 30 años para extraer, diferenciadamente, ADN de células de esperma y de células epiteliales vaginales, haciendo posible generar perfiles de ADN de las dos fuentes. Cuando los dos tipos de células son iguales y uno de los contribuyentes es conocido, los alelos del individuo conocido pueden sustraerse del conjunto de alelos identificados en la mezcla<sup>179</sup>.

### 5.1.4. Validez de los fundamentos

Para evaluar la validez de los fundamentos de un método objetivo (tales como los análisis de una única fuente y de mezclas simples), podemos examinar la fiabilidad de cada uno de los pasos individuales en lugar de tener que confiar en estudios de caja negra.

### 5.1.5. Muestras de una única fuente

Cada paso en el análisis es objetivo y conlleva poco o ningún juicio humano.

(1) *Identificación de características.* En contraste con otros métodos tratados en este informe, las características utilizadas en los análisis de ADN (las longitudes de los marcadores) están definidas *de antemano*.

(2) *Medida de las características y comparación.* La amplificación PCR, inventada en 1983, se usa en decenas de miles de laboratorios de biología molecular, incluyéndose muchas aplicaciones médicas para las cuales ha sido rigurosamente validada. Los kits PCR multiplex diseñados por casas comerciales para su uso en laboratorios forenses deben validarse tanto externamente (a través de estudios de validación de su desarrollo, publicados en revistas revisadas por pares) como internamente (por cada

---

<sup>178</sup> Las probabilidades de coincidencia aleatoria pueden también expresarse en términos de relaciones de verosimilitud (LR), que es la relación de (1) la probabilidad de observar el perfil de ADN si el individuo en cuestión es la fuente de la muestra de ADN y (2) la probabilidad de observar el perfil de ADN si el individuo en cuestión *no* es la fuente de la muestra de ADN. En la situación de una muestra procedente de una única fuente, el LR debe ser simplemente el recíproco de la probabilidad de coincidencia aleatoria (porque la primera probabilidad en el LR es 1 y la segunda es la probabilidad de coincidencia aleatoria).

<sup>179</sup> En muchos casos, el ADN estará presente en la mezcla en diferentes pero suficientes cantidades de forma que las alturas de los picos en el electroferograma procedente de las dos fuentes puedan distinguirse, permitiendo al analista separar más fácilmente las fuentes.

laboratorio que desee utilizar el kit), antes de que puedan utilizarse<sup>180</sup>. Los tamaños de los fragmentos se miden mediante un procedimiento automatizado cuya variabilidad está bien caracterizada y es pequeña; la desviación estándar es de, aproximadamente, 0.05 pares de bases, lo cual proporciona mediciones altamente fiables<sup>181, 182</sup>. Se han llevado a cabo estudios de validación del desarrollo —incluyendo al FBI— para verificar la precisión, repetibilidad y reproducibilidad del procedimiento<sup>183, 184</sup>.

(3) *Comparación de características*. En el caso de muestras procedentes de una única fuente, existen «reglas de coincidencia» claras y bien especificadas para declarar si los perfiles de ADN coinciden. Cuando se buscan en NDIS perfiles de ADN completos con criterios de «elevada rigurosidad», se encuentran «coincidencias» únicamente cuando cada alelo del perfil desconocido coincide con un alelo del perfil conocido, y *viceversa*. Cuando se buscan perfiles parciales obtenidos de muestras parcialmente degradadas o contaminadas con criterios de «moderada rigurosidad», se muestran candidatos si uno de los alelos del perfil desconocido coincide con un alelo del perfil conocido<sup>185, 186</sup>.

<sup>180</sup> Se requiere que los laboratorios que realizan análisis de ADN con fines forenses sigan los Estándares de Aseguramiento de la Calidad del FBI para Laboratorios de Pruebas de ADN como condición para participar en el Sistema Indexado de ADN Nacional: [www.fbi.gov/about-us/lab/biometricanalysis/codis/qas-standards-for-forensic-dna-testing-laboratories-effective-9-1-2011](http://www.fbi.gov/about-us/lab/biometricanalysis/codis/qas-standards-for-forensic-dna-testing-laboratories-effective-9-1-2011). El FBI's Working Group on DNA Analysis Methods (SWGDM) (Grupo de Trabajo Científico del FBI sobre Métodos de Análisis de ADN) ha publicado directrices para laboratorios relativas a los procesos de validación de acuerdo con los Estándares de Aseguramiento de la Calidad del FBI (QAS). SWGDAM Validation Guidelines for DNA Analysis Methods (Directrices para la Validación de Métodos de Análisis de ADN del Grupo de Trabajo SWGDAM), diciembre de 2012. Véase: [www.media.wix.com/ugd/4344b0\\_cbc27d16dcb-64fd88cb36ab2a2a25e4c.pdf](http://www.media.wix.com/ugd/4344b0_cbc27d16dcb-64fd88cb36ab2a2a25e4c.pdf).

<sup>181</sup> Los laboratorios forenses generalmente utilizan analizadores genéticos desarrollados por el grupo Applied Biosystems de Thermo Fisher Scientific (ABI 310, 3130 o 3500).

<sup>182</sup> Para estimar incorrectamente la longitud de un fragmento por un par de bases (la mínima diferencia en tamaño), se requiere un error de medida de 0.5 pares de bases, que se corresponde con 10 desviaciones estándar. Además, ordinariamente los alelos se diferencian por, al menos, 4 pares de bases (aunque algunos marcadores STR tengan alelos bastante comunes que difieren en uno o dos nucleótidos).

<sup>183</sup> Como ejemplos de esos estudios véanse: BUDOWLE, MORETTI, KEYS, KOONS y SMERICK, 1997: 701-707; KIMPTON, OLDROYD, WATSON, ET AL., 1996: 1283-1293; LYGO, JOHNSON, HOLDAWAY, ET AL., 1994: 77-89; FREGEAU, BOWEN y FOURNEY, 1999: 133-166.

<sup>184</sup> Por ejemplo, en un estudio de 2001 se compararon las características de funcionamiento de distintos kits de STR disponibles comercialmente, comprobándose la consistencia y la reproducibilidad de los resultados utilizando muestras de casos previamente tipados, muestras influenciadas ambientalmente y muestras de fluidos corporales depositadas sobre varios sustratos. El estudio halló que todos los kits se podían utilizar para amplificar y tipar marcadores STR satisfactoriamente, así como que todos los procedimientos utilizados para cada uno de los kits eran robustos y válidos. No se hallaron evidencias de resultados de falsos positivos o falsos negativos, así como tampoco hubo evidencia sustancial de amplificación preferente dentro de un marcador en ninguno de los kits para las pruebas. MORETTI, BAUMSTARK, DEFENBAUGH, ET AL., 2001: 647-660.

<sup>185</sup> Véase: FBI's Frequently Asked Questions (FAQs) on the CODIS Program and the National DNA Index System. [www.fbi.gov/about-us/lab/biometric-analysis/codis/codis-and-ndis-fact-sheet](http://www.fbi.gov/about-us/lab/biometric-analysis/codis/codis-and-ndis-fact-sheet).

<sup>186</sup> Las muestras contaminadas no se introducen en NDIS.

(4) *Estimación de la probabilidad de coincidencia aleatoria.* El proceso de cálculo de la probabilidad de coincidencia aleatoria (es decir, la probabilidad de una coincidencia por azar) se fundamenta en principios bien establecidos en la genética poblacional y estadística. El FBI ha obtenido las frecuencias de los alelos individuales a partir de perfiles de aproximadamente 200 individuos no emparentados entre sí, pertenecientes a 6 grupos poblacionales y fueron evaluadas antes de que fueran utilizadas<sup>187</sup>. La frecuencia de un patrón completo de alelos —es decir, la probabilidad de coincidencia aleatoria— se estima generalmente multiplicando las frecuencias de los marcadores individuales, asumiendo que los alelos son independientes entre sí—<sup>188</sup>. La probabilidad resultante suele ser ordinariamente menor a 1 en 10 billones\*, excluyendo la posibilidad de parientes cercanos<sup>189</sup>. (Nota: La multiplicación de la frecuencia de los alelos puede sobreestimar la rareza de un patrón porque los alelos pueden no ser completamente independientes debido a la subestructura de población. Un informe del NRC de 1996 concluyó que el efecto de la subestructura poblacional sobre el valor calculado venía a estar probablemente dentro de un factor de 10 (por ejemplo, para una probabilidad de frecuencia aleatoria de 1 entre 10 millones, es altamente probable que la verdadera probabilidad esté entre 1 en 1 millón y 100 millones)<sup>190</sup>. Sin embargo, un estudio reciente realizado por científicos del NIST sugiere que la variación puede ser sustancialmente más grande que 10 veces más<sup>191</sup>.

<sup>187</sup> Los datos de población iniciales generados por el FBI incluían datos de 6 poblaciones étnicas distintas, con tamaños muestrales de 200 individuos. Véase: BUDOWLE, MORETTI, BAUMSTARK, ET AL., 1999: 1277-1286; BUDOWLE, SHEA, NIEZGODA, ET AL., 2001: 453-89. En julio de 2015 se informó de errores encontrados en la base de datos original. (Erratum, *Journal of Forensic Sciences*, Vol. 60, No. 4 (2015): 1114-6, se valoró el impacto de estas discrepancias sobre los cálculos de probabilidad de los perfiles (y se encontró ser menor que un factor de 2 en un perfil completo) y las frecuencias alélicas estimadas fueron correspondientemente corregidas. Al mismo tiempo que se corrigió el conjunto de datos original, el Laboratorio del FBI también publicó conjuntos de datos adicionales tras retirar las muestras originales para incluir marcadores adicionales. Además, las muestras poblacionales que habían sido estudiadas originalmente en otros laboratorios fueron tipadas con marcadores adicionales, de tal forma que la base de datos completa incluye nueve poblaciones. Estos conjuntos de datos «expandidos» están en uso en el Laboratorio del FBI y pueden encontrarse en [www.fbi.gov/about-us/lab/biometric-analysis/codis/expanded-fbi-str-final-6-16-15.pdf](http://www.fbi.gov/about-us/lab/biometric-analysis/codis/expanded-fbi-str-final-6-16-15.pdf).

<sup>188</sup> De forma más precisa, primero se calcula la frecuencia genotípica de cada marcador. Si el marcador tiene dos copias del mismo alelo con frecuencia  $p$ , la frecuencia se calcula como  $p^2$ . Si el marcador tiene dos alelos diferentes con frecuencias respectivas  $p$  y  $q$ , la frecuencia genotípica se calcula como  $2pq$ . La frecuencia del patrón completo se calcula multiplicando los valores obtenidos para cada marcador individual.

\* N. del T.: el billón anglosajón equivale a 1000 millones o 10<sup>9</sup> en el sistema decimal; el billón continental como millón de millones es también representado como 10<sup>12</sup>).

<sup>189</sup> La probabilidad de coincidencia aleatoria es más elevada entre parientes cercanos. Se espera que los perfiles de ADN de dos gemelos idénticos coincidan perfectamente. Para parientes en primer grado, la probabilidad de coincidencia aleatoria es del orden de 1 entre 100,000 cuando se analizan los 13 marcadores STR del núcleo de CODIS. Véase: BUTLER, 2015.

<sup>190</sup> National Research Council, 1996; GOODE, 2002: 45-77.

<sup>191</sup> GITTELSON y BUCKLETON, 2016. Véase: [www.cstl.nist.gov/strbase/pub\\_pres/Gittelson-AA-FS2016-Factor-of-10.pdf](http://www.cstl.nist.gov/strbase/pub_pres/Gittelson-AA-FS2016-Factor-of-10.pdf)

La probabilidad de coincidencia aleatoria debe calcularse utilizando una fórmula estadística adecuada que tenga en cuenta la subestructura de población<sup>192</sup>.

#### 5.1.6. Mezclas simples

Las etapas para analizar mezclas simples son las mismas que para analizar muestras de una única fuente, hasta incluso la interpretación. Los perfiles de ADN que contienen una mezcla de dos contribuyentes, donde uno de ellos es conocido, pueden interpretarse del mismo modo que las muestras procedentes de una única fuente. Esto ocurre frecuentemente en casos de agresiones sexuales, donde un perfil genético contiene una mezcla de ADN de la víctima y del agresor. Los métodos que se utilizan para extraer ADN diferenciadamente de las células de esperma y de las células epiteliales vaginales en agresiones sexuales están bien establecidos<sup>193</sup>. Cuando los dos tipos celulares son el mismo, uno de los contribuyentes a la mezcla de ADN puede ser dominante, produciéndose un distintivo contraste entre las alturas de los picos de los dos contribuyentes; en estos casos, los alelos del contribuyente mayoritario (correspondiente al de los picos alélicos más grandes) y los del contribuyente minoritario pueden ser, por lo general, fiablemente interpretados, cuando la proporción del contribuyente minoritario no sea demasiado baja<sup>194</sup>.

#### 5.1.7. Validez en la aplicación

Aunque el análisis de ADN de muestras de una única fuente y el de mezclas simples es un método válido en sus fundamentos y fiable, no es infalible en la práctica. Pueden ocurrir y ocurren errores en las pruebas de ADN. Aunque la probabilidad de que dos muestras de diferentes fuentes tengan el mismo perfil de ADN es pequeña, la probabilidad de que ocurra un error humano es mucho más alta. Tales errores pueden derivarse de mezclas de muestras, contaminación, interpretación incorrecta y errores en los informes<sup>195</sup>.

Para minimizar el error humano, el FBI requiere, como una condición para participar en NDIS, que los laboratorios sigan los Estándares de Aseguramiento de la

---

<sup>192</sup> BALDING y NICHOLS, 1994: 125-140.

<sup>193</sup> GILL, JEFFREYS y WERRETT, 1985: 577-579.

<sup>194</sup> CLAYTON, WHITAKER, SPARKES, ET AL., 1998: 55-70.

<sup>195</sup> KRIMSKY y SIMONCELLI, 2011. Quizá, el más espectacular error humano hasta la fecha está relacionado con la investigación del gobierno alemán del caso «*Phantom of Heilbronn*», una mujer cuyo ADN apareció en la escena del crimen en más de 40 delitos de tres países, incluyendo 6 asesinatos, varios atracos y docenas de allanamiento de morada en el transcurso de más de una década. Después de un esfuerzo que incluyó el análisis de muestras de ADN de más de 3000 mujeres de cuatro países y que costó 18 millones de dólares, las autoridades descubrieron que la mujer de interés era una trabajadora de una fábrica austríaca que fabricaba hisopos utilizados para la recogida de ADN. La mujer había contaminado inadvertidamente un gran número de hisopos con su propio ADN y, por eso, se encontró en muchos análisis de ADN.

Calidad del FBI (QAS)<sup>196</sup>. Antes de que los resultados de los análisis de ADN puedan ser comparados, el analista tiene que someterse a una serie de controles que comprueban si ha podido producirse una posible contaminación y que aseguran que el proceso PCR ha transcurrido apropiadamente. El QAS también requiere que todos los analistas que realizan pruebas de ADN en casos forenses se sometan a una prueba de aptitud cada semestre. Los resultados de las pruebas no tienen por qué publicarse, pero el laboratorio debe guardarlos, así como las discrepancias, errores cometidos y acciones correctivas efectuadas<sup>197</sup>.

Los analistas que practican la ciencia forense en los Estados Unidos generalmente no informan de cuestiones de calidad que surjan durante los análisis forenses. Por contraste, las tasas de error en pruebas médicas generalmente se miden y se informa sobre ellas<sup>198</sup>. Merece destacarse un artículo del año 2014 dimanante del Instituto Forense Holandés (NFI), una agencia gubernamental, que realizó un amplio análisis de todas las «notificaciones sobre cuestiones de calidad» halladas en casos forenses, categorizadas por tipología, fuente e impacto<sup>199,200</sup>. Los autores piden mayor «transparencia» y un «cambio cultural» escribiendo que:

El análisis de ADN con fines forenses se realiza en un gran número de laboratorios de todo el mundo, tanto pertenecientes a empresas privadas como a institutos gubernamentales. Los procedimientos de calidad están implantados en todos los laboratorios, pero la naturaleza del sistema de calidad varía mucho entre ellos. En particular, hay muchos laboratorios de ADN forense que operan sin un sistema de notificación de incidencias de calidad como el descrito en este artículo. Desde nuestra experiencia, ese sistema es extremadamente importante para la detección y la gestión apropiada de los errores. Esto es crucial en casos forenses que puedan tener un mayor impacto en la vida de las personas. Por consiguiente, proponemos que la implementación del sistema de notificación de incidencias de calidad es necesario para cualquier laboratorio que esté involucrado en resolver casos de ADN forenses.

Un sistema así solo puede trabajar en óptimas condiciones, cuando exista una cultura en el laboratorio no centrada en la culpabilización y que se extienda a la policía y al sistema judicial. La gente tiene tendencia natural a ocultar sus errores y es esencial crear una atmósfera donde no haya consecuencias personales adversas cuando se cometan errores en los informes. La dirección debe tomar la iniciativa en este cambio cultural...

Hasta lo que sabemos, el NFI es el primer laboratorio forense de ADN del mundo en revelar detalladamente tales datos e informes. Demuestra que eso es posible sin que ocurra ningún desastre o abuso, y no existen razones para la no divulgación. Como se mencionó en la introducción, la pub-

<sup>196</sup> FBI, 2011. Véase: [www.fbi.gov/aboutus/lab/biometric-analysis/codis/qas-standards-for-forensic-dna-testing-laboratories-effective-9-1-2011](http://www.fbi.gov/aboutus/lab/biometric-analysis/codis/qas-standards-for-forensic-dna-testing-laboratories-effective-9-1-2011).

<sup>197</sup> *Ibid.*, Secciones 12, 13, y 14.

<sup>198</sup> Véase, por ejemplo: PLEBANI y CARRORO, 1997: 1348-1351; STAHL, LUND y BRANDSLUND, 1998: 2195-2197; HOFGARTNER y TAIT, 1999: 14-21; CARRORO y PLEBANI, 2007: 1338-1342.

<sup>199</sup> KLOOSTERMAN, SJERPS y QUAK, 2014: 77-85 y BUTLER 2015. [www.cstl.nist.gov/strbase/pub\\_pres/Butler-ErrorManagement-DNA-Error.pdf](http://www.cstl.nist.gov/strbase/pub_pres/Butler-ErrorManagement-DNA-Error.pdf).

<sup>200</sup> Holanda tiene un sistema de justicia penal «inquisitorial» no un sistema adversarial como el utilizado en los Estados Unidos. La preocupación por tener que explicar cuestiones de calidad ante los Tribunales puede explicar en parte por qué los laboratorios en los Estados Unidos no informan rutinariamente sobre ellas.

licación de datos sobre tasas de error en el uso de medicamentos de laboratorio se ha convertido en una práctica estándar. Las tasas de fallos de calidad en este dominio son comparables a las nuestras.

Finalmente, resaltamos que es necesario mejorar las pruebas de aptitud. No hay actualmente requerimientos sobre el nivel de dificultad que deben tener estos tests. Las pruebas deben ser representativas del rango completo de situaciones que es probable encontrar en los casos reales.

#### **HALLAZGO 2: ANÁLISIS DE ADN**

*Validez de los fundamentos.* El PCAST encuentra que el análisis de muestras procedentes de una única fuente o de mezclas simples de dos individuos, tales como las que dan en casos de violación, es un método objetivo que ha sido establecido como válido en sus fundamentos.

*Validez en la aplicación.* Como los errores por fallos humanos son dominantes en la probabilidad de coincidencias aleatorias, los criterios científicos para la validez en la aplicación requieren que un analista (1) deba haber realizado una prueba de aptitud rigurosa y relevante para demostrar su capacidad para aplicar fielmente el método, (2) deba revelar rutinariamente en informes y testimonios si, cuando llevó a cabo el examen, era consciente de algunos hechos del caso que quizá influyeran en la conclusión y (3) deba revelar, previa solicitud, toda la información sobre las pruebas de calidad y las incidencias de calidad en su laboratorio.

### 5.2. *Análisis de ADN en muestras con mezclas complejas*

Algunas investigaciones conllevan análisis de ADN de mezclas complejas de muestras biológicas procedentes de múltiples individuos desconocidos en proporciones desconocidas. Tales muestras quizá se produzcan, por ejemplo, en manchas de sangre mezcladas. A medida que los kits de pruebas de ADN han adquirido mayor sensibilidad, ha habido un creciente interés en el «ADN por contacto» —por ejemplo, pequeñas cantidades de ADN dejadas por múltiples individuos en el volante de un coche—.

#### 5.2.1. Metodología

La diferencia fundamental entre el análisis de ADN de muestras con mezclas complejas y el análisis de ADN de una única fuente o de una mezcla simple no está en el proceso de laboratorio sino en la interpretación del perfil de ADN resultante.

El análisis de ADN de mezclas complejas —definido como mezclas con más de dos contribuyentes— es inherentemente difícil e incluso más para pequeñas cantida-

des de ADN<sup>201</sup>. Esas muestras producen un perfil de ADN que superpone múltiples perfiles individuales. La interpretación de un perfil mezcla es diferente por múltiples motivos: cada individuo puede contribuir con dos, uno o cero alelos en cada marcador; los alelos pueden superponerse entre sí; las alturas de los picos pueden diferir considerablemente, debido a las diferencias en la cantidad y en el estado de preservación del ADN de cada fuente; y los «pequeños picos» que rodean a los alelos (comunes artefactos del proceso de amplificación del ADN) pueden dificultar la detección de los alelos presentes o sugerir alelos que no están<sup>202</sup>. Es frecuentemente imposible decir con certeza qué alelos están presentes en la mezcla o cómo distintos individuos contribuyen a la mezcla, por no hablar de la precisión para inferir el perfil de ADN de cada individuo<sup>203</sup>.

En cambio, los analistas se deben preguntar: «¿podría el perfil de ADN del sospechoso estar presente dentro del perfil mezcla? Y, ¿cuál es la probabilidad de que esa observación pueda ocurrir por casualidad?». Esas preguntas son difíciles por las razones anteriormente dadas. Como consecuencia de que muchos perfiles genéticos diferentes pueden encajar dentro de algunos perfiles mezcla, la probabilidad de que un sospechoso «no pueda ser excluido» como posible contribuyente a una mezcla compleja puede ser *mucho más alta* (en algunos casos, millones de veces más alta) que las probabilidades halladas en coincidencias de perfiles procedentes de una única fuente. Como resultado, el cálculo apropiado del peso estadístico es crítico a la hora de presentar información precisa al tribunal.

### 5.2.2. Interpretación subjetiva de mezclas complejas

Los planteamientos iniciales para interpretar mezclas complejas se apoyan en juicios subjetivos de los analistas, así como en el uso de métodos estadísticos simplificados como la «*Combined Probability of Inclusion*» (CPI) (Probabilidad Combinada de Inclusión). Estos planteamientos son problemáticos porque las elecciones subjetivas hechas por los analistas, tales como qué alelos se incluyen en el cálculo, pueden alterar sustancialmente el resultado y conducir a respuestas imprecisas.

El problema existente con el análisis subjetivo de mezclas complejas se ilustra bien con el caso de doble homicidio *Winston v. Commonwealth* en 2003<sup>204</sup>. Un perito de la acusación informó que el acusado no podía ser excluido como posible contribuyente al ADN encontrado en un guante desechado que contenía perfiles de ADN mezclados de al menos tres contribuyentes; el acusado fue condenado y sentenciado

---

<sup>201</sup> Véase, por ejemplo, el documento SWGDAM sobre interpretación de mezclas de ADN. [www.swgdam.org/#!public-comments/c1t82](http://www.swgdam.org/#!public-comments/c1t82).

<sup>202</sup> Los retos en el análisis de ADN «low-template» (con baja cantidad de ADN molde) se describen en este reciente artículo, BUTLER, 2015.

<sup>203</sup> Véanse: BUCKLETON, CURRAN y GILL, 2007: 20-28; COBLE, BRIGHT, BUCKLETON, ET AL., 2015: 207-11.

<sup>204</sup> *Winston v. Commonwealth*, 604 S.E.2d 21 (Va. 2004).

a muerte. El fiscal dijo al jurado que la probabilidad de que sucediera el cotejo positivo por azar era de de 1 entre 1.1 billones (anglosajones). Un artículo publicado en el año 2009, sin embargo, muestra un caso científicamente razonable en el que la probabilidad es próxima a 1 entre 2 —es decir, el 50% de la población relevante no podía ser excluido—<sup>205</sup>. Esa enorme discrepancia no es aceptable, especialmente en casos en los que el acusado fue sentenciado a muerte.

Hay dos artículos que claramente demuestran que las aproximaciones utilizadas generalmente en el análisis de ADN de mezclas complejas pueden ser problemáticas. En un estudio de 2011, Dror y Hampikian probaron si información de contexto irrelevante pudiera sesgar las conclusiones de los analistas, para ello utilizaron una prueba de ADN de un caso penal ya juzgado (un caso de violación en grupo en Georgia)<sup>206</sup>. En este caso, uno de los sospechosos implicó a otro mediante un acuerdo de culpabilidad. Los dos analistas que examinaron la prueba de la escena del crimen eran conscientes de este testimonio contra el sospechoso y sabían que el testimonio basado en el acuerdo de culpabilidad podía ser utilizado ante los tribunales sólo si la evidencia de ADN lo corroboraba. Debido a la compleja naturaleza de la mezcla de ADN recogida en la escena del crimen, el análisis de esta prueba requirió juicio e interpretación por parte de los analistas. Ambos analistas concluyeron que el sospechoso no podía ser excluido como contribuyente.

Dror y Hampikian presentaron la prueba de ADN original de este delito a 17 analistas en ADN, pero sin facilitarles ninguna información de contexto irrelevante. Encontraron que sólo 1 de los 17 analistas estuvo de acuerdo con los analistas originales que fueron expuestos a la información sesgadora (de hecho, 12 de los analistas *excluyeron* al sospechoso como posible contribuyente).

En otro artículo, De Keijser y sus colegas, se presentó a 19 analistas en ADN un caso simulado de un supuesto robo con violencia fuera de un bar:

Hay un sospechoso varón que niega haber hecho nada malo. Las piezas de convicción que fueron muestreadas para el análisis de ADN son una camisa de una (supuesta) víctima que es mujer (que sostiene haber sido sujeta por su asaltante), una colilla recogida por la policía y cuyo cigarrillo fue supuestamente fumado por la víctima /o el sospechoso y pedazos de uñas de la víctima que dice haber arañado al agresor<sup>207</sup>.

Aunque a todos los analistas se les proporcionaron los mismos perfiles de ADN (preparados a partir de las tres muestras y las dos personas mencionadas), sus conclusiones variaron radicalmente. Un analista excluyó al sospechoso como posible contribuyente, mientras que otro declaró una coincidencia entre el perfil del sospechoso y algunos picos menores en el perfil mezcla de las uñas —informando de una probabilidad de coincidencia aleatoria de aproximadamente 1 en 209 millones—. Otros analistas incluso declararon la prueba como no concluyente.

<sup>205</sup> THOMPSON, 2009: 257-276.

<sup>206</sup> DROR Y HAMPKIAN, 2011: 204-208.

<sup>207</sup> DE KEIJSER, MALSCH, LUINING, ET AL., 2016: 71-82.

En el verano de 2015, una notable cadena de sucesos en Texas reveló que los problemas con los análisis subjetivos de mezclas complejas de ADN no se limitaban a unos pocos casos: eran sistemáticos<sup>208</sup>. El Departamento de Seguridad Pública de Texas (TX-DPS) emitió una declaración pública el 30 de junio de 2015 a la comunidad de justicia penal de Texas observando que (1) el FBI había informado recientemente de que había identificado y corregido errores menores en sus bases de datos poblacionales utilizadas para cálculos estadísticos de casos de ADN; (2) se espera que los errores no tengan un efecto significativo en los resultados; y, (3) que el Sistema de Laboratorios de Criminalística de Texas, previa solicitud, recalcularía estadísticas previamente notificadas en casos individuales.

Cuando algunos fiscales presentaron solicitudes de recálculo a TX-DPS y otros laboratorios, se quedaron sorprendidos al descubrir que algunos resultados estadísticos habían cambiado considerablemente —*por ejemplo, de 1 en 1.4 billones (anglosajones) a 1 en 36, en un caso; de 1 en 4000 a no concluyente, en otro*—. Estos fiscales buscaron la asistencia de la Comisión de Ciencia Forense de Texas (TFSC) para entender las razones del cambio y el alcance de los casos potencialmente afectados.

Tras consulta con analistas en ADN forense, el TFSC determinó que los grandes cambios observados en algunos casos no estaban relacionados con las correcciones menores en la base de datos poblacional del FBI, sino que más bien eran debidos al hecho de que los laboratorios forenses habían cambiado la forma en que calculaban la probabilidad combinada de inclusión (CPI estadística) —especialmente en cómo habían tratado fenómenos como la falta de alelos (drop-out) en determinados marcadores de ADN—.

El TFSC puso en marcha un *DNA Mixture Notification Subcommittee* (Subcomité estatal de Notificación de Mezclas de perfiles de ADN), que incluyó representantes de unidades de integridad de condenas, fiscales de distrito y condado, abogados defensores, representantes de ciertos Proyectos Inocencia, el fiscal general del estado y el gobernador de Texas. Por septiembre de 2015, el TX-DSP había generado una lista, condado por condado, de más de 24.000 casos de mezclas de perfiles de ADN analizados entre 1999 y 2015. Debido a que TX-DPS es responsable de aproximadamente la mitad de los casos en el estado, el número total de casos de ADN de Texas que requieren revisión puede exceder de los 50.000 (Aunque no se han realizado esfuerzos comparables en otros estados, es probable que el problema tenga alcance nacional y no serlo específicamente solo en los laboratorios forenses en Texas).

TFSC convocó también un panel internacional de analistas científicos —de la *Harvard Medical School* (facultad de medicina de Harvard), el *University of North Texas Health Science Center* (Centro de ciencias de la salud de la Universidad del Norte de Texas), la *New Zealand's Forensic Research Unit* (Unidad de Investigación

---

<sup>208</sup> Pueden encontrarse documentos relevantes y más detalles en: [www.fsc.texas.gov/texas-dna-mixture-interpretation-casereview](http://www.fsc.texas.gov/texas-dna-mixture-interpretation-casereview). Lynn Garcia, Consejera General de la Comisión de Ciencia Forense de Texas, también proporcionó un resumen útil a PCAST.

Forense de Nueva Zelanda) y el NIST— para clarificar el uso apropiado de la CPI. Estos científicos presentaron observaciones en un congreso público, en el que muchos abogados aprendieron por primera vez hasta qué punto los análisis de mezclas de ADN conllevaba interpretaciones subjetivas. Salieron a colación muchos de los problemas de la CPI, porque las directrices existentes no especificaron de manera clara, adecuada o correcta el uso apropiado o las limitaciones de esa aproximación estadística.

En resumen, la interpretación de las mezclas complejas de perfiles de ADN con la CPI estadística ha resultado ser un método inadecuadamente especificado —y de este modo, inapropiadamente subjetivo—. Como tal, el método carece claramente de validez en sus fundamentos.

En un intento de colmar este vacío, los analistas convocados por TFSC escribieron un documento científico conjunto, publicado online el 31 de agosto de 2016<sup>209</sup>. El documento subraya la «necesidad perentoria ... de una estandarización del procedimiento, formación y de pruebas (de aptitud) continuas de los analistas de ADN». Los autores proponen un conjunto de normas específicas para el uso de la CPI estadística.

Las normas propuestas son claramente *necesarias* para un método científicamente válido que aplique la CPI. Como el artículo vio luz cuando este informe se estaba terminando de redactar, el PCAST no ha dispuesto del tiempo adecuado para valorar si las normas son también suficientes para definir un método válido, objetiva y científicamente, para la aplicación de la CPI.

### 5.2.3. Esfuerzos actuales para desarrollar métodos objetivos

Dada la existencia de esos problemas, algunos grupos han realizado esfuerzos para desarrollar programas computerizados de «genotipado probabilístico» que aplican diversos algoritmos para interpretar las mezclas complejas. Hasta marzo de 2014, se habían desarrollado al menos 8 programas de genotipado probabilístico (con el nombre de LRmix, Lab Retriever, likeLTD, FST, Armed Xpert, TrueAllele y DNA View Mixture Solution), siendo algunos de ellos de código abierto y otros productos comerciales<sup>210</sup>. El Laboratorio del FBI empezó a utilizar el software STRmix hace menos de un año, diciembre de 2015, y aún sigue en proceso de publicación su propia validación interna de desarrollo.

Estos programas informáticos de software genotipado probabilístico representan claramente una mejora importante sobre la interpretación subjetiva pura. Sin embargo, aún se requiere un cuidadoso escrutinio para determinar (1) si los métodos son

---

<sup>209</sup> BIEBER, BUCKLETON, BUDOWLE, ET AL., «Evaluation of forensic DNA mixture evidence: protocol for evaluation, interpretation, and statistical calculations using the combined probability of inclusion». *BMC Genetics*. [www.bmcgenet.biomedcentral.com/articles/10.1186/s12863-016-0429-7](http://www.bmcgenet.biomedcentral.com/articles/10.1186/s12863-016-0429-7).

<sup>210</sup> El tema se trata en BUTLER, 2015: 333-48.

científicamente válidos, incluyendo la definición de las limitaciones de su fiabilidad (es decir, las circunstancias en las cuales puedan ofrecer resultados poco fiables) y (2) si el software implementa correctamente los métodos. Esto es particularmente importante porque los programas emplean diferentes algoritmos matemáticos y pueden dar diferentes resultados para el mismo perfil mezcla<sup>211</sup>.

Cualquier evaluación apropiada de los métodos propuestos ha de consistir en estudios realizados por múltiples grupos, *no asociados con los desarrolladores del software*, que investiguen el funcionamiento y definan las limitaciones de los programas, probándolos sobre un amplio rango de mezclas con diferentes propiedades. En particular, es importante abordar los siguientes problemas:

(1) ¿Cuán bien funciona el método en función del número de contribuyentes a la mezcla? ¿Qué nivel de funcionamiento tiene el método cuando el número de contribuyentes a la mezcla es *desconocido*?

(2) ¿Cuán bien funciona el método en función del número de alelos compartidos entre individuos en la mezcla? Y con relación a esto, ¿cómo funciona cuando las mezclas incluyen individuos emparentados?

(3) ¿Cuán bien funciona el método —y cómo se degrada la precisión— en función de cantidades absolutas y relativas de ADN procedente de varios contribuyentes? Por ejemplo, puede resultar difícil determinar si un pequeño pico en el perfil de la mezcla representa un verdadero alelo procedente de un contribuyente minoritario o tratarse de un pico secundario de un alelo procedente de otro contribuyente diferente. (En particular, esta cuestión está en el fondo de un caso actual que ha recibido una atención considerable<sup>212</sup>).

(4) ¿En qué circunstancias —y por qué— el método produce resultados (probabilidades de inclusión aleatorias) que difieren sustancialmente de las producidas por otros métodos?

---

<sup>211</sup> Algunos programas utilizan métodos discretos (semicontinuos), que usan solo información alélica junto con probabilidades de pérdida y ganancia de alelos, mientras que otros programas utilizan métodos continuos, que también incorporan información sobre altura de picos y de otro tipo. Dentro de esas dos clases, los programas difieren con respecto a cómo utilizan la información. Algunos de los métodos hacen asunciones sobre el número de individuos que contribuyen a un perfil de ADN y utilizan esa información para eliminar ruido (como los picos «stutter» en un perfil de ADN).

<sup>212</sup> En este caso, los analistas utilizaron dos programas informáticos distintos (STRMix y TrueAllele) y obtuvieron conclusiones diferentes sobre si el ADN del acusado pudiera estar incluido dentro de un perfil mezcla de baja cantidad de ADN obtenido de una muestra recogida en una de las uñas de la víctima. El juez dictaminó que la muestra de ADN que implicaba al acusado era inadmisibile. MCKINLEY, 2016. Véase: [www.nytimes.com/2016/07/25/nyregion/potsdam-boys-murder-case-mayhinge-on-statistical-analysis.html](http://www.nytimes.com/2016/07/25/nyregion/potsdam-boys-murder-case-mayhinge-on-statistical-analysis.html) (última consulta el 22 de agosto de 2016). SOMMERSTEIN, «DNA results will not be allowed in Hillary murder trail.» North Country Public Radio (última consulta el 1 de septiembre de 2016). La decisión puede encontrarse aquí: [www.northcountrypublicradio.org/assets/files/08-26-16DecisionandOrder-DNAAnalysisAdmissibility.pdf](http://www.northcountrypublicradio.org/assets/files/08-26-16DecisionandOrder-DNAAnalysisAdmissibility.pdf)

Se han publicado una serie de artículos dirigidos a abordar algunas de estas cuestiones que analizan mezclas conocidas<sup>213</sup>. Hay que tener en cuenta dos cosas sobre estos estudios. La primera cuestión es que la mayoría de los estudios que evalúan paquetes de software han sido realizados por los propios desarrolladores del software. Mientras que resulta plenamente adecuado que los propios desarrolladores de un método lo evalúen, el establecimiento de la validez científica también requiere que otros grupos de científicos que no desarrollaron el método lo hagan. Y la segunda cuestión es que ha habido pocos estudios comparativos entre los métodos para evaluar las diferencias entre ellos —y hasta donde llega nuestro conocimiento, no se han realizado estudios comparativos dirigidos por grupos independientes—<sup>214</sup>.

Y lo más importante, los actuales estudios han explorado adecuadamente sólo un rango limitado de tipos de mezclas (con respecto al número de contribuyentes, la proporción de contribuyentes minoritarios, y la cantidad total de ADN). Los dos métodos más usados (STRMix y TrueAllele) aparecen como fiables dentro de un cierto rango, fundamentándose en la evidencia disponible y en la dificultad inherente al problema<sup>215</sup>. Específicamente, estos métodos aparecen como fiables en mezclas de tres personas, en las cuales, el contribuyente minoritario constituye al menos el 20% del ADN intacto en la mezcla y en el cual, la cantidad de ADN excede el mínimo nivel requerido por el método<sup>216</sup>.

Para mezclas más complejas (por ejemplo, mayor número de contribuyentes o proporciones menores), hay relativamente poca evidencia publicada<sup>217</sup>. En genética

---

<sup>213</sup> Por ejemplo: PERLIN, HORNYAK, SUGIMOTO y MILLER, 2015: 857-868; GREENSPOON, SCHIERMEIER-WOOD, ET AL., 2015:1263-1276; BRIGHT, TAYLOR, MCGOVERN, ET AL., 2016: 226-239; BRIGHT, TAYLOR, CURRAN, ET AL., 2014:102-110; TAYLOR, BUCKLETON y EVETT, 2015: 165-171; TAYLOR y BUCKLETON, 2015: 13-16.

<sup>214</sup> BILLE, WEITZ, COBLE, ET AL., 2014: 3125-3133.

<sup>215</sup> La interpretación de las mezclas de ADN incrementa su dificultad a medida que crece el número de contribuyentes. Véanse, por ejemplo: TAYLOR, BUCKLETON, EVETT, 2015: 165-171; BRIGHT, TAYLOR, MCGOVERN, ET AL., 2016: 226-39; BRIGHT, TAYLOR, CURRAN y BUCKLETON, 2014:102-110; BIEBER, BUCKLETON, BUDOWLE, ET AL., «Evaluation of forensic DNA mixture evidence: protocol for evaluation, interpretation, and statistical calculations using the combined probability of inclusion». *BMC Genetics*. [bmcgenet.biomedcentral.com/articles/10.1186/s12863-016-0429-7](https://doi.org/10.1186/s12863-016-0429-7).

<sup>216</sup> Las muestras mezcladas de tres personas que tienen proporciones similares son más fáciles de interpretar debido al número limitado de alelos y a la relativa similitud de las alturas de los picos. Los métodos pueden ser aplicados fiablemente a muestras procedentes de una única fuente y a mezclas simples, con tal que, en los casos en que dos contribuyentes no puedan separarse por extracción diferencial, la proporción del contribuyente minoritario no sea demasiado baja (por ejemplo, al menos un 10%).

<sup>217</sup> Para mezclas de cuatro personas, por ejemplo, hay artículos que describen validaciones experimentales con mezclas conocidas utilizando TrueAllele que incluyen 7 y 17 mezclas diferentes, respectivamente, con relativamente grandes cantidades de ADN (al menos 200 picogramos), mientras que aquellos artículos en los que se usa STRMix incluyen 2 y 3 mezclas diferentes, respectivamente, aunque usan mucha menos cantidad de ADN (en el rango de 10 picogramos). GREENSPOON, SCHIERMEIER-WOOD y JENKINS, 2015:1263-1276; PERLIN, HORNYAK, SUGIMOTO y MILLER, 2015: 857-868; TAYLOR, 2014: 144-153; TAYLOR, BUCKLETON y EVETT, 2015: 165-171; TAYLOR y BUCKLETON, 2015: 13-16; BRIGHT, TAYLOR, MCGOVERN, COOPER, RUSSELL, ABARNO y BUCKLETON, 2016: 226-239.

molecular humana, la validación experimental de un método diagnóstico importante necesitaría generalmente cientos de muestras distintas<sup>218</sup>. Un científico forense dijo al PCAST que han sido analizadas, de hecho, muchas más muestras diferentes, pero que los datos no se han recopilado y publicado<sup>219</sup>. Como consecuencia de que la evidencia empírica es esencial para establecer la validez de los fundamentos de un método, el PCAST urge a los científicos forenses a que presenten y publiquen en revistas científicas estudios de validación de alta calidad que propiamente establezcan el rango de fiabilidad de los métodos para el análisis de mezclas de ADN complejas.

Una vez que se publiquen más estudios, es probable que sea posible ampliar el rango en el que la validez científica ha sido establecida para incluir muestras más desafiantes. Como se señaló anteriormente, dichos estudios deben realizarse por grupos de investigación independientes que no estén relacionados con los desarrolladores de los métodos y sin intereses por el resultado.

#### 5.2.4. Conclusión

El PCAST, fundamentándose en su evaluación de la literatura publicada hasta la fecha, llegó a una serie de conclusiones sobre la validez de los fundamentos de los métodos para el análisis de mezclas de ADN complejas. Resaltamos que la validez de los fundamentos debe establecerse con respecto a un método específico que se aplica en un determinado rango. Además de formar su propio juicio, el PCAST también realizó una consulta al Sr. Jonn Butler, asistente especial del Director de Ciencias Forenses en el NIST y Vicepresidente del NCFS. Butler estuvo de acuerdo con el resultado de la investigación del PCAST.

#### **HALLAZGO 3: ANÁLISIS DE ADN DE MUESTRAS DE MEZCLAS COMPLEJAS**

*Validez de los fundamentos.* El PCAST ha averiguado que:

(1) Métodos basados en la Probabilidad Combinada de Inclusión (CPI). El análisis de ADN de mezclas complejas basado en aproximaciones fundamentadas en la CPI ha sido un método inadecuadamente especificado y subjetivo que tiene potencial para conducir a resultados erróneos. Como tal, no es válido en sus fundamentos.

Un artículo reciente ha propuesto normas específicas que abordan una serie de problemas con el uso de la CPI. Esas normas son claramente *necesarias*. Sin embargo, el PCAST no ha tenido tiempo suficiente para valorar si son también suficientes para definir un método científicamente válido y objetivo. Si, por un tiempo limitado, los tribunales deciden admitir resultados basados en la aplicación de la CPI, la validez de su

<sup>218</sup> La preparación y la realización de la amplificación PCR de cientos de mezclas de ADN es sencilla; puede conseguirse en unas pocas semanas o en menos tiempo.

<sup>219</sup> Entrevista del PCAST a John Buckleton, científico principal del Instituto de Ciencia Medioambiental y de Investigación en Nueva Zelanda y codesarrollador de STRMix.

aplicación requeriría que, como mínimo, fuera consistente con las normas especificadas en el artículo.

El análisis de ADN de mezclas complejas debe dirigirse sin demora hacia métodos más apropiados basados en genotipado probabilístico.

(2) Genotipado probabilístico. El análisis objetivo de mezclas complejas de ADN con software de genotipado probabilístico es un enfoque relativamente nuevo y prometedor. Se requiere evidencia empírica para establecer la validez de los fundamentos de cada método dentro de rangos específicos. Hasta el momento, hay evidencia publicada que apoya la validez de los fundamentos del análisis, con algunos programas, de mezclas de ADN de tres individuos, en las que el menor contribuyente constituye, al menos, el 20% del ADN intacto en la mezcla y en las que la cantidad de ADN excede del mínimo requerido por el método. El rango en el que ha quedado establecida la validez de los fundamentos es probable que crezca en la medida en que se obtenga evidencia adecuada para mezclas más complejas y se publique.

Validez en la aplicación. Para métodos que son válidos en sus fundamentos, la validez en la aplicación conlleva consideraciones similares a las realizadas para análisis de ADN procedente de una única fuente y para mezclas simples, con un énfasis especial en asegurar que el método se aplique correctamente y dentro de su rango empíricamente establecido.

### 5.2.5. El camino a seguir

Existe un camino claro para ampliar el rango de métodos objetivos establecidos como válidos en sus fundamentos —específicamente, a través de la publicación de los estudios científicos pertinentes—.

Esos esfuerzos se verán favorecidos por la creación y difusión (bajo restricciones relacionadas con el uso apropiado de los datos y su privacidad) de grandes colecciones de cientos de perfiles de ADN a partir de mezclas conocidas —que representen la amplia variedad de la complejidad con respecto (1) al número de contribuyentes, (2) a las relaciones entre los contribuyentes, (3) las cantidades absolutas y relativas de materiales y (4) al estado de conservación de los materiales —que pueden utilizarse por grupos independientes para evaluar y comparar los métodos—. De forma destacada, el denominado *PROVEDIt Initiative (Project Research Openness for Validation with Experimental Data)* (Proyecto de transparencia de la investigación para la validación con datos experimentales) de la Universidad de Boston ha logrado que pueda disponerse de 25.000 perfiles de mezclas de ADN para la investigación<sup>220, 221</sup>.

<sup>220</sup> Véase: [www.bu.edu/dnamixtures](http://www.bu.edu/dnamixtures).

<sup>221</sup> La colección contiene muestras con mezclas de ADN de 1 a 5 personas, amplificadas con ADN molde (*targets ranging*) en el rango de 1 a 0.007 nanogramos. En las mezclas multi-personas, la relación de los contribuyentes varía de 1:1 a 1:19. Adicionalmente, los perfiles se generaron utilizando una variedad de condiciones de laboratorio a partir de muestras que contenían ADN prístino; ADN dañado por rayos UV; ADN degradado enzimática o sónica; y ADN con inhibidores.

Además de estudios científicos sobre conjuntos comunes de muestras con el propósito de evaluar la validez de los fundamentos, los laboratorios forenses querrán conducir sus propios estudios internos de validación del desarrollo de modo individual, para valorar la validez del método por sí mismos<sup>222</sup>.

El NIST debe ser líder en este proceso, asegurando la creación y diseminación de los materiales y estimulando estudios por grupos independientes mediante subvenciones, contratos y premios; y evaluando los resultados de esos estudios.

### 5.3. *Análisis de marcas mordeduras*

#### 5.3.1. Metodología

El análisis de marcas de mordedura es un método subjetivo. Generalmente conlleva examinar marcas dejadas sobre la víctima o un objeto en la escena del crimen y comparar esas marcas con impresiones dentales tomadas al sospechoso<sup>223</sup>. La comparación de marcas de mordedura se fundamenta en las premisas: (1) las características dentales, particularmente la disposición de los dientes frontales, difieren sustancialmente entre las personas y (2) la piel (o alguna otra superficie marcada en la escena del crimen) puede capturar fiablemente esas características distintivas.

El análisis de marcas de mordedura comienza con la decisión del analista sobre si una herida es una marca causada por dientes humanos<sup>224</sup>. Si es así, el analista toma fotografías o impresiones de la mordedura en cuestión y de la dentición del sospechoso, compara la marca de mordedura con la dentición y determina si la dentición: (1) no puede excluirse como la que ha dejado la marca, (2) cabe excluirla de haber dejado la marca o (3) es no concluyente. Los estándares de marcas de mordedura no están bien definidos respecto el grado de similitud que debe identificarse para apoyar una conclusión fiable sobre si la marca pudo producirse o no por la dentición en cuestión. Las conclusiones sobre estos asuntos se dejan a juicio del analista.

#### 5.3.2- Estudios sobre cuestiones de fondo

Antes de adentrarnos en la cuestión de la validez de los fundamentos, abordamos algunos estudios sobre cuestiones de fondo (concernientes a temas como la unicidad y la consistencia) que arrojan algo de luz sobre el asunto. Estos estudios ponen en serias dudas las premisas de esta disciplina.

---

<sup>222</sup> El Laboratorio del FBI ha completado recientemente un estudio de validación de desarrollo y lo está preparando para su publicación.

<sup>223</sup> Menos frecuentemente se encuentran marcas sobre el sospechoso que pueden proceder de la víctima.

<sup>224</sup> AFBO Bitemark Methodology Standards and Guidelines (Estándares y Guías de la Metodología de Marcas de Mordedura de AFBO, el Consejo Americano de Odontología Forense), [abfo.org/wp-content/uploads/2016/03/ABFO-BitemarkStandards-03162016.pdf](http://abfo.org/wp-content/uploads/2016/03/ABFO-BitemarkStandards-03162016.pdf) (última consulta el 6 de julio de 2016).

Un artículo ampliamente citado de 1984 afirmaba que «la dentición humana era única más allá de toda duda razonable»<sup>225</sup>. El estudio examinó cuidadosamente 397 marcas de mordedura en una oblea de cera, midió 12 parámetros en cada una, y asumiendo, sin evidencia alguna, que los parámetros no estaban correlacionados —sugirió que la probabilidad de que dos marcas de mordedura tuvieran los mismos parámetros era menor a 1 entre 6 trillones (anglosajones)—. El artículo fue teórico en vez de empírico: no intentó realmente comparar las marcas de mordedura entre sí.

Un artículo en 2010 desacreditó esas afirmaciones<sup>226</sup>. Tras estudiar 344 moldes dentales humanos y medirlos con un escáner láser tridimensional, estos autores demostraron que las coincidencias ocurrían mucho más frecuentemente que lo esperado por el modelo teórico. Por ejemplo, el modelo teórico predijo que la probabilidad de encontrar *incluso una simple coincidencia entre cinco dientes* en la colección de marcas de mordedura era menor a 1 entre 1 millón; aunque la comparación empírica reveló que se encontraron 32.

En particular, estos estudios examinaron patrones de dentición humana medidos en condiciones idealizadas. Por contra, se ha demostrado que la piel es un medio poco fiable para registrar patrones precisos de los dientes. Los estudios de marcas de mordedura realizados en cerdos vivos<sup>227</sup> (utilizados como un modelo de piel humana), o en cadáveres humanos<sup>228</sup>, han demostrado una significativa distorsión en todas las direcciones. Un estudio de 2010 sobre marcas de mordedura creadas experimentalmente y producidas por personas identificadas concluyó que la deformación de la piel distorsiona las marcas de mordedura de forma tan sustancial y con tal variabilidad que los actuales procedimientos para comparar marcas de mordedura son incapaces de excluir o incluir fiablemente a un sospechoso como potencial mordedor («Los datos derivados no mostraban correlación alguna y no eran reproducibles, es decir, la misma dentición podría no producir una impresión medible que fuera consistente en todos los parámetros en cualquiera de las circunstancias de la prueba»)<sup>229</sup>. Tal distorsión es aún más complicada en el contexto de los casos penales, en los que la mordedura tiene lugar con frecuencia durante forcejeos y en los que la piel puede estirarse y contorsionarse en el momento en que se produce la marca de la mordedura.

La investigación empírica sugiere que los odontólogos forenses no están de acuerdo consistentemente incluso en la determinación de si una herida es una mordedura humana. Un estudio del Consejo Nortamericano de Odontología Forense (AFBO)<sup>230</sup>

---

<sup>225</sup> RAWSON, OMMEN, KINARD, ET AL., 1984: 245-53.

<sup>226</sup> BUSH, M.A., BUSH, P.J. Y SHEETS, 2011: 118-23.

<sup>227</sup> DORION, ED., 2011.

<sup>228</sup> SHEETS, BUSH, P.J. Y BUSH M.A., 2012: 202-207; BUSH, M.A., MILLER, BUSH, P.J, ET AL., 2009: 167-76.

<sup>229</sup> BUSH M.A., COOPER Y DORION, 2010: 976-983.

<sup>230</sup> FREEMAN Y PRETTY, «Construct validity of bitemark assessments using the ABFO decision tree», presentación en el Congreso Anual de la Academia Americana de Ciencias Forenses. Véase: [online.wsj.com/public/resources/documents/ConstructValidBMdecisiontreePRETTYFREEMAN.pdf](http://online.wsj.com/public/resources/documents/ConstructValidBMdecisiontreePRETTYFREEMAN.pdf).

supuso la observación de 100 fotos de patrones de lesiones por parte de analistas de marcas de mordedura certificados por la AFBO y se les pidió que contestaran a tres cuestiones básicas sobre (1) la existencia de pruebas suficientes para emitir una opinión sobre si el patrón de la lesión era el de una mordedura humana; (2) si la marca era una mordedura humana, si era posible sugerir esa posibilidad o si no lo era; y (3) si las características distintivas (arcos y marcas dentales) eran identificables<sup>231</sup>. De los 38 analistas que completaron el estudio, se informó que hubo acuerdo unánime en la primera pregunta en solo 4 de los 100 casos y acuerdo en, al menos, el 90% en solo 20 de los 100 casos. En las tres preguntas hubo un acuerdo de al menos el 90% en solo 8 de los 100 casos.

En un estudio similar en Australia, a 15 odontólogos se les mostró una serie de seis marcas de mordedura de casos contemporáneos, cinco de los cuales habían sido reconocidas como producidas por dientes por víctimas sobrevivientes, y se les pidió que explicaran de forma narrativa si las heridas eran realmente marcas de mordedura<sup>232</sup>. El estudio mostró amplia variabilidad entre los analistas en sus conclusiones sobre el origen, circunstancias y características de los patrones de las heridas en las seis imágenes. Sorprendentemente, los más experimentados (21 o más años) tendían a tener el más amplio rango de opiniones sobre si la marca era de origen dental humano o no<sup>233</sup>. Las opiniones de los analistas sobre si pensaban que una marca era apta para su análisis variaron considerablemente y los profesionales demostraron poca consistencia en sus enfoques en el análisis de una marca desde una mordedura a otra. El estudio concluyó que «esta inconsistencia indica un defecto fundamental en la metodología del análisis de las marcas de mordedura y debe conducir a preguntarnos, con preocupación, sobre la fiabilidad de las conclusiones alcanzadas sobre las coincidencias entre una marca de mordedura y una dentición»<sup>234</sup>.

### 5.3.3. Estudios de validez científica y fiabilidad

Como se dijo más arriba, la validez de los fundamentos de un método subjetivo solo puede establecerse mediante múltiples estudios de caja negra independientes.

---

<sup>231</sup> Los datos están a disposición de los interesados previa petición a los autores. Fueron revisados por la Profesora Karen Kafadar, miembro del panel de consejeros *senior* para este estudio.

<sup>232</sup> PAGE, TAYLOR y BLENKIN, 2013: 664-672.

<sup>233</sup> Por ejemplo, un analista manifestó que con certeza una de las imágenes era de una marca de mordedura, diciendo: «Conozco, desde la experiencia, que es de dientes porque tuve un caso al comienzo del año que cuando miré por primera vez las imágenes, a causa de la severidad de las heridas, no pensé que fueran de dientes. Pero cuando vi los modelos y los observé atentamente, me parecieron que lo eran». Otro expresó dudas de que la misma imagen fuera una marca de mordedura, también basándose en su experiencia: «Honestamente, no creo que sea una marca de mordedura... hay muchas cosas que podrían haber causado eso. Dudo de que haya aquí marcas de dientes individuales. Nunca he visto nada como esto». *Ibid.*, 666.

<sup>234</sup> *Ibid.*, 670.

El informe del NRC (2009) aseveró que la validez científica de los análisis de marcas de mordedura no había sido establecida<sup>235</sup>. El PCAST, en su propia revisión de la literatura, encontró pocos estudios empíricos que intentasen estudiar la validez y fiabilidad de los métodos para identificar los orígenes de una marca de mordedura.

En un artículo de 1975, a dos analistas les pidieron que emparejasen fotografías de marcas de mordedura realizadas por 24 voluntarios en la piel de cerdos recién sacrificados con modelos dentales de esos mismos voluntarios<sup>236</sup>. Las fotografías se tomaron entre 0,1 y 24 horas después de que se produjeran las marcas de mordedura. El rendimiento de los analistas fue pobre, deteriorándose conforme el transcurrir del tiempo. La proporción de fotografías incorrectamente atribuidas fue del 28%, 65% y 84% tras 0, 1 y 24 horas.

En un artículo de 1999, a 29 forenses analistas dentales —y a 80 personas más, incluyendo dentistas en general, estudiantes de odontología y otros participantes legos en la materia— les mostraron impresiones en color de mordeduras humanas de 50 casos judiciales y les pidieron que decidiesen si cada marca de mordedura era de un adulto o de un niño<sup>237</sup>. Las decisiones fueron comparadas con los veredictos de todos los casos. Todos los grupos realizaron su función deficientemente<sup>238</sup>.

En un artículo del año 2001, a 32 diplomados certificados por AFBO se les pidió que informaran sobre su certeza respecto si 4 marcas de mordedura específicas procedían de 7 modelos dentales, consistentes en 4 fuentes correctas y 3 de muestras no relacionadas<sup>239,240</sup>. Tal diseño de «conjunto cerrado» (donde la fuente correcta está presente para cada una de las muestras cuestionadas) es inapropiado para valorar la fiabilidad, porque tenderá a subestimar la tasa de falsos positivos<sup>241</sup>. Incluso con el

---

<sup>235</sup> «Hay una discusión continua sobre el valor y validez científica de la comparación e identificación de las marcas de mordedura». National Research Council, 2009: 151.

<sup>236</sup> WHITTAKER, 1975: 166–71.

<sup>237</sup> WHITTAKER, BRICKLEY y EVANS, 1998: 11–20.

<sup>238</sup> Los autores preguntaron a los observadores que indicasen hasta qué punto tenían certeza de que la marca de mordedura había sido realizada por un adulto utilizando una escala de 6 puntos. Se calcularon y representaron curvas ROC (*Receiver-Operator Characteristic*) a partir de los datos. El área bajo la curva (AUC) se calculó para cada grupo (donde AUC = 1 representa una clasificación perfecta y AUC = 0.5 es equivalente a una decisión aleatoria). El área bajo la curva (AUC) quedó en el intervalo 0.62–0.69, que representa una ejecución pobre.

<sup>239</sup> ARHEART y PRETTY, 2001: 104–111.

<sup>240</sup> De las cuatro marcas de mordedura, tres eran de casos penales y una producida por un individuo que deliberadamente mordió un trozo de queso. Los siete modelos dentales procedían de tres acusados en casos penales en los que fueron declarados culpables (presumiblemente fueron los que hicieron las mordeduras), el individuo que mordió el queso y tres individuos no relacionados.

<sup>241</sup> En tests de conjunto cerrado, los analistas realizarán bien su cometido siempre y cuando elijan el modelo dental más parecido. En un diseño de conjunto abierto, en el que puede que ninguno de los modelos sea correcto, la probabilidad de falsos positivos es más alta. El diseño de conjunto abierto se parece a la aplicación en casos reales. Véase la extensa discusión sobre diseños en conjunto cerrado en análisis de armas de fuego (Sección 5.5).

diseño de conjunto cerrado, el 11% de las comparaciones con fuente incorrecta fueron declaradas coincidencias «probables», «posibles», o de «razonable certeza médica».

En otro artículo de 2001, a 10 diplomados con certificado AFBO les dieron 10 tests independientes, cada uno consistente en evidencia de marca de mordedura y dos fuentes posibles. La evidencia se produjo aplicando un modelo dental a cerdos recién sacrificados, confirmando subjetivamente que «fueron registrados suficientes detalles» y fotografiando la marca de la mordedura. La fuente correcta estuvo presente en todos menos en dos de los tests (un diseño en conjunto cerrado mayor). La tasa media de falsos positivos fue del 15.9% —es decir, aproximadamente 1 de cada 6—.

En un artículo del año 2010, a 29 analistas con distintos niveles de entrenamiento (incluidos 9 diplomados con certificado AFBO) se les proporcionaron 18 fotografías de marcas de mordedura y la dentición humana de tres individuos (A, B y C) y se les pidió que decidieran si las marcas de mordedura procedían de A, B o C, o de ninguno de los tres. Las marcas de mordedura habían sido producidas en cerdos vivos, utilizando una máquina de morder con dentición de los individuos A, B y D (del cual no se proporcionó su dentición a los analistas). De las marcas de mordedura producidas por D, los diplomados declararon erróneamente una coincidencia con A, B o C en el 17% de los casos —de nuevo, aproximadamente 1 de cada 6—.

#### 5.3.4. Conclusión

Se han realizado pocos estudios empíricos para estudiar la capacidad de los analistas en identificar con precisión la fuente de una marca de mordedura. Entre esos estudios que se han llevado a cabo, las tasas de falsos positivos observadas fueron tan altas que el método claramente no es científicamente fiable en la actualidad. (Más aún, algunos de esos estudios emplean diseños de conjunto cerrado inapropiados que probablemente subestiman la tasa de falsos positivos).

#### **HALLAZGO 4: ANÁLISIS DE MARCAS DE MORDEDURA**

*Validez de los fundamentos.* El PCAST encuentra que el análisis de marcas de mordedura no cumple los estándares científicos para la validez de los fundamentos, y está lejos de cumplir tales estándares. Al contrario, la evidencia científica disponible sugiere fuertemente que los analistas no pueden estar de acuerdo consistentemente sobre si una herida es de mordedura humana y no pueden identificar la fuente de una marca de mordedura con precisión razonable.

#### 5.3.5. El camino a seguir

Algunos profesionales han expresado su preocupación por el hecho de que la exclusión de las marcas de mordedura en los tribunales pueda obstaculizar los esfuerzos

para conseguir la condena de los acusados en algunos casos<sup>242</sup>. Si así fuera el caso, la solución correcta, desde una perspectiva científica, no sería la de admitir testimonio pericial basado en métodos inválidos y no fiables, sino más bien en intentar desarrollar métodos fiables desde un punto de vista científico.

Sin embargo, el PCAST considera que las perspectivas de desarrollar el análisis de marcas de mordedura como un método científicamente válido son bajas. Aconsejamos no dedicar recursos significativos a tales esfuerzos.

#### 5.4. *Análisis de huellas dactilares latentes*

Los análisis de huellas latentes se propusieron primeramente para su uso en identificación criminal en el siglo XIX y han sido utilizados durante más de un siglo. El método fue aclamado durante mucho tiempo como infalible, a pesar de la falta de estudios apropiados para valorar su tasa de error. Como se dijo anteriormente, esta escasez de pruebas empíricas manifestaba una grave debilidad en la cultura científica de la ciencia forense —donde la validez fue asumida en lugar de probada—. Citando anteriores guías que ahora se reconoce que han sido inapropiadas<sup>243</sup>, el DOJ recientemente señaló,

[H]istóricamente, fue práctica común que un analista testificara que cuando la ... metodología era correctamente aplicada, siempre produciría la conclusión correcta. De este modo, cualquier error que ocurriera sería un error humano y la tasa de error resultante de la metodología sería cero. Este enfoque fue descrito por el Departamento de Justicia en 1984 en la publicación titulada *The Science of Fingerprints* (La Ciencia de las Huellas Dactilares), donde se afirma que «de todos los métodos de identificación, sólo el de las huellas dactilares ha resultado ser infalible y factible<sup>244</sup>.

En respuesta al informe del NRC (2009), la disciplina del análisis de las huellas latentes ha hecho progresos en el reconocimiento de la necesidad de realizar estudios empíricos para valorar la validez de los fundamentos y medir la fiabilidad. Merece especial reconocimiento el laboratorio del FBI, que ha liderado el camino en la realización tanto de estudios de caja negra, diseñados para medir la fiabilidad, como de «estudios de caja blanca», diseñados para comprender los factores que afectan a las decisiones de los analistas<sup>245</sup>. El PCAST aplaude los esfuerzos del FBI. También existen esfuerzos incipientes para empezar a transformar la disciplina desde un método puramente subjetivo hacia un método objetivo, aunque existe aún un considerable camino que recorrer para lograr ese importante objetivo.

---

<sup>242</sup> La proporción precisa de casos en los que las marcas de mordedura juegan un papel clave no está clara, aunque es claramente pequeña.

<sup>243</sup> Federal Bureau of Investigation, 1984: iv.

<sup>244</sup> Véase: [www.justice.gov/olp/file/861906/download](http://www.justice.gov/olp/file/861906/download).

<sup>245</sup> Véanse: HICKLIN, BUSCAGLIA, ROBERTS, ET AL., 2011: 385-419; HICKLIN, BUSCAGLIA y ROBERTS, 2013: 106-17; ULERY, HICKLIN, KIEBUZINSKI, ET AL., 2013: 99-106; ULERY, HICKLIN y BUSCAGLIA, 2012; ULERY, HICKLIN, ROBERTS, ET AL., 2015: 54-61.

### 5.4.1. Metodología

El análisis de huellas dactilares conlleva generalmente (1) comparar una «huella latente» (una impresión con crestas papilares completas o parciales de un sujeto desconocido) que ha sido revelada u observada sobre un vestigio con (2) una o más «impresiones conocidas» (huellas dactilares deliberadamente coleccionadas en un entorno controlado de sujetos conocidos; también referidas como «decadactilares»), para valorar si las dos pueden haberse originado a partir de la misma fuente. (También puede conllevar la comparación de huellas latentes entre sí).

Es importante distinguir las huellas latentes de las huellas conocidas. Una impresión de huellas conocidas contiene imágenes de huellas dactilares hasta un total de diez, que se toman en un entorno controlado, como en una detención o en una verificación de antecedentes<sup>246</sup>. Como las impresiones conocidas tienden a ser de alta calidad, pueden buscarse automática y fiablemente en grandes bases de datos. Por contra, las huellas latentes reveladas en casos penales a menudo son incompletas y de calidad variable (manchadas o distorsionadas de otro modo), con calidad y claridad dependiente de factores como la superficie tocada y la mecánica del tacto.

Puede requerirse un analista para (1) comparar una huella latente con las huellas de una impresión decadactilar de un sospechoso conocido que ha sido identificado por otros medios («sospechoso identificado») o para (2) buscar en una gran base de datos de huellas dactilares para identificar a un sospechoso («búsqueda en base de datos»).

Los analistas siguen ordinariamente un proceso denominado «ACE» o «ACE-V», para el Análisis, Comparación, Evaluación y Verificación<sup>247, 248</sup>. El proceso pide a los analistas que realicen una serie de valoraciones subjetivas. Un analista utiliza juicio subjetivo para seleccionar regiones particulares de una huella latente para su análi-

---

<sup>246</sup> Véase: Committee on Science, Subcommittee on Forensic Science of the National Science and Technology Council, 2014. [www.whitehouse.gov/sites/default/files/microsites/ostp/NSTC/afis\\_10-20-2014\\_draftforcomment.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/NSTC/afis_10-20-2014_draftforcomment.pdf).

<sup>247</sup> «Un examen de una huella latente utilizando el proceso ACE-V consiste en lo siguiente: *Análisis* se refiere a una fase de recogida de información inicial en la que el analista estudia la impresión desconocida para valorar la calidad y cantidad de detalles discriminantes presentes. El analista considera aspectos como el sustrato, el método de desarrollo, diversos niveles de detalle de las crestas y distorsiones debida a la presión. Luego realiza un análisis separado de la huella impresa de autoría conocida. La *comparación* consiste en la observación simultánea del detalle de las crestas de fricción en las dos impresiones para determinar el acuerdo o desacuerdo en los detalles. En la fase de *Evaluación*, el analista valora el acuerdo o desacuerdo de la información observada durante el Análisis y la Comparación y redacta una conclusión. La *Verificación* en algunas unidades policiales o instituciones forenses es una revisión de las conclusiones del analista con conocimiento de esas conclusiones; en otras entidades es un nuevo examen independiente realizado por un segundo analista que no conoce el resultado del primer examen». National Institute of Standards and Technology, 2012. Disponible en: [www.nist.gov/oles/upload/latent.pdf](http://www.nist.gov/oles/upload/latent.pdf).

<sup>248</sup> REZNICEK, RUTH y SCHILENS, 2010: 87-103.

sis. Si no hay personas identificadas en las que se pueda estar interesado, el analista someterá la huella latente a una búsqueda en el *Automated Fingerprint Identification System (AFIS)* (Sistema Automático de Identificación Dactilar)<sup>249</sup>, que contiene grandes números de huellas conocidas, que utiliza algoritmos de reconocimiento de imágenes patentadas (no públicas)<sup>250</sup> para generar una lista de potenciales candidatos que comparten características dactilares similares<sup>251</sup>. El analista, después, compara manualmente la huella latente con las huellas dactilares de la persona específica de interés o con las huellas de los candidatos con mayores coincidencias generados por el ordenador, mediante el estudio de las características seleccionadas<sup>252</sup>, y luego toma una decisión de forma subjetiva sobre si son lo suficientemente similares como para hacer una propuesta de identificación.

ACE-V añade una etapa de verificación. Con respecto a la verificación, la implementación varía ampliamente<sup>253</sup>. En muchos laboratorios solo se verifican las identificaciones, porque se considera demasiado gravoso, en términos de tiempo y coste, realizar exámenes independientes en todos los casos (por ejemplo, con exclusiones). Este procedimiento es problemático porque no es ciego: el segundo analista conoce que el primero hizo una propuesta de identificación, ello genera un potencial sesgo de confirmación. Tras el caso de la identificación errónea en el atentado con bombas en los trenes de Madrid (véase más adelante), el Laboratorio del FBI adoptó medidas para llevar a cabo, en ciertos casos, «la aplicación independiente de ACE a una impre-

<sup>249</sup> El estado y las jurisdicciones locales comenzaron a comprar sistemas *AFIS* en las décadas de los 70 y 80 a proveedores privados, cada uno con su propio software y algoritmos de búsqueda patentado. En 1999, el FBI puso en marcha el Sistema Automático de Identificación de Huellas Dactilares Integrado (*IAFIS*), una base de datos nacional que contiene huellas dactilares y antecedentes policiales de más de 70 millones de individuos introducidos por agencias policiales estatales, locales y federales (recientemente reemplazado por el *Next Generation Identification (NGI) System* (Sistema de Identificación de la Próxima Generación). Algunas agencias de justicia penal tienen la capacidad de buscar huellas latentes no solo en su propia base de datos dactilar, sino también en una jerarquía de bases de datos locales, estatales y federales. Sin embargo, todavía no se ha alcanzado la total interoperabilidad del sistema. Véase: Committee on Science, Subcommittee on Forensic Science of the National Science and Technology Council, 2014. [www.whitehouse.gov/sites/default/files/microsites/ostp/NSTC/afis\\_10-20-2014\\_draftforcomment.pdf](http://www.whitehouse.gov/sites/default/files/microsites/ostp/NSTC/afis_10-20-2014_draftforcomment.pdf).

<sup>250</sup> Los algoritmos utilizados para generar candidatos de potenciales coincidencias están patentados y no están públicamente disponibles.

<sup>251</sup> El Laboratorio del FBI requiere a los analistas que completen y documenten sus análisis de las huellas dactilares latentes antes de revisar cualquier huella conocida o moverse a la fase de comparación y evaluación. Este requisito no es compartido por todos los laboratorios.

<sup>252</sup> Las características dactilares se comparan en tres niveles de detalle: —nivel 1 («flujo de las crestas»), nivel 2 («camino de las crestas») y nivel 3 («características de las crestas» o «formas»). «El flujo de las crestas» se refiere a las clases de patrones compartidos por muchos individuos como formaciones en bucle o espiral; este nivel es solo suficiente para realizar exclusiones, no para declarar identificaciones. «El camino de las crestas» se refiere a los puntos característicos que pueden ser utilizados para declarar identificaciones como las bifurcaciones o los puntos. «Las formas de las crestas» incluyen las terminaciones de las crestas y la localización de poros. Véase: National Institute of Standards and Technology, 2012. Disponible en: [www.nist.gov/oles/upload/latent.pdf](http://www.nist.gov/oles/upload/latent.pdf).

<sup>253</sup> BLACK, 2012: 80-100.

sión de crestas de fricción por otro analista cualificado, que no conociera la conclusión del primer analista»<sup>254</sup>. En particular, el Laboratorio del FBI utiliza verificación ciega en los casos considerados hasta el presente como los que tienen mayor riesgo de error, tal como cuando una única huella es identificada, excluida o se considera no concluyente<sup>255</sup>.

Como se resaltó en la sección 2, las anteriores preocupaciones sobre la fiabilidad de los análisis de huellas latentes<sup>256</sup>, se incrementaron sustancialmente tras el destacado error de identificación de una huella latente recuperada en el atentado con explosivos en el año 2004 en el sistema de trenes de cercanías de Madrid. Un analista del FBI concluyó que con «una certeza del 100%» la huella coincidía con la de Brandon Mayfield, un norteamericano de Portland (Oregón), incluso aunque las autoridades españolas no pudieron confirmar la identificación. Los revisores creen que la identificación errónea se debió, en parte, al «sesgo de confirmación» o «razonamiento inverso» —es decir, el ir de la impresión conocida (la del sospechoso) a la imagen latente, de forma que llevó a confiar exacerbadamente en las aparentes similitudes y prestar inadecuada atención a las diferencias—<sup>257</sup>. Como se describió en un reciente artículo de científicos del Laboratorio del FBI,

[U]n notable ejemplo del problema del sesgo a partir de una huella de autoría conocida que da lugar a un razonamiento circular ocurrió en el error de identificación del caso del atentado en Madrid, en el que el primer analista reinterpretó cinco de los originales siete puntos de análisis como más consistentes con la huella conocida (incorrecta). «Habiendo encontrado hasta diez puntos característicos de inusual similitud, los analistas del FBI comenzaron a “encontrar” adicionales características en LPF 17 [la huella latente] que realmente no estaban allí, sino que fueron sugeridas a los analistas por las características en las impresiones de Mayfield»<sup>258</sup>.

A diferencia de los análisis de ADN, las reglas para declarar una identificación que históricamente se utilizaron en los análisis de huellas dactilares no fueron establecidas de antemano ni fueron uniformes entre los analistas. Como se describió en un informe realizado por un Grupo de Trabajo de Analistas fechado en febrero de 2012, encargado por el NIST y el NIJ:

Los umbrales para estas decisiones pueden variar entre los analistas y entre los proveedores de servicios forenses. Algunos analistas afirman que declaran una identificación si encuentran un

<sup>254</sup> U.S. Department of Justice, Office of the Inspector General, 2011. [www.oig.justice.gov/special/s1105.pdf](http://www.oig.justice.gov/special/s1105.pdf). Véase también: Federal Bureau of Investigation. Laboratory Division, 2007 (modificada el 24 de mayo de 2011).

<sup>255</sup> Federal Bureau of Investigation. Laboratory Division, 2007 (modificado el 24 de mayo de 2011).

<sup>256</sup> FAIGMAN, KAYE, SAKS, ET AL. (EDS.), 2016; SAKS, 1994.

<sup>257</sup> Una revisión del tratamiento del FBI del caso Brandon Mayfield. U.S. Department of Justice, Office of the Inspector General, 2006. [oig.justice.gov/special/s0601/final.pdf](http://oig.justice.gov/special/s0601/final.pdf).

<sup>258</sup> ULERY, HICKLIN, ROBERTS, ET. AL., 2015: 54-61. La cita interna es del U.S. Department of Justice, Office of the Inspector General. Una revisión del tratamiento del FBI del caso Brandon Mayfield (marzo 2006), [www.justice.gov/oig/special/s0601/PDF\\_list.htm](http://www.justice.gov/oig/special/s0601/PDF_list.htm). US Department of Justice Office of the Inspector General: A review of the FBI's handling of the Brandon Mayfield case (March 2006), [www.justice.gov/oig/special/s0601/PDF\\_list.htm](http://www.justice.gov/oig/special/s0601/PDF_list.htm).

número particular de relativamente raras concurrentes características, por ejemplo, ocho o doce. Otros no utilizan ningún número fijo estándar. Algunos analistas descartan detalles aparentemente diferentes siempre que encuentren suficientes similitudes entre las dos impresiones. Otros analistas practican la regla de una disimilitud, excluyendo una impresión si existe una única disimilitud no atribuible a una distorsión perceptible. Si el analista decide que el grado de similitud se queda corto para satisfacer el estándar, puede declarar un resultado no concluyente. Si la conclusión es que el grado de similitud satisface el estándar, el analista declara una identificación<sup>259</sup>.

En septiembre de 2011, el *Scientific Working Group on Friction Ridge Analysis, Study and Technology (SWGFAST)* (Grupo de Trabajo Científico sobre Análisis, Estudio y Tecnologías de Crestas de Fricción) emitió el documento «*Standards for Examining Friction Ridge Impressions and Resulting Conclusions (Latent/Tenprint)*» (Estándares para el Examen de Impresiones de Crestas de Fricción y Conclusiones Resultantes (Latentes/Decadactilares)) que comienza a transformar el análisis de las huellas latentes en la dirección de un marco objetivo. En particular, sugiere criterios sobre qué combinación de calidad de imagen y cantidad de características (por ejemplo, número de puntos característicos compartidos entre dos huellas dactilares) serían suficientes para declarar una identificación. Los criterios no son todavía plenamente objetivos, pero suponen un paso en la dirección correcta. El Subcomité de Crestas de Fricción de OSAC ha reconocido la necesidad de contar con criterios objetivos al identificar las «Necesidades de Investigación»<sup>260</sup>. Observamos que los estudios de caja negra que describimos a continuación no se propusieron probar estos criterios específicos, por lo que aún no han sido científicamente validados.

#### 5.4.2. Estudios de validez científica y fiabilidad

Como se dijo previamente, la validez de los fundamentos de los métodos subjetivos puede *solo* establecerse a través de múltiples estudios independientes de caja negra apropiadamente diseñados para evaluar la validez y la fiabilidad.

Más abajo explicamos diversos estudios de análisis de huellas latentes. Los primeros cinco no se idearon como estudios de validación, aunque proporcionan alguna información incidental sobre el rendimiento. Notablemente, solo ha habido dos estudios de caja negra que fueron intencionada y apropiadamente diseñados para valorar la validez y la fiabilidad —el primero fue publicado por el Laboratorio del FBI en 2011; el segundo se terminó en 2014, pero no se ha publicado todavía—. Las conclusiones sobre la validez de los fundamentos deben apoyarse sobre esos dos estudios recientes.

Para resumir esos estudios, aplicamos la guía escrita anteriormente en este informe (véase Sección 4 y Apéndice A). Primero, aunque observamos tanto (1) la tasa estimada de falsos positivos como (2) el límite superior de confianza del 95% de falsos positivos, nos centramos en el segundo al considerarla como la tasa más apropiada

<sup>259</sup> Véase: NIST, 2012. Disponible en: [www.nist.gov/oles/upload/latent.pdf](http://www.nist.gov/oles/upload/latent.pdf).

<sup>260</sup> Véase: [workspace.forensicosac.org/kws/groups/fric\\_ridge/documents](http://workspace.forensicosac.org/kws/groups/fric_ridge/documents).

desde el punto de vista científico para informar a un jurado —porque su principal preocupación debería ser subestimar la tasa de falsos positivos, pudiendo llegar a ser la tasa real razonablemente tan alta como este valor—<sup>261</sup>. Segundo, aunque notamos tanto (1) la tasa de falsos positivos en las examinaciones *concluyentes* (identificaciones o exclusiones) como (2) *todas* las examinaciones (incluyendo las no concluyentes) que sean relevantes, nos centramos en la primera al ser la tasa desde el punto de vista científico más apropiada para informar a un jurado —porque la prueba de huellas dactilares utilizada ante un tribunal contra un acusado será generalmente el resultado de un examen concluyente—.

#### 5.4.2.1. Evett y Williams (1996)

Este artículo es un ensayo de revisión histórica discursiva que contiene una breve descripción de un pequeño «estudio de colaboración» relevante para la precisión de los análisis de huellas dactilares<sup>262</sup>. En este estudio, a 130 analistas de Inglaterra y Gales muy experimentados, cada uno con al menos diez años de experiencia en análisis de huellas dactilares forenses, se les presentaron diez parejas de huellas latentes conocidas. Nueve de las parejas procedían de casos realizados en el pasado por la Scotland Yard, suponiéndose «correctamente emparejadas» (es decir, procedentes de la misma fuente). La décima pareja no estaba realmente emparejada (procedían de fuentes diferentes), incluyendo una latente deliberadamente producida en un vaso de cerveza con hoyuelos. Sobre esta única pareja dispar, los analistas no hicieron identificaciones falsas. Como el artículo no distinguió entre exclusiones y examinaciones no concluyentes (y los autores no disponen ya de los datos)<sup>263</sup>, no es posible inferir el límite superior del 95% de confianza<sup>264</sup>.

#### 5.4.2.2. Langeburg (2009a)

En un pequeño estudio piloto, el autor examinó la ejecutoria de seis analistas sobre 60 tests cada uno<sup>265</sup>. Solo hubo 15 examinaciones concluyentes relacionados con parejas dispares en origen (véase la Tabla 1 del artículo). Hubo un falso positivo que el autor excluyó porque pareció ser un error administrativo que no se repitió en una nueva prueba posterior. Incluso si ese error se excluye, el pequeño tamaño de la muestra da lugar a un enorme intervalo de confianza (límite superior del 95% de confianza del 19%), con ese límite superior correspondiendo a 1 error cada 5 casos.

---

<sup>261</sup> Por convención, el límite superior del nivel de confianza del 95% es el más utilizado en estadística para reflejar el rango de valores posibles (véase Apéndice A).

<sup>262</sup> EVETT y WILLIAMS, 1996: 49-73.

<sup>263</sup> I.W. Evett, información dada en persona.

<sup>264</sup> Por ejemplo, el límite superior del 95% de confianza sería 1 cada 44 si los 130 exámenes fueran concluyentes y 1 cada 22 si lo fueran la mitad.

<sup>265</sup> LANGENBURG, 2009: 219-257.

#### 5.4.2.3. Langeburg (2009b)

En este pequeño estudio piloto para el siguiente artículo, el autor comprobó la destreza de los analistas en una sala de conferencias dentro de una convención de analistas en identificación forense<sup>266</sup>. Los analistas fueron divididos en tres grupos: alto-sesgo (n=16), bajo-sesgo (n=12), y control (n=15). A cada grupo se le presentaron 6 parejas de huellas latentes conocidas, 3 emparejadas realmente y otras tres no. Los dos primeros grupos recibieron información diseñada para sesgar su juicio al aumentar su atención, mientras que el grupo de control recibió una descripción genérica. Con respecto a las no emparejadas, el grupo de control tuvo 1 falso positivo entre 43 exámenes concluyentes. La tasa de falsos positivos fue del 2.3% (límite superior del 95% de confianza en el 11%), con un límite superior correspondiente a 1 error cada 9 casos<sup>267, 268</sup>.

#### 5.4.2.4. Langeburg, Champod y Genessay (2012)

Este estudio no fue diseñado para valorar la precisión de los análisis de huellas dactilares latentes, sino más bien para explorar cómo los analistas en huellas dactilares incorporarían información proveniente de nuevas herramientas desarrolladas (tales como una herramienta de calidad para ayudar a la evaluación de la claridad de los detalles de las crestas de fricción; una herramienta estadística para proporcionar relaciones de verosimilitudes que representen la fuerza de las correspondientes características entre las huellas comparadas; e información de consenso de un grupo de analistas en huellas entrenados) en sus procesos de toma de decisiones<sup>269</sup>. Sin embargo, el estudio provee alguna información sobre la precisión del análisis de huella dactilares latentes. Brevemente, se les pidió a 158 analistas (y a algunos en capacitación) que cotejaran 12 parejas de huellas latentes indubitadas, 7 emparejadas y 5 no emparejadas realmente. Con las no emparejadas hubo 17 falsos positivos entre 711 exámenes concluyentes de los analistas<sup>270</sup>. La tasa de falsos positivos fue del 2.4% (límite superior del 95% de confianza del 3.5%). El error estimado se corresponde con 1 error cada 42 casos, con un límite superior correspondiente a 1 error cada 28 casos<sup>271</sup>.

---

<sup>266</sup> LANGENBURG, CHAMPOD y WERTHEIM, 2009: 571-82.

<sup>267</sup> Si se incluyen los dos exámenes inconclusos, los valores son solo ligeramente diferentes: 2.2 % (límite superior del 95% de confianza del 10.1%), con una apuesta de 1 a 10.

<sup>268</sup> Los grupos sesgados no cometieron errores entre los 69 exámenes concluyentes.

<sup>269</sup> LANGENBURG, CHAMPOD y GENESSAY, 2012: 183-98.

<sup>270</sup> Agradecemos a G. Langeburg por proporcionarnos los datos de los analistas individualizados.

<sup>271</sup> Si se incluyen los 79 exámenes inconclusos, la tasa de falsos positivos fue 2.15% (límite superior del 95% de confianza en el 3.2%). La tasa de falsos positivos estimada se corresponde con 1 error cada 47 casos, con un límite superior de confianza correspondiente a 1 cada 31.

#### 5.4.2.5. Tangen et al. (2011)

Este estudio australiano fue diseñado para estudiar la fiabilidad de los análisis de huellas latentes por parte de los analistas en huellas dactilares<sup>272</sup>. Los autores pidieron a 37 analistas en huellas dactilares, así como a 37 novatos, que examinaran 36 parejas de huellas latentes conocidas —consistentes en 12 pares emparejados, 12 no emparejadas, elegidas por su «similitud» (las huellas indubitadas procedentes de fuentes diferentes mejor situadas en la clasificación en el Sistema Automático Nacional de Identificación Dactilar de Australia) y 12 no emparejadas y «no similares» (elegidas aleatoriamente a partir de otras impresiones)—. A los analistas se les pidió que midieran la probabilidad de que vinieran de la misma fuente en una escala de 1 a 12. Los autores eligieron definir las puntuaciones de 1-6 como identificaciones y de 7-12 como exclusiones<sup>273</sup>. Esta aproximación no se corresponde con los procedimientos utilizados en el examen de huellas dactilares convencional.

Con las parejas «similares» pero no emparejadas, los analistas cometieron 3 errores en 444 comparaciones; la tasa de falsos positivos fue del 0.68% (límite superior del 95% de confianza en el 1.7%), con el límite superior correspondiendo a 1 error cada 58 casos. Con los pares «no similares» no emparejados, los analistas no cometieron error alguno en las 444 comparaciones; la tasa de falsos positivos fue, de este modo, del 0% (límite superior del 95% de confianza en el 0.62%), con el límite superior correspondiendo a 1 error cada 148 casos. Los analistas superaron sustancialmente a los novatos.

Aunque es interesante, el estudio no constituye un estudio de validación de caja negra de análisis de huellas latentes porque su diseño no se asemeja a los procedimientos utilizados en la práctica forense (en particular, el proceso de asignación de calificación en una escala de 12 puntos que los autores posteriormente convirtieron en identificaciones y exclusiones).

#### 5.4.3. Estudios del FBI

El primer estudio diseñado para comprobar la validez de los fundamentos y medir la fiabilidad de los análisis de huellas dactilares latentes fue un estudio de caja negra realizado por científicos y colaboradores del FBI. Emprendido como respuesta al informe del NRC (2009), el estudio se publicó en 2011 en una revista científica líder internacional, *Proceedings of the National Academy of Sciences*<sup>274</sup>. Los autores recogieron una selección de 744 parejas de huellas latentes conocidas, consistente en 520 emparejamientos y 224 no emparejadas. Para intentar asegurar que los pares no

---

<sup>272</sup> TANGEN, THOMPSON y MCCARTHY, 2011: 995-7.

<sup>273</sup> No hubo, de este modo, resultados inconclusos en este estudio.

<sup>274</sup> ULERY, HICKLIN, BUSCAGLIA, ET AL., 2011: 7733-7338.

emparejados fuesen representativos del tipo de coincidencias que pueden presentarse cuando la policía identifica a un sospechoso buscado en la base de datos de huellas dactilares, las huellas conocidas fueron seleccionadas buscando las huellas latentes en 58 millones de huellas dactilares de la base de datos AFIS y seleccionando uno de los resultados coincidentes más cercanos. A cada uno de los 169 analistas de huellas dactilares se les mostraron 100 parejas y se les pidió que las clasificaran como una identificación, una exclusión o una inconclusión. El estudio informó que se habían producido 6 falsas identificaciones entre los 3.628 pares de huellas no emparejadas y que los analistas juzgaron que tenían «valor para la identificación». La tasa de falsos positivos fue, de este modo, del 0.17% (límite superior del 95% de confianza del 0.33%). La tasa estimada se corresponde con 1 error cada 604 casos, con el límite superior indicando que la tasa podría ser tan alta como 1 error cada 306 casos<sup>275, 276</sup>.

En 2012, los mismos autores informaron de un estudio de seguimiento que probaba la repetibilidad y la reproducibilidad. Después de un periodo de 7 meses, 75 de los analistas del estudio anterior reexaminaron un subconjunto de comparaciones de huellas latentes conocidas de ese estudio precedente. Entre los 476 pares de huellas latentes no emparejadas que condujeron a exámenes concluyentes (incluyendo 4 de las parejas que condujeron a falsos positivos en el estudio inicial y que fueron reasignadas al analista que había cometido la decisión errónea), no hubo falsos positivos. Esos resultados (límite superior del 95% de confianza en el 0.63%), correspondiente a 1 error cada 160) son ampliamente consistentes con la tasa de falsos positivos medida en el estudio previo<sup>277</sup>.

#### 5.4.3.1. Estudio Miami-Dade (Pacheco et al.)

La Oficina de Servicios Forenses del Departamento de Policía de Miami-Dade, con financiación del NIJ, llevó a cabo un estudio de caja negra diseñado para valorar la validez de los fundamentos y medir la fiabilidad; los resultados fueron entregados al patrocinador y publicado en internet, pero aún no se han publicado en una revista científica con revisión por pares<sup>278</sup>. El estudio difiere significativamente del estudio de caja negra del FBI en 2011 en importantes aspectos, incluido el hecho de que las huellas conocidas no fueron seleccionadas mediante una búsqueda en una gran base de datos para que fueran similares a las huellas latentes (que debía, en principio, haber hecho más fácil declarar exclusiones en los pares no emparejados). El estudio

---

<sup>275</sup> Si incluimos los 455 resultados inconclusos dentro de las huellas latentes juzgadas como que tienen «valor para la identificación», la tasa de falsos positivos es del 0.15% (límite superior del 95% de confianza del 0.29%). La tasa de falsos positivos estimada se corresponde con 1 error cada 681 casos, con límite superior correspondiente de 1 cada 344.

<sup>276</sup> La sensibilidad (proporción de muestras emparejadas que fueron correctamente declaradas como coincidentes) fue del 92.5%.

<sup>277</sup> En general, entre el 85-90% de los resultados concluyentes no se modificaron, con aproximadamente el 30% de falsas exclusiones repetidas.

<sup>278</sup> PACHECO, CERCHIAI y STOILOFF, 2014. [www.ncjrs.gov/pdffiles1/nij/grants/248534.pdf](http://www.ncjrs.gov/pdffiles1/nij/grants/248534.pdf).

informó de 42 falsos positivos entre 995 examinaciones concluyentes. La tasa de falsos positivos fue del 4.2% (límite superior del 95% de confianza en el 5.4%). La tasa estimada se corresponde con 1 error cada 24 casos, con un límite superior indicando que la tasa podría llegar a ser tan alta como 1 error en 18 casos<sup>279</sup>. (Observación: El artículo dice que «en 35 de las identificaciones erróneas, los participantes parecían haber cometido un error administrativo, pero los autores no pudieron determinarlo con certeza»). En estudios de validación es inapropiado excluir errores de manera *post hoc* (véase Cuadro 4). Sin embargo, si esos errores fueran excluidos, la tasa de falsos positivos sería del 0.7% (intervalo de confianza del 1.4%), con límite superior correspondiente a 1 error cada 73 casos.

#### 5.4.4. Conclusiones de los estudios

Si bien es descorazonador que los estudios significativos para valorar la validez de los fundamentos y la fiabilidad no comenzaran hasta época reciente, nos alienta que ahora se estén haciendo serios esfuerzos para tratar de dotar a la disciplina de un fundamento científico sólido —incluso midiendo la precisión, definiendo la calidad de las huellas latentes, estudiando las razones de los errores, etc.—. Gran parte del crédito pertenece al Laboratorio del FBI, así como a académicos que habían presionado acerca de la necesidad de la investigación. De manera significativa, el FBI es responsable del único estudio de caja negra hasta la fecha que ha sido publicado en una revista científica con revisión por pares.

Los estudios mencionados no pueden compararse directamente entre sí por muchas razones —incluyendo diferencias en el diseño experimental, la selección y el nivel de dificultad de los pares de huellas latentes, así como el grado en el cual representan las circunstancias, procedimientos y presiones que se encuentran en los casos reales—. Sin embargo, pueden emitirse ciertas conclusiones a partir de los resultados de los estudios (resumidos en la Tabla 1 más abajo):

(1) Los estudios colectivamente demuestran que muchos analistas pueden, bajo *algunas* circunstancias, producir respuestas correctas con *algún* nivel de precisión.

(2) Las tasas de falsos positivos empíricamente estimadas son mucho más altas de lo que el público en general (y, por extensión, la mayoría de los miembros del jurado) probablemente creerían, basándose en las afirmaciones largamente sostenidas en el tiempo sobre la seguridad de los análisis de huellas dactilares<sup>280, 281</sup>.

<sup>279</sup> Si se incluyen los 403 exámenes inconclusos, la tasa de falsos positivos fue del 3% (límite superior del 95% de confianza del 3.9%). La tasa de falsos positivos estimada se corresponde con 1 error cada 33 casos, con límite superior correspondiente a 1 cada 26.

<sup>280</sup> La conclusión se mantiene con independencia de si las tasas están basadas en estimaciones puntuales o en límites de confianza del 95%, así como sobre exámenes concluyentes o sobre todos los exámenes.

<sup>281</sup> Esas afirmaciones incluyen la afirmación anterior del Departamento de Justicia, durante mucho tiempo mantenida, de que el análisis de huellas dactilares es «infalible». ([www.justice.gov/olp/](http://www.justice.gov/olp/))

(3) De los dos estudios de caja negra apropiadamente diseñados, el mayor estudio (el del FBI de 2011) arrojó una tasa de falsos positivos que es improbable que exceda de 1 cada 306 exámenes concluyentes, mientras que el otro (el de Miami-Dade de 2014) arrojó una tasa de falsos positivos considerablemente más alta, 1 cada 18<sup>282</sup>. (Los estudios anteriores, que no fueron diseñados como estudios de validación, arrojaron también altas tasas de falsos positivos).

En general, sería apropiado informar a los miembros del jurado que (1) solo se han realizado dos estudios correctamente diseñados sobre la precisión de los análisis de huellas latentes y que (2) esos estudios informaron de tasas de falsos positivos que pudieran llegar a ser tan altas como de 1 cada 306 casos en un estudio y 1 cada 18 en el otro. Esto informaría apropiadamente a los miembros del jurado que los errores ocurren con una determinada frecuencia, permitiéndoles sopesar el valor probatorio de la evidencia.

Es probable que un programa de verificación sistemática y ciega apropiadamente diseñado disminuyera la tasa de falsos positivos, porque los analistas en los estudios tienden a cometer errores *diferentes*<sup>283</sup>. Sin embargo, no se han realizado pruebas empíricas para obtener una estimación cuantitativa de la tasa de falsos positivos que se hubieran producido con un programa como el mencionado<sup>284</sup>. Y no sería apropiado simplemente *inferir* el impacto de la verificación independiente basándose en el supuesto teórico de que los errores de los analistas no están correlacionados<sup>285</sup>.

---

file/861906/download); el testimonio de un anterior jefe de la unidad de huellas dactilares del FBI era que «el FBI tenía una tasa de error de 1 cada 11 millones de casos» (véase p. 68); y un estudio con miembros del jurado simulados que estimaban que la tasa de falsos positivos en análisis de huellas latentes era de 1 cada 5.5 millones (véase p. 60). KOEHLER, 2016. Disponible en: [www.papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2817443](http://www.papers.ssrn.com/sol3/papers.cfm?abstract_id=2817443).

<sup>282</sup> Como se observó anteriormente, la tasa es de 1 cada 73 si ignoramos los errores supuestamente administrativos —aunque este ajuste *post hoc* no es apropiado en estudios de validación—.

<sup>283</sup> Los autores del estudio de caja negra del FBI observan que cinco de los falsos positivos ocurrieron en problemas de prueba en los que una gran mayoría de los analistas declararon correctamente una exclusión, mientras que uno ocurrió en un problema de prueba en el que la mayoría de los analistas tomaron decisiones de inconclusión. Afirman que «esto sugiere que estos errores de individualización hubieran sido detectados si se hubiese realizado una verificación ciega de manera rutinaria». ULERY, HICKLIN, BUSCAGLIA, ET AL., 2011: 7733-7738.

<sup>284</sup> El estudio Miami-Dade incluyó un pequeño test sobre el paso de la verificación, lo que supuso la verificación de 15 de los 42 falsos positivos. De esos 15 casos, el segundo analista declaró 13 casos de exclusión y 2 de inconclusión. El tamaño muestral es demasiado pequeño como para proporcionar una conclusión significativa. Y el artículo no informa sobre resultados de verificación sobre los otros 27 casos positivos.

<sup>285</sup> El DOJ ha propuesto al PCAST que «la teoría básica de la probabilidad establece que, dada una tasa de error para un analista, la probabilidad de que un segundo analista cometa exactamente el mismo error (verificación/verificación ciega) conlleva que las tasas deberían multiplicarse». Sin embargo, tal modelo teórico asume que los errores de los diferentes analistas no están correlacionados; pero ellos pueden depender de la dificultad del problema y, por tanto, estar correlacionados. Se necesitan estudios empíricos que estimen las tasas de error bajo verificación ciega.

Es importante observar que para que un programa de verificación sea completamente ciego y, por consiguiente, evite el sesgo cognitivo, los analistas no pueden solo verificar individualizaciones. Como proponen los autores del estudio de caja negra del FBI, «esto puede asegurarse realizando verificaciones sobre una mezcla de tipos de conclusiones, no meramente sobre individualizaciones» —es decir, una mezcla que asegure que quienes verifican no puedan hacer inferencias sobre las conclusiones que están siendo verificadas<sup>286</sup>—. No tenemos conocimiento de ningún programa de verificación ciego que siga actualmente esta práctica.

Hasta el presente, el testimonio que asevere cualquier nivel específico de mayor precisión (más allá de lo medido en los estudios) debido a la verificación ciega independiente, sería científicamente inapropiado, en tanto que especulación no está apoyado por evidencia empírica.

Resaltamos que el DOJ opina que la alta tasa de falsos positivos en el estudio Miami-Dade (1 cada 24, con límite superior de confianza de 1 cada 18) es improbable que sea aplicable a los casos del Laboratorio del FBI, porque cree que tan alta tasa de error hubiera sido detectada por los procedimientos de verificación del laboratorio. Una evaluación independiente de los protocolos de verificación podría arrojar luz sobre hasta qué punto tales inferencias podrían realizarse partiendo procedimientos de verificación actuales del laboratorio.

También resaltamos que es posible que la tasa de falsos positivos en casos reales pueda ser más alta que la observada en los estudios experimentales, debido a la exposición a información potencialmente susceptible de producir sesgos en el transcurso de un caso. La introducción de muestras de prueba ciega en el flujo de casos podría proporcionar información valiosa sobre las tasas de error reales en los casos en curso.

En conclusión, el estudio de caja negra del Laboratorio del FBI ha contribuido a un significativo avance en la disciplina. Es necesario el estudio continuo sobre la fiabilidad de los análisis de huellas latentes, basándose en el estudio de su diseño. Los estudios deben, idealmente, estimar las tasas de error en huellas latentes de diversos niveles de «calidad», utilizando medidas bien definidas (idealmente, medidas objetivas implementadas mediante software automático<sup>287</sup>). Como se señaló anteriormente, los estudios deben ser diseñados y dirigidos junto con terceras partes que no

---

<sup>286</sup> ULERY, HICKLIN, BUSCAGLIA, ET. AL., 2011: 7733-7738.

<sup>287</sup> Un ejemplo es *Latent Quality Assessment (LQAS)* (Valoración de la calidad de la latente), diseñado como una herramienta de prueba de un concepto para evaluar la claridad de las impresiones. Hay estudios que demuestran que las tasas de error están correlacionadas con la calidad de las impresiones. El software proporciona un manual y definiciones automatizadas de mapas de claridad, funciones para procesar mapas de claridad y la anotación de los puntos correspondientes que proporcionan un método para superponer áreas de impresión. HICKLIN, BUSCAGLIA y ROBERTS, 2013: 106-117. Otro ejemplo es *Picture Annotation System (PiAnoS)*, desarrollado por la Universidad de Lausana, que está siendo probado como una métrica de calidad y una herramienta de valoración estadística para analistas. Esta plataforma utiliza herramientas que (1) valoran la calidad de los detalles de las crestas de fricción, (2) proporciona relaciones de verosimilitud que representan la fuerza de las características coincidentes

posean interés en el resultado. Este importante rasgo no estuvo presente en el estudio del FBI.

**TABLA 1: TASAS DE ERROR EN ESTUDIOS DE ANÁLISIS DE HUELLAS LATENTES<sup>o</sup>**

ESTUDIO	FALSOS POSITIVOS			
	Datos originales	Frecuencia (límite de confianza)	Tasa estimada	Límite en tasa estimada
Estudios anteriores				
Langeburg (2009a)	0/14	0% (19%)	1 en $\infty$	1 en 5
Langeburg (2009b)	1/43	2.3% (11%)	1 en 43	1 en 9
Langeburg et al. (2012)	17/711	2.4% (3.5%)	1 en 42	1 en 28
Tangen et al. (2011) («pares similares»)	3/444	0.68% (1.7%)	1 en 148	1 en 58
Tangen et al. (2011) («pares diferentes»)	0/444	0% (0.67%)	1 en $\infty$	1 en 148
Estudios de caja negra				
Ulery et al. 2011 (FBI) <sup>oo</sup>	6/3628	0.17% (0.33%)	1 en 604	1 en 306
Pacheco et al. 2014 (Miami-Dade)	42/995	4.2% (5.4%)	1 en 24	1 en 18
Pacheco et al. 2014 (Miami-Dade) (excluyendo errores administrativos)	7/960	0.7% (1.4%)	1 en 137	1 en 73

<sup>o</sup> «Datos originales»: Número de falsos positivos dividido por el número de examinaciones concluyentes incluyendo pares no coincidentes. «Frecuencia (límite de confianza)»: Estimación puntual de la frecuencia de falsos positivos, y límite superior del 95% de confianza. «Tasa estimada»: la apuesta de que ocurra un falso positivo, basada en la proporción observada de falsos positivos. «Límite de la tasa estimada»: la apuesta de ocurrencia de un falso positivo, basado en el límite superior del 95% de confianza—es decir, la tasa podría ser razonablemente tan alta como este valor—.

<sup>oo</sup> Si los exámenes inconclusos se incluyen en el estudio del FBI, las tasas son 1 cada 681 y 1 cada 344, respectivamente.

#### 5.4.5. Estudios científicos sobre cómo los analistas de huellas latentes alcanzan sus conclusiones

Complementando los estudios de caja negra, algunos estudios han arrojado luz de forma importante sobre cómo los analistas de huellas latentes alcanzan sus conclusiones y cómo esas conclusiones pueden verse influenciadas por factores externos. Esos estudios subrayan los serios riesgos que pueden suscitarse con métodos subjetivos.

entre las huellas y (3) brinda información de consenso procedente de un grupo de analistas en huellas dactilares con experiencia. *PiAnoS* es un software de fuente abierta disponible en: [ipslabs.unil.ch/pianos](http://ipslabs.unil.ch/pianos).

#### 5.4.5.1. Estudios sobre sesgos cognitivos

Itiel Dror y sus colegas han realizado un trabajo pionero sobre el potencial papel de los sesgos cognitivos en los análisis de huellas latentes<sup>288</sup>. En un estudio exploratorio de 2006, demostraron que el juicio de los analistas podía verse influenciado por las decisiones de otros analistas forenses (una forma de «sesgo de confirmación»)<sup>289</sup>. A cinco analistas de huellas dactilares se les facilitó pares de huellas dactilares que ellos habían estudiado cinco años antes en casos reales y que habían declarado «coincidentes». Se les pidió que reexaminaran las impresiones, pero se les indujo a creer que ellas eran un par de impresiones que habían sido declaradas erróneamente coincidentes por el FBI en un caso con notoriedad. Aunque fueron instruidos para que ignorasen esta información, cuatro de los cinco analistas dejaron de considerar a las huellas como «coincidentes». Aunque estos estudios son demasiado reducidos como para proporcionar estimaciones precisas del impacto de los sesgos cognitivos, han sido fundamentales para llamar la atención sobre la cuestión.

Se han propuesto diversas estrategias para mitigar los sesgos cognitivos en los laboratorios forenses, incluyendo la gestión del flujo de información en un laboratorio de criminalística para minimizar la exposición del analista forense a información de contexto irrelevante (tales como confesiones o identificaciones de testigos oculares) y asegurar que los analistas trabajen de un modo lineal, documentando sus hallazgos sobre la evidencia recogida en la escena del crimen *antes* de que realicen comparaciones con muestras de un sospechoso<sup>290, 291</sup>.

#### 5.4.5.2. Estudios de caja blanca del FBI

En los últimos años, científicos del FBI y sus colaboradores también han llevado a cabo una serie de estudios de «caja blanca» para comprender los factores subyacentes en el proceso de análisis de huellas dactilares latentes. Esos estudios incluyen análisis de la calidad de las huellas<sup>292, 293</sup>, los procesos de los analistas para determinar el valor de una huella latente para su identificación o su exclusión<sup>294</sup>, la suficiencia de

---

<sup>288</sup> DROR, CHARLTON y PERON, 2006: 74-878; DROR y CHARLTON, 2006: 600-616.

<sup>289</sup> DROR, CHARLTON y PERON, 2006: 74-878.

<sup>290</sup> KASSIN, DROR y KAKUCKA, 2013: 42-52. Véase también: KRANE, FORD, GILDER, *et al.*, 2008: 1006-7.

<sup>291</sup> La información de contexto irrelevante podría, dependiendo de su naturaleza, sesgar la opinión de un analista conduciéndole hacia una incorrecta identificación o a una incorrecta exclusión. Ambos resultados son indeseables.

<sup>292</sup> HICKLIN, BUSCAGLIA, ROBERTS, *et al.*, 2011: 385-419.

<sup>293</sup> HICKLIN, BUSCAGLIA y ROBERTS, 2013: 106-17.

<sup>294</sup> ULERY, HICKLIN, KIEBUZINSKI, *et al.*, 2013: 99-106.

información para las identificaciones<sup>295</sup> y cómo cambia la valoración de los analistas sobre una huella latente cuando la comparan con una posible huella coincidente<sup>296</sup>.

Entre el trabajo realizado sobre métodos subjetivos de comparación de características, esta serie de artículos es única en su amplitud, rigor y voluntad de explorar temas desafiantes. No pudimos encontrar análisis autorreflexivos similares para otras disciplinas subjetivas.

Los dos artículos más recientes son particularmente notables porque abordan el grave asunto del sesgo de confirmación.

En un artículo de 2014, los científicos del FBI escribieron:

ACE distingue entre la fase de Comparación (valoración de las características) y la fase de Evaluación (determinación), implicando que las determinaciones están basadas en la valoración de las características. Sin embargo, nuestros resultados sugieren que no se trata de una simple relación causal: las marcas de los analistas están también influenciadas por sus determinaciones. No es obvio el modo en que se produce esta influencia inversa. Los analistas pueden alcanzar subconscientemente una determinación preliminar rápidamente y esto influye en su comportamiento durante la Comparación (por ejemplo, en el nivel de esfuerzo realizado, la manera de tratar las características ambiguas). Después de tomar una decisión, los analistas pueden luego revisar sus anotaciones como ayuda para documentar esa decisión y pueden estar más motivados a proporcionar un marcado metódico y cuidadoso en apoyo de las individualizaciones más que otras determinaciones. Como evidencia en apoyo de nuestra conjetura, observamos en particular las distribuciones de los recuentos de puntos característicos, que muestran un aumento escalonado asociado con los umbrales de decisión: este paso ocurrió en torno a los siete puntos característicos para la mayoría de los analistas, aunque en torno a 12 para aquellos que siguen el estándar de los 12 puntos<sup>297</sup>.

Dror *et al.* han hecho observaciones similares al notar que el número de puntos característicos marcados en una huella latente era mayor cuando estaba presente una huella indubitada<sup>298</sup>. Además, Evett y Williams describieron cómo los analistas británicos, que utilizaban el estándar de 16 puntos para declarar una identificación, utilizaban una huella indubitada (con la que se compara la latente) para «descartar los puntos discordantes» de la huella latente después de que hubieran alcanzado la «íntima convicción» de que las huellas eran coincidentes<sup>299</sup>.

En un informe complementario del año 2015, los científicos del FBI estudiaron cuidadosamente cómo los analistas analizaban las impresiones y confirmaron que, en la gran mayoría de las decisiones de identificación (> 90%), los analistas modificaron los puntos característicos marcados en la huella latente como respuesta a una aparente coincidencia con una huella dactilar conocida (más a menudo añadiendo que

---

<sup>295</sup> ULERY, HICKLIN y BUSCAGLIA, 2012.

<sup>296</sup> ULERY, HICKLIN, ROBERTS, *et al.*, 2015: 54-61.

<sup>297</sup> ULERY, HICKLIN, ROBERTS y BUSCAGLIA, 2014.

<sup>298</sup> DROR, CHAMPOD, LANGENBURG, *et al.*, 2011: 10-17.

<sup>299</sup> EVETT y WILLIAMS, 1996: 49-73.

sustrayendo puntos característicos)<sup>300</sup>. (El único falso positivo en su estudio fue un caso extremo en el que la conclusión estuvo casi enteramente basada en el marcado subsiguiente de puntos característicos que no habían sido inicialmente encontrados y en el borrado de puntos catacterísticos que habían sido previamente marcados).

Los autores concluyeron que «hay necesidad de que los analistas tengan algún medio de documentar sin ambigüedad lo que ven durante el análisis y la comparación (en el proceso ACE-V)» y que «métodos rigurosamente definidos y consistentemente aplicados para realizar y documentar ACE-V mejoraría la transparencia del proceso de examen de las huellas latentes».

El PCAST felicita a los científicos del FBI por llamar la atención sobre el riesgo de sesgo de confirmación derivado de un razonamiento circular. Como una cuestión de validez científica, a los analistas se les debe requerir que «completen y documenten sus análisis de una huella latente antes de mirar cualquier huella conocida» y «deben documentar por separado cualquier dato en el que se basaron durante la comparación o la evaluación que difiera de la información en la que se basaron durante el análisis»<sup>301</sup>. El FBI adoptó estas reglas tras el error de identificación relacionado con el atentado con explosivos de los trenes de Madrid; tales reglas deben ser universalmente adoptadas por todos los laboratorios.

#### 5.4.6. Validez en la aplicación

La validez de los fundamentos significa que un gran grupo de analistas examinando un específico tipo de muestra puede, bajo las condiciones de la prueba, generar respuestas correctas con una frecuencia conocida y útil. Eso no significa que un analista particular tenga la capacidad de aplicar fiablemente el método, que las muestras en los estudios de los fundamentos sean representativas de la evidencia real en los casos o que las circunstancias del estudio de los fundamentos representen una aproximación razonable a las circunstancias del caso.

Para abordar estos asuntos, los tribunales deben tener en cuenta algunas consideraciones clave:

(1) Como consecuencia de que el análisis de huellas latentes, tal y como se practica actualmente, depende de un juicio subjetivo, no está científicamente justificado concluir que un analista particular es capaz de aplicar fiablemente el método a menos que el analista haya sido sometido a pruebas de aptitud regulares y rigurosas. Desafortunadamente, no es posible asegurar que las actuales pruebas de aptitud sean apropiadas porque los problemas que se evalúan en esas pruebas no están disponibles públicamente (Como se enfatizó previamente, ni el entrenamiento ni la experiencia

<sup>300</sup> ULERY, HICKLIN, ROBERTS, ET. AL., 2015: 54-61.

<sup>301</sup> U.S. Department of Justice, Office of the Inspector General, 2011: 5, 27. [www.oig.justice.gov/special/s1105.pdf](http://www.oig.justice.gov/special/s1105.pdf).

son sustitutos, porque ninguno proporciona garantía alguna de que el analista pueda aplicar el método fiablemente).

(2) En cualquier caso, debe establecerse que las huellas latentes poseen la calidad y completitud representada en los estudios de validez de los fundamentos.

(3) Como consecuencia de que el sesgo contextual puede tener un impacto sobre las decisiones de los analistas, los tribunales deben valorar las medidas tomadas para mitigar los sesgos durante la realización de un caso —por ejemplo, asegurándose de que los analistas no están expuestos a información potencialmente sesgada y asegurarse de que los analistas documenten las características de las crestas de una huella desconocida antes de referirse a la impresión conocida (un procedimiento conocido como «ACE-V lineal»)»<sup>302</sup>.

#### HALLAZGO 5: ANÁLISIS DE HUELLAS DACTILARES LATENTES

*Validez de los fundamentos.* Basándose principalmente en dos estudios recientes de caja negra apropiadamente diseñados, el PCAST concluye que el análisis de huellas dactilares latentes es una metodología subjetiva válida en sus fundamentos —aunque con una tasa de falsos positivos importante y probablemente más alta que la esperada por muchos miembros del jurado considerando afirmaciones que han sido mantenidas durante mucho tiempo sobre la infalibilidad de los análisis de huellas dactilares.

Las conclusiones de una propuesta de identificación pueden ser científicamente válidas si están acompañadas de información precisa sobre las limitaciones de la fiabilidad de la conclusión —específicamente, que (1) solo se han llevado a cabo dos estudios apropiadamente diseñados sobre la validez de los fundamentos y la precisión de los análisis de huellas latentes; (2) estos estudios encontraron tasas de falsos positivos que podrían ser tan altas como de 1 error cada 306 casos, en uno de ellos, y de 1 error cada 18 en el otro; y (3) como consecuencia de que los analistas eran conscientes de que estaban siendo evaluados, la tasa real de falsos positivos en casos reales puede ser más alta. Hasta el presente, afirmaciones sobre una mayor precisión no están garantizadas o científicamente justificadas. Se requieren estudios de caja negra adicionales para aclarar la fiabilidad del método.

*Validez en la aplicación.* Aunque concluimos que el método es válido en sus fundamentos, hay una serie de cuestiones importantes relacionadas con la validez de su aplicación.

(1) *Sesgo de confirmación.* El trabajo realizado por científicos del FBI ha mostrado que los analistas suelen alterar las características que marcan inicialmente en una huella latente basándose en las comparaciones realizadas con una huella indubitada aparentemente coincidente. Ese razonamiento circular introduce un serio riesgo de sesgo de confirmación. A los analistas se les debe requerir que completen y documenten sus análisis de una huella latente *antes* de mirar una huella conocida y deben documentar,

<sup>302</sup> U.S. Department of Justice, Office of the Inspector General, 2011: 27. [www.oig.justice.gov/special/s1105.pdf](http://www.oig.justice.gov/special/s1105.pdf).

separadamente, cualquier dato adicional utilizado durante su comparación y evaluación.

(2) *Sesgo contextual.* El trabajo realizado por miembros de la comunidad académica ha demostrado que el juicio de los analistas puede estar influenciado por información irrelevante sobre los hechos de un caso. Deben realizarse esfuerzos para asegurar que los analistas no están expuestos a información potencialmente sesgada.

(3) *Pruebas de aptitud.* La prueba de aptitud es esencial para valorar la capacidad de un analista y su rendimiento en la emisión de juicios precisos. Como se trató sobre el particular en otro lugar de este informe, las pruebas de aptitud necesitan mejorarse haciéndolas más rigurosas, incorporándolas dentro del flujo de trabajo, y divulgando las pruebas para que ellas sean evaluadas por la comunidad científica.

Desde el punto de vista científico, la validez en la aplicación requiere que un analista: (1) haya realizado pruebas de aptitud apropiadas para asegurar que es capaz de analizar todo el rango de huellas latentes que es posible encontrar en los casos reales y que informe de los resultados de las pruebas de aptitud; (2) revele si documentó las características halladas en la huella latente por escrito antes de compararla con la impresión conocida; (3) proporcione un análisis escrito explicando la selección y comparación de las características; (4) revele si, al realizar la examinación, tenía conocimiento de otros hechos del caso que pudieran haber influido en su conclusión; y (5) verifique que la huella latente en el caso en cuestión es similar en calidad al rango de huellas latentes consideradas en los estudios de los fundamentos.

#### 5.4.7. El camino a seguir

Hacen falta continuados esfuerzos para mejorar el estado de los análisis de huellas latentes —y esos esfuerzos darán claramente dividendos para el sistema de justicia penal—.

Una primera dirección es continuar mejorando el análisis de huellas latentes como método subjetivo. Con solo dos estudios de caja negra hasta la fecha (con tasas de error muy diferentes), hay necesidad de que se realicen estudios de caja negra adicionales sobre la base del diseño del estudio de caja negra del FBI. Los estudios deben estimar las tasas de error de las huellas latentes en función de su variabilidad en calidad y completitud, utilizando mediciones bien definidas. Como se señaló anteriormente, los estudios deben diseñarse y dirigirse junto a terceros sin interés en los resultados.

Una segunda dirección —y más importante— es la de convertir los análisis de huellas latentes de un método subjetivo a un método objetivo. En las décadas pasadas se han visto avances extraordinarios en análisis de imagen automatizados basados en *machine learning* y otras aproximaciones— que han conducido a mejoras espectacu-

lares como las conseguidas en reconocimiento facial<sup>303</sup>,<sup>304</sup>. En medicina, por ejemplo, se espera que el análisis de imágenes automatizado conducirá hacia un estándar de oro para muchas aplicaciones que precisan de una interpretación de imágenes en radiografías, resonancias magnéticas, oftalmoscopías y dermatológicas<sup>305</sup>.

Los métodos objetivos basados en análisis de imagen automatizados podrían producir grandes beneficios —incluida una mayor eficiencia y menores tasas de error—; podrían permitir también la estimación de las tasas de error de millones de comparaciones por parejas. Los esfuerzos iniciales para desarrollar sistemas automatizados no pudieron superar a los seres humanos<sup>306</sup>. Sin embargo, dado el ritmo de los progresos en análisis de imágenes y de *machine learning*, creemos que es probable que el análisis de huellas latentes totalmente automatizado sea posible en un futuro próximo. Se han dado ya pasos iniciales en esa dirección, tanto en el ámbito académico como en la industria<sup>307</sup>.

El recurso más importante para impulsar el desarrollo de los métodos objetivos sería la creación de enormes bases de datos que contengan impresiones conocidas, cada una con muchas impresiones latentes «simuladas» de calidad e integridad muy diversa, que pudieran ponerse a disposición de muchos investigadores científicamente capacitados del ámbito académico y de la industria. Las impresiones latentes simuladas podrían crearse «modelando» (*morphing*) las impresiones conocidas, basándose en transformaciones derivadas de la recolección de pares de impresión latente real—impresión registrada<sup>308</sup>.

---

<sup>303</sup> Véase: [cs.stanford.edu/people/karpathy/cvpr2015.pdf](https://cs.stanford.edu/people/karpathy/cvpr2015.pdf).

<sup>304</sup> LU Y TANG, «Surpassing human-level face verification performance on LFW with Gaussian-Face.» [arxiv.org/abs/1404.3840](https://arxiv.org/abs/1404.3840) (última consulta el 2 de julio de 2016). TAIGMAN, YANG, RANZATO Y WOLF, «Deepface: Closing the gap to human-level performance in face verification» .[www.cs.toronto.edu/~ranzato/publications/taigman\\_cvpr14.pdf](https://www.cs.toronto.edu/~ranzato/publications/taigman_cvpr14.pdf) (última consulta el 2 de julio de 2016); SCHROFF, KALENICHENKO Y PHILBIN, «FaceNet: A unified embedding for face recognition and clustering». [arxiv.org/abs/1503.03832](https://arxiv.org/abs/1503.03832) (última consulta el 2 de julio de 2016).

<sup>305</sup> DOI, 2007: 198-211; SHIRAIISHI, LI, APPELBAUM, ET AL. 2011: 449-462.

<sup>306</sup> Por ejemplo, un estudio de 2010 concluyó que los seres humanos mejoraban los resultados de los programas automatizados de comparación de marcas de herramientas. Véase: CHUMBLEY, MORRIS, ET AL., 2010: 953-961.

<sup>307</sup> ARUNALATHA, TEJASWI, SHAILA, ET AL., 2015: 482-490; SRIHARI, 2013. [www.crime-sceneinvestigator.net/QuantitativeMeasuresinSupportofLatentPrint.pdf](http://www.crime-sceneinvestigator.net/QuantitativeMeasuresinSupportofLatentPrint.pdf). Además, el grupo de Christophe Champod en la Universidad de Lausana mantiene un programa activo en esta área.

<sup>308</sup> Por razones de privacidad, podrían utilizarse huellas dactilares de fallecidos.

## 5.5. *Análisis de armas de fuego*

### 5.5.1. Metodología

En los análisis de armas de fuego, los analistas intentan determinar si la munición está o no asociada con una *específica* arma de fuego basándose en las marcas de herramientas producidas por las armas sobre la munición<sup>309, 310</sup>. (Brevemente, los cañones de las armas usualmente están estriados para mejorar la precisión, lo que conlleva que se corten ranuras en espiral en el interior del cañón para dar un giro sobre el proyectil). Las imperfecciones individuales aleatorias que se producen durante el proceso del corte de la herramienta y mediante el «desgaste» por el uso del arma dejan marcas de herramienta sobre proyectiles y casquillos cuando salen del arma de fuego. Las partes del arma de fuego que entran en contacto con los casquillos tienen otros métodos mecánicos de funcionamiento).

Esta disciplina se fundamenta en la idea de que las marcas de herramienta producidas por diferentes armas de fuego varían sustancialmente lo suficiente (debido a las variaciones en su fabricación y uso) como para permitir que los componentes de los cartuchos disparados sean identificados como provenientes de armas de fuego específicas. Por ejemplo, los analistas pueden comparar los cartuchos «en cuestión» de un arma recuperados en una escena del crimen con los disparos de prueba de un arma sospechosa.

Resumidamente, el examen comienza con una evaluación de características de clase de los proyectiles y casquillos, que son características permanentes y predeterminadas anteriores a su fabricación. Si esas características de clase son diferentes, se emite una conclusión de exclusión. Si son similares, el examen procede a identificar y comparar características individuales, tales como las estrías que surgen durante el disparo de un arma en particular. De acuerdo con la *Association of Firearm and Tool Mark Examiners (AFTE)* (Asociación de Analistas de Armas de Fuego y de Marcas de Herramientas), «el método más ampliamente aceptado que se utiliza para llevar a cabo un examen de marcas de herramientas es el de una comparación microscópica conjunta entre las marcas sobre una pieza de material cuestionada y las marcas de origen conocido realizadas por una herramienta»<sup>311</sup>.

---

<sup>309</sup> Los analistas pueden también emprender otras clases de análisis, tales como determinaciones de distancias, operatividad de las armas y restauraciones de números de serie, así como el análisis de residuos de disparo para determinar si alguien utilizó el arma recientemente.

<sup>310</sup> Para descripciones más completas, véase, por ejemplo, National Research Council, 2009; y [www.archives.fbi.gov/archives/aboutus/lab/forensic-science-communications/fsc/july2009/review/2009\\_07\\_review01.htm](http://www.archives.fbi.gov/archives/aboutus/lab/forensic-science-communications/fsc/july2009/review/2009_07_review01.htm).

<sup>311</sup> Véase: Foundational Overview of Firearm/Toolmark Identification *tab* (descripción general de la pestaña sobre los fundamentos de la identificación de armas de fuego/marcas de herramientas) en [afte.org/resources/swgun-ark](http://afte.org/resources/swgun-ark) (última consulta el 12 de mayo de 2016).

### 5.5.2. *Antecedentes*

En la sección anterior, el PCAST expresó su preocupación sobre ciertos documentos fundamentales que subyacen a la disciplina científica sobre el análisis de armas de fuego y de marcas de herramienta. En particular, observamos que la «Teoría de la identificación en cuanto relacionada con las marcas de herramienta» de AFTE —que define los criterios para declarar una identificación— es circular<sup>312</sup>. La «teoría» establece que un analista puede concluir que dos vestigios tienen un origen común si sus marcas son «suficientemente concordes», donde «suficientemente concordes» se define cuando el analista está convencido de que es extremadamente improbable que los dos vestigios tengan diferente origen. Además, la «teoría» establece, explícitamente, que las conclusiones son subjetivas.

En esta disciplina científica se ha prestado mucha atención en tratar de probar la noción de que cada arma produce marcas de herramienta «únicas». En el año 2004, el NIJ pidió a NRC que estudiase la viabilidad, precisión, fiabilidad y conveniencia de desarrollar una base de datos nacional de balística integral de imágenes de proyectiles disparados por todas o casi todas las armas de nueva fabricación o importadas, con el propósito de permitir la comparación balística de una escena del crimen con un arma y con la información de su propietario inicial.

En su informe de 2008, un comité del NRC, respondiendo la solicitud del NIJ, dijo que «la validez de los presupuestos fundamentales de unicidad y reproducibilidad de las marcas de herramientas relacionadas con armas de fuego» aún no se habían demostrado y que, dados los métodos de comparación actuales, una búsqueda en una base de datos «probablemente produciría un subconjunto de candidatos a coincidencias demasiado grande como para ser útil a efectos prácticos con fines de investigación»<sup>313</sup>.

Desde luego, no es necesario que las marcas de herramientas sean únicas para que provean de información útil sobre si un proyectil ha sido disparado por un arma en particular. Sin embargo, es *esencial* que la precisión del método para compararlos se conozca mediante estudios empíricos.

Los analistas de armas de fuego llevan mucho tiempo sosteniendo que su disciplina es de una precisión cercana a la perfección. En un artículo del año 2009, el jefe de la unidad de armas de fuego-marcas de herramientas del laboratorio del FBI afirmó que «un analista cualificado, raramente, si es que ocurre alguna vez, cometerá un falso positivo (falsa identificación)», para ello cita su revisión, haciendo una declaración jurada, de estudios empíricos que mostraban virtualmente ausencia de errores<sup>314</sup>.

<sup>312</sup> ASSOCIATION OF FIREARM AND TOOL MARK EXAMINERS, 2011: 287.

<sup>313</sup> National Research Council, 2008, 3-4.

<sup>314</sup> Véase: [www.fbi.gov/about-us/lab/forensic-science-communications/fsc/july2009/review/2009\\_07\\_review01.htm](http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/july2009/review/2009_07_review01.htm)

Con respecto al análisis de armas de fuego, el informe del NRC (2009) concluyó que «no se han realizado suficientes estudios para comprender la fiabilidad y reproducibilidad de los métodos» —es decir, que la validez de los fundamentos de la disciplina no había sido establecida—<sup>315</sup>.

El *Scientific Working Group on Firearms Analysis (SWGUN)* (grupo de trabajo científico sobre análisis de armas de fuego) respondió a las críticas del informe del NRC (2009) diciendo que:

SWGUN desde hace tiempo ha sido consciente de las cuestiones científicas y sistémicas identificadas en este informe y ha estado trabajando diligentemente para abordarlas... [el informe del NRC] identifica las áreas en las que fundamentalmente debemos mejorar nuestros procedimientos para mejorar la calidad y fiabilidad de nuestros resultados científicos, así como para articular mejor la base de nuestra ciencia<sup>316</sup>.

### 5.5.3. Estudios de análisis de armas de fuego distintos a los de caja negra: análisis basados en conjuntos

Como consecuencia de que el análisis de armas de fuego es, hasta el presente, un método subjetivo de comparación de características, la validez de sus fundamentos, tal como se explicó anteriormente, *solo* puede ser establecida mediante múltiples e independientes estudios de caja negra.

Aunque el análisis de armas de fuego se ha utilizado durante muchas décadas, solo recientemente su validez ha sido sometida a pruebas empíricas significativas. En los últimos 15 años, la disciplina ha emprendido una serie de estudios con el objetivo de estimar la precisión de las conclusiones de los analistas. Aunque los resultados demuestran que los analistas pueden, en ciertas circunstancias, identificar la fuente de la munición disparada, muchos de los estudios no fueron apropiados para valorar la validez científica y estimar la fiabilidad, porque emplearon diseños artificiales que difieren en aspectos importantes de los problemas a los que se enfrentan los analistas en casos reales.

Específicamente, muchos de los estudios emplean análisis «basados en conjuntos», en los que se pide a los analistas que realicen todas las comparaciones por pares dentro de un pequeño conjunto de muestras o entre pequeños conjuntos de muestras. Por ejemplo, un análisis «dentro del conjunto» que involucra  $n$  objetos pide a los analistas que rellenen una matriz  $n \times n$  que indique cuál de los  $n(n-1)/2$  posibles

<sup>315</sup> El informe afirma que «los análisis de marcas de herramientas y de armas de fuego padecen las mismas limitaciones discutidas anteriormente que la prueba de impresiones. A causa de que no resulta suficientemente conocida la variabilidad entre las herramientas y armas individuales, no es posible especificar cuántos puntos de similitud son necesarios para alcanzar un determinado nivel de confianza en el resultado. No se han realizado suficientes estudios para comprender la fiabilidad y repetibilidad de los métodos. El comité está de acuerdo en que las características de clase son útiles para estrechar la lista de herramientas que pueden haber dejado una marca distintiva». National Research Council, 2009: 154.

<sup>316</sup> Véase: [www.swgun.org/index.php?option=com\\_content&view=article&id=37&Itemid=22](http://www.swgun.org/index.php?option=com_content&view=article&id=37&Itemid=22).

pares coinciden. Algunos científicos forenses han favorecido los diseños basados en conjuntos porque un pequeño número de objetos da lugar a un gran número de comparaciones. Sin embargo, el diseño del estudio tiene un grave defecto: las comparaciones no son *independientes* entre sí. Más bien, implican dependencias internas que (1) restringen y, por consiguiente, informan las respuestas de los analistas y (2), en algunos casos, permite a los analistas hacer inferencias sobre el diseño del estudio. (El primer punto se ilustra mediante la observación de que, si A y B se juzgan coincidentes, entonces cada elemento adicional C ha de coincidir con *ambos* o con *ninguno* de ellos, reduciendo el espacio de posibles respuestas a la mitad. Si A y B coinciden entre sí, pero no coinciden con C, esto crea dependencias adicionales. Y así sucesivamente. El segundo punto se ilustra mediante diseños de «conjunto cerrado», que se describen más abajo).

Como consecuencia de las complejas dependencias entre las respuestas, los estudios basados en conjuntos no son estudios de caja negra de diseño apropiado a partir de los cuales podamos obtener estimaciones adecuadas de precisión. Es más, el análisis de los resultados empíricos de, al menos, algunos estudios basados en el diseño de conjunto (los diseños de «conjunto cerrado»), sugiere que pueden subestimar la tasa de falsos positivos.

El director del Centro de Ciencias Forenses de la Defensa (DFSC en inglés) hizo una analogía de los estudios de conjunto cerrado con la resolución de un «sudoku», donde las respuestas iniciales se pueden utilizar para ayudar a completar las respuestas subsiguientes<sup>317</sup>. Como se explica a continuación, la disconformidad de DFSC con los estudios basados en conjuntos le condujo a financiar el primer estudio (y hasta la fecha, el único) de caja negra apropiadamente diseñado para el análisis de armas de fuego.

Seguidamente analizamos los estudios más citados basados en conjuntos. Adoptamos el mismo marco de referencia que con las huellas latentes, tomando en consideración como medidas apropiadas para mencionarlas en los informes (véase p.115) principalmente: (1) el límite superior del 95% de confianza de la tasa de falsos positivos y (2) la tasa de falsos positivos basada en la proporción de exámenes concluyentes.

#### 5.5.4. Comparación dentro de un conjunto

Algunos estudios incluyen comparaciones dentro de un conjunto, en el que se presenta a los analistas, por ejemplo, una selección de muestras y se les pide que determinen qué muestras fueron disparadas con la misma arma de fuego. Revisa-

---

<sup>317</sup> Entrevista del PCAST a Jeff Salyards, director del DFSC.

mos dos estudios frecuentemente citados con este diseño<sup>318, 319</sup>. En estos estudios, la mayoría de las muestras procedían de distintas fuentes, con solo 2 o 3 muestras procedentes del mismo origen. A lo largo de los dos estudios, los analistas identificaron 55 de las 61 coincidencias y no cometieron falsos positivos. En el primer estudio, la gran mayoría de las muestras procedentes de fuentes diferentes (97%) se declararon inconclusas; solo hubo 18 exámenes concluyentes de casquillos procedentes de fuentes diferentes y ninguna examinación concluyente de proyectiles procedentes de fuentes diferentes<sup>320</sup>. En el segundo estudio, los resultados se describen solamente en breves párrafos y el número de exámenes concluyentes para muestras procedentes de fuentes diferentes no fue reflejado. De este modo, resulta imposible estimar la tasa de falsos positivos entre las exámenes concluyentes, que es la medida clave que hay que tener en cuenta (como se explicó anteriormente).

#### 5.5.4.1. Comparación del tipo conjunto a conjunto/conjunto cerrado

Otro diseño común ha sido la comparación entre conjuntos implicando un «conjunto cerrado». En este caso, a los analistas se les entrega un conjunto de muestras cuestionadas y se les pide que las comparen con un conjunto de estándares conocidos, que representan las posibles armas de las cuales se dispararon las municiones cuestionadas. En un diseño de «conjunto cerrado», el arma fuente siempre está presente. Analizamos cuatro de esos estudios en detalle<sup>321, 322, 323, 324</sup>. En estos estudios, a los analistas

<sup>318</sup> SMITH, E., 2005: 130-135. En este estudio del FBI, se dispararon casquillos y proyectiles de 9 pistolas Ruger P89 incautadas de casos reales. A los analistas se les dio cajas (de casquillos o proyectiles) con muestras disparadas por cada una de las 9 pistolas y una muestra adicional disparada con una de las pistolas; se les pidió que determinaran qué muestras habían sido disparadas con la misma arma. De las 16 comparaciones procedentes de la misma fuente, 13 (muestras) fueron identificadas y 3 se declararon inconclusas. De las 704 comparaciones procedentes de distintas fuentes, el 97% se declararon inconclusas, un 2.5% exclusiones y un 0% falsos positivos.

<sup>319</sup> DEFRANCE Y VAN ARSDALE, 2003: 35-37. En este estudio del FBI, se dispararon 5 proyectiles de 5 cañones de barril de revólveres Smith & Wesson, de calibre 357 Magnum, fabricados consecutivamente. Cada uno de los 9 analistas recibió dos cajas para las pruebas, cada una conteniendo un proyectil de cada uno de los 5 revólveres y dos proyectiles adicionales (de diferentes revólveres en una de las cajas y del mismo revólver en la otra); se les pidió que realizaran las 42 posibles comparaciones por pares, que incluían 37 comparaciones con muestras de diferentes fuentes. Del total de 45 comparaciones de muestras procedentes de la misma fuente, hubo 42 identificaciones y 3 exámenes inconclusas. Del total de 333 comparaciones de muestras de diferentes fuentes, el artículo afirma que no hubo falsos positivos, pero no informa sobre el número de exámenes inconclusas.

<sup>320</sup> Algunas políticas de laboratorio exigen un listón muy alto para declarar exclusiones.

<sup>321</sup> STROMAN, 2014: 157-175. En este estudio, los proyectiles fueron disparados por tres revólveres Smith & Wesson. Cada uno de los 25 analistas recibió un conjunto de prueba con 3 casquillos cuestionados y 3 casquillos conocidos procedentes de cada revólver. De las 75 respuestas, hubo 74 asignaciones correctas y un examen inconcluso.

<sup>322</sup> BRUNDAGE, 1998: 438-444. En este estudio, los proyectiles fueron disparados desde 10 cañones de pistola semiautomática de 9 milímetros, de la marca Ruger P-85, consecutivamente fabricados. Cada uno de los 30 analistas recibió un conjunto de prueba de 20 proyectiles cuestionados para su comparación. Nota 323 y 324 en página siguiente

se les entregó una colección de proyectiles cuestionados y/o casquillos disparados desde un pequeño número de armas consecutivamente fabricadas de la misma marca (3, 10, 10 y 10 armas, respectivamente) y una colección de proyectiles (o casquillos) conocidos por haber sido disparados por esas mismas armas. Se les pidió que realizaran un ejercicio comparativo —asignando los proyectiles (o casquillos) de un conjunto a los proyectiles (o casquillos) del otro—.

Este diseño de «conjunto cerrado» es más simple que los problemas que se encuentran en los casos reales, porque la respuesta correcta está siempre presente en la colección. En tales estudios, los analistas pueden ejercer perfectamente su función con tal que simplemente emparejen cada proyectil con el estándar que sea *más próximo*. Por contraste, en un estudio de conjunto abierto (como en un caso real), no hay garantía de que la respuesta correcta esté presente —y, de este modo, no hay garantía de que la coincidencia más próxima sea la correcta—. De las comparaciones de conjunto cerrado cabe esperar, por tanto, que subestimen la tasa de falsos positivos.

Es importante reseñar que no es necesario que a los analistas se les informe, explícitamente, que el diseño del estudio es de conjunto cerrado. Como se observó en uno de los estudios:

A los participantes no se les dijo si los casquillos constituían un conjunto abierto o cerrado. Sin embargo, a partir del cuestionario/ hoja de respuestas, los participantes podrían haber asumido que se trataba de un conjunto cerrado y que cada casquillo cuestionado debía asociarse con una de las diez correderas<sup>325</sup>.

Además, a medida que los participantes encuentren que muchos de los casquillos cuestionados tienen fuertes similitudes con los casquillos conocidos, su suposición de que los conocidos coincidentes siempre están presentes tenderá a confirmarse.

El problema que se suscita con este diseño de estudio no es solo una posibilidad teórica: es evidente en los propios resultados. Específicamente, los estudios de conjunto cerrado tienen tasas de exámenes inconclusas y de falsos positivos sustan-

---

ración con un conjunto de 15 proyectiles estándares, conteniendo al menos un proyectil disparado con cada una de las diez pistolas. De las 300 respuestas dadas, no hubo asignaciones incorrectas y una examinación inconclusa.

<sup>323</sup> FADUL, HERNANDEZ, STOILOFF, *et al.*, 2013: 376-393. Un estudio empírico para mejorar la fundamentación científica de la identificación de armas de fuego y marcas de herramienta utilizando 10 pistolas con corredera fabricadas consecutivamente. En este estudio, los proyectiles fueron disparados desde 10 cañones de pistola semiautomática de 9 milímetros con corredera, de la marca Ruger P-85, consecutivamente fabricados. Cada uno de los 217 analistas recibió un conjunto de prueba de 15 casquillos cuestionados y 2 conocidos de cada una de las 10 pistolas. De las 3255 respuestas recibidas, hubo 3239 asignaciones correctas, 14 exámenes inconclusos y 2 falsos positivos.

<sup>324</sup> HAMBY, BRUNDAGE y THORPE, 2009: 99-110. En este estudio, se dispararon proyectiles de 10 cañones consecutivamente estriados de la marca Ruger P-85. Cada uno de los 440 analistas recibió un conjunto de prueba consistente en 15 proyectiles cuestionados y 2 conocidos estandarizados de cada una de las 10 pistolas. De las 6600 respuestas recibidas, hubo 6593 asignaciones correctas, siete exámenes inconclusos y ningún falso positivo.

<sup>325</sup> FADUL, HERNANDEZ, STOILOFF, *et al.*, 2013: 376-93.

cialmente más bajas (en más de 100 veces) que los estudios de diseño parcialmente abierto (estudio Miami-Dade) o totalmente abierto, los diseños de caja negra (Laboratorio Ames) descritos a continuación (Tabla 2)<sup>326</sup>.

En resumen, el diseño en conjunto cerrado es problemático por principio y parece subestimar la tasa de falsos positivos en la práctica<sup>327</sup>. El diseño no es apropiado para valorar la validez científica y medir la fiabilidad.

#### 5.5.4.2. Comparación del tipo conjunto a conjunto/conjunto parcialmente abierto ('Estudio Miami-Dade')

Un estudio involucró una comparación, conjunto a conjunto, en el que unas pocas de las muestras cuestionadas carecían de un estándar conocido coincidente<sup>328</sup>. A los 165 analistas en el estudio se les pidió que asignaran una colección de 15 muestras cuestionadas, disparadas desde 10 pistolas, a una colección de estándares conocidos; dos de las 15 muestras cuestionadas procedían de un arma de la que no se proporcionaron estándares conocidos. Respecto a estas dos muestras, hubo 188 eliminaciones, 138 exámenes inconclusos y 4 falsos positivos. La tasa de inconclusión fue del 41.8% y la tasa de falsos positivos entre los exámenes concluyentes fue del 2.1% (intervalo de confianza del 0.6% al 5.25%). La tasa de falsos positivos se corresponde con una tasa estimada de 1 error cada 48 casos, con un límite superior de 1 error cada 19.

Como se indicó anteriormente, los resultados del estudio Miami-Dade son marcadamente diferentes de los estudios de conjunto cerrado: (1) la proporción de resultados no concluyentes fue 200 veces mayor y (2) la tasa de falsos positivos 100 veces más alta, aproximadamente.

#### 5.5.5. Estudios recientes de análisis caja negra sobre armas de fuego

En 2011, el Comité de Investigación Forense de la *American Society of Crime Lab Directors* (Sociedad estadounidense de directores de laboratorios de criminalística) identificó, entre las necesidades de mayor rango en la ciencia forense, la importancia

<sup>326</sup> De las 10.230 respuestas dadas entre los tres estudios, hubo 10.205 asignaciones correctas, 23 exámenes inconclusos y 2 falsos positivos.

<sup>327</sup> STROMAN (2014) reconoce que, aunque las instrucciones de la prueba no indicaron explícitamente que el estudio era cerrado, su estudio podría mejorarse si «se utilizaran armas de fuego adicionales y se conociera que solo una parte de esas armas fueron usadas en los kits de prueba, presentando así un conjunto abierto de (armas) desconocidas para los participantes. Si bien esto podría aumentar la probabilidad de resultados no concluyentes, sería un reflejo más preciso de los tipos de evidencia que se reciben en los casos reales».

<sup>328</sup> FADUL, HERNANDEZ, STOILOFF, *et al.*, 2013, [www.ncjrs.gov/pdffiles1/nij/grants/244232.pdf](http://www.ncjrs.gov/pdffiles1/nij/grants/244232.pdf).

de acometer un estudio de caja negra en el análisis de armas de fuego análogo al estudio de caja negra del FBI sobre huellas dactilares latentes. El DFSC, insatisfecho con el diseño de estudios previos de análisis de armas de fuego, concluyó que era necesario un estudio de caja negra y que debía ser dirigido por un laboratorio de prueba independiente no afiliado a las fuerzas de seguridad pública que involucraría a los analistas forenses como participantes en el estudio. El DFSC y la *Defense Forensics and Biometrics Agency* (Agencia de ciencias forenses y biométricas de la defensa), conjuntamente financiaron un estudio realizado por el Laboratorio Ames, un laboratorio nacional del Departamento de energía afiliado a la Universidad estatal de Iowa<sup>329</sup>.

#### 5.5.6. Pruebas independientes/abiertas (‘Estudio del laboratorio Ames’)

El estudio empleó un diseño similar al estudio de caja negra del FBI sobre huellas dactilares latentes, con muchos analistas tomando una serie de decisiones de comparaciones *independientes* entre una muestra cuestionada y una o más muestras conocidas que pueden o no contener la fuente. Todas las muestras procedieron de 25 pistolas Ruger, de calibre 9 milímetros, compradas nuevas<sup>330</sup>. A cada uno de los 218 analistas<sup>331</sup> se les presentó 15 problemas de comparación *separados* —cada uno consistente en una muestra cuestionada y tres disparos de prueba conocidos realizados con una misma arma conocida, que podría o no haber sido la fuente—<sup>332</sup>. Sin que los analistas lo supieran, había cinco comparaciones de la misma fuente y diez de fuentes diferentes. (En un diseño ideal, la proporción de las comparaciones de la misma fuente y de distintas fuentes diferiría entre los analistas).

En las 2.178 comparaciones de diferentes fuentes hubo 1.421 eliminaciones, 735 no concluyentes y 22 falsos positivos. La tasa de no concluyentes fue del 33.7% y la de falsos positivos entre eliminaciones concluyentes fue del 1.5% (límite superior del 95% de confianza del 2.2%). La tasa de falsos positivos se corresponde con una tasa estimada de 1 error cada 22 casos, con límite superior de 1 cada 46. (Debe advertirse que 20 de los 22 falsos positivos fueron realizados por solo 5 de los 218 analistas — sugiriendo fuertemente que la tasa de falsos positivos es altamente heterogénea entre los analistas—).

---

<sup>329</sup> BALDWIN, BAJIC, MORRIS, *et al.*, 2014.

<sup>330</sup> Hubo una crítica, realizada por un científico forense, consistente en decir que el estudio no se llevó a cabo con armas *consecutivamente fabricadas*.

<sup>331</sup> Los participantes fueron miembros de AFTE que eran analistas en ejercicio, empleados o retirados, de una agencia nacional o internacional de seguridad pública, con adecuado entrenamiento.

<sup>332</sup> Los casos reales pueden ser situaciones más complejas (por ejemplo, muchos proyectiles diferentes en la escena del crimen). Sin embargo, una apropiada valoración de la validez de los fundamentos debe *comenzar* con la pregunta sobre con cuánta frecuencia puede un analista determinar si un proyectil cuestionado procede de una fuente específica conocida.

Los resultados de los distintos estudios se muestran en la Tabla 2. Las tablas muestran una diferencia llamativa entre los estudios de conjunto cerrado (donde por diseño siempre existe un estándar de coincidencia) y los estudios que no lo son (donde no hay garantía de que alguno de los estándares conocidos coincida). Específicamente, los estudios de conjunto cerrado muestran una tasa sustancialmente menor de exámenes no concluyentes y de falsos positivos. Con este diseño inusual, los analistas logran tener éxito en responder a todas las preguntas y alcanzan esencialmente puntuaciones perfectas. En los diseños abiertos, más realistas, esas tasas son mucho más altas.

**TABLA 2: RESULTADOS DE LOS ESTUDIOS DE ARMAS DE FUEGO<sup>◊</sup>**

TIPO DE ESTUDIO	RESULTADOS DE COMPARACIONES CON DIFERENTES FUENTES				
	Datos originales	No concluyentes	Falsos positivos entre exámenes concluyentes <sup>333</sup>		
	Exclusiones / Inconclusiones / Falsos positivos		Frecuencia (Límite de confianza)	Tasa estimada	Límite sobre tasa estimada
Conjunto a conjunto/cerrado (4 estudios)	10.205/23/2	0.2%	0.02% (0.06%)	1 en 5.103	1 en 1.612
Conjunto a conjunto/parcialmente abierto (estudio Miami-Dade)	188/134/4	41.8%	2% (2.7%)	1 en 49	1 en 21
Estudio de caja negra (Estudio del Laboratorio Ames)	1.421/735/22	33.7%	1.5% (2.2%)	1 en 66	1 en 46

<sup>◊</sup> «No concluyentes»: proporción del total de exámenes que se calificaron como inconclusas. «Datos originales»: Número de falsos positivos dividido por el número de exámenes concluyentes, incluyendo vestigios cuestionados sin un correspondiente conocido (para conjunto a conjunto/abierto ligeramente) o parejas no realmente emparejadas (para independiente/abierto). «Frecuencia (límite de confianza)»: Estimación puntual de frecuencia de falsos positivos, y límite superior del 95% de confianza. «Tasa estimada»: la apuesta de ocurrencia de un falso positivo, basada en la proporción observada de falsos positivos. «Límite en tasa estimada»: la apuesta de ocurrencia de un falso positivo, basada en el límite superior del intervalo de confianza del 95% —es decir, la tasa podría ser razonablemente tan alta como este valor—.

### 5.5.7. Conclusiones

Los primeros estudios indican que los analistas pueden, bajo ciertas circunstancias, asociar la munición con el arma con la que fue disparada. Sin embargo, como

<sup>333</sup> Las tasas para *todos* los exámenes son, leyendo entre filas: 1 en 5.115; 1 en 1.416; 1 en 83; 1 en 33; 1 en 99; y 1 en 66.

se describió más arriba, la mayoría de esos estudios se realizaron con diseños que no son apropiados para valorar la validez científica o estimar la fiabilidad del método en la práctica. En efecto, la comparación de los estudios sugiere que, debido a su diseño, muchos estudios citados con frecuencia subestiman seriamente la tasa de falsos positivos.

Hasta la fecha, hay solo un estudio que fue apropiadamente diseñado para evaluar la validez de los fundamentos y estimar la fiabilidad (el estudio del Laboratorio Ames). Es importante señalar que el estudio fue dirigido por un grupo independiente, no afiliado a un laboratorio de criminalística. Aunque el informe está disponible en la web, no ha sido todavía sometido a revisión por pares y a publicación.

Los criterios científicos para la validez de los fundamentos requieren estudios apropiadamente diseñados *por más de un grupo* para asegurar la reproducibilidad. Dado que solo ha habido un estudio apropiadamente diseñado, la evidencia actual no cumple los criterios científicos para la validez de los fundamentos<sup>334</sup>. Hay, de este

---

<sup>334</sup> El DOJ pidió al PCAST que revisara un reciente artículo, publicado en julio de 2016, y juzgara si constituye un estudio de caja negra adicional apropiadamente diseñado sobre el análisis de armas de fuego (es decir, la capacidad de asociar munición con un arma *en particular*). El PCAST revisó cuidadosamente el artículo, incluyendo una entrevista con los tres autores sobre el diseño del estudio. SMITH, T.P., SMITH, G.A. Y J.B. SNIPES, 2016: 939-946.

El artículo incluye un novedoso y complejo diseño que es diferente a cualquier estudio previo. Brevemente, el diseño del estudio fue el siguiente: (1) se dispararon seis tipos diferentes de munición desde 8 pistolas del calibre 40 de cuatro fabricantes (2 Taurus, 2 Sig Sauer, 2 Smith and Wesson y 2 Glock) que habían estado en uso en la población en general y obtenido por el Departamento de policía de San Francisco; (2) se crearon kits de pruebas seleccionando aleatoriamente 12 muestras (proyectiles o casquillos); (3) a 31 analistas se les dijo que la munición fue recuperada de una única escena del crimen y se les pidió que preparasen notas describiendo sus conclusiones sobre qué conjunto de muestras habían sido disparadas con la misma arma; y (4) basándose en las notas de cada analista, los autores buscaron recrear la ruta lógica de las comparaciones seguida por cada analista y calcular las estadísticas basadas en este número inferido de comparaciones realizadas por cada analista.

Aunque es interesante, el artículo no es claramente un estudio de caja negra para valorar la fiabilidad de los análisis de armas de fuego para asociar munición con un arma en particular y sus resultados no pueden compararse con estudios previos. Específicamente: (1) El estudio emplea un diseño de *comparación dentro de un conjunto* (comparaciones interdependientes dentro de un conjunto) en vez de un diseño de *caja negra* (muchas comparaciones independientes); (2) el estudio involucra a un pequeño número de analistas; (3) la pregunta central con respecto al análisis de armas de fuego es si los analistas pueden asociar munición consumida con un arma *en particular*, no simplemente con una particular *marca* de arma. Para responder a este interrogante, los estudios han de valorar el desempeño del analista con munición disparada de diferentes armas de la *misma marca* («comparaciones dentro de la misma clase») en lugar de con armas de *diferentes marcas* («comparación entre clases»); la última comparación es mucho más simple porque las armas de distintas marcas producen marcas con distintas características «de clase» (debido al diseño del arma), mientras que las armas de la misma marca deben distinguirse por características «adquiridas aleatoriamente» por cada arma (adquiridas durante el estriado o por el uso). De acuerdo con esto, los estudios previos utilizaron solo comparaciones intra-clase. Por contraste, el reciente estudio consiste en una mezcla de comparaciones intra-clase contra entre-clases, siendo una sustancial mayoría las comparaciones más simples entre-clases. Para estimar la tasa de falsos positivos para comparaciones *intra-clase* (la cantidad relevante), necesitamos saber el número de pruebas inde-

modo, necesidad de estudios de caja negra adicionales, apropiadamente diseñados, para proporcionar estimaciones de su fiabilidad.

**HALLAZGO 6:  
ANÁLISIS DE ARMAS DE FUEGO**

*Validez de los fundamentos.* El PCAST encuentra que el análisis de armas de fuego no cumple en la actualidad los criterios para la validez de los fundamentos, porque solo hay un único estudio apropiadamente diseñado para medir la validez y estimar la fiabilidad. Los criterios científicos para la validez de los fundamentos requieren más de un estudio de ese tipo para demostrar la reproducibilidad.

Pertenece a los tribunales decidir si el análisis de armas de fuego debe considerarse admisible basándose en la actual evidencia.

Si el análisis de armas de fuego se permite en los tribunales, los criterios científicos para la validez de su aplicación requieren que se informe claramente sobre las tasas de error vistas en estudios de caja negra apropiadamente diseñados (estimadas en 1 error cada 66 comparaciones, con un límite superior del 95% de confianza que se corresponde con 1 error cada 46 comparaciones, en el único estudio realizado hasta la fecha).

*Validez en la aplicación.* Si el análisis de armas de fuego es permitido en los tribunales, la validez en su aplicación requiere, desde el punto de vista científico, que el analista:

(1) haya superado una prueba de aptitud rigurosa sobre un gran número de problemas de prueba que permitan evaluar su capacidad y rendimiento, y que revele los resultados de las pruebas de aptitud; y

(2) revele si, cuando llevó a cabo la examinación, era consciente de algunos otros hechos del caso que pudieran influir en la conclusión.

### 5.5.8. El camino a seguir

Se necesitan continuos esfuerzos para mejorar el estado de los análisis de armas de fuego —y esos esfuerzos aportarán claramente dividendos para el sistema de justicia penal—.

pendientes que conllevan comparaciones intra-clase de distintas fuentes que den como resultado examinaciones concluyentes (identificación o eliminación). El artículo no distingue entre comparaciones intra-clase de entre-classes y los autores reconocieron que no hicieron esos análisis.

Los comentarios del PCAST no tratan de hacer una crítica del artículo reciente, que es un proyecto de investigación novedoso y valioso. Simplemente responden a la pregunta específica del DOJ: el reciente artículo no representa un estudio de caja negra adecuado para valorar la validez científica o estimar la precisión de los analistas para asociar munición con un arma *en particular*.

Una dirección es la de continuar mejorando el análisis de armas de fuego como método subjetivo. Con solo un estudio de caja negra hasta la fecha, hay necesidad de estudios de caja negra adicionales basados en el diseño de estudio de caja negra del Laboratorio Ames. Como se señaló anteriormente, los estudios deben ser diseñados y dirigidos en conjunto con terceras partes que no tengan interés en los resultados (tales como el Laboratorio Ames o centros de investigación como el *Center for Statistics and Applications in Forensic Evidence [CSAFE]* [Centro de estadística y aplicaciones en pruebas forenses]). Hay necesidad de pruebas de aptitud más rigurosas, utilizando problemas que sean apropiadamente difíciles y que puedan ser públicamente accesibles después de las pruebas.

Una segunda —y más importante— dirección es (al igual que con los análisis de huellas latentes) convertir los análisis de armas de fuego de un método subjetivo a un método objetivo.

Esto implicaría el desarrollo y la comprobación de funcionamiento de los algoritmos de análisis de imagen para comparar la similitud de las marcas de herramienta sobre los proyectiles. Ya ha habido pasos alentadores hacia ese objetivo<sup>335</sup>. Recientes esfuerzos para caracterizar imágenes 3D de proyectiles han utilizado métodos estadísticos y de *machine learning* para construir una «firma» cuantitativa de cada proyectil que pueda ser utilizada en comparaciones entre muestras. Un artículo reciente analiza la posibilidad de utilizar métodos topográficos de superficie en balística y se sugieren enfoques para el uso de esos métodos en los exámenes de armas de fuego<sup>336</sup>. Los autores señalan que el desarrollo de métodos ópticos ha mejorado la velocidad y la precisión en captar la topografía de superficie, conduciendo a una cuantificación mejorada de los grados de similitud.

En un estudio reciente, los investigadores utilizaron imágenes de un estudio anterior para desarrollar un sistema asistido por ordenador para comparar proyectiles que minimiza la intervención humana<sup>337</sup>. El algoritmo del grupo genera una firma cuantitativa de la imagen 3D del proyectil, compara la firma con dos o más muestras y produce una «puntuación de coincidencias» que refleja la fuerza de la coincidencia. En un conjunto de datos de prueba reducido, el algoritmo tuvo una tasa de error muy baja.

Existen esfuerzos adicionales en el sector privado que se centran en desarrollar representaciones precisas de casquillos en alta resolución para mejorar la precisión y

---

<sup>335</sup> Por ejemplo, un estudio reciente utilizó datos de microscopía confocal 3-D de munición para desarrollar una métrica de similitud para comparar imágenes. Tras realizar todas las comparaciones por pares entre un total de 90 casquillos disparados desde 10 pistolas con corredera, los autores encontraron que la distribución de la métrica para los pares de las mismas armas no se solapaba con la distribución de la métrica para los pares de distintas armas. Aunque es un estudio reducido, es alentador. WELER, ZHENG, THOMPSON, ET AL., 2012: 912-917.

<sup>336</sup> VORBURGER, SONG y PETRACO, 2016, 013002.

<sup>337</sup> HARE, HOFMANN y CARRIQUIRY, «Automatic matching of bullet lands.» Unpublished paper, available at: [arxiv.org/pdf/1601.05788v2.pdf](https://arxiv.org/pdf/1601.05788v2.pdf).

permitir funciones de similitud de mayor calidad que mejoren y asignen niveles de confianza de las coincidencias en las búsquedas en bases de datos. La actual base de datos NIBIN utiliza tecnología más antigua (no 3D) y no proporciona una función de puntuación o una asignación de confianza a cada candidato a una coincidencia. Se ha sugerido que podría utilizarse una función de puntuación para una verificación ciega por parte de analistas humanos.

Dado el enorme progreso del análisis de imagen en otros campos en la pasada década, creemos que es probable que el análisis de las armas de fuego completamente automático sea posible en un futuro próximo. Sin embargo, los esfuerzos se ven actualmente obstaculizados por la falta de acceso a bases de datos realmente grandes y complejas que puedan utilizarse para continuar con el desarrollo de estos métodos y validar las propuestas iniciales.

El NIST, en coordinación con el laboratorio del FBI, debería desempeñar un papel de liderazgo en impulsar esta transformación creando y difundiendo bases de datos apropiadamente grandes. Estos organismos también deberían proporcionar subvenciones y contratos para sustentar el trabajo y los procesos sistemáticos de evaluación de los métodos. En particular, creemos que los concursos «premiados» basados en grandes colecciones de imágenes accesibles al público<sup>338</sup> podrían atraer un notable interés del mundo académico y de la industria.

## 5.6. *Análisis de huellas de calzado: identificando características*

### 5.6.1. Metodología

El análisis de huella de calzado es un proceso que generalmente implica la comparación de un objeto conocido, tal como un zapato, con una impresión completa o parcial encontrada en la escena del crimen, para valorar la probabilidad de que el objeto sea la fuente de la impresión. El proceso se desarrolla de forma escalonada, comenzando con una comparación de «características de clase» (tales como diseño, tamaño físico y desgaste, en general) y pasando después a «identificar características» o «características aleatoriamente adquiridas (RACs, por las siglas en inglés de «*randomly acquired characteristics*») (tales como las marcas de un zapato causadas por cortes, muescas y roturas en el transcurso de su uso)<sup>339</sup>.

<sup>338</sup> El 7 de julio de 2016, el NIST publicó la base de datos de investigación de código abierto *NIST Ballistics Toolmark Research Database (NBTRD)* como una base de datos de investigación de acceso libre de datos de marcas de herramientas de proyectiles y casquillos ([tsapps.nist.gov/NRBTD](http://tsapps.nist.gov/NRBTD)). La base de datos contiene imágenes de microscopía de reflectancia y datos topográficos de superficie tridimensionales adquiridos por el NIST o aportados por los usuarios.

<sup>339</sup> Véanse: Rango de conclusiones estándares para las pruebas de examen de calzado e impresiones de neumáticos de SWGTREAD, 2013. Guía para las pruebas de examen de calzado e impresiones de

En este informe, no afrontamos la cuestión de si los analistas pueden fiablemente determinar las características de clase —por ejemplo, si una impresión de calzado concreta fue dejada por un calzado talla 12 de una marca en particular—. Aunque es importante que haya estudios que estimen la fiabilidad de los análisis de calzado dirigidos a determinar las características de clase, el PCAST no eligió poner su atención en este aspecto del examen del calzado porque la determinación de las características de clase la estimación de la frecuencia con que una característica de clase particular se presenta en el calzado o (para miembros del jurado) la comprensión de la naturaleza de las características en cuestión no son *inherentemente* un problema que suponga un reto de medición.

En su lugar, el PCAST centró su atención en la fiabilidad de las conclusiones, basadas en *RACs*, sobre que una impresión provenga probablemente de un calzado específico. Esto es un problema mucho más difícil, porque requiere conocer con qué precisión los analistas identifican las características específicas compartidas entre un calzado y una impresión, con qué frecuencia fallan al identificar las características que los distinguirían y qué valor probatorio puede otorgarse a un *RAC* particular.

A pesar de la ausencia de estudios empíricos que midan la precisión de los analistas, las autoridades en la disciplina del examen de huella de calzado expresan su confianza en que pueden identificar la fuente de una impresión basándose en un único *RAC*.

Como se describió en un artículo de 2009 escrito por un analista del FBI y publicado en la revista *Forensic Science Communications* del FBI:

Un analista primero determina si existe una correspondencia de características de clase entre la impresión del calzado cuestionada y el calzado conocido. Si el analista considera que no existen inconsistencias entre las características de clase, el examen progresa hacia cualquier característica identificativa en la impresión cuestionada. El analista compara estas características con cualquier característica identificativa observada en el calzado conocido. Aunque sean impredecibles en cuanto a su ocurrencia, el tamaño, la forma y la posición de esas características tienen una baja probabilidad de ocurrencia de la misma manera en diferentes calzados. De este modo, combinadas con las características de clase, incluso una sola característica identificativa constituye una prueba extremadamente poderosa en apoyo de la conclusión de la identificación<sup>340</sup>.

Como apoyo a esta tesis, el artículo cita un manual líder en identificación de huellas de calzado:

De acuerdo con William J. Bodziak (2000), «pueden hacerse identificaciones positivas con tan solo una característica identificativa aleatoria, pero solo si esa característica es confirmable, tiene suficiente definición, claridad y propiedades, está en la misma posición y orientación sobre la suela del calzado y, en opinión del analista experimentado, no se haría de nuevo por otra suela de calzado»<sup>341</sup>.

---

neumáticos de SWGTREAD, 2006 y BODZIAK, 2000: p 347.

<sup>340</sup> SMITH, M.B., *The Forensic Analysis of Footwear Impression Evidence*. [www.fbi.gov/about-us/lab/forensic-sciencecommunications/fsc/july2009/review/2009\\_07\\_review02.htm](http://www.fbi.gov/about-us/lab/forensic-sciencecommunications/fsc/july2009/review/2009_07_review02.htm).

<sup>341</sup> BODZIAK, 2000.

El artículo se refiere al modelo matemático de Stone que afirma que la probabilidad de que dos calzados compartan una misma característica identificativa es de 1 en 16.000 y la probabilidad de que compartan tres características es de 1 en 683 billones (anglosajones)<sup>342</sup>.

Tales afirmaciones sobre «identificación» basadas en el análisis de calzado son impresionantes, pero carecen de fundamento científico.

La afirmación de Bodziak tiene dos componentes: (1) que el analista observe, consistentemente, un *RAC* demostrable en un conjunto de impresiones y (2) que el analista esté convencido de que el *RAC* no ocurriría en otro calzado. La primera parte es razonable, pero la segunda es profundamente problemática: requiere que el analista confíe en recuerdos y conjeturas sobre la frecuencia de las características.

El modelo de Stone es completamente teórico: realiza muchas suposiciones no respaldadas (sobre la frecuencia y la independencia estadística de las marcas) que no pone a prueba de ninguna manera.

Todo el proceso —desde la elección de características que deben considerarse (o ignorarse) y la determinación de la rareza— se basa enteramente en el juicio subjetivo de un analista. En tales circunstancias, es esencial que la validez del método y las estimaciones de su fiabilidad se establezcan mediante múltiples y apropiados estudios de caja negra<sup>343</sup>.

### 5.6.2. Antecedentes

El informe del NRC (2009) mencionado citó algunos artículos que siembran dudas sobre si los analistas de huellas de calzado alcanzan conclusiones consistentes cuando se presentan con la misma prueba. Por ejemplo, el informe contiene una detallada explicación sobre un trabajo europeo de 1996 que presentó a los analistas frente a seis casos simulados —dos con calzado desgastado utilizado en la escena del crimen, cuatro con calzado nuevo cuyas características identificativas habían sido añadidas deliberadamente; el artículo informó de una considerable variabilidad en las respuestas<sup>344</sup>. El PCAST también destaca un estudio israelí que trata de dos casos en la escena del crimen y que llega a similares conclusiones<sup>345</sup>.

Como respuesta al informe del NRC (2009), un artículo de 2013 afirmó demostrar que analistas de huellas de calzado estadounidenses y canadienses presentaban una mayor consistencia que la observada en el estudio europeo de 1996<sup>346</sup>. Sin embargo, este estudio difirió sustancialmente del primero porque los analistas no reali-

---

<sup>342</sup> STONE, 2006: 577-99.

<sup>343</sup> Además de los estudios de caja negra, también son valiosos los estudios de caja blanca para identificar las fuentes de los errores.

<sup>344</sup> MAJAMMA e YTTI, 1996: 109-20.

<sup>345</sup> SHOR y WEISNER, 999: 380-4384.

<sup>346</sup> HAMMER, DUFFY, FRASER, ET AL., 2013: 205-18.

zaron los exámenes por sí mismos. Por ejemplo, las fotografías fueron previamente anotadas para señalar todas las características relevantes para la comparación —es decir, a los analistas no se les pidió que identificaran las características—<sup>347</sup>. De este modo, el estudio, en virtud de su diseño, no puede abordar la consistencia en el proceso de examinación.

Además, el tema fundamental no es la *consistencia* (si los analistas dieron la *misma* respuesta) sino la *precisión* (si dieron la respuesta *correcta*). La precisión puede ser evaluada solamente con grandes estudios de caja negra apropiadamente diseñados.

### 5.6.3. Estudios de validez científica y fiabilidad

El PCAST no pudo encontrar estudios de caja negra apropiadamente diseñados para establecer la validez de los fundamentos de la identificación basada en análisis de huellas de calzado.

#### **HALLAZGO 7: ANÁLISIS DE HUELLAS DE CALZADO**

*Validez de los fundamentos.* El PCAST encuentra que no existen estudios empíricos apropiados que sostengan la validez de los fundamentos de los análisis de huellas de calzado para asociar impresiones de esas huellas con un calzado en particular con base en marcas identificativas específicas (en ocasiones denominadas “características adquiridas aleatoriamente”). Tales conclusiones no están respaldadas por ninguna evidencia significativa o estimaciones de su precisión y, por ello, no son científicamente válidas.

El PCAST no ha evaluado la validez de los fundamentos de los análisis de las huellas de calzado para identificar características de clase (por ejemplo, la talla del calzado o la marca).

Mostrando acuerdo con nuestra conclusión, el Subcomité de huellas de calzado y de neumáticos de la OSAC, recientemente identificó la necesidad de realizar estudios de caja negra y de caja blanca sobre la fiabilidad de los analistas —citándola como una «brecha importante en el conocimiento actual» en la que «no se están llevando a cabo una investigación en la actualidad o se hace limitadamente—»<sup>348</sup>.

<sup>347</sup> El artículo afirma que «Todas las características y observaciones que fueron consideradas por los analistas durante la comparación fueron claramente identificadas y etiquetadas en cada impresión».

<sup>348</sup> Véase: [www.nist.gov/forensics/osac/upload/SAC-Phy-Footwear-Tire-Sub-R-D-001-Examiner-ReliabilityStudy\\_Revision\\_Feb\\_2016.pdf](http://www.nist.gov/forensics/osac/upload/SAC-Phy-Footwear-Tire-Sub-R-D-001-Examiner-ReliabilityStudy_Revision_Feb_2016.pdf) (último acceso 12 de mayo de 2016).

#### 5.6.4. El camino a seguir

En contraste con los análisis de huellas latentes y de armas de fuego, existe escasa investigación sobre la que fundamentar las conclusiones que buscan asociar una impresión de calzado a un calzado en particular (conclusiones de identificación).

Se necesitan nuevos enfoques para desarrollar paradigmas. Como paso inicial, el laboratorio del FBI está involucrado en un estudio en el que se examinan un conjunto de 700 botas similares que fueron utilizadas por los cadetes de agentes especiales del FBI durante las 16 semanas de su programa de formación. El estudio tiene como objetivo evaluar si los *RACs* se observan en calzados de diferentes individuos. Si bien tales estudios sobre «unicidad» (es decir, demostraciones de que muchos objetos tienen características distintivas) no pueden establecer la validez de los fundamentos (véase p. 51), las impresiones generadas a partir del calzado podrían proporcionar un conjunto de datos iniciales para (1) un estudio de caja negra piloto y (2) una base de datos piloto de frecuencias de características. Es importante destacar que el NIST está comenzando un estudio para determinar si es posible cuantificar el proceso de examinación de las huellas de calzado o, como mínimo, algún aspecto del proceso, en un esfuerzo por incrementar la objetividad de los análisis de huellas de calzado.

Separadamente, deben llevarse a cabo evaluaciones sobre la precisión y fiabilidad de la determinación de características de clase, un tema que no se ha abordado en este informe.

#### 5.7. *Análisis de cabellos*

El análisis de cabellos es un proceso mediante el cual los analistas comparan propiedades microscópicas del cabello para determinar si una persona en particular pudiera ser la fuente de un cabello determinado. Mientras el PCAST estaba completando este informe, el DOJ publicó, para que pudieran hacerse comentarios, directrices sobre el testimonio en exámenes de cabello que incluían documentos de respaldo que abordan la validez y la fiabilidad de la disciplina<sup>349</sup>. Aunque el PCAST no ha llevado a cabo una revisión integral de la disciplina, realizó una revisión del documento de apoyo con el fin de arrojar más luz sobre los estándares para conducir una evaluación científica de una disciplina forense de comparación de características.

El documento de apoyo afirma que «la comparación de cabellos microscópica ha demostrado ser una metodología científica válida y fiable», aunque resalta que «la

---

<sup>349</sup> Véanse: Propuesta del Departamento de Justicia para uniformizar el lenguaje en el testimonio y en los informes de la disciplina del examen forense de cabellos, disponible en: [www.justice.gov/dag/file/877736/download](http://www.justice.gov/dag/file/877736/download) y documentación en apoyo de la propuesta del Departamento de Justicia para uniformizar el lenguaje en el testimonio y en los informes de la disciplina del examen forense de cabellos, disponible en: [www.justice.gov/dag/file/877741/download](http://www.justice.gov/dag/file/877741/download).

comparación de cabellos microscópica por sí sola no puede conducir a una identificación personal y es crucial que esta limitación se transmita, tanto en el informe escrito como en el testimonio en juicio oral».

### 5.7.1. Estudios de los fundamentos del examen microscópico de cabellos

En apoyo a su conclusión de que el examen de cabellos es válido y fiable, el documento de apoyo del DOJ diserta sobre cinco estudios de comparación de cabello humano. El principal apoyo es una serie de tres estudios realizados por Gaudette en 1974, 1976 y 1978<sup>350</sup>. En 1974 y 1976, los estudios se centraron, respectivamente, sobre cabello de la cabeza y vello púbico. Dado que el diseño y los resultados son similares, nos centramos en el estudio del cabello de la cabeza.

El documento de apoyo del DOJ afirma que «del total de 370.230 intercomparaciones realizadas en los estudios del cabello de la cabeza, solo nueve pares de cabellos no pudieron distinguirse» —correspondiéndose con una tasa de falsos positivos de 1 en 40.000—. Más específicamente, el diseño de este estudio de 1974 fue como sigue: un único analista (1) puntuó entre 6 y 11 cabellos de cabeza de cada uno de 100 individuos (con un total de 861 cabellos) con respecto de 23 categorías distintas (con un total de 96 valores posibles); (2) comparó los cabellos de *diferentes* individuos para identificar pares de cabellos con menos de 4 diferencias; y (3) comparó estos pares de cabellos microscópicamente para ver si podían distinguirse.

El documento de apoyo del DOJ no aclara que estos estudios fueron fuertemente criticados por otros científicos por metodología defectuosa<sup>351</sup>. La crítica más seria fue que Gaudette solo comparó cabellos de *diferentes* individuos, y no se fijó en los cabellos de un *mismo* individuo. Como señalan en un artículo de 1990 dos autores de la *Hair and Fibre Unit of the Royal Canadian Mounted Police Forensic Laboratory* (Unidad de cabello y fibras del laboratorio forense de la policía montada del Canadá) (así como en otros artículos), la aparentemente baja tasa de falsos positivos podría ser el resultado de un sesgo del analista, es decir, el analista explícitamente conocía que todos los cabellos examinados procedían de *diferentes* individuos y, por tanto, pudo estar inclinado, consciente o inconscientemente, a buscar diferencias<sup>352</sup>. En resumen, no se puede valorar apropiadamente la tasa de falsos positivos de un método sin, simultáneamente, valorar su tasa de *verdaderos* positivos (sensibilidad). En el artículo de 1990, los autores utilizaron un diseño de estudio similar, pero se empleó a *dos* analistas que examinaron *todos* los pares de cabellos. Encontraron no repetibilidad en

<sup>350</sup> GAUDETTE y KEEPING, 1974: 599-606; GAUDETTE, 1976: 514-517; GAUDETTE, 1978: 758-763.

<sup>351</sup> WICKENHEISER y HEPWORTH, 1990: 1323-29. Véanse también BARNETT y OGLE, 1982: 272-278; GAUDETTE, 1982: 279-289.

<sup>352</sup> Además, la inconsistencia en las características de puntuación añadiría ruido aleatorio a cualquier estructura de los datos (por ejemplo, correlaciones entre las características) y, por tanto, disminuiría la frecuencia de las coincidencias producidas por casualidad.

cada uno de los analistas («cada analista tuvo una considerable variación día a día en la clasificación de las características de los cabellos») y no reproducibilidad entre los analistas («en muchos casos, los analistas clasificaron los mismos cabellos de forma diferente»). En particular, observaron que, si bien los analistas no encontraron coincidencias entre los cabellos de diferentes individuos, tampoco encontraron casi ninguna coincidencia entre los cabellos de una *misma* persona. De 15 parejas de cabellos de una misma fuente que los autores determinaron que *deberían* haber sido declaradas coincidentes, solo dos fueron correctamente clasificados por ambos analistas.

En el estudio de Gaudette de 1978, el autor dio un cabello diferente a cada uno de tres aspirantes a analista en formación, que habían completado un año de entrenamiento, y se les pidió que identificaran cualquier muestra coincidente en un conjunto de referencia de 100 cabellos (que, sin que los aspirantes a analista en formación lo supieran, procedían de 100 personas diferentes, incluida las fuentes de los cabellos). Los tres aspirantes a analista en formación informaron de 1, 1 y 4 coincidencias, lo cual supone 3 respuestas correctas y 3 incorrectas. De las coincidencias declaradas, el 50% eran falsas asociaciones positivas. Entre las 300 comparaciones en total, la tasa general de falsos positivos fue del 1%, que es 400 veces más alta que la tasa estimada en el estudio de 1974.

Curiosamente, observamos que el documento de apoyo del DOJ informa erróneamente de los resultados del estudio, porque dice que el tercer analista en formación cometió solo un error en lugar de 3 errores. La explicación de esta discrepancia se encuentra en un pasaje notablemente franco del texto, que ilustra sobre la necesidad de emplear protocolos rigurosos en la evaluación de los resultados de los experimentos:

Dos aspirantes a analista en formación identificaron correctamente un cabello y solo uno como similar al estándar. Al principio, el tercer aspirante a analista en formación concluyó que había cuatro cabellos similares al estándar. Tras un examen más detenido y una consulta con los otros aspirantes a analista, fue capaz de identificar fácilmente una de sus elecciones como incorrecta. Sin embargo, aún estaba convencido de que había tres cabellos similares al estándar, el correcto y otros dos. El examen del autor trajo la opinión de que uno de esos dos podía ser eliminado, pero que el restante era indistinguible de los cabellos del estándar. Otro analista experimentado estudió luego los cabellos y también llegó a la conclusión de que uno de los otros dos podía ser eliminado. Esta vez, sin embargo, ¡fue el contrario del escogido por el autor!<sup>353</sup>

La reclasificación *post-facto* de los errores generalmente no es aconsejable en estudios relativos a la validez y la fiabilidad.

Los otros dos estudios de cabellos humanos que se tratan en el documento de apoyo del DOJ son también problemáticos. Un artículo de 1983 se refirió a muestras de cabello de 100 individuos, clasificados en tres grupos raciales<sup>354</sup>. Después de que la autora hubiera estudiado exhaustivamente los cabellos, pidió a una tercera parte

<sup>353</sup> GAUDETTE, 1978: 758–763.

<sup>354</sup> STRAUSS, 1983: 15–29.

neutral que configurase siete problemas difíciles y «ciegos» para ella, seleccionando 10 cabellos cuestionados y 10 conocidos (entre-grupos en 3 casos e intra-grupo en 4 casos)<sup>355</sup>. Los resultados consisten en una sola frase en la que la autora simplemente afirma que actuó con una «precisión del 100%». El rendimiento autoinformado en una prueba generalmente no se considera como una metodología científica apropiada.

Un artículo de 1984 estudió cabellos de 17 pares de gemelos (9 mellizos, 6 gemelos univitelinos y 2 de cigosidad desconocida) y un conjunto de trillizos idénticos<sup>356</sup>. Curiosamente, los cabellos de gemelos idénticos no mostraban mayor similitud que los de gemelos fraternales. En la única prueba diseñada para simular casos reales, a dos analistas se les proporcionó siete problemas difíciles, cada uno consistente en comparar un cabello cuestionado con entre cinco y diez cabellos conocidos. La tasa de falsos positivos fue de 1 cada 12, que es aproximadamente 3.300 veces más alta que la del estudio de cabellos de individuos no emparentados de Gaudette en 1974<sup>357</sup>.

El PCAST, tomando en cuenta su metodología y sus resultados, encuentra que los artículos descritos en el documento de apoyo del DOJ no proporcionan una base científica para concluir que el examen microscópico del cabello sea un proceso válido y fiable.

Después de describir los artículos científicos, el documento del DOJ pasa a discutir las conclusiones que pueden aducirse sobre la comparación de cabellos:

Estos estudios han demostrado también que la comparación microscópica de los cabellos por sí sola no puede conducir a una identificación personal y que es crucial que esta limitación sea transmitida tanto en el informe escrito como en el testimonio.

La ciencia de la comparación de cabellos por microscopía reconoce que las características microscópicas exhibidas en un cabello cuestionado pueden ser abarcadas por la gama de características exhibidas por muestras de cabello conocidas de más de una persona. Si un cabello cuestionado se asocia con una muestra de cabello conocido que realmente no es la fuente, eso no significa que la asociación microscópica del cabello sea errónea. Más bien, resalta la limitación de la ciencia en el sentido de que existe un grupo desconocido de personas que podrían haber contribuido como fuentes del cabello cuestionado. Sin embargo, los estudios no han determinado el número de personas que comparten cabellos con las mismas o similares características.

Este pasaje viola principios científicos fundamentales de dos maneras importantes. El primer problema es que utiliza el hecho de que la precisión del método no sea *perfecta* para descartar la necesidad de conocer la precisión del método *por completo*. De acuerdo con el documento de respaldo, no es un «error» sino simplemente una «limitación de la ciencia» que un analista asocie un cabello con un individuo que

---

<sup>355</sup> El documento de apoyo del DOJ, erróneamente informa que la prueba de microscopía comparativa consistió en comparar 100 cabellos cuestionados con 100 cabellos conocidos.

<sup>356</sup> BISBING y WOLNER, 1984: 780-786.

<sup>357</sup> El documento de apoyo del DOJ describe los resultados en términos positivos: «En las siete pruebas, un analista excluyó, correctamente, 47 de 52 muestras, y un segundo analista excluyó correctamente 49 de 52 muestras». Eso no especifica si los resultados restantes son inconclusos o falsos positivos.

no era realmente la fuente del cabello. Esto es falso. Cuando un perito dice a los miembros de un jurado que el cabello encontrado en la escena del crimen es microscópicamente indistinguible del cabello del acusado, el analista y la acusación pretenden que la declaración tenga peso. Sin embargo, el documento continúa diciendo que no existe información disponible sobre la proporción de individuos que tienen similares características. Como en la sección 4 se deja claro, esto es científicamente inaceptable. Sin estimaciones apropiadas de precisión, la afirmación de un analista de que dos muestras son similares —o, incluso, indistinguibles— carece de significado científico: no tiene valor probatorio y sí posee un considerable potencial impacto perjudicial. En resumen, si el análisis científico del cabello *significa* algo, debe haber *evidencia empírica* real sobre su significado.

El segundo problema del pasaje es su implicación de que no existen pruebas empíricas relevantes sobre la precisión del análisis del cabello. De hecho, sin embargo, tales pruebas fueron realizadas por el laboratorio del FBI. Tratamos sobre este punto seguidamente.

#### 5.7.2. Estudio del FBI comparando exámenes microscópicos de cabello y análisis de ADN

Un aspecto particularmente preocupante del documento de apoyo del DOJ es cómo trata el estudio del FBI sobre el examen de cabello que se aborda en la sección 2. En ese estudio del año 2002, miembros del FBI utilizaron análisis de ADN mitocondrial para reexaminar 170 muestras de casos anteriores en los que el laboratorio del FBI había realizado un examen microscópico del cabello. Los autores encontraron que en 9 de 80 casos (11%) en los que el Laboratorio del FBI había hallado que los cabellos eran microscópicamente indistinguibles, el análisis de ADN mostraba que los cabellos procedían en realidad de individuos *diferentes*.

El estudio del FBI del año 2002 es un hito en la historia de la ciencia forense porque fue el primer estudio que analizó sistemática y completamente una gran colección de casos anteriores para medir la frecuencia de las asociaciones de falsos positivos. Su conclusión es de enorme importancia para la ciencia forense, para la policía, para los tribunales y para los miembros del jurado: *Cuando los analistas del cabello concluyen que dos muestras de cabello son microscópicamente indistinguibles, los cabellos proceden a menudo de distintas fuentes (1 de cada 9 veces)*.

Sorprendentemente, el documento del DOJ ignora completamente este hallazgo clave. En su lugar, hace referencia al estudio del FBI para apoyar la proposición de que el análisis de ADN «puede realizarse conjuntamente con el análisis microscópico del cabello», citando «un estudio del año 2002 que indicó que, de 80 asociaciones microscópicas, aproximadamente el 88% también fue incluido mediante prueba adicional de ADN mitocondrial». El documento no reconoce que el resto de asociaciones fueron asociaciones falsas, es decir, resultados que, si se presentaran como

pruebas en un juicio contra un acusado, confundirían a los miembros del jurado sobre los orígenes de los cabellos<sup>358</sup>.

### 5.7.3. Conclusión

Nuestra breve revisión pretende simplemente ilustrar sobre los potenciales escollos en las evaluaciones de la validez de los fundamentos y de la fiabilidad de un método. El PCAST es consciente de las limitaciones a las que se enfrenta el DOJ para llevar a cabo evaluaciones científicas sobre la validez y fiabilidad de los métodos forenses, ya que las evaluaciones críticas del Departamento de Justicia podrían interpretarse como admisiones que podrían utilizarse para impugnar condenas pasadas o procesamientos actuales.

Estas cuestiones ponen de relieve por qué es importante que las evaluaciones sobre la validez científica y la fiabilidad sean realizadas por una agencia científica que no participe, por sí misma, en la aplicación de la ciencia forense dentro del sistema jurídico (véase Sección 6.1).

También subrayan por qué es importante que la información cuantitativa sobre la fiabilidad de los métodos (por ejemplo, la frecuencia de falsas asociaciones en los análisis de cabellos) se indique claramente en el testimonio de los analistas. Volveremos a esta cuestión en la sección 8, en el que consideraremos las directrices propuestas por el Departamento de Justicia que impedirían a los analistas dar información sobre el peso estadístico o la probabilidad de una conclusión sobre si un cabello cuestionado procede de una fuente en particular.

### 5.8. Aplicación a métodos adicionales

Aunque hemos emprendido detalladas evaluaciones de sólo seis métodos específicos e incluido una discusión sobre el séptimo método, el análisis básico puede aplicarse para valorar la validez de los fundamentos de cualquier método forense de comparación de características —incluyendo disciplinas forenses tradicionales (tales como el examen de documentos) así como métodos pendientes de desarrollo (tales como el análisis microbiano o de patrones de navegación por internet)—.

Queremos resaltar que la evaluación de la validez científica se fundamenta en la evidencia científica disponible en un momento dado. Algunos métodos cuya validez en sus fundamentos no se haya demostrado, pueden serlo en última instancia, aun-

---

<sup>358</sup> En una nota al pie de página, el documento también se esforzó en resaltar que el artículo no podía considerarse como el que proporcionaba una estimación de la *tasa de falsos positivos* en comparación microscópica de cabellos, porque no contenía datos sobre el número de comparaciones de fuentes diferentes que los analistas excluyeron correctamente. Aunque esta afirmación es correcta, es engañosa —porque el artículo proporciona una estimación de una cantidad mucho más importante— concretamente la frecuencia de falsas asociaciones acaecidas en casos reales.

que es posible que requieran modificaciones significativas para lograr ese objetivo. Otros métodos quizá no sean recuperables, como sucedió con el caso del análisis composicional del plomo de los proyectiles o es probable que tampoco lo sean las marcas de mordedura. Sin embargo, otros métodos pueden ser subsumidos por métodos diferentes, pero más fiables, al igual que ocurre con el análisis de ADN que ha reemplazado a otros métodos en muchos casos.

### 5.9. Conclusión

Como se deja claro en la sección anterior, históricamente muchos métodos forenses de comparación de características se han *asumido* como válidos en sus fundamentos, en lugar de haber sido *establecidos* como tales basándose en evidencia empírica apropiada. Solo dentro de la pasada década, la comunidad de ciencias forenses empezó a reconocer la necesidad de probar empíricamente si sus métodos específicos cumplían los criterios de validez científica. Solo en los últimos cinco años, por ejemplo, se han realizado estudios apropiados para establecer la validez de los fundamentos y medir la fiabilidad de los análisis de huellas dactilares latentes. Para la mayoría de los métodos subjetivos no hay estudios apropiados de caja negra, con el resultado de que no hay evidencia apropiada de la validez de sus fundamentos o estimaciones de su fiabilidad.

El análisis científico y los hallazgos de las Secciones 4 y 5 van dirigidos a ayudar a los actores relevantes sobre cómo garantizar la validez científica, tanto para las tecnologías existentes como para las pendientes de desarrollo.

El PCAST espera que algunos métodos forenses de comparación de características puedan ser rechazados por los tribunales como inadmisibles porque carecen de pruebas adecuadas de validez científica. Resaltamos que las decisiones de excluir métodos no fiables históricamente ha ayudado a impulsar mejoras importantes en la ciencia forense —como sucedió en los primeros días de la prueba de ADN— con el resultado de que algunos métodos llegan a establecerse como científicamente válidos (posiblemente tras una revisión) y otros son descartados.

En las Secciones restantes, ofrecemos recomendaciones sobre acciones específicas que podrían llevarse a cabo por el Gobierno Federal —incluidas las agencias científicas (NIST y OSTP), el Laboratorio del FBI, el Fiscal General y el Poder Judicial Federal— para garantizar la validez científica y la fiabilidad de los métodos forenses de comparación de características y promover su uso más riguroso en las salas de los tribunales.

## 6. Acciones para asegurar la validez científica en la ciencia forense: recomendaciones al NIST y a la OSTP

Basados en los hallazgos científicos de las Secciones 4 y 5, el PCAST ha identificado acciones que creemos deberían ser tomadas por las agencias científicas federales,

específicamente el NIST y la OSTP, para asegurar la validez científica de los métodos de comparación de características.

### 6.1. *El papel del NIST en la evaluación continuada de la validez de los fundamentos*

Existe la necesidad urgente de una evaluación continua de la validez de los fundamentos de métodos importantes, para proporcionar orientación a los tribunales, al DOJ y a la comunidad forense. Deben acometerse evaluaciones tanto de metodologías existentes que aún no hayan cumplido los estándares científicos para la validez de sus fundamentos como de nuevas metodologías que estén siendo desarrolladas o lo estarán en años venideros. Para asegurar que las evaluaciones científicas no estén sesgadas y sean independientes, tales evaluaciones deben estar claramente dirigidas por una agencia científica sin interés por los resultados<sup>359</sup>.

Esta responsabilidad debe recaer en el NIST. El NIST es el laboratorio metrológico líder mundial, con una larga y distinguida historia en la ciencia y tecnología de la medición. Tiene enorme experiencia en el diseño y puesta en práctica de estudios de validación, así como en valorar la validez de los fundamentos y la fiabilidad de técnicas y prácticas de laboratorio. La misión del NIST de promover la ciencia, la tecnología y los estándares de la medición se ha ampliado, a partir de los estándares tradicionales de la medición física, para dar respuesta a otras muchas necesidades sociales importantes, incluidas las de las ciencias forenses, en la que el NIST tiene programas de calado<sup>360</sup>. Como se dijo anteriormente, el NIST ha comenzado a liderar una serie de esfuerzos importantes para fortalecer las ciencias forenses, incluyendo sus funciones con el NCFS y OSAC.

El PCAST recomienda que al NIST se le asigne la responsabilidad de preparar un informe anual que evalúe la validez de los fundamentos de los métodos forenses de comparación de características clave, basados en estudios empíricos publicados y disponibles. Estas evaluaciones deben llevarse a cabo bajo los auspicios del NIST, con aportaciones de analistas adicionales, ajenos a la ciencia forense, que se consideren necesarios, y supervisados por un panel de revisión apropiado. Los informes deben seguir, como mínimo, las líneas sobre los métodos expuestas en este informe, actualizándolas cuando sea apropiado. Nuestra intención no es que el NIST tenga un papel de regulador formal con respecto a la ciencia forense, sino más bien que las evaluaciones del NIST ayuden para informar a los tribunales, el DOJ y a la comunidad de ciencia forense.

---

<sup>359</sup> Por ejemplo, las agencias que apliquen métodos de comparación de características dentro del sistema jurídico tienen un claro interés en los resultados de tales evaluaciones.

<sup>360</sup> Véase: [www.nist.gov/forensics](http://www.nist.gov/forensics).

No esperamos que el NIST asuma la responsabilidad de *realizar* los estudios de validación necesarios. Sin embargo, el NIST debería asesorar sobre el diseño y la ejecución de dichos estudios. El NIST podría realizar algunos estudios mediante su propio programa de investigación interno y a través del CSAFE. Sin embargo, la mayoría de los estudios probablemente se llevarán a cabo por otros grupos —como los Centros de investigación de cooperación entre la industria y la Universidad previstos por la NSF, el laboratorio del FBI, los laboratorios nacionales de los Estados Unidos, otras agencias federales, los laboratorios estatales y los académicos—.

Resaltamos que el NCFs ha respaldado recientemente la necesidad de una revisión científica independiente de los métodos de la ciencia forense. Un documento aprobado abrumadoramente por la comisión en junio de 2016, afirma que «todas las metodologías de la ciencia forense deben ser evaluadas por un organismo científico independiente para caracterizar sus capacidades y limitaciones, con el fin de responder de una manera precisa y fiable a una clara y bien definida cuestión forense» y que el «Instituto Nacional de Estándares y Tecnología (NIST), debe asumir el papel de evaluador científico independiente dentro del sistema de justicia para este propósito»<sup>361</sup>.

Por último, creemos que el estado de la ciencia forense mejoraría si se publicasen artículos sobre la validez de los fundamentos de los métodos de comparación de características forenses en revistas científicas líderes en lugar de en revistas de ciencias forenses, donde, debido a las debilidades en la cultura de la investigación de la comunidad de ciencia forense abordadas en este informe, los estándares de revisión por pares son menos rigurosos. Es encomiable que los científicos del FBI publicaran su estudio de caja negra de huellas dactilares latentes en *Proceedings of the National Academy of Sciences*. Sugerimos que el NIST explore, con una o más revistas científicas líderes, la posibilidad de crear un proceso de revisión rigurosa y publicación online de estudios importantes sobre la validez de los fundamentos en las ciencias forenses. Las revistas apropiadas podrían incluir la revista *Metrología*, una revista internacional líder en metrología pura y aplicada, y *Proceedings of the National Academy of Sciences*.

## 6.2. *Acelerar el desarrollo de los métodos objetivos*

Como se describe a lo largo del informe, los métodos objetivos son generalmente preferibles a los métodos subjetivos. Las razones incluidas en esa preferencia son una mayor precisión, mayor eficiencia, menor riesgo de error humano, menor riesgo de sesgos cognitivos y mayor facilidad para establecer la validez de los fundamentos y

---

<sup>361</sup> Opiniones de la Comisión: Technical Merit Evaluation of Forensic Science Methods and Practices (Evaluación de la Calidad Técnica de los Métodos y Prácticas de la Ciencia Forense). [www.justice.gov/ncfs/file/881796/download](http://www.justice.gov/ncfs/file/881796/download).

estimar la fiabilidad. Donde sea posible, deberían realizarse vigorosos esfuerzos para transformar los métodos subjetivos en métodos objetivos.

Hay dos métodos de comparación de características forenses —los análisis de huellas dactilares latentes y los de armas de fuego— que están listos para esa transformación. Como se trató en el anterior Sección, existen fuertes razones para creer que tales métodos pueden convertirse en objetivos a través del análisis de imagen automatizado. Además, el análisis de mezclas complejas de ADN se ha convertido recientemente en un método objetivo válido en sus fundamentos para un rango limitado de mezclas, aunque será necesario trabajo adicional para expandir los límites del rango.

El NIST, junto al Laboratorio del FBI, debe desempeñar un papel de liderazgo en impulsar esa transformación, mediante (1) la creación y difusión de grandes bases de datos que sustenten el desarrollo y la prueba de los métodos tanto por parte de las empresas como por académicos, (2) subvenciones y ayudas a la contratación, y (3) procesos de patrocinio, como concursos mediando algún premio, para la evaluación de los métodos.

### 6.3. *Mejorar la organización de los Comités de Áreas Científicas*

La creación por el NIST del OSAC fue un paso importante para el fortalecimiento de la práctica de la ciencia forense. El diseño organizativo —que alberga a todas las comunidades de un área temática bajo una única estructura y alienta la comunicación y coordinación interdisciplinar— es una mejora significativa respecto a los anteriores *Scientific Working Groups (SWGs)* (Grupos de trabajo científicos), que funcionaban menos formalmente como comités independientes.

Sin embargo, las lecciones iniciales tras los primeros años de su funcionamiento han revelado algunas deficiencias importantes. Entre los miembros de OSAC hay relativamente pocos científicos independientes: está dominado por profesionales forenses que suman más de dos tercios de sus miembros. Del mismo modo, tiene pocos estadísticos independientes: mientras que prácticamente todos los estándares y directrices evaluados por este organismo necesitan la consideración de principios estadísticos, solo hay 14 estadísticos repartidos por los cuatro *Scientific Area Committees* (Comités de áreas científicas) y 23 subcomités entre los 600 miembros de OSAC.

#### 6.3.1. *Reestructurando*

El PCAST concluye que el OSAC carece de suficiente experiencia científica y supervisora independientes para superar los graves defectos de la ciencia forense. Se necesita cierta reestructuración para garantizar que científicos y estadísticos independientes tengan más voz en el proceso de desarrollo de los estándares, un requisito

para la validez científica significativa. Lo más importante es que el OSAC debe tener un comité formal —un Comité de recursos de metrología al nivel de los otros tres Comités de recursos (el de recursos jurídicos, el de factores humanos y el de infraestructuras de calidad)—. Este Comité debería estar compuesto por científicos de laboratorio y estadísticos ajenos a la comunidad de las ciencias forenses, encargándose de revisar cada estándar y directriz que los Comités de áreas científicas recomienden para la aprobación del registro, antes de su envío para una revisión final a la *Forensic Science Standards Board* (FSSB) (Junta de estándares para la ciencia forense).

### 6.3.2. Disponibilidad de los estándares de OSAC

El OSAC no es un organismo formal de normalización. Revisa y evalúa los estándares relevantes para la ciencia forense desarrollados por organizaciones de desarrollo de estándares como *ASTM International*, la *National Fire Protection Association (NFPA)* (Asociación nacional de protección contra incendios) y la *International Organization for Standardization (ISO)* (Organización internacional para la estandarización), para su inclusión en los *OSAC Registries of Standards and Guidelines* (Registros de estándares y directrices de OSAC). El proceso de evaluación de OSAC incluye un periodo de comentarios públicos. OSAC, trabajando con los desarrolladores de estándares, ha dispuesto que el contenido de los estándares bajo consideración sea accesible al público durante el periodo de comentarios públicos. Una vez aprobado por OSAC, un estándar se enumera listándolo por título en un registro público mantenido por NIST. Es habitual que algunas organizaciones de desarrollo de estándares, incluida *ASTM International*, cobre una cuota por una copia con licencia de cada estándar protegido por derechos de autor y restrinja a los usuarios la distribución de esos estándares<sup>362, 363</sup>.

El NIST recientemente negoció un acuerdo de licencia con *ASTM International* por el que mediante una tarifa se permite a los empleados del gobierno federal, estatal y local el acceso online al Comité de estándares *ASTM E30*<sup>364</sup>. Sin embargo, esta lista no incluye a los acusados indigentes, abogados defensores privados o el enorme número de potenciales interesados de la comunidad académica. Hasta el presente, se han negociado contratos con otras *Standards Developing Organizations (SDOs)* (Organizaciones de desarrollo de estándares) que tienen, en la actualidad, estándares

---

<sup>362</sup> Para una lista de estándares de ciencia forense de *ASTM*, consúltese: [www.astm.org/DIGITAL\\_LIBRARY/COMMIT/PAGES/E30.htm](http://www.astm.org/DIGITAL_LIBRARY/COMMIT/PAGES/E30.htm),

<sup>363</sup> La American Academy of Forensic Sciences (AAFD) (Academia Norteamericana de Ciencias Forenses) llegará también a ser una Organización de Desarrollo de Estándares acreditada (SDO) y podría, en el futuro, desarrollar estándares para su revisión e inclusión en la lista por OSAC.

<sup>364</sup> De acuerdo con el contrato revisado, *ASTM* proporcionará acceso ilimitado a su web de todos los Estándares de Ciencia Forense del Comité E30 de *ASTM*: a miembros y afiliados de OSAC; a Laboratorios de criminalística federales, estatales y locales y de NIST; a Oficinas de Abogados Defensores públicos; Oficinas de la Fiscalía; Agencias de Seguridad Públicas; y a Médicos forenses.

bajo revisión por la OSAC. El PCAST cree que es importante que los estándares destinados a ser usados en la administración de justicia penal estén ampliamente disponibles para todos los que puedan necesitar su acceso. Es importante que los estándares estén fácilmente accesibles tanto a los acusados como a los observadores externos, que tienen un papel muy importante que desempeñar para garantizar la calidad de la justicia penal<sup>365</sup>.

El NIST debe garantizar que el contenido de los estándares y directrices registradas por la OSAC esté libremente accesible a cualquier parte que pueda necesitarlas en relación con un caso jurídico o para su evaluación e investigación, incluyendo su alineación con las políticas relacionadas con la razonable disponibilidad de los estándares de la *Office of Management and Budget Circular A-119* (Oficina de gestión y circular presupuestaria A-119), en la *Federal Participation in the Development and Use of Voluntary Consensus Standards and Conformity Assessment Activities* (Participación federal en el desarrollo y uso de estándares de consenso voluntario y actividades de evaluación de conformidad), y la *Office of Federal Register* (Oficina del registro federal), *IBR Handbook* (Manual IBR) (incorporación por referencia).

#### 6.4. Necesidad de una estrategia de I+D para la ciencia forense

El informe del NRC (2009) expuso que existe una necesidad urgente de fortalecer la ciencia forense, resaltando que «la investigación en ciencia forense no está bien sustentada y no existe una estrategia unificada para desarrollar un plan de investigación en ciencia forense entre las agencias federales»<sup>366</sup>.

Es especialmente importante crear y sustentar una comunidad de investigación académica vibrante enraizada en la cultura científica de las universidades. Esto requerirá una financiación significativa para apoyar a los grupos de investigación académica, pero redundará en obtener dividendos para impulsar la calidad y la innovación, tanto en métodos existentes como en otros completamente nuevos.

Tanto el NIST como NSF han tomado medidas iniciales, recientemente, para ayudar a cerrar las brechas de calado existentes entre los analistas forenses y las comunidades académicas a través de centros de investigación multidisciplinarios. Estos centros prometen involucrar a la más amplia comunidad investigadora en el avance de la ciencia forense, así como crear vínculos necesarios entre la comunidad de ciencias forenses y una amplia base de universidades que pueda ayudar a impulsar la investigación básica crítica.

---

<sup>365</sup> El PCAST no expresa opinión alguna sobre la idoneidad de las tasas de los estándares en áreas diferentes a la justicia penal.

<sup>366</sup> National Research Council, 2009: 78.

No obstante, como se señala en la sección 2, el nivel total de fondos federales del NIJ, NIST y NSF destinado a la comunidad académica para la investigación básica en ciencias forenses es extremadamente reducido. Se precisa una financiación sustancialmente mayor para desarrollar una comunidad de investigación sólida y sustentar el desarrollo y la evaluación de nuevas tecnologías prometedoras.

Se necesita una mejor coordinación de los esfuerzos federales en I+D en ciencias forenses, tanto dentro como fuera de su ámbito natural. Ninguna agencia, individualmente, posee la responsabilidad de asegurar que las ciencias forenses estén adecuadamente sustentadas. Se necesita una mayor coordinación entre las agencias federales y los laboratorios relevantes para garantizar que la financiación se destine a las principales prioridades y que el trabajo sea de la máxima calidad.

OSTP debe convocar a las agencias federales, laboratorios y partes interesadas relevantes para desarrollar una estrategia de investigación nacional y un plan quinquenal para asegurar que la investigación básica en apoyo de las ciencias forenses esté bien coordinada, consolidar los compromisos contraídos hasta la fecha por las agencias federales, e impulsar nuevas acciones y fondos que pudieran adoptarse para fomentar la investigación básica adicional, mejorar los métodos forenses actuales, apoyar la creación de nuevas bases de datos de investigación y supervisar la revisión y priorización periódicas de la investigación.

### 6.5. Recomendaciones

#### **RECOMENDACIÓN 1. VALORACIÓN DE LA VALIDEZ DE LOS FUNDAMENTOS**

Es importante que las evaluaciones científicas de la validez de los fundamentos se realicen de forma continua para poder valorar la validez de los fundamentos de tecnologías de comparación de características forenses actuales y recientemente desarrolladas. Para asegurar que los juicios científicos no sean sesgados y sean independientes, tales evaluaciones deben estar dirigidas por un organismo científico que no tenga interés en el resultado.

(A) El Instituto Nacional de Estándares y Tecnología (NIST) debe realizar esas evaluaciones y debe emitir un informe público anual evaluando la validez de los fundamentos de los métodos clave de comparación de características.

(j) Las evaluaciones deben (a) valorar si cada método revisado ha sido adecuadamente definido, si ha sido adecuadamente establecida la validez de sus fundamentos y si el nivel estimado de precisión se basa en evidencia empírica; (b) basarse en estudios publicados en la literatura científica de laboratorios y organismos de Estados Unidos y de otros países, así como cualquier trabajo dirigido por la propia plantilla y concesionarios del NIST; (c) como mínimo, elaborar evaluaciones sobre la línea del presente informe, actualizándolas según corresponda; (d) y que se lleven a cabo bajo los auspicios del

NIST, con asesoramiento adicional de analistas ajenos a la ciencia forense cuando se considere necesario.

(ii) El NIST debe establecer un comité asesor de científicos experimentales y estadísticos ajenos a la comunidad de ciencias forenses que proporcione asesoramiento sobre las evaluaciones y debe asegurarse de que sean rigurosos e independientes. Los miembros del comité consultivo deben seleccionarse conjuntamente por el NIST y la Oficina de políticas científicas y tecnológicas.

(iii) El NIST debe priorizar aquellos métodos forenses de comparación de características que más necesiten de una evaluación, incluyendo los que están actualmente en uso y los que estén en desarrollo en etapas avanzadas, partiendo de las aportaciones del Departamento de Justicia y de la comunidad científica.

Sobre la base de sus hallazgos científicos, el PCAST hace las siguientes recomendaciones.

(iv) Cuando el NIST evalúe que un método ha sido establecido como válido en sus fundamentos, debe (a) aportar estimaciones apropiadas de tasas de error basadas en estudios sobre los fundamentos y (b) identificar cualquier problema relevante para la validez de su aplicación.

(v) Cuando el NIST evalúe que un método no ha sido establecido como válido en sus fundamentos, debe sugerir qué medidas, si las hubiere, podrían adoptarse para establecer la validez del método.

(vi) El NIST no debe tener responsabilidades regulatorias con respecto a la ciencia forense.

(vii) El NIST debe alentar a una o más revistas científicas líderes ajenas a la comunidad forense a que desarrolle mecanismos para promover la rigurosa revisión por pares y la publicación de artículos relacionados con la validez de los fundamentos de los métodos forenses de comparación de características.

(B) El Presidente debe solicitar y el Congreso debe proporcionar un aumento de crédito al NIST de (a) 4 millones de dólares para sustentar las actividades de evaluación anteriormente descritas y (b) de 10 millones de dólares para sustentar el aumento de las actividades de investigación en ciencia forense, incluyendo las mezclas complejas de ADN, las huellas dactilares latentes, el reconocimiento de voz y del hablante y la biometría facial y del iris.

**RECOMENDACIÓN 2.**  
**DESARROLLO DE MÉTODOS OBJETIVOS PARA EL ANÁLISIS DE ADN**  
**DE MUESTRAS DE MEZCLAS COMPLEJAS,**  
**ANÁLISIS DE HUELLAS DACTILARES LATENTES Y ANÁLISIS DE ARMAS DE FUEGO.**

El Instituto Nacional de Estándares y Tecnología (NIST) debe desempeñar un papel líder en la transformación de tres importantes métodos de comparación de características que son actualmente subjetivos —análisis de huellas dactilares latentes, análisis de

armas de fuego y, en algunas circunstancias, análisis de ADN de mezclas complejas— en métodos objetivos.

(A) El NIST debe coordinar esos esfuerzos con el Laboratorio de la Oficina Federal de Investigaciones (FBI), el Centro de Ciencias Forenses de la Defensa, el Instituto Nacional de Justicia y otras agencias relevantes.

(B) Esos esfuerzos deben incluir (i) la creación y difusión de grandes bases de datos y materiales de pruebas (tales como mezclas de ADN complejas) que permitan el desarrollo y la prueba de los métodos, tanto por el sector privado como por académicos, (ii) subvenciones y contratos y (iii) procesos de patrocinio, tales como competencias premiadas, para evaluar métodos.

### **RECOMENDACIÓN 3. MEJORA DEL PROCESO DE ORGANIZACIÓN PARA LOS COMITÉS DE ÁREAS CIENTÍFICAS**

(A) El Instituto Nacional de Estándares y Tecnología (NIST) debería mejorar la organización para los Comités de Áreas Científicas (OSAC), que fue establecida para desarrollar y promulgar estándares y directrices que permitan las mejores prácticas en la comunidad de ciencia forense.

(i) El NIST debe establecer un Comité de recursos de metrología, compuesto por metrologos, estadísticos y otros científicos ajenos a la comunidad de ciencia forense. Un representante del Comité de recursos de metrología debe prestar servicio en cada uno de los Comités de áreas científicas (SACs) para proporcionar orientación sobre la aplicación de los principios estadísticos y de medición de los estándares documentados en desarrollo.

(ii) El Comité de recursos de metrología, en su conjunto, debe revisar y aprobar o desaprobado públicamente todos los estándares propuestos por los Comités de áreas científicas antes de que sean transmitidos a la Junta Directiva de estándares de ciencia forense.

(B) El NIST debe asegurar que el contenido de los estándares y directrices registrados por la OSAC estén disponibles libremente para cualquier parte que los desee en relación con un caso jurídico, o para evaluación o investigación, incluso alineándose con las políticas relacionadas con la disponibilidad razonable de estándares en la Oficina de gestión y circular presupuestaria A-119, en la Participación federal en el desarrollo y uso de estándares de consenso voluntario y en las actividades de evaluación de conformidad y la Oficina del Registro Federal, Manual IBR (incorporación por referencia).

**RECOMENDACIÓN 4.  
STRATEGIA DE I+D PARA LA CIENCIA FORENSE**

(A) La Oficina de política de la ciencia y de la tecnología (OSTP) debe coordinar la creación de una estrategia nacional de investigación y desarrollo de la ciencia forense. La estrategia debe abordar los planes y las necesidades de financiación para:

(i) una gran expansión y fortalecimiento de la comunidad académica que investiga en las ciencias forenses, incluyendo un aumento sustancial de la financiación tanto para la investigación como para la formación;

(ii) estudios de validez de los fundamentos de los métodos forenses de comparación de características;

(iii) mejora de los métodos forenses actuales, incluyendo la transformación de métodos subjetivos en métodos objetivos y el desarrollo de nuevos métodos forenses;

(iv) el desarrollo de bases de datos de características forenses, con las adecuadas protecciones de la privacidad, que puedan utilizarse para la investigación;

(v) salvar la brecha entre los investigadores científicos y los profesionales forenses;

(vi) supervisión y revisión periódica de la investigación en ciencias forenses.

(B) Al preparar la estrategia, OSTP debe solicitar la opinión de las agencias federales apropiadas, incluyendo especialmente el Departamento de Justicia, el Departamento de Defensa, a la Fundación Nacional de la Ciencia, y el Instituto Nacional de Estándares y Tecnología; a profesionales de la ciencia forense de ámbito federal y estatal; a investigadores científicos forenses y no forenses; y a otras partes interesadas.

## 7. Acciones para garantizar la validez científica en la ciencia forense: recomendaciones para el Laboratorio del FBI

Sobre la base de los hallazgos científicos de las Secciones 4 y 5, el PCAST ha identificado acciones que creemos que han de adoptarse en el Laboratorio del FBI para garantizar la validez científica de los métodos de comparación de características.

Resaltamos que el Laboratorio del FBI ha desempeñado un importante papel en los últimos años en la realización de estudios científicos de alta calidad en análisis de huellas dactilares latentes. El PCAST aplaude esos esfuerzos e insta al Laboratorio del FBI a ampliarlos.

### 7.1. *Papel del Laboratorio del FBI*

El Laboratorio del FBI es un centro de excelencia en el estado del arte de última generación que trabaja para aplicar la ciencia de vanguardia con el fin de resolver

casos y prevenir el delito. Su misión consiste en aplicar capacidades científicas y servicios técnicos para la recogida, procesado y explotación de pruebas para el Laboratorio y otras agencias de seguridad y de inteligencia debidamente constituidas en apoyo de las prioridades de investigación e inteligencia. Actualmente, el Laboratorio tiene, aproximadamente, 750 empleados y más de 300 proveedores para cumplir con el amplio alcance de esta misión.

### 7.1.1. Capacidades y servicios del Laboratorio

El FBI dispone de capacidades y personal especializado para responder a incidentes, recoger pruebas en su campo, llevar a cabo análisis forenses y proporcionar testimonios de analistas en el proceso judicial. El FBI apoya a los equipos de recolección de pruebas en las 56 oficinas existentes en todo EUA y Puerto Rico y cuenta con personal especializado en pruebas de naturaleza peligrosa<sup>367</sup> y documentación y recopilación de datos en la escena del crimen. El Laboratorio es responsable de capacitar y proporcionar estas actividades de respuestas para el personal del FBI en los Estados Unidos<sup>368</sup>. El Laboratorio también gestiona el Centro de análisis de dispositivos explosivos terroristas (TEDAC), que recibió cerca de 1000 solicitudes de pruebas en el ejercicio fiscal 2015 y comunicó unos 2000 informes de inteligencia.

El laboratorio del FBI dispone de analistas forenses que realizan análisis en una serie de disciplinas que incluyen la química, criptoanálisis, ADN, análisis de armas de fuego y de marcas de herramientas, huellas latentes, documentos cuestionados y rastreo de pruebas. El Laboratorio del FBI recibió más de 3.875 solicitudes de informes y confeccionó más de 4.850 informes en el año fiscal 2015.

Además de llevar a cabo el análisis de casos en el ámbito federal, el Laboratorio proporciona apoyo a laboratorios estatales y locales y realiza pruebas en casos de ámbito estatal y local en algunas disciplinas.

### 7.1.2. Actividades de investigación y desarrollo

Además de sus servicios, el Laboratorio del FBI desarrolla importantes actividades de investigación y desarrollo. Esas actividades son fundamentales para proporcionar al Laboratorio las herramientas más punteras para avanzar en su misión. Un programa y una cultura de investigación sólida son también importantes para que el Laboratorio pueda mantener la excelencia y atraer a personal altamente cualificado.

Debido al ámbito expansivo y a los muchos requerimientos de sus operaciones, solo alrededor del 5% de los 100 millones de dólares del presupuesto anual del Labo-

<sup>367</sup> \* N.del T.: por ejemplo, los incendios y entornos NBQ.

<sup>368</sup> El Laboratorio del FBI apoyó 162 despliegues y 168 intervenciones de respuesta, así como la impartición de 239 cursos de formación en el ejercicio fiscal 2015.

torio del FBI está disponible para actividades de I+D<sup>369</sup>. El presupuesto en I+D se extiende a través de una serie de actividades de investigación aplicada, incluyéndose los estudios de validación (para nuevos métodos o productos comerciales tales como nuevos analizadores de ADN). Para sus actividades de investigación interna, el Laboratorio depende, en gran medida, de su Programa de científicos visitantes, que atrae a aproximadamente 25 estudiantes postdoctorales, de maestrías y de licenciatura cada año. El Laboratorio se ha asociado con otros organismos gubernamentales para proporcionar más recursos a sus prioridades de investigación mediante iniciativa compartida y también ha sido capaz de ampliar los presupuestos disponibles con estudios de investigación en cuestiones cruciales de forma creciente con el paso de los años.

La serie de estudios del Laboratorio del FBI en el examen de huellas latentes es un ejemplo de la importante investigación básica que fue capaz de desarrollar de forma creciente en un periodo de cinco años. El trabajo incluye estudios de «caja negra» que evalúan la precisión y la fiabilidad de las conclusiones de los analistas con huellas latentes, así como estudios de «caja blanca» que evalúan cómo la cantidad y calidad de las características se relacionan con las decisiones de los analistas en huellas latentes. Estos estudios han dado lugar a una serie de publicaciones importantes que han ayudado a cuantificar las tasas de error en la práctica de esas pericias y evalúan la repetibilidad y reproducibilidad de las decisiones de los analistas en huellas dactilares latentes. De hecho, la conclusión del PCAST de que el análisis de huellas dactilares latentes es válido en sus fundamentos descansa fuertemente en el estudio de caja negra del FBI. En algunas otras disciplinas se están llevando a cabo similares líneas de investigación, como en el examen de armas de fuego y documentos cuestionados.

Desafortunadamente, la limitada financiación disponible para esos estudios —y para el programa de investigación intramuros en general— ha obstaculizado el progreso en la prueba de la validez de los fundamentos de los métodos de las ciencias forenses y su fortalecimiento. PCAST cree que el presupuesto para el Laboratorio del FBI debe aumentarse significativamente y que debe orientarse a fin de permitir que el presupuesto de I+D se incremente en un total de 20 millones de dólares.

### 7.1.3. Accesos a bases de datos

El FBI tiene también un importante papel que jugar en fomentar la investigación entre científicos externos, facilitándoles el acceso, bajo apropiadas condiciones, a grandes bases de datos forenses. La mayoría de las bases de datos que se utilizan habitualmente en los análisis forenses no son accesibles a los investigadores y la falta de acceso obstaculiza el progreso en la mejora de la ciencia forense. Por ejemplo,

---

<sup>369</sup> En 2014, el Laboratorio del FBI invirtió 10.9 millones de dólares en I+D en ciencia forense, aproximadamente la mitad con su propio presupuesto y la otra mitad con subvenciones de NIST y el Departamento de Seguridad Nacional. Véase: National Academies of Sciences, Engineering, and Medicine, 2015: 31.

sistemas de bases de datos balísticas, como el Sistema Nacional Integrado de Información Balística de la Oficina de alcohol, tabaco, armas de fuego y explosivos (NIBIN), donde los analistas en armas de fuego buscan identificar un arma de fuego o un cartucho, no pueden ser evaluados para estudiar su completitud, relevancia o calidad y el algoritmo de búsqueda utilizado para identificar potenciales coincidencias tampoco puede ser evaluado.

El sistema NGI (formalmente IAFIS)<sup>370</sup> que actualmente alberga más de 70 millones de entradas de huellas dactilares ampliaría enormemente los datos disponibles para su estudio; en la actualidad, solo existe una base de datos de huellas dactilares disponible al público con 258 pares de huellas latentes e impresiones decadactilares<sup>371</sup>. Y el sistema NDIS del FBI, que actualmente contiene más de 14 millones de perfiles de ADN de delincuentes y detenidos. El NIST ha desarrollado un inventario de todas las bases de datos forenses que son más utilizadas por las fuerzas de seguridad y científicos forenses con información sobre su accesibilidad.

Se necesitan esfuerzos sustanciales para lograr que las bases de datos forenses existentes sean más accesibles a la comunidad investigadora, garantizando la apropiada protección de la privacidad como la eliminación de información de identificación personal y las restricciones al uso de los datos.

Para algunas disciplinas como los análisis de armas de fuego y las marcas de rodadura no existen preocupaciones significativas de privacidad.

Con respecto a las huellas latentes, podrían mejorarse las preocupaciones referentes a la privacidad de diversas formas. Por ejemplo, podría evitarse esa preocupación (1) generando grandes colecciones de pares de impresiones latentes conocidas con calidad y cantidad de información variables a través del tacto y del manejo de objetos naturales en una amplia variedad de circunstancias (superficies, presión, distorsión, etc.); (2) utilizando software para generar automáticamente «*morphing transformation*» de las impresiones conocidas y las impresiones latentes; y (3) aplicar esas transformaciones a impresiones de fallecidos para crear millones de pares de impresiones latentes conocidas<sup>372</sup>.

En ADN se han desarrollado protocolos en investigación genómica humana, lo que plantea similares o mayores preocupaciones de privacidad, para permitir el acceso a información a los investigadores de buena fe<sup>373</sup>. Estas políticas también deberían

---

<sup>370</sup> NGI estandariza la «Identificación en la Siguiete Generación» y combina múltiples sistemas de información biométrica, incluyendo IAFIS, sistemas de reconocimiento de iris y rostro, y otros.

<sup>371</sup> Base de Datos Especial de NIST 27A, accesible en: [www.nist.gov/itl/iad/image-group/nist-special-database-27a-sd-27a](http://www.nist.gov/itl/iad/image-group/nist-special-database-27a-sd-27a).

<sup>372</sup> Las oficinas de medicina forense toman rutinariamente las huellas dactilares a los individuos fallecidos como parte del proceso de autopsia; estas huellas dactilares podrían coleccionarse y utilizarse para crear grandes bases de datos con fines de investigación.

<sup>373</sup> Podrían aplicarse aquí una serie de modelos que han sido desarrollados en el contexto de la investigación biomédica para permitir el acceso a datos sensibles, al tiempo que proporcionan una

ser viables para las bases de datos forenses de ADN. Observamos que la ley que autoriza al FBI a mantener una base de datos nacional forense de ADN contempla, explícitamente, que se permite el acceso a las muestras de ADN y a los análisis de ADN si «se elimina la información de identificación personal ... para fines de investigación, de identificación y desarrollo de protocolos»<sup>374</sup>. Aunque la ley no contiene una declaración explícita sobre este particular, el DOJ interpreta que la ley permite el uso para este propósito solo por parte de los organismos de justicia penal. Es reactivo, a falta de aclaraciones jurídicas, incluso a un acceso controlado a otros investigadores. Este tema merece atención.

El PCAST cree que la disponibilidad de datos acelerará el desarrollo de métodos, herramientas y software que mejorarán la ciencia forense. Para las bases de datos bajo su control, el Laboratorio del FBI debe desarrollar programas para hacer que las bases de datos forenses (o subconjuntos de esas bases) sean accesibles a investigadores en condiciones que protejan la privacidad. Con respecto a bases de datos propiedad de terceros, el Laboratorio del FBI y el NIST deben trabajar con las agencias y empresas que controlan esas bases de datos para desarrollar programas que les permitan un acceso apropiado.

## 7.2. Recomendaciones

Sobre la base de sus hallazgos científicos, el PCAST emite la siguiente recomendación:

**RECOMENDACIÓN 5.  
AGENDA PARA LA AMPLIACIÓN DE LAS CIENCIAS FORENSES  
EN EL LABORATORIO DEL FBI**

(A) *Programas de investigación.* El Laboratorio de la Oficina Federal de Investigación (FBI) debería emprender un vigoroso programa de investigación para mejorar la ciencia forense, sobre la base de su reciente e importante trabajo en análisis de huellas dactilares latentes. El programa debe incluir:

- (i) la realización de estudios sobre la fiabilidad de los métodos de comparación de características, junto a terceros independientes sin interés en el resultado;
- (ii) el desarrollo de nuevos enfoques que mejoren la fiabilidad de los métodos de comparación de características;
- (iii) la ampliación de los programas de colaboración con científicos externos; y

adecuada protección de la privacidad. Podría exigirse a los investigadores la firma de un acuerdo de no divulgación (NDA) o que accedan bajo un acuerdo de uso limitado. Podría exigirse a los investigadores que accedan a los datos en un determinado modo y lugar, de forma que los datos no puedan descargarse o compartirse, o podría permitirse la descarga de forma agregada o resumida.

<sup>374</sup> Federal DNA Identification Act, 42 U.S.C. §14132(b)(3)(D)).

(iv) la garantía de que los científicos externos tengan acceso adecuado a los conjuntos de datos y colecciones de muestras para que puedan realizar estudios independientes.

(B) *Estudios de caja negra.* Basándose en su experiencia en la investigación científica forense, el Laboratorio del FBI debería ayudar en el diseño y ejecución de estudios adicionales de caja negra para métodos subjetivos, incluido los análisis de huellas dactilares latentes y de armas de fuego. Esos estudios deben ser realizados por o en conjunto con terceros independientes sin interés en el resultado.

(C) *Desarrollo de métodos objetivos.* El Laboratorio del FBI debe trabajar con el Instituto Nacional de Estándares y Tecnología para transformar tres importantes métodos de comparación de características que actualmente son subjetivos —análisis de huellas dactilares latentes, análisis de armas de fuego y, en algunas circunstancias, análisis de ADN de mezclas complejas—, en métodos objetivos. Estos esfuerzos deben incluir (i) la creación y difusión de grandes bases de datos que sustenten el desarrollo y las pruebas de métodos, tanto por parte de las empresas como de académicos, (ii) becas y apoyo contractual y (iii) patrocinio de competiciones premiadas para evaluar los métodos.

(D) *Pruebas de aptitud.* El Laboratorio del FBI debe promover un mayor rigor en las pruebas de aptitud (i) dentro de los próximos cuatro años, instituyendo pruebas de aptitud ciegas de rutina dentro del flujo de casos en su propio laboratorio; (ii) ayudando a otros laboratorios federales, estatales y locales a hacerlo igualmente; y (iii) fomentar el acceso rutinario y la evaluación utilizadas en las pruebas comerciales de aptitud.

Recomendación 5. Agenda para la ampliación de las ciencias forenses en el Laboratorio del FBI (continuación)

(E) *Análisis de huellas dactilares latentes.* El Laboratorio del FBI debe promover vigorosamente la adopción, por todos los laboratorios que realizan análisis de huellas dactilares latentes, de las normas que requieren un proceso de “Análisis, comparación y evaluación lineales” —en el que los analistas deben completar y documentar sus análisis de huellas dactilares latentes antes de mirar a una impresión dactilar conocida y deben, separadamente, documentar cualquier dato adicional utilizado en la evaluación y en la comparación—.

(F) *Transparencia en materia de calidad en casos.* El Laboratorio del FBI, así como otros laboratorios forenses federales, debe informar públicamente y regular problemas de calidad en el análisis de casos (de forma similar a las prácticas empleadas por el Instituto forense holandés, descritas en la sección 5), como un medio para mejorar la calidad y promover la transparencia.

(G) *Presupuesto.* El Presidente debe solicitar y el Congreso debe proporcionar un incremento de la partida presupuestaria del FBI que restaure el presupuesto del Laboratorio del FBI para las actividades de investigación en ciencias forenses partiendo de su nivel actual con 30 millones de dólares adicionales y debe evaluar la necesidad de aumentar la financiación para otras actividades de investigación en ciencias forenses en el Departamento de Justicia.

## 8. Acciones para garantizar la validez científica en la ciencia forense: recomendaciones al Fiscal General

Sobre la base de los hallazgos científicos referidos en las Secciones 4 y 5, el PCAST ha identificado acciones que creemos que deberían ser llevadas a efecto por el Fiscal General para asegurar la validez científica de los métodos de comparación de características y promover su uso más riguroso en las salas de audiencias.

### 8.1. *Garantizar el uso de métodos científicamente válidos en procesos judiciales*

El Gobierno Federal tiene el profundo compromiso de garantizar que los procesos judiciales penales no solo sean justos en su procedimiento, sino correctos en sus resultados —es decir, que los culpables sean condenados, mientras que los inocentes no lo sean—.

Con este fin, el Departamento de Justicia debe asegurarse que el testimonio sobre las pruebas forenses presentadas ante un tribunal sea científicamente válido. Este informe proporciona orientación al Departamento de Justicia acerca de los criterios científicos tanto para la validez de los fundamentos como para la validez de su aplicación, así como evaluaciones de seis métodos forenses específicos y una discusión sobre un séptimo. A largo plazo, el DOJ debe inspeccionar las evaluaciones en curso de los métodos forenses que deben ser realizadas por el NIST (como se describe en la sección 6).

Mientras tanto, el DOJ debería llevar a cabo una revisión de los métodos de comparación de características (más allá de los revisados en el este informe) para identificar qué métodos utilizados por el DOJ carecen de los apropiados estudios de caja negra necesarios para evaluar la validez de los fundamentos. Como consecuencia de que de tales métodos subjetivos no están presuntamente establecidos como válidos en sus fundamentos, el DOJ debe evaluar (1) si debe presentar ante los tribunales conclusiones basadas en tales métodos y (2) si deben iniciarse estudios de caja negra para evaluar esos métodos.

### 8.2. *Revisión de las directrices propuestas recientemente por el Departamento de Justicia sobre el testimonio de los analistas*

El 3 de junio de 2016, el Departamento de Justicia publicó, para recepción de comentarios, un primer conjunto de directrices propuestas, junto con documentos de apoyo, sobre un «Lenguaje Uniforme Propuesto para Testimonio e Informes» en varias ciencias forenses, incluyendo el análisis de huellas dactilares latentes y el análisis de

impresiones de calzado y neumáticos forense<sup>375</sup>. El 21 de julio de 2016, el DOJ publicó, para recepción de comentarios, un segundo conjunto de directrices propuestas y documentos de apoyo para diversas ciencias forenses adicionales, incluyendo análisis microscópico del cabello, ciertos tipos de análisis de ADN y otras disciplinas.

Las directrices representan un importante paso hacia adelante porque instruyen a los analistas del Departamento de Justicia a no realizar afirmaciones categóricas como que pueden identificar la fuente de una huella dactilar o de una impresión de calzado excluyendo cualquier otra fuente posible. El PCAST aplaude la intención y los esfuerzos del DOJ para lograr uniformidad y evitar testimonios inexactos en comparaciones de características.

Algunos aspectos de las directrices, sin embargo, no son científicamente apropiadas y encarnan puntos de vista heterodoxos del tipo de aquellos discutidos en la Sección 4.7. A modo de ilustración, nos centramos en las directrices para el análisis de impresiones de calzado y neumáticos y en las de análisis del cabello.

### 8.2.1. Análisis de impresiones de calzado y neumáticos

En el Cuadro 6 se muestran partes relevantes de las directrices para el testimonio y la redacción de informes sobre impresiones de calzado y de ruedas de neumáticos.

**CUADRO 6.**  
**EXTRACTO DE LA PROPUESTA DEL DOJ SOBRE UN LENGUAJE UNIFORME**  
**PARA TESTIMONIOS E INFORMES EN LAS DISCIPLINAS DE IMPRESIONES DE CALZADO**  
**Y DE NEUMÁTICOS FORENSES<sup>376</sup>.**

Afirmaciones aprobadas para su uso en informes de los laboratorios y en las declaraciones de los analistas con respecto al examen forense de la evidencia de huellas de calzado y de neumáticos.

*Identificación*

El analista puede afirmar que, en su opinión, el calzado o el neumático es la fuente de la impresión porque hay suficiente cantidad y calidad de características coincidentes como para esperar que no pudiera encontrarse la misma combinación de características repetida por esa fuente. Se trata del mayor grado de asociación entre una impresión cuestionada y una fuente conocida. La opinión requiere que la impresión cuestionada y la conocida coincidan en las características de clase y que también compartan una o más características aleatoriamente adquiridas. Esta opinión reconoce

<sup>375</sup> Véase: [www.justice.gov/dag/proposed-language-regarding-expert-testimony-and-lab-reports-forensic-science](http://www.justice.gov/dag/proposed-language-regarding-expert-testimony-and-lab-reports-forensic-science).

<sup>376</sup> El 21 de julio de 2016 se liberó un segundo conjunto de directrices propuestas incluyendo análisis de cabello, ADN mitocondrial y tipado de cromosoma Y. ([www.justice.gov/dag/proposed-uniform-language-documents-anthropology-explosive-chemistry-explosive-devicesgeology](http://www.justice.gov/dag/proposed-uniform-language-documents-anthropology-explosive-chemistry-explosive-devicesgeology)).

<sup>376</sup> Véase: [www.justice.gov/olp/file/861936/download](http://www.justice.gov/olp/file/861936/download)

que una identificación que excluya todas las demás fuentes posibles no puede nunca probarse empíricamente.

Afirmaciones no aprobadas para su uso en informes de los laboratorios y en las declaraciones de los peritos con respecto al examen forense de la evidencia de huellas de calzado y de neumáticos.

*Exclusión de todas las demás fuentes potenciales*

El analista no puede afirmar que un calzado o un neumático sea la fuente de una impresión cuestionada excluyendo todos los demás calzados/neumáticos porque los demás calzados/neumáticos no han sido examinados. Examinar todos los calzados/neumáticos existentes en el mundo es imposible en la práctica.

*Tasa de error*

El analista no puede establecer un valor numérico o un porcentaje con respecto a la tasa de error asociada, tanto con respecto a la metodología utilizada para realizar el examen pericial como con respecto a la ejecución del analista que realiza los análisis.

*Valoración estadística*

El analista no puede establecer un valor numérico o una probabilidad asociada a su opinión. No existen en la actualidad datos precisos y fiables o modelos estadísticos que permitan realizar una determinación cuantitativa en un examen forense de la evidencia de una impresión de calzado o rueda de neumático.

Estas directrices propuestas presentan problemas serios.

Un analista puede opinar que un calzado es la fuente de una impresión, pero no que el calzado es la fuente de la impresión *excluyendo todos los otros calzados posibles*. Pero, como una cuestión de lógica, no hay diferencia entre esas dos afirmaciones. Si un analista cree que X es la fuente de Y, cree necesariamente que nada más es la fuente de Y. Cualquier miembro de un jurado que sea sensato debe comprender esta equivalencia.

Entonces, ¿cuál es la finalidad de las directrices? Parece ser la de reconocer la posibilidad de error. En efecto, los analistas deben decir, «Yo creo que X es la fuente de Y, aunque podría equivocarme sobre ello».

Esto es apropiado. Pero la cuestión crucial es entonces: ¿con qué probabilidad se equivoca el analista?

He aquí el problema: las directrices impiden al perito disertar sobre la probabilidad de error porque no hay información precisa o fiable sobre la precisión (del método). En efecto, se instruye a los analistas para que digan: «Creo que X es la fuente de Y, aunque podría estar equivocado al respecto. Pero no sé con qué frecuencia puedo equivocarme al no disponer de información fiable sobre eso».

Tal afirmación no supera prueba plausible alguna sobre su validez científica. Como escribió el Juez Easterly en *Williams v. los Estados Unidos*, una afirmación de identificación en tales circunstancias:

tiene el mismo valor probatorio que la visión de un vidente: no refleja nada más que la infundada creencia individual de creer estar en lo cierto. Esta es no es la evidencia en la que podemos confiar en buena conciencia, particularmente en casos penales, donde demandamos pruebas -pruebas reales- más allá de toda duda razonable, precisamente porque lo que está en juego vale mucho<sup>377</sup>.

### 8.2.2. Análisis de cabellos

En el Cuadro 7 se muestran partes relevantes de las directrices para el testimonio y la redacción de informes sobre el examen forense del cabello.

<b>CUADRO 7. EXTRACTO DE LA PROPUESTA DEL DOJ SOBRE UN LENGUAJE UNIFORME PARA TESTI- MONIOS E INFORMES EN LA DISCIPLINA DEL EXAMEN FORENSE DEL CABELLO<sup>378</sup></b>
Afirmaciones no aprobadas para su uso en informes de los laboratorios y en las declaraciones de los analistas con respecto al examen forense del cabello.
<i>Comparaciones de cabello humano</i>
El analista puede afirmar o implicar que el cabello humano cuestionado es microscópicamente consistente con la muestra de cabello conocida y, consecuentemente, la fuente de la muestra de cabello conocida pudiera incluirse en las posibles fuentes del cabello cuestionado.
Afirmaciones no aprobadas para su uso en informes de los laboratorios y en las declaraciones de los analistas con respecto al examen forense del cabello.
<i>Individualización</i>
El analista no puede afirmar o implicar que el cabello humano cuestionado proceda de una fuente en particular excluyendo todas las demás.
<i>Valoración estadística</i>
El analista no puede afirmar o implicar un peso estadístico o una probabilidad a una conclusión, o una verosimilitud sobre si el cabello cuestionado procedió de una fuente en particular.
<i>Tasa de error</i>
El analista no puede afirmar o implicar que el método utilizado en realizar el examen microscópico del cabello tenga una tasa de error igual a cero o que sea infalible.

<sup>377</sup> *Williams v. United States*, DC Tribunal de Apelación, resuelto el 21 de junio de 2016, (voto concurrente de Easterly). Citamos la analogía por su valor expositivo sobre la cuestión científica; no expresamos posición alguna sobre el caso en cuanto autoridad jurídica.

<sup>378</sup> Department of Justice Proposed Uniform Language for Testimony and Reports for the Forensic Hair Examination Discipline (Lenguaje Uniforme Propuesto por el Departamento de Justicia para el Testimonio y los Informes en la Disciplina Forense de Examinación de Cabellos), disponible en: [www.justice.gov/dag/file/877736/download](http://www.justice.gov/dag/file/877736/download)

Las directrices afirman apropiadamente que los analistas no pueden defender que pueden individualizar la fuente del cabello o que tengan una tasa de error igual a cero. Sin embargo, mientras que los analistas pueden «afirmar o implicar que el cabello humano cuestionado es microscópicamente consistente con la muestra de cabello conocida y, consecuentemente, la fuente de la muestra de cabello conocida pudiera incluirse en las posibles fuentes del cabello cuestionado», se les impide proporcionar información precisa sobre la fiabilidad de tales conclusiones. Esto es contrario al requerimiento científico de que los métodos de comparación de características deben sustentarse y acompañarse de estimaciones empíricas apropiadas de su fiabilidad.

En particular, como se discutió en la Sección 5.7, un estudio emblemático realizado por científicos del Laboratorio del FBI en el año 2002 demostró que de las 80 ocasiones con casos reales en los que los analistas concluyeron que un cabello cuestionado era microscópicamente consistente con la muestra de cabello conocida, el análisis de ADN demostró que el cabello procedía de una fuente diferente en el 11% de los casos. El hecho de que una proporción tan significativa de conclusiones fueran falsas asociaciones es de enorme importancia en la interpretación de las conclusiones de los analistas del cabello.

En casos de análisis de cabellos no acompañados por análisis de ADN, se debe solicitar a los peritos que revelen la alta frecuencia de falsas asociaciones observadas en el estudio del FBI, de forma que los miembros del jurado puedan apropiadamente sopesar las conclusiones.

### 8.2.3. Conclusión

El DOJ debería revisar las directrices propuestas a fin de armonizarlas con los estándares científicos de validez científica. También debe revisarse la documentación justificativa, como se expone en la Sección 5.7.

## 8.3. Recomendaciones

Sobre la base de sus hallazgos científicos, el PCAST hace las siguientes recomendaciones.

**RECOMENDACIÓN 6.  
USO DE LOS MÉTODOS DE COMPARACIÓN DE CARACTERÍSTICAS  
EN PROCESOS JUDICIALES FEDERALES**

*(A) El Fiscal General debe ordenar a los fiscales que comparezcan en nombre del Departamento de Justicia (DOJ) que aseguren que el testimonio de los analistas ante los tribunales sobre los métodos de comparación de características cumpla los estándares científicos de validez científica.*

Si bien las investigaciones previas al juicio pueden basarse en una gama más amplia de métodos, el testimonio de los analistas ante los tribunales sobre métodos forenses de comparación de características en casos penales — que pueden ser muy influenciados y que ha conducido a muchas condenas injustas — debe cumplir un estándar más alto. Particularmente, los fiscales que comparezcan en nombre del Departamento de Justicia deben asegurarse de que:

(i) los métodos forenses de comparación de características sobre los que se basa el testimonio hayan sido establecidos como válidos en sus fundamentos, como lo demuestran los estudios empíricos apropiados y la consistencia con las evaluaciones realizadas por el Instituto Nacional de Estándares y Tecnología (NIST), cuando estén disponibles; y

(ii) el testimonio sea científicamente válido, con las declaraciones del analista sobre la precisión de los métodos y el valor probatorio de las identificaciones propuestas constreñidas por la evidencia empírica que les sirve de apoyo, sin implicar un mayor grado de certeza.

(B) *El Departamento de Justicia debe llevar a cabo una revisión inicial, con la asistencia de NIST, de los métodos de comparación de características subjetivos utilizados por el DOJ, con el fin de identificar qué métodos (más allá de los revisados en este informe) carecen de estudios apropiados de caja negra necesarios para evaluar la validez de los fundamentos.* Como consecuencia de que se presume que tales métodos subjetivos no son válidos en sus fundamentos, el Departamento de Justicia debe evaluar si es apropiado que se presenten ante los tribunales conclusiones basadas en tales métodos.

(C) *Cuando existan métodos relevantes que aún no se hayan establecido como válidos en sus fundamentos, el DOJ debe fomentar y proporcionar apoyo para que se realicen estudios apropiados de caja negra para evaluar la validez de los fundamentos y medir la fiabilidad.* El diseño y ejecución de estos estudios debe llevarse a cabo por o en unión de terceros independientes sin participación alguna en el resultado.

## 9. Acciones para garantizar la validez científica en la ciencia forense: recomendaciones al poder judicial

Sobre la base de los hallazgos referidos en las Secciones 4 y 5, el PCAST ha identificado acciones que creemos que deberían ser tenidas en cuenta por el Poder Judicial para garantizar la validez científica de la evidencia fundamentada en métodos de comparación de características y promover su uso más riguroso en las salas de audiencias.

### 9.1. *La validez científica como fundamento del testimonio experto*

En los tribunales federales, a los jueces se les asigna el papel crítico de «porteros» encargados de garantizar que el testimonio experto «descanse sobre una base

fiable»<sup>379</sup>. Específicamente, la regla 702 (c,d) de las Reglas Federales de la Prueba requiere que (1) el testimonio experto debe ser el producto de «principios y métodos fiables» y (2) los expertos deben haber «aplicado fiablemente» los métodos a los hechos del caso<sup>380</sup>. La Corte Suprema ha dictaminado que los jueces deben determinar «si el razonamiento o la metodología subyacente al testimonio es científicamente válida»<sup>381</sup>.

Como se explica en la sección 3, este entorno de trabajo establece un importante diálogo entre el poder judicial y la comunidad científica. La admisibilidad del testimonio experto depende de un umbral de cumplimiento de ciertos criterios *jurídicos* de fiabilidad probatoria, que son exclusiva potestad del poder judicial. Sin embargo, en los casos que entrañen pruebas científicas, esos estándares jurídicos deben «basarse en la validez científica»<sup>382</sup>.

El PCAST no opina sobre los criterios jurídicos, pero tiene como objetivo en este informe esclarecer los estándares *científicos* que subyacen a ellos. Para garantizar que la distinción entre conceptos científicos y jurídicos sea clara, hemos adoptado términos específicos para referirnos a los conceptos *científicos* (*validez de los fundamentos* y *validez en la aplicación*) con la intención de establecer un paralelismo con los conceptos jurídicos expresados en la Regla 702 (c,d).

Como ha señalado la Corte Suprema, la investigación del juez en virtud de la Regla 702 es flexible: no existe una simple prueba única que pueda aplicarse uniformemente a todas las disciplinas científicas<sup>383</sup>. Más bien, la evaluación de la validez científica debe basarse en los criterios científicos adecuados para el campo científico. Más aún, el campo científico apropiado debe ser la disciplina científica más amplia a la que pertenece<sup>384</sup>.

En este informe, el PCAST se ha centrado sobre métodos forenses de comparación de características —que pertenece al campo de la metrología, la ciencia de la medición y de su aplicación—<sup>385</sup>. Hemos buscado —en una forma utilizable por los tribunales, así como por científicos y otros que buscan mejorar la ciencia forense— establecer criterios científicos para la validez de los fundamentos y la validez en la

<sup>379</sup> *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579 (1993) en 597.

<sup>380</sup> Véase: [www.uscourts.gov/file/rules-evidence](http://www.uscourts.gov/file/rules-evidence).

<sup>381</sup> *Daubert v. Merrell Dow Pharmaceuticals*, 509 U.S. 579 (1993) en 592.

<sup>382</sup> *Daubert*, en FN9 («en un caso que conlleve prueba científica, la *fiabilidad de la prueba* estará basada en la validez científica.» [énfasis en el texto original]).

<sup>383</sup> *Daubert*, en 594.

<sup>384</sup> Por ejemplo, en el caso *Frye*, el tribunal valoró si un detector de mentiras propuesto había conseguido «reconocimiento permanente y científico entre las autoridades fisiológicas y psicológicas» en lugar de entre analistas en detectores de mentiras. *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923). De forma similar, el hecho de que los analistas en mordeduras crean que el examen de las mordeduras es válido, tiene escaso valor.

<sup>385</sup> Véase nota 93.

aplicación (Sección 4) e ilustrar su aplicación a métodos de comparación de características específicos (Sección 5).

Los criterios científicos se describen en el Hallazgo 1. Las conclusiones del PCAST se pueden resumir del siguiente modo:

La validez y fiabilidad científicas requieren que un método haya sido sometido a prueba empírica, bajo condiciones apropiadas a su pretendido uso, que proporcione estimaciones válidas sobre la frecuencia con la que el método alcanza una conclusión incorrecta. Para los métodos subjetivos de comparación de características, se precisan estudios de caja negra apropiadamente diseñados, en los que muchos analistas toman decisiones sobre muchas pruebas independientes (generalmente involucrando muestras «cuestionadas» y una o más muestras «conocidas») y en los que se determinan las tasas de error. Sin estimaciones apropiadas de la precisión, una declaración de un perito que consista en decir que dos muestras son similares —o, incluso, indistinguibles— carece de significado científico: no tiene valor probatorio, y sí un considerable potencial de producir un impacto perjudicial. Nada —ni la experiencia personal ni las prácticas profesionales— puede sustituir una adecuada demostración empírica de la precisión.

En los hallazgos 2-7 se describen las aplicaciones a métodos de comparación de características específicos. El conjunto completo de hallazgos científicos se recoge en la sección 10.

Por último, resaltamos que la Corte Suprema en *Daubert* sugirió que los jueces deberían tener en cuenta la Regla 706, que permite a un tribunal, bajo su propia discreción, obtener la asistencia de un experto de su elección<sup>386</sup>. Tales expertos pueden proporcionar valoraciones independientes concernientes, entre otras cosas, a la validez de los métodos científicos y sus aplicaciones.

## 9.2. *El papel de precedentes*

Un tema importante que surgió en nuestras deliberaciones fue el del papel de los precedentes.

Como es explico en la sección 5, nuestra revisión científica encontró que la mayoría de los métodos forenses de comparación de características (con la notable excepción del análisis de ADN de muestras procedentes de una sola fuente y de mezclas simples) históricamente se han *asumido* antes que *establecido* como válidos en sus fundamentos. Solo después de que se hiciera evidente en los últimos años (basándose en el ADN y en otros análisis) la existencia de problemas en la fundamentación de la fiabilidad de algunos de estos métodos, la comunidad de ciencias forenses comenzó a reconocer la necesidad de probar *empíricamente* si ciertos métodos específicos cumplían los criterios de validez científica.

---

<sup>386</sup> *Daubert*, en 595.

Esto crea una tensión obvia, porque muchos tribunales con fundamento en precedentes que fueron establecidos antes de que estos problemas fundamentales se descubrieran siguen admitiendo métodos de comparación de características.

Desde un punto de vista puramente *científico*, la resolución es clara. Cuando hechos nuevos falsean antiguas suposiciones, los tribunales no deben verse obligados a citar los precedentes del pasado: deben mirar con nuevos ojos las cuestiones científicas. ¿Cómo se resuelven esas tensiones desde el punto de vista jurídico? La Corte Suprema ha dejado claro que un tribunal puede anular un precedente si constata que un caso anterior fue «erróneamente resuelto y que los acontecimientos posteriores han socavado su validez»<sup>387</sup>.

El PCAST no expresa opinión sobre la cuestión jurídica de si algunos de los casos del pasado fueron «erróneamente decididos». Sin embargo, el PCAST advierte que, desde un punto de vista *científico*, acontecimientos posteriores han socavado la validez de las conclusiones que no se basaron en pruebas empíricas apropiadas. Estos acontecimientos incluyen (1) el reconocimiento de problemas sistémicos con algunos métodos forenses de comparación de características, que se muestran mediante los estudios de las causas de cientos de condenas erróneas descubiertas a través del ADN y otros análisis; (2) el informe del NRC (2009) de la Academia Nacional de Ciencias, el principal órgano consultivo científico establecido por el poder legislativo<sup>388</sup>, que encontró que algunos métodos forenses de comparación de características carecen de fundamento científico; y (3) la revisión científica realizada por el PCAST en el presente informe, el principal órgano consultivo científico establecido por el poder ejecutivo<sup>389</sup>, encuentra que algunos métodos forenses de comparación de características carecen de validez en sus fundamentos.

### 9.3. *Medios para los jueces*

Otra cuestión importante que surgió frecuentemente en nuestras conversaciones con analistas fue la necesidad de ofrecer a los jueces mejores medios para la evaluación de los métodos forenses de comparación de características para su uso en los tribunales.

<sup>387</sup> *Boys Markets, Inc. v. Retailers Clerks Union*, 398 U.S. 235, 238 (1970). Véase también: *Patterson v. McLean Credit Union*, 485 U.S. 617, 618 (1988) (resaltando que el Tribunal ha «anulado los precedentes legales en una serie de casos»). El PCAST solicitó asesoramiento sobre esta cuestión a su panel de asesores superiores.

<sup>388</sup> La Academia Nacional de Ciencias fue creada por el Congreso en 1863 para asesorar al gobierno federal en asuntos de ciencia (U.S. Code, Section 36, Title 1503).

<sup>389</sup> El Presidente estableció, formalmente, un consejo científico permanente consultivo poco después del lanzamiento del Sputnik en 1957. Actualmente se denomina Consejo de Asesores de Ciencia y Tecnología del Presidente (que opera bajo la Orden Ejecutiva 13539, modificada por la Orden Ejecutiva 13596).

Los órganos más apropiados para proporcionar esos recursos son la Conferencia Judicial de los Estados Unidos y el Centro Judicial Federal.

La Conferencia Judicial de los Estados Unidos es el órgano nacional de formulación de políticas para los tribunales federales<sup>390</sup>. Su responsabilidad estatutaria incluye el estudio del funcionamiento y el efecto de las normas generales que regulan la práctica y el procedimiento en los tribunales federales. La Conferencia Judicial desarrolla manuales de buenas prácticas y emite notas del Comité Consultivo para ayudar a los jueces respecto a temas específicos, incluso a través de su Comité Consultivo Permanente sobre las Reglas Federales de la Prueba.

El Centro Judicial Federal es la agencia de investigación y educación del sistema judicial federal<sup>391</sup>. Sus deberes estatutarios incluyen (1) dirigir y promover la investigación sobre procedimientos judiciales federales y el funcionamiento de los tribunales y (2) dirigir y promover la formación, la educación y la capacitación continua de los jueces federales, empleados judiciales y otros.

El PCAST recomienda que la Conferencia Judicial de los Estados Unidos, a través de su Subcomité de las Reglas Federales de la Prueba, desarrolle manuales de buenas prácticas y una nota sobre la interpretación adecuada de las reglas relevantes por parte del Comité Consultivo y que el Centro Judicial Federal desarrolle programas educativos relacionados con los procedimientos para evaluar la validez científica de los métodos forenses de comparación de características.

#### 9.4. Recomendación

Sobre la base de sus hallazgos científicos, el PCAST realiza la siguiente recomendación.

#### **RECOMENDACIÓN 8. LA VALIDEZ CIENTÍFICA COMO FUNDAMENTO DEL TESTIMONIO EXPERTO**

(A) Al decidir sobre la admisibilidad de testimonios de analistas, los jueces federales deben tener en cuenta los criterios científicos apropiados para evaluar la validez científica, que incluyen:

- (i) *la validez de los fundamentos*, con respecto al requisito de la Regla 702(c) de que el testimonio sea el producto de principios y métodos fiables; y
- (ii) *la validez en la aplicación*, con respecto al requisito de la Regla 702(d) de que el analista haya aplicado fiablemente los principios y métodos a los hechos del caso.

<sup>390</sup> Creado en 1922 bajo la denominación de Conferencia de Jueces de Circuito Superior, la Conferencia Judicial de los Estados Unidos está actualmente establecida mediante la 28 U.S.C. § 331.

<sup>391</sup> El Centro Judicial Federal fue creado por el Congreso en 1967 (28 U.S.C. §§ 620-629), tras la recomendación de la Conferencia Judicial de los Estados Unidos.

Estos criterios científicos se describen en el Hallazgo 1.

(B) Los jueces federales, cuando permitan que un analista testifique sobre un método de comparación de características válido en sus fundamentos, deben asegurarse de que el testimonio sobre la precisión del método y el valor probatorio de las identificaciones propuestas sea científicamente válido en el sentido de que esté limitado a lo que la evidencia empírica sustente. Las declaraciones que sugieran o impliquen una mayor certeza no son científicamente válidas y no deben permitirse. En particular, los tribunales nunca deben permitir aseveraciones científicamente indefendibles como: tasa de error “cero”, “muy pequeña”, “esencialmente cero”, “despreciable”, “mínima” o “microscópica”; “certeza del 100%” o prueba “hasta un grado razonable de certeza científica”; identificación “hasta la exclusión de todas las demás fuentes posibles”; o una probabilidad de error tan remota como para que sea una “imposibilidad práctica”.

(C) Para ayudar a los jueces, la Conferencia Judicial de los Estados Unidos a través de su Comité Consultivo Permanente sobre las Reglas Federales de la Prueba, debe preparar, con el asesoramiento de la comunidad científica, un manual de buenas prácticas y una nota sobre la interpretación adecuada del Comité Consultivo, proporcionando orientación a los jueces federales sobre la admisibilidad bajo la Regla 702 de testimonios expertos basados en métodos forenses de comparación de características.

(D) Para ayudar a los jueces, el Centro Judicial Federal debe desarrollar programas sobre criterios científicos para la validez científica de los métodos forenses de comparación de características.

## 10. Hallazgos científicos

Los hallazgos científicos del PCAST en este informe se recogen a continuación. El Hallazgo 1, que se refiere a los criterios científicos para la validez científica, se fundamenta en las explicaciones de la sección 4. Los hallazgos 2-6, sobre la validez de los fundamentos de 6 métodos forenses de comparación de características, se basan en las evaluaciones de la sección 5.

### **HALLAZGO 1: CRITERIOS CIENTÍFICOS PARA LA VALIDEZ CIENTÍFICA DE UN MÉTODO DE COMPARACIÓN DE CARACTERÍSTICAS FORENSE**

(1) *Validez de los fundamentos.* Para establecer la validez de los fundamentos de un método de comparación de características, se requieren los siguientes elementos:

(a) un procedimiento reproducible y consistente para (i) la identificación de las propiedades en muestras que son ofrecidas como prueba; (ii) la comparación de las propiedades entre dos muestras; (iii) la determinación, basándose en la similitud de las

propiedades entre dos conjuntos de características, sobre si se debe declarar que las muestras proceden probablemente del mismo origen (“regla de coincidencia”); y

(b) estimaciones empíricas, tras estudios apropiadamente diseñados a partir de múltiples grupos, que establezcan: (i) la tasa de falsos positivos del método —es decir, la probabilidad de una propuesta de identificación entre muestras que realmente proceden de diferentes fuentes— y (ii) la sensibilidad del método —es decir, la probabilidad de una propuesta de identificación entre muestras que realmente proceden de la misma fuente—.

Como se describió en el Cuadro 4, la validación científica debe satisfacer una serie de criterios: (a) debe basarse en conjuntos suficientemente grandes de muestras conocidas y representativas de poblaciones relevantes; (b) debe conducirse de tal modo que los analistas no tengan información sobre las respuestas correctas; el diseño del estudio y el plan de análisis deben especificarse por adelantado y no modificarse posteriormente después de alcanzar los resultados; (d) el estudio debe ser conducido o supervisado por individuos u organizaciones sin beneficio en los resultados; (e) los datos, el software y los resultados deben estar disponibles para permitir que otros científicos puedan revisar las conclusiones; y (f) para asegurar que los resultados sean robustos y reproducibles, debe haber diversos estudios independientes realizados por grupos separados que alcance resultados similares.

Una vez que un método ha sido establecido como válido en sus fundamentos mediante adecuados estudios empíricos, las afirmaciones sobre la precisión y el valor probatorio de las propuestas de identificación, para que sean válidas, deben basarse sobre tales estudios empíricos.

Para métodos objetivos, la validez de los fundamentos puede establecerse demostrando la fiabilidad de cada uno de los pasos individuales (identificación de características, comparación de propiedades, regla de coincidencia, probabilidad de coincidencia falsa y sensibilidad).

Para métodos subjetivos, la validez de los fundamentos *solo* puede establecerse mediante estudios de caja negra que midan la frecuencia en que muchos examinadores alcancen conclusiones precisas en muchos problemas de comparación de características que contengan muestras representativas del uso pretendido. En ausencia de tales estudios, un método subjetivo de comparación de características no puede considerarse científicamente válido.

La validez de los fundamentos es un *sine qua non*, que solo puede demostrarse mediante estudios empíricos. Es muy importante resaltar que las buenas prácticas profesionales —tales como la existencia de sociedades profesionales, programas de certificación, programas de acreditación, artículos revisados por pares, protocolos estandarizados, pruebas de aptitud y códigos éticos— no pueden sustituir a la evidencia empírica de la validez y fiabilidad científica.

(2) *Validez en la aplicación.* Una vez que un método de comparación de características forense ha sido considerado como válido en sus fundamentos, es necesario establecer su validez en la aplicación en un caso dado.

Como se describió en el Cuadro 5, la validez en la aplicación requiere que: (a) el examinador forense debe haber demostrado que es *capaz* de aplicar el método fielmente, lo que se hace sometiéndose a tests de aptitud apropiados (véase Sección 4.6) y debe

*de hecho* haberlo llevado a cabo, como se ha demostrado en el procedimiento *de hecho* utilizado en el caso, los resultados obtenidos y las notas de laboratorio, que deben estar disponibles para que otros puedan revisarlas científicamente; y (b) las aserciones sobre el valor probatorio de las propuestas de identificación deben ser científicamente válidas —incluyendo que los examinadores deben informar sobre la tasa completa de falsos positivos y la sensibilidad del método establecidas en los estudios de validez de los fundamentos—; demostrar que las muestras utilizadas en los estudios de los fundamentos son relevantes para los hechos del caso; y cuando sea aplicable, demostrar el valor probatorio de la coincidencia observada a partir de las características específicas observadas en el caso; además de no realizar afirmaciones o implicaciones que vayan más allá de la evidencia empírica.

### **HALLAZGO 2: ANÁLISIS DE ADN**

*Validez de los fundamentos.* El PCAST encuentra que el análisis de muestras procedentes de una única fuente o de mezclas simples de dos individuos, tales como las que dan en casos de violación, es un método objetivo que ha sido establecido como válido en sus fundamentos.

*Validez en la aplicación.* Como los errores por fallos humanos son dominantes en la probabilidad de coincidencias aleatorias, los criterios científicos para la validez en la aplicación requieren que un analista (1) deba haber realizado una prueba de aptitud rigurosa y relevante para demostrar su capacidad para aplicar fiablemente el método, (2) deba revelar rutinariamente en informes y testimonios si cuando llevó a cabo el examen era consciente de algunos hechos del caso que podrían influenciar en la conclusión y (3) deba revelar, previa solicitud, toda la información sobre las pruebas de calidad y las incidencias de calidad en su laboratorio.

### **HALLAZGO 3: ANÁLISIS DE ADN DE MUESTRAS DE MEZCLAS COMPLEJAS**

*Validez de los fundamentos.* El PCAST encuentra que:

(1) Métodos basados en la Probabilidad Combinada de Inclusión (CPI). El análisis de ADN de mezclas complejas basado en aproximaciones fundamentadas en la CPI ha sido un método inadecuadamente especificado y subjetivo que tiene potencial para conducir a resultados erróneos. Como tal, no es válido en sus fundamentos.

Un artículo reciente ha propuesto normas específicas que abordan una serie de problemas con el uso de la CPI. Esas normas son claramente *necesarias*. Sin embargo, el PCAST no ha tenido tiempo necesario para valorar si son también suficientes para definir un método científicamente válido y objetivo. Si, por un tiempo limitado, los tribunales deciden admitir resultados basados en la aplicación de la CPI, la validez de su

aplicación requeriría que como mínimo fuera consistente con las normas especificadas en el artículo.

El análisis de ADN de mezclas complejas debe dirigirse sin demora hacia métodos más apropiados basados en genotipado probabilístico.

(2) Genotipado probabilístico. El análisis objetivo de mezclas complejas de ADN con software de genotipado probabilístico es un enfoque relativamente nuevo y prometedor. Se requiere evidencia empírica para establecer la validez de los fundamentos de cada método dentro de rangos específicos. Hasta el momento, hay evidencia publicada que apoya la validez de los fundamentos del análisis, con algunos programas, de mezclas de ADN de tres individuos, en las que el menor contribuyente constituye, al menos, el 20% del ADN intacto en la mezcla, y en las que la cantidad de ADN excede del mínimo requerido por el método. El rango en el que ha quedado establecida la validez de los fundamentos es probable que crezca en la medida en que se obtenga evidencia adecuada para mezclas más complejas y se publique.

*Validez en la aplicación.* Para métodos que son válidos en sus fundamentos, la validez en la aplicación conlleva consideraciones similares a las realizadas para análisis de ADN procedente de una única fuente y para mezclas simples, con un énfasis especial en asegurar que el método se aplique correctamente y dentro de su rango empíricamente establecido.

#### **HALLAZGO 4: ANÁLISIS DE MARCAS DE MORDEDURA**

*Validez de los fundamentos.* El PCAST encuentra que el análisis de marcas de mordedura no cumple los estándares científicos para la validez de los fundamentos y está lejos de cumplir tales estándares. Al contrario, la evidencia científica disponible sugiere fuertemente que los analistas no pueden estar de acuerdo consistentemente sobre si una herida es de mordedura humana y no pueden identificar la fuente de una marca de mordedura con precisión razonable.

#### **HALLAZGO 5: ANÁLISIS DE HUELLAS DACTILARES LATENTES**

*Validez de los fundamentos.* Basándose principalmente en dos estudios recientes de caja negra apropiadamente diseñados, el PCAST encuentra que el análisis de huellas dactilares latentes es una metodología subjetiva válida en sus fundamentos —aunque con una tasa de falsos positivos que es sustancial y probablemente más alta que la esperada por muchos miembros del jurado que está basada en afirmaciones largamente mantenidas en el tiempo sobre la infalibilidad de los análisis de huellas dactilares.

Las conclusiones de una propuesta de identificación pueden ser científicamente válidas con tal que estén acompañadas de información precisa sobre las limitaciones de

la fiabilidad de la conclusión —específicamente, que (1) solo se han llevado a cabo dos estudios apropiadamente diseñados sobre la validez de los fundamentos y la precisión de los análisis de huellas latentes; (2) estos estudios encontraron tasas de falsos positivos que podrían ser tan altas como de 1 error cada 306 casos, en uno de ellos, y de 1 error cada 18 en el otro; y (3) como consecuencia de que los analistas eran conscientes de que estaban siendo evaluados, la tasa real de falsos positivos en casos reales puede ser más alta. Hasta el presente, afirmaciones sobre una mayor precisión no están garantizadas o científicamente justificadas. Se requieren estudios de caja negra adicionales para aclarar la fiabilidad del método.

*Validez en la aplicación.* Aunque concluimos que el método es válido en sus fundamentos, hay una serie de cuestiones importantes relacionadas con la validez en su aplicación.

(1) Sesgo de confirmación. El trabajo realizado por científicos del FBI ha mostrado que los analistas suelen alterar las características que marcan inicialmente en una huella latente basándose en las comparaciones realizadas con una huella indubitada aparentemente coincidente. Ese razonamiento circular introduce un serio riesgo de sesgo de confirmación. A los analistas se les debe requerir que completen y documenten sus análisis de una huella latente *antes* de mirar a una huella conocida y deben documentar, separadamente, cualquier dato adicional utilizado durante su comparación y evaluación.

(2) Sesgo contextual. El trabajo realizado por miembros de la comunidad académica ha demostrado que el juicio de los analistas puede estar influenciado por información irrelevante sobre los hechos de un caso. Deben realizarse esfuerzos para asegurar que los analistas no están expuestos a información potencialmente sesgadora.

(3) Pruebas de aptitud. La prueba de aptitud es esencial para valorar la capacidad de un analista y su rendimiento en la emisión de juicios precisos. Como se trató sobre el particular en otro lugar de este informe, las pruebas de aptitud necesitan mejorarse haciéndolas más rigurosas, incorporándolas dentro del flujo de trabajo, y divulgando las pruebas para que ellas sean evaluadas por la comunidad científica.

#### **HALLAZGO 6: ANÁLISIS DE ARMAS DE FUEGO**

*Validez de los fundamentos.* El PCAST encuentra que el análisis de armas de fuego no cumple en la actualidad los criterios para la validez de los fundamentos, porque solo hay un único estudio apropiadamente diseñado para medir la validez y estimar la fiabilidad. Los criterios científicos para la validez de los fundamentos requieren más de un estudio de ese tipo para demostrar la reproducibilidad.

Es competencia de los tribunales decidir si el análisis de armas de fuego debe considerarse admisible basándose en la evidencia actual.

Si el análisis de armas de fuego se permite en los tribunales, los criterios científicos para la validez de su aplicación requieren que se informe claramente sobre las tasas

de error vistas en estudios de caja negra apropiadamente diseñados (estimadas en 1 error cada 66 comparaciones, con un límite superior del 95% de confianza que se corresponde con 1 error cada 46 comparaciones, en el único estudio realizado hasta la fecha).

*Validez en la aplicación.* Si el análisis de armas de fuego es permitido en los tribunales, la validez de su aplicación requiere, desde el punto de vista científico, que el analista:

(1) haya superado una prueba de aptitud rigurosa sobre un gran número de problemas de prueba que permitan evaluar su capacidad y rendimiento y que revele los resultados de las pruebas de aptitud; y

(2) revele si, cuando llevó a cabo la examinación, era consciente de algunos otros hechos del caso que pudieran influir en la conclusión.

#### **HALLAZGO 7: ANÁLISIS DE HUELLAS DE CALZADO**

*Validez de los fundamentos.* El PCAST encuentra que no existen estudios empíricos apropiados que sostengan la validez de los fundamentos de los análisis de huellas de calzado para asociar impresiones de esas huellas con un calzado en particular basándose en marcas identificativas específicas (en ocasiones denominadas “características adquiridas aleatoriamente”). Tales conclusiones no están respaldadas por ninguna evidencia significativa o estimaciones de su precisión y, por ello, no son científicamente válidas.

El PCAST no ha evaluado la validez de los fundamentos de los análisis de las huellas de calzado para identificar características de clase (por ejemplo, la talla del calzado o la marca).

## APÉNDICE A: CUESTIONES ESTADÍSTICAS

Para mejorar su accesibilidad a un público amplio, el texto principal de este informe evita, en la medida de lo posible, el uso de terminología matemática y estadística. Sin embargo, para la implementación efectiva de alguno de los principios reflejados en el informe, se necesitan descripciones algo más precisas. Este apéndice resume los conceptos relevantes desde la estadística básica<sup>392</sup>.

<sup>392</sup> Véanse, por ejemplo: AMITAGE, BERRY y MATTHEWS, 2002; SNEDECOR y COCHRAN, 1989; VAN BELLE, FISHER, HEAGERTY, ET AL., 2004; AGRESTI y COULL, 1998: 119-126; HOGG, TANIS y ZIMMERMAN, 2015; FREEDMAN, PISANI y PURVES, 2007; MOSES, 1986; MOORE, McCABE y CRAIG, 2009.

## 1. Sensibilidad y tasa de falsos positivos

Los métodos forenses de comparación de características generalmente tratan de determinar con qué probabilidad dos muestras proceden de una misma fuente, dado el resultado de una prueba forense sobre las muestras. Se consideran dos posibilidades: la hipótesis nula ( $H_0$ ) consistente en que las muestras proceden de fuentes diferentes y la hipótesis alternativa ( $H_1$ ) de que proceden de una misma fuente. El resultado de la prueba puede resumirse como una declaración de coincidencia (M) o de no coincidencia (O).

Existen dos caracterizaciones necesarias de la precisión de un método: sensibilidad (abreviado por SEN) y la tasa de falsos positivos (FPR).

La sensibilidad se define como la probabilidad de que el método declare una coincidencia entre dos muestras cuando se conoce que proceden de una misma fuente (extraída de una población apropiada), es decir,  $SEN = P(M|H_1)$ . Por ejemplo, un valor de  $SEN = 0.95$  indicaría que las dos muestras procedentes de una misma fuente serían declaradas como coincidentes el 95% de las veces. En la literatura estadística, SEN es a veces llamada «tasa de verdaderos positivos», «TPR» o «el complementario del error tipo II»<sup>393</sup>.

La tasa de falsos positivos (abreviado FPR) se define como la probabilidad de que el método declare una coincidencia entre dos muestras procedentes de fuentes diferentes (de nuevo, en una población apropiada), es decir,  $FPR = P(M|H_0)$ . Por ejemplo, un valor de  $FPR = 0.01$  indicaría que dos muestras procedentes de fuentes diferentes serían (erróneamente) denominadas coincidentes un 1% de las veces<sup>394</sup>. Los métodos con alta FPR no son científicamente fiables como para emitir juicios importantes sobre la fuente de una muestra ante un tribunal. Para ser considerados fiables, el FPR debería ser menor al 5% y pudiera ser apropiado que sea considerablemente menor dependiendo de la finalidad de la aplicación.

Los resultados de un estudio empírico dado pueden resumirse en cuatro valores: el número de ocurrencias en el estudio de verdaderos positivos (TP), de falsos positivos (FP), de falsos negativos (FN) y de verdaderos negativos (TN). (La matriz de estos valores se conoce, curiosamente, como la «matriz de confusión»).

---

<sup>393</sup> El término «tasa de falso negativo» se utiliza, a veces, como el complementario de SEN, es decir,  $FNR = 1 - SEN$ .

<sup>394</sup> Los estadísticos pueden referirse a la especificidad de un método (SPC) en vez de a su tasa de falsos positivos (FPR). Los dos están relacionados mediante la fórmula  $FPR = 1 - SPC$ . En el ejemplo dado,  $FPR = 0.01$  (1%) y  $SPC = 0.99$  (99%).

	Resultado de la prueba	
	Coincidencia	No coincidencia
H1: verdadera procedencia de la misma fuente	TP	FN
H0: verdadera procedencia de fuentes diferentes	FP	TN

En esta terminología estándar, aunque confusa, «verdadero» y «falso» se refieren al acuerdo o no con la verdad (ya sea H0 o H1), mientras que «positivo» y «negativo» se refieren a los resultados de la prueba (es decir, resultan M y O, respectivamente).

Una estimación ampliamente utilizada de SEN, denominada estimación de *máxima verosimilitud*, viene dado por  $TP/(TP+FN)$ , la fracción de sucesos H1 verdaderos (misma fuente) que se declaran correctamente como coincidentes M (match).

Puesto que la tasa de falsos positivos será frecuentemente el factor matemático determinante para el valor probatorio de un método en un caso particular (se explica más adelante), resulta particularmente importante que FPR sea medido empíricamente.

Además, las pruebas con muy baja sensibilidad deben ponerse bajo sospecha porque los escasos resultados positivos de las pruebas pueden ser coincidentes o ser superados por la ocurrencia de falsos positivos<sup>395</sup>.

## 2. Intervalos de confianza

Como se explicó en el texto principal, para ser válidas las medidas empíricas de SEN y FPR han de basarse en grandes colecciones de muestras conocidas y representativas de cada población relevante, de manera que reflejen la frecuencia con la que ocurre una determinada característica o una combinación de características. (En el texto principal también se explican otros requisitos para la validez).

Como las medidas empíricas se basan en un limitado número de muestras, SEN y FPR no pueden ser medidos con exactitud, sino solamente estimados. Dado los tamaños finitos de las muestras, las estimaciones de máxima verosimilitud no cuentan la historia completa. Por tanto, es necesario y apropiado citar los límites de confianza dentro de los cuales es muy probable que se encuentren SEN y FPR.

Debido a que uno debería estar preocupado principalmente de sobreestimar SEN o de subestimar FPR, resulta apropiado utilizar un límite de confianza *unilateral*. Por convención, un nivel de confianza del 95% es el que más se utiliza —lo que sig-

<sup>395</sup> El argumento a favor de una prueba que dice algo como «esta prueba tiene éxito solo ocasionalmente, pero en este caso tuvo éxito» es por tanto falaz.

nifica que hay un 5% de probabilidad de que el valor verdadero supere el límite—. Se deben utilizar, por tanto, límites de confianza unilaterales superiores al 95% para estimar las tasas de error y las cantidades asociadas que caracterizan los métodos forenses de comparación de características (El uso de valores más bajos puede verse, con razón, como un intento de ofuscación).

El límite de confianza para proporciones depende del tamaño muestral en el estudio empírico. Cuando es pequeño, las estimaciones pueden quedar lejos del valor verdadero. Por ejemplo, si un estudio empírico no encontró falsos positivos en 25 pruebas individuales existe todavía una probabilidad razonable (de menos del 5%) de que la verdadera tasa de error pueda llegar a ser tan alta como de 1 entre 9, aproximadamente.

Por razones técnicas, no existe un método único, universalmente aceptado, para calcular estos intervalos de confianza (un problema conocido como «intervalo de confianza de la proporción binomial»). Sin embargo, los diversos métodos generalmente dan resultados muy similares y todos deben considerarse aceptables: el método Clopper-Pearson/Binomial exacto, el intervalo de puntuación Wilson, el intervalo Agresti-Coull (Wald ajustado) y el intervalo de Jeffreys<sup>396</sup>. De todos estos métodos hay calculadoras disponibles en Internet<sup>397</sup>. Por ejemplo, si un estudio encuentra cero falsos positivos en 100 pruebas, los cuatro métodos mencionados dan, respectivamente, los valores 0.030, 0.026, 0.032 y 0.019 para el límite superior de confianza del 95%. Desde el punto de vista científico, cualquiera de ellos podría ser apropiadamente referido en un informe dirigido a un jurado significando que «la tasa de falsos positivos podría ser tan alta como» (En este informe utilizamos el método Clopper-Pearson/Binomial exacto).

### 3. Calculando resultados para pruebas concluyentes

En muchas pruebas forenses, los peritos pueden alcanzar una conclusión (por ejemplo, se ha producido una coincidencia o no se ha producido) o declarar que la prueba es inconclusa. Por tanto, SEN y FPR pueden calcularse sobre la base de pruebas *concluyentes* o sobre *todas* las pruebas realizadas. Aunque ambas tasas tienen interés desde el punto de vista científico, la primera ha de utilizarse para informar de FPR al jurado. Esto resulta apropiado porque las pruebas utilizadas contra un acusado se basarán generalmente en pruebas *concluyentes* y no en pruebas inconclusas. Para ilustrar este punto, consideremos el caso extremo en el que un método ha sido probado 1000 veces y se encontró 990 resultados inconclusos, 10 falsos positivos y ningún resultado correcto. Sería engañoso informar que la tasa de falsos positivos

<sup>396</sup> BROWN, 2001: 101-133.

<sup>397</sup> Por ejemplo, véase: [epitools.ausvet.com.au/content.php?page=CIPProportion](http://epitools.ausvet.com.au/content.php?page=CIPProportion).

fue del 1% (10 sobre 1000 pruebas). Por el contrario, se debería informar de que el 100% de los resultados concluyentes fueron falsos positivos (10 sobre 10 pruebas).

#### 4. Análisis bayesiano

En este apéndice, hemos centrado la atención en la sensibilidad y en la tasa de falsos positivos ( $SEN = P(M|H1)$  y  $FPR = P(M|H0)$ ). La cantidad de mayor interés en un juicio penal es  $P(H1|M)$ , es decir, «la probabilidad de que las muestras procedan de la misma fuente *dado* que se ha declarado la coincidencia»\*. Esta cantidad es frecuentemente denominada *valor predictivo positivo* (PPV) de la prueba.

El cálculo de PPV depende de dos cantidades: del «factor de Bayes»  $BF = SEN/FPR$  y de una segunda cantidad denominada «apuesta *a priori*» (POR). Esta última cantidad se define, matemáticamente, como  $POR = P(H0)/P(H1)$ , donde  $P(H0)$  y  $P(H1)$  son las probabilidades *a priori* (es decir, antes de realizar la prueba) de las hipótesis  $H0$  y  $H1$ <sup>398</sup>. La fórmula de PPV en términos de BF y POR es:  $PPV = BF / (BF + POR)$ , una fórmula que se deriva del principio estadístico conocido como el Teorema de Bayes<sup>399</sup>.

El Teorema de Bayes ofrece una forma matemática de combinar el resultado de la prueba con información independiente —tales como (1) la propia probabilidad *a priori* de que dos muestras procedan de una misma fuente y (2) el número de muestras buscadas—. Algunos estadísticos bayesianos elegirían un *a priori*  $POR = 1$  en el supuesto de una coincidencia con una única muestra (implicando que es igualmente probable *a priori* que las muestras proceden de una misma fuente que de fuentes diferentes) y una apuesta *a priori*  $(POR) = 100.000$  para una coincidencia tras una búsqueda en una base de datos de 100.000 registros. Otros establecerían  $POR = (1-p)/p$ , donde  $p$  es la probabilidad *a priori* de la identidad de la misma fuente en la población relevante, dados los otros hechos del caso.

El enfoque bayesiano es matemáticamente elegante. Sin embargo, plantea retos para su uso en los tribunales: (1) diferentes personas pueden tener diferentes creencias sobre POR y (2) muchos miembros del jurado podrían no entender cómo las creencias sobre POR afectan al cálculo matemático de PPV. (Además, como se ha señalado anteriormente, las estimaciones empíricas de SEN y FPR son inciertas, por lo que el factor de Bayes estimado,  $BF = SEN/FPR$ , también lo es).

\* N. del T.: la coincidencia se expresa con el término «match» en inglés en la literatura especializada.

<sup>398</sup> Es decir, si  $p$  es la probabilidad *a priori* de la identidad de la misma fuente en la población bajo examen, entonces  $POR = (1-p)/p$ .

<sup>399</sup> En el texto principal, la frase «apropiadamente correcto para el tamaño del grupo que se exploró para la identificación de un sospechoso» se refiere al uso de esta fórmula con un apropiado valor de POR.

Por tanto, algunos analistas, por consiguiente, prefieren simplemente informar de las cantidades medidas empíricamente (la sensibilidad, la tasa de falsos positivos de la prueba y la probabilidad de que una coincidencia sea falso positivo dado el número de muestras contra las que la muestra buscada ha sido comparada), permitiendo al jurado que las incorpore a sus propios juicios bayesianos intuitivos. (Por ejemplo, «*Sí, la prueba tiene una tasa de falsos positivos de solo 1 cada 100, pero dos testigos ubican al acusado a 1.000 millas de la escena del crimen, por lo que el resultado de la prueba fue probablemente uno de esos 1 de cada 100 falsos positivos*»).

## APÉNDICE B: EXPERTOS ADICIONALES QUE REALIZAN APORTACIONES

El PCAST buscó aportaciones de un grupo adicional de expertos y sectores interesados. El PCAST expresa su gratitud a los que aquí se listan por compartir su experiencia. No tuvieron la oportunidad de revisar los borradores del informe, y su consentimiento para participar con el PCAST en puntos específicos no implica que respalden los puntos de vista expresados en este informe. La responsabilidad de las opiniones, hallazgos y recomendaciones en este informe, así como cualquier error de hecho o de interpretación, descansa únicamente sobre el PCAST.

Richard Alpert  
Asistente del Fiscal de Distrito Penal Tarrant  
Oficina del Fiscal de Distrito Criminal del Condado

Kareem Belt  
Analista de Política Forense  
Proyecto Inocencia

William Bodziak  
Consultor  
Bodziak Forensics

John Buckleton  
Director Científico  
Instituto de Medio Ambiente e Investigación Científica  
Nueva Zelanda

Bruce Budowle  
Catedrático, Director Ejecutivo del Instituto de Genética Aplicada  
Centro de Ciencias de la Salud de la Universidad del Norte de Texas

Mary A. Bush  
Profesora Asociada  
Departamento de Odontología Restaurativa  
Facultad de Medicina Dental de la Universidad de Búfalo

Peter Bush

Profesor de Investigación  
Director del Centro de Instrumentos del Campus Sur  
Facultad de Medicina Dental de la Universidad de Búfalo

John Butler

Adjunto Especial del Director de Ciencias Forenses  
Oficina de Programas Especiales  
Instituto Nacional de Estándares y Tecnología

Arturo Casadevall

Catedrático  
Departamento de Microbiología e Inmunología y Departamento de Medicina  
Facultad de Medicina Albert Einstein

Alicia Carriquiry

Catedrática Distinguida en el Estado de Iowa y Directora  
Centro de Estadística y Aplicaciones en Pruebas Forenses  
Universidad del Estado de Iowa

Richard Cavanagh

Director  
Oficina de Programas Especiales  
Instituto Nacional de Estándares y Tecnología

Eleanor Celeste

Analista Política  
Ciencias Médicas y Forenses  
Oficina de Política Científica y Tecnológica

Christophe Champod

Catedrático de Derecho, Ciencias Penales y Administración Pública  
Universidad de Lausana

Sarah Chu

Abogada Senior de Política Forense  
Proyecto Inocencia

Simon A. Cole

Catedrático de Criminología, Derecho y Sociedad  
Facultad de Ecología Social  
Universidad de California en Irvine

Kelsey Cook

Director de Programa  
Medición e Imagen Químicas  
National Science Foundation

Patricia Cummings  
Jefe de la Oficina de Sectores Especiales  
Oficina del Fiscal de Distrito del Condado de Dallas

Christopher Czyryca  
Presidente  
Collaborative Testing Services

Dana Delger  
Abogado  
Proyecto Inocencia

Shari Diamond  
Cátedra de Derecho Howard J. Trienens  
Catedrático de Psicología  
Facultad de Derecho Pritzker  
Northwestern University

Itiel Dror  
Investigador Senior en Neurociencia Cognitiva  
University College London

Meredith Drosback  
Directora Adjunta  
Ciencias de la Educación y de la Física  
Oficina de Política en Ciencia y Tecnología

Kimberley Edwards  
Científico en Física  
Analista Forense  
Laboratorio del FBI

Ian Evett  
Estadístico Forense  
Principal Forensic Services

Chris Fabricant  
Director, Litigación Estratégica  
Proyecto Inocencia

Kenneth Feinberg  
Profesora Visitante Steven and Mauren Klinsky de Prácticas en Liderazgo y Progreso  
Facultad de Derecho de Harvard

Rebecca Ferrell  
Director de Programa  
Antropología Biológica  
National Science Foundation

Jennifer Friedman  
Coordinadora de Ciencia Forense  
Defensora Pública del Condado de Los Ángeles

Lynn Garcia  
Departamento Jurídico  
Comisión de Ciencia Forense de Texas

Daniel Garner  
Director Ejecutivo y Presidente  
Centro de Ciencia Forense de Houston

Constantine A. Gatsonis  
Cátedra Henry Ledyard Goddard de Bioestadística  
Catedrático de Bioestadística  
Director del Centro de Ciencias Estadísticas  
Brown University

Eric Gilkerson  
Analista Forense  
Laboratorio del FBI

Brandon Giroux  
Presidente  
Giroux Forensics, L.C.C.  
Presidente  
Forensic Assurance

Catherine Grgicak  
Profesora Asociada  
Anatomía y Neurobiología  
Facultad de Medicina de la Universidad de Boston

Austin Hicklin  
Miembro  
Fellow

Cindy Homer  
Científico Forense  
Laboratorio de Criminalística de la Policía del Estado de Maine

Alice Isenberg  
Subdirectora Adjunta  
Laboratorio del FBI

Matt Johnson  
Especialista Forense Senior  
Departamento del Sheriff del Condado de Orange

Jonathan Koehler  
Cátedra de Derecho Beatrice Kuhn  
Facultad de Derecho Pritzker  
Northwestern University

Glenn Langenburg  
Supervisor de Ciencia Forense  
Oficina de Detención Penal de Minnesota

Gerald LaPorte  
Director  
Oficina de Ciencias de la Investigación y Forenses  
Instituto Nacional de Justicia

Julia Leighton  
Departamento Jurídico  
Servicio de Defensor Público  
Distrito de Columbia

Alan I. Leshner  
Director Ejecutivo, Emérito  
Asociación (Norte)americana para el Avance de la Ciencia y  
Editor Ejecutivo de la revista *Science*

Ryan Lilien  
Director Científico  
Cadre Research Labs

Elizabeth Mansfield  
Subdirectora  
Medicina Personalizada  
Food and Drug Administration

Anne-Marie Mazza  
Directora  
Comité sobre Ciencia, Tecnología y Derecho  
Academias Nacionales de Ciencias, Ingeniería y Medicina

Willie E. May  
Director  
Instituto Nacional de Estándares y Tecnología

Daniel MacArthur  
Profesor Adjunto  
Facultad de Medicina de Harvard  
Co-Director de Genética Médica y Poblacional  
Instituto Broad de Harvard y del MIT

Brian McVicker  
Analista Forense  
Laboratorio del FBI

Stephen Mercer  
Director  
Grupo de Apoyo en la Litigación  
Oficina del Defensor Público  
Estado de Maryland

Melissa Mourges  
Jefe  
Ciencias Forenses/Unidad de Casos Abiertos  
Oficina del Fiscal de Distrito del Condado de Nueva York

Peter Neufeld  
Codirector y Cofundador  
Proyecto Inocencia

Steven O'Dell  
Director  
División de Servicios Forenses  
Departamento de Policía de Baltimore

Lynn Overmann  
Consejero de Política Senior  
Oficina de Política de Ciencia y Tecnología

Skip Palenik  
Fundador  
Microtrace

Matthew Redle  
Fiscal de la Acusación y del Condado  
Oficina del Fiscal del Condado de Sheridan

Maria Antonia Roberts  
Directora de Programa de Investigación  
Unidad de Apoyo a la Impresión de Latentes  
Laboratorio del FBI

Walter F. Rowe  
Catedrático de Ciencias Forenses  
Universidad George Washington

Norah Rudin  
Presidente y CEO  
Grupo de Colaboración Científica, Innovación y Educación

Jeff Salyards

Director

Centro de Ciencias Forenses de la Defensa  
Agencia Forense y Biométrica de la Defensa

Rodney Schenck

Centro de Ciencias Forenses de la Defensa

Antropología Biológica

Agencia Forense y Biométrica de la Defensa

David Senn

Director

Centro para la Educación e Investigación en Ciencias Forenses

Simposio del Suroeste en Odontología Forense

Centro de Ciencias de la Salud en San Antonio de la Universidad de Texas

Stephen Shaw

Analista de Trazas

Laboratorio del FBI

Andrew Smith

Supervisor de la Unidad de Armas de Fuego y Marcas de Herramienta

Departamento de Policía de San Francisco

Erich Smith

Científico en Física

Unidad de Armas de Fuego y Marcas de Herramienta

Laboratorio del FBI

Tasha Smith

Unidad de Armas de Fuego y Marcas de Herramienta

Laboratorio de Criminalística

Departamento de Policía de San Francisco

Jeffrey Snipes

Profesor Adjunto

Estudios de Justicia Penal

Universidad del Estado de San Francisco

Jill Spriggs

Director de Laboratorio

Oficina del Fiscal de Distrito del Condado de Sacramento

Harry Swofford

Jefe, Área de Impresión de Latentes

Centro de Ciencias Forenses de la Defensa

Agencia Forense y Biométrica de la Defensa

Robert Thompson

Director de Programas de Sistemas de Datos Forenses  
Oficina de Estándares para el Cumplimiento de la Ley  
Instituto Nacional de Estándares y Tecnología

William Thompson

Catedrático de Criminología, Derecho y Sociedad, y de Psicología y Comportamiento Social  
Facultad de Derecho de Ecología Social  
Universidad de California en Irvine

Rick Tontarski

Científico Jefe  
Centro de Ciencias Forenses de la Defensa

Jeremy Triplett

Supervisor de Laboratorio  
Laboratorio Forense Central de la Policía del Estado de Kentucky

Richard Vorder Bruegge

Técnico fotográfico senior  
FBI

Victor Weedn

Cátedra de Ciencias Forenses  
Departamento de Ciencias Forenses  
Universidad George Whashington

Robert Wood

Profesor Adjunto y Jefe  
Departamento de Oncología Dental  
Prótesis Dentales, Oculares y Maxilofaciales  
Centro (de investigación) sobre el Cáncer Princess Margaret  
Universidad de Toronto

Xiaoyu Alan Zheng

Ingeniero Mecánico  
Instituto Nacional de Estándares y Tecnología

## BIBLIOGRAFÍA

### 1. Informes

FDA GUIDANCE (2016), *Adaptive Designs for Medical Device Clinical Studies*. Disponible en: [www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm446729.pdf](http://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm446729.pdf).

FORENSIC SCIENCE COMMISSION (2016), *Forensic bite mark comparison complaint filed by National Innocence Project on behalf of Steven Mark Chaney – Final Report*. [www.fsc.texas.gov/sites/default/files/FinalBiteMarkReport.pdf](http://www.fsc.texas.gov/sites/default/files/FinalBiteMarkReport.pdf).

- INNOCENCE PROJECT, *DNA Exonerations in the United States*, [www.innocenceproject.org/dna-exonerations-in-the-united-states](http://www.innocenceproject.org/dna-exonerations-in-the-united-states).
- MINISTERIO DEL INTERIOR ESPAÑOL (2018), *La relevancia del título oficial del perito criminalístico nombrado por el juez en la jurisdicción penal española*
- NATIONAL ACADEMIES OF SCIENCES, ENGINEERING, and Medicine (2015), *Support for Forensic Science Research: Improving the Scientific Role of the National Institute of Justice*. The National Academies Press. Washington DC.
- NATIONAL COMMISSION ON FORENSIC SCIENCE (2015), *Defining forensic science and related terms*. [www.justice.gov/ncfs/file/786571/download](http://www.justice.gov/ncfs/file/786571/download).
- NATIONAL COMMISSION ON FORENSIC SCIENCE (2016), «Recommendations to the Attorney General Regarding Use of the Term ‘Reasonable Scientific Certainty’». [www.justice.gov/ncfs/file/839726/download](http://www.justice.gov/ncfs/file/839726/download).
- NATIONAL INSTITUTE OF JUSTICE (2006), *Status and Needs of Forensic Science Service Providers: A Report to Congress*. [www.ojp.usdoj.gov/nij/pubs-sum/213420.htm](http://www.ojp.usdoj.gov/nij/pubs-sum/213420.htm).
- NATIONAL PHYSICAL LABORATORY (2010), *A Beginner’s Guide to Measurement*. Disponible en: [www.npl.co.uk/upload/pdf/NPL-Beginners-Guide-to-Measurement.pdf](http://www.npl.co.uk/upload/pdf/NPL-Beginners-Guide-to-Measurement.pdf).
- NATIONAL RESEARCH COUNCIL (1996), *The Evaluation of Forensic DNA Evidence*, Washington DC, The National Academies Press.
- NATIONAL RESEARCH COUNCIL (2004), *Forensic Analysis: Weighing Bullet Lead Evidence*, Washington DC, The National Academies Press.
- (2008), *Ballistic Imaging*, Washington DC, The National Academies Press.
- (2009), *Strengthening Forensic Science in the United States: A Path Forward*. The National Academies Press. Washington DC, The National Academies Press.
- (2010), *Biometric Recognition: Challenges and Opportunities*, Washington DC, The National Academies Press.
- ORGANIZATION FOR SCIENTIFIC AREA COMMITTEES (OSAC) (2015), *Research Needs Assessment Form*, «Study to Assess the Accuracy and Reliability of Firearm and Toolmark». Disponible en: [www.nist.gov/forensics/osac/upload/FATM-Research-NeedsAssessment\\_Blackbox.pdf](http://www.nist.gov/forensics/osac/upload/FATM-Research-NeedsAssessment_Blackbox.pdf).
- U.S. DEPARTMENT OF JUSTICE, Office of Justice Programs, National Institute of Justice (1996), *Convicted by Juries, Exonerated by Science: Case Studies in the Use of DNA Evidence to Establish Innocence after Trial*. xxviii.
- U.S. DEPARTMENT OF JUSTICE (2006), *Office of the Inspector General (Oficina del Inspector General). A review of the FBI’s handling of the Brandon Mayfield case*, [oig.justice.special/s0601/final.pdf](http://oig.justice.special/s0601/final.pdf).
- (2015), *New NIST Center of Excellence to Improve Statistical Analysis of Forensic Evidence*, [www.nist.gov/forensics/center-excellence-forensic052615.cfm](http://www.nist.gov/forensics/center-excellence-forensic052615.cfm).
- U.S. DEPARTMENT OF JUSTICE. National Institute of Justice (2016), *Report Forensic Science: Fiscal Year 2015 Funding for DNA Analysis, Capacity Enhancement and Other Forensic Activities*.

## 2. Libros/Artículos

- ALOSH, M., FRITSCH, K., HUQUE, M., MAHJOOB, K., PENNELLO, G., ROTHMANN, M., RUSSEK-COHEN, E., SMITH, F., WILSON, S. y YUE, L., 2015: «Statistical considerations on subgroup analysis in clinical trials», *Statistics in Biopharmaceutical Research*, vol. 7: 286-303.
- ASSOCIATION OF FIREARM AND TOOL MARK EXAMINERS, 2011: «Theory of Identification as it Relates to Tool Marks: Revised», *AFTE Journal*, vol. 43(4).
- BALDING, D.J. y NICHOLS, R.A., 1994: «DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands», *Forensic Science International*, vol. 64: 125-140.
- BIEBER, P., 2014: «Fire investigation and cognitive bias», *Wiley Encyclopedia of Forensic Science*. Disponible en [onlinelibrary.wiley.com/doi/10.1002/9780470061589.fsa1119/abstract](http://onlinelibrary.wiley.com/doi/10.1002/9780470061589.fsa1119/abstract).

- BIEBER, F.R., BUCKLETON, J.S., BUDOWLE, B., BUTLER, J.M. y COBLE M.D., «Evaluation of forensic DNA mixture evidence: protocol for evaluation, interpretation, and statistical calculations using the combined probability of inclusion», *BMC Genetics*. [www.bmcgenet.biomedcentral.com/articles/10.1186/s12863-016-0429-7](http://www.bmcgenet.biomedcentral.com/articles/10.1186/s12863-016-0429-7).
- BOUCHET, C., GUILLEMIN, F. y BRAINCON, S., 1996: «Nonspecific effects in longitudinal studies: impact on quality of life measures», *Journal of Clinical Epidemiology*, vol. 49(1).
- BORODITSKY, L., 2007: «Comparison and the development of knowledge», *Cognition*, vol. 102.
- BODZIAK, W. J., 2000: *Footwear Impression Evidence: Detection, Recovery, and Examination*, 2nd ed., Boca Raton, Florida: CRC Press-Taylor & Francis.
- BUDOWLE, B., BUSCAGLIA, J. y PERLMAN, R.S., 2006: «Review of the scientific basis for friction ridge comparisons as a means of identification: committee findings and recommendations», *Forensic Science Communications*, vol. 8.
- BUDOWLE, B., MORETTI, T.R., KEYS, K.M., KOONS, B.W. y SMERICK, J.B. 1997: «Validation studies of the CTT STR multiplex system», *Journal of Forensic Sciences*, vol. 42, No. 4.
- BUDOWLE, B., MORETTI, T.R., BAUMSTARK, A.L., DEFENBAUGH, D.A. y KEYS, K.M., 1999: «Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians», *Journal of Forensic Sciences*, vol. 44(6).
- BUDOWLE, B., SHEA, B., NIEZGODA, S. y CHAKRABORTY, R., 2001: «CODIS STR loci data from 41 sample populations», *Journal of Forensic Sciences*, vol. 46 (3): 453-489. En julio de 2015 se informó de errores encontrados en la base de datos original. (Erratum, *Journal of Forensic Sciences*, vol. 60, No. 4 (2015): 1114-6, pueden encontrarse en [www.fbi.gov/about-us/lab/biometric-analysis/codis/expanded-fbi-str-final-6-16-15.pdf](http://www.fbi.gov/about-us/lab/biometric-analysis/codis/expanded-fbi-str-final-6-16-15.pdf)).
- BUCKLETON, J.S., CURRAN, J.M. y GILL, P., 2007: «Towards understanding the effect of uncertainty in the number of contributors to DNA stains», *Forensic Science International Genetics*, vol. 1(1).
- BUTLER, J.M., 2015: *Advanced Topics in Forensic DNA Typing: Interpretation*, Waltham, MA: Elsevier/Academic.
- BUTLER, J.M., 2015: «The future of forensic DNA analysis», *Philosophical Transactions of the Royal Society B*, 370: 20140252.
- BUTLER, J.M., 2015: «DNA Error Rates», presentation at the International Forensics Symposium, Washington, D.C. (2015). [www.cstl.nist.gov/strbase/pub\\_pres/Butler-ErrorManagement-DNA-Error.pdf](http://www.cstl.nist.gov/strbase/pub_pres/Butler-ErrorManagement-DNA-Error.pdf).
- BRACHT, G.H. y GLASS, G.V., 1968: «The external validity of experiments», *American Educational Research Journal*, vol. 5(4): 437-474.
- BUSH, M.A., COOPER, H.I. y DORION, R.B., 2010: «Inquiry into the scientific basis for bitemark profiling and arbitrary distortion compensation», *Journal of Forensic Sciences*, vol. 55(4).
- BUSH, M.A., MILLER, R.G., BUSH, P.J. y DORION, R.B., 2009: «Biomechanical factors in human dermal bitemarks in a cadaver model», *Journal of Forensic Sciences*, vol. 54(1).
- CARRORO, P. y PLEBANI M., 2007: «Errors in a stat laboratory: types and frequencies 10 years later», *Clinical Chemistry*, vol. 53.
- CHAMPOD, C., 2014: «Research focused mainly on bias will paralyse forensic science», *Science & Justice*, vol. 54.
- CLAYTON, T.M., WHITAKER, J.P., SPARKES, R. y GILL, P., 1998: «Analysis and interpretation of mixed forensic stains using DNA STR profiling», *Forensic Science International*, vol. 91, No. 1: 55-70.
- COBLE, M.D., BRIGHT, J.A., BUCKLETON, J.S. y CURRAN, J.M., 2015: «Uncertainty in the number of contributors in the proposed new CODIS set», *Forensic Science International Genetics*, vol. 19.
- COLE, S.A., 2004: «Grandfathering evidence: Fingerprint admissibility rulings from Jennings to Llera Plaza and back again», *American Criminal Law Review*, 41(1189).
- COLE, S.A., 2005: «More than zero: Accounting for error in latent fingerprint identification», *Journal of Criminal Law and Criminology*, vol. 95(3): 985-1078.
- CONNAY, L., 2010: «Latent Print Training to Competency: Is it Time for a Universal Training Program?», *Journal of Forensic Identification*, vol. 60: 223-58.

- DROR, I.E., CHARLTON, D. y PERON A.E., 2006: «Contextual information renders experts vulnerable to making erroneous identifications», *Forensic Science International*, vol. 156.
- DROR, I.E. y HAMPIKIAN G., 2011: «Subjectivity and bias in forensic DNA mixture interpretation», *Science & Justice*, Vol. 51(4).
- DROR, I.E., 2016: «A hierarchy of expert performance», *Journal of Applied Research in Memory and Cognition*, Vol. 5: 121-127.
- FAIGMAN, D.L., CHENG, E.K., MNOOKIN, J.L., MURPHY, E.E., SANDER, J. y SLOBOGIN, C. (Eds.), 2016: *Modern Scientific Evidence: The Law and Science of Expert Testimony, 2015-2016*, Thomson/ West Publishing.
- FREGEAU, C.J., BOWEN, K.L. y FOURNEY, R.M., 1999: «Validation of highly polymorphic fluorescent multiplex short tandem repeat systems using two generations of DNA sequencers», *Journal of Forensic Sciences*, vol. 44(1): 133-166.
- GAUDETTE, B.D. y KEEPING, 1975: «An attempt at determining probabilities in human scalp hair comparisons», *Journal of Forensic Sciences*, vol. 19.
- GARRETT, B.L. y NEUFELD, P.J., 2009: «Invalid forensic science testimony and wrongful convictions», *Virginia Law Review*, vol. 91(1): 1-97.
- GIANNELLI, P.C., 1980: «The admissibility of novel scientific evidence: Frye v. United States, a half-century later», *Columbus Law Review*, vol. 80(6).
- GIANELLI, P.C. 2003: «The Supreme Court's 'Criminal' Daubert Cases», *Seton Hall Law Review*, vol. 33(1096).
- GIANNELLI, P.G., 2010: «Independent crime laboratories: The problem of motivational and cognitive bias», *Utah Law Review*, 247-266.
- GILL, P., JEFFREYS, A.J. y WERRETT, D.J., 1985: «Forensic application of DNA 'fingerprints'», *Nature*, vol. 318(6046).
- GOODE, M., 2002: «Some observations on evidence of DNA frequency», *Adelaide Law Review*, vol. 23: 45-77.
- GOLDSTONE, R. L., 1994: «The role of similarity in categorization: Providing a groundwork», *Cognition*, vol. 52: 125-157.
- GROSS S.R., y SHAFFER, M., 2012: «Exonerations in the United States, 1989-2012», National Registry of Exonerations. Disponible en: [www.law.umich.edu/special/exoneration/Documents/exonerations\\_us\\_1989\\_2012\\_full\\_report.pdf](http://www.law.umich.edu/special/exoneration/Documents/exonerations_us_1989_2012_full_report.pdf).
- HASSIN, R., 2001: «Making features similar: comparison processes affect perception», *Psychonomic Bulletin & Review*, vol. 8.
- HOFGARTNER, W.T. y TAIT, J.F., 1999: «Frequency of problems during clinical molecular-genetic testing», *American Journal of Clinical Pathology*, vol. 112.
- HOUCK, M.M., y B. BUDOWLE, 2002: «Correlation of microscopic and mitochondrial DNA hair comparisons», *Journal of Forensic Sciences*, vol. 47(5).
- KASSIN, S.M., DROR I.E. y KAKUCKA, J., 2013: «The forensic confirmation bias: Problems, perspectives, and proposed solutions», *Journal of Applied Research in Memory and Cognition*, vol. 2 (1): 42-52.
- KAYE D.H., 1993: «DNA Evidence: Probability, Population Genetics, and the Courts», *Harv. J.L. & Tech*, vol 7: 101-172.
- KEIJSER, J.W., MALSCH, M., LUINING, E.T., KRANENBARG, M.W. y LENSSEN, D.J.H.M., 2016: «Differential reporting of mixed DNA profiles and its impact on jurists' evaluation of evidence: An international analysis», *Forensic Science International: Genetics*, vol. 23.
- KIM, J., NOVEMSKY, N. y DHAR, R.2012: «Adding small differences can increase similarity and choice», *Psychological Science*, vol. 24.
- KIESER, J.A., BERNAL, V., NEIL WADDELL, J. y S. RAJU, 2007: «The uniqueness of the human anterior dentition: a geometric morphometric analysis», *Journal of Forensic Sciences*, vol. 52.
- KOEHLER, J.J., 2016: «Intuitive error rate estimates for the forensic sciences». Disponible en: [papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2817443](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2817443)
- KOEHLER, J.J., 2016: «Forensics or fauxrensicis? Ascertaining accuracy in the forensic sciences», [papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2773255](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2773255)

- KIMPTON, C.P., OLDROYD, N.J., WATSON, S.K., FRAZIER, R.R., JOHNSON, P.E., MILLICAN, E.S., UR-GUHART, A., SPARKES, B.L. y GILL, P., 1996: «Validation of highly discriminating multiplex short tandem repeat amplification systems for individual identification», *Electrophoresis*, vol. 17(8).
- KLOOSTERMAN, A., SJERPES, M. y QUAK, A., 2014: «Error rates in forensic DNA analysis: Definition, numbers, impact and communication», *Forensic Science International: Genetics*, vol. 12.
- KOPPL, R. y KRANE, D., 2016: «Minimizing and leveraging bias in forensic science.» En ROBERTSON C.T. y A.S. KESSELHEIM (Eds.), *Blinding as a solution to bias: Strengthening biomedical science, forensic science, and law*. Atlanta, GA: Elsevier.
- KRANE, D.E., FORD, S., GILDER, J., IMAN, K., JAMIESON, A., TAYLOR, M.S. y THOMPSON, W.C., 2008: «Sequential unmasking: A means of minimizing observer effects in forensic DNA interpretation», *Journal of Forensic Sciences*, vol. 53(4).
- KRIMSKY, S. y T. SIMONCELLI, 2011: *Genetic Justice: DNA Data Banks, Criminal Investigations, and Civil Liberties*, Columbia University Press.
- LANDER, E.S., 1994: «DNA fingerprinting on trial», *Nature*, vol. 339: 501-505.
- LANDER, E.S. y B. BUDOWLE, 1994: «DNA fingerprinting dispute laid to rest», *Nature*, vol. 371: 735-738.
- LANGLEBEN, D.D. y MORIARTY, J.C., 2013: «Using brain imaging for lie detection: Where science, law, and policy collide», *Psychology, Public Policy, and Law*, vol. 19(2).
- LANGENBURG, G., 2009: «A performance study of the ACE-V process: A pilot study to measure the accuracy, precision, reproducibility, repeatability, and biasability of conclusions resulting from the ACE-V process», *Journal of Forensic Identification*, vol. 59(2).
- LARKEY, L.B., y MARKMAN, A.B., 2005: «Processes of similarity judgment», *Cognitive Science*, vol. 29.
- LYGO, J.E., JOHNSON, P.E., HOLDAWAY, D.J., WOODROFFE, S., WHITAKER, J.P., CLAYTON, T.M., KIMPTON, C.P. y GILL, P.: 1994: «The validation of short tandem repeat (STR) loci for use in forensic casework», *International Journal of Legal Medicine*, vol. 107(2).
- MANGIONE-SMITH, R., ELLIOTT, M.N., McDONALD, L. y MCGLYNN, E.A., 2002: «An observational study of antibiotic prescribing behavior and the Hawthorne Effect», *Health Services Research*, vol. 37(6): 1603-1623.
- MCCABE, J., 1996: «DNA fingerprinting: The failings of Frye», *Northern Illinois University Law Review*, vol. 16: 455-82.
- MCCARNEY, R., WARNER, J., ILIFFE, S., VAN HASELEN, R., GRIFFIN, M. y FISHER, P., 2007: «The Hawthorne Effect: a randomized, controlled trial», *BMC Medical Research Methodology*, vol. 7(30).
- MEDIN, D.L., GOLDSTONE, R.L. y GENTNER, D., 1993: «Respects for similarity», *Psychological Review*, vol. 100.
- MEDIN, D.L., GOLDSTONE, R.L. y MARKMAN, A.B., 1995: «Comparison and choice: Relations between similarity processes and decision processes», *Psychonomic Bulletin and Review*, vol. 2: 1-19.
- MILLER, L.S., 1987: «Procedural bias in forensic examinations of human hair», *Law and Human Behavior*, vol. 11.
- MORETTI, T.R., BAUMSTARK, A.L., DEFENBAUGH, D.A., KEYS, K.M., SMERICK, J.B. y BUDOWLE, B., 2001: «Validation of Short Tandem Repeats (STRs) for forensic usage: performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples», *Journal of Forensic Sciences*, vol. 46(3): 647-660.
- MORRISON, G.S., ZHANG, C. y ROSE, P., 2011: «An empirical estimate of the precision of likelihood ratios from a forensic-voicecomparison system», *Forensic Science International*, vol. 208: 59-65.
- MUJIS, D., 2006: «Measuring teacher effectiveness: Some methodological reflections», *Educational Research and Evaluation*, vol. 12(1): 53-74.
- NOSOFSKY, R. M., 1986: «Attention, similarity, and the identification categorization relation», *Journal of Experimental Psychology*, General, vol. 115: 39-57.
- MNOOKIN, J.L., COLE, S.A., DROR, I.E., FISHER, B.A.J., HOUCK, M.M., INMAN, K., KAYE, D.H., KOEHLER, J.J., LANGENBURG, G., RISINGER, D.M., RUDIN, N., SIEGEL, J. y STONEY D.A., 2011: «The need for a research culture in the forensic sciences», *UCLA Law Review*, vol. 725.

- NEUFELD, P.J. y COLMAN, N., 1991: «When science takes the witness stand», *Scientific American*, vol. 262: 46-53.
- PAGE, M., TAYLOR, J. y BLENKIN, M., 2011: «Forensic identification science evidence since Daubert: Part II—judicial reasoning in decisions to exclude forensic identification evidence on grounds of reliability», *Journal of Forensic Sciences*, vol. 56(4).
- PAVESE, F., 2009: «An Introduction to Data Modelling Principles in Metrology and Testing», en PAVESE, F. y FORBES A.B. (Eds.), *Data Modelling for Metrology and Testing in Measurement Science*, Birkhäuser.
- PERLIN, M.W., HORNYAK, J.M., SUGIMOTO, G. y MILLER, K.W.P., 2015: «TrueAllele genotype identification on DNA mixtures containing up to five unknown contributors», *Journal of Forensic Sciences*, vol. 60(4).
- PETERSON, J.L., LIN, G., HO, M., CHEN, Y. y GAENSSLEN, R.E., 2003: «The feasibility of external blind DNA proficiency testing. II. Experience with actual blind tests», *Journal of Forensic Science*, vol. 48(1).
- PETRACO, N.D., SHENKIN, P., SPEIR, J., DIACZUK, P., PIZZOLA, P.A., GAMBINO, C. y PETRACO, N., 2012: «Addressing the National Academy of Sciences' Challenge: A Method for Statistical Pattern Comparison of Striated Tool Marks», *Journal of Forensic Sciences*, vol. 57.
- PLEBANI, M. y CARRORO, P., 1997: «Mistakes in a stat laboratory: types and frequency», *Clinical Chemistry*, vol. 43.
- POCOCK, S.J., 1983: *Clinical trials: a practical approach*. Wiley, Chichester.
- PRETTY, I.A., 2011: «Resolving Issues in Bitemark Analysis», en *Bitemark Evidence: A Color Atlas*, R.B.J Dorian, Chicago: CRC Press.
- PRETTY, I.A. y SWEET, D., 2010: «A paradigm shift in the analysis of bitemarks», *Forensic Science International*, vol. 201.
- RISINGER, D.M., THOMPSON, W.C., JAMIESON, A., KOPPL, R., KORNFIELD, I., KRANE, D., MNOOKIN, J.L., ROSENTHAL, R., SAKS, M.J. y ZABELL, S.L., 2014: «Regarding Champod, editorial: «Research focused mainly on bias will paralyse forensic science», *Science and Justice*, vol. 54.
- RIVA, F. y C. CHRISTOPE, 2014: «Automatic comparison and evaluation of impressions left by a firearm on fired cartridge cases», *Journal of Forensic Sciences*, vol. 59.
- ROBERTS, L., 1991: «Fight erupts over DNA fingerprinting», *Science*, vol. 254: 1721-1723.
- SAKS, M.J. y KOEHLER, J.J., 2005: «The coming paradigm shift in forensic identification science», *Science*, vol. 309 (5736): 892-895.
- SAKS, M.J. y KOEHLER, J.J., 2008: «The individualization fallacy in forensic science evidence», *Vanderbilt Law Review*, vol. 61(1): 199-218.
- STACEY, R.B., 2005: «Report on the erroneous fingerprint individualization in the Madrid train bombing case», *Forensic Science Communications*, vol. 7.
- STAHL, M., LUND, E.D. y BRANDSLUND, I., 1998: «Reasons for a laboratory's inability to report results for requested analytical tests», *Clinical Chemistry*, vol. 44.
- TANGEN, J.M., THOMPSON, M.B. y MCCARTHY, D.J., 2011: «Identifying fingerprint expertise», *Psychological Science*, vol. 22(8).
- THOMPSON, W.C. y FORD, S., 1990: «Is DNA fingerprinting ready for the courts?», *New Scientist*, vol. 125: 38-43.
- THOMPSON, W.C., TARONI F. y AITKEN, C.G.G., 2003: «How the Probability of a False Positive Affects the Value of DNA Evidence», *J Forensic Sci*, vol. 48(1).
- THOMPSON, W.C., 2009: «Painting the target around the matching profile: the Texas sharpshooter fallacy in forensic DNA interpretation», *Law, Probability and Risk*, vol. 8(3).
- THOMPSON, W.C., 2013: «The Myth of Infallibility», en SHELDON KRIMSKY & JEREMY GRUBER (Eds.), *Genetic Explanations: Sense and Nonsense*, Harvard University Press.
- THOMPSON, S.G., 2015: *Cops in Lab Coats: Curbing Wrongful Convictions through Independent Forensic Laboratories*. Carolina Academic Press.
- TVERSKY, A., 1977: «Features of similarity», *Psychological Review*, vol. 84: 327-52.

- ULERY, B.T., HICKLIN, R.A., BUSCAGLIA, J. y ROBERTS, M.A., 2011: «Accuracy and reliability of forensic latent fingerprint decisions», *Proceedings of the National Academy of Sciences*, vol. 108(19).
- WEECH, T.L. y GOLDHOR, H., 1982: «Obtrusive versus unobtrusive evaluation of reference service in five Illinois public libraries: A pilot study», *Library Quarterly: Information, Community, Policy*, vol. 52, No. 4: 305-324.
- WERTHEIM, KASEY, 2002: «Letter re: ACE-V: Is it scientifically reliable and accurate?», *Journal of Forensic Identification*, vol. 52.
- WILSON, H.D., 2012: «Comparison of the individual characteristics in the outsoles of thirty-nine pairs of Adidas Supernova Classic shoes», *Journal of Forensic Identification*, vol. 62(3).

## FRECIENTES SIGLAS EN INGLÉS TRADUCIDAS AL ESPAÑOL

ABFO	<i>American Board of Forensic Odontology</i> (Junta Directiva Norteamericana de Odontología Forense).
DENL	<i>Department of Energy National Laboratories</i> (Laboratorios nacionales del Departamento de Energía).
DOD	<i>Department of Defense</i> (Departamento de Defensa).
DOJ	<i>Department of Justice</i> (Departamento de Justicia).
FBIL	<i>Federal Bureau of Investigation Laboratory</i> (Laboratorio de la Oficina Federal de Investigación).
FJC	<i>Federal Judicial Center</i> (Centro Judicial Federal).
FSC	<i>Defense Department's Forensic Science Center</i> (Centro de Ciencias Forenses del Departamento de Defensa).
FSSB	<i>Forensic Science Standards Board</i> (Junta Directiva de Estándares de Ciencia Forense).
FRE	<i>Federal Rules of Evidence</i> (Reglas Federales de la Prueba).
JCUS	<i>Judicial Conference of the United States</i> (Conferencia Judicial de los Estados Unidos).
MRC	<i>Metrology Resource Committee</i> (Comité de Recursos de Metrología).
NASEM	<i>National Academies of Science, Engineering, and Medicine</i> (Academias Nacionales de Ciencias, Ingeniería y Medicina).
NDIS	<i>National DNA Index System</i> (Sistema de Reseña de ADN Nacional).
NCFS	<i>National Commission on Forensic Science</i> (Comisión Nacional sobre Ciencia Forense).
NIJ	<i>National Institute of Justice</i> (Instituto Nacional de Justicia).
NIST	<i>National Institute of Standards and Technology</i> (Instituto Nacional de Estándares y Tecnología).
NSF	<i>National Science Foundation</i> (Fundación de la Ciencia Nacional).
NRC	<i>National Research Council</i> (Consejo de Investigación Nacional).
NSTC	<i>National Science and Technology Council</i> (Consejo de Ciencia y Tecnología Nacional).
OSAC	<i>Organization for Scientific Area Committees</i> (Organización de Comités de Áreas Científicas).
OSTP	<i>Office of Science and Technology Policy</i> (Oficina de Políticas Científicas y Tecnológicas).
PCAST	<i>President's Council of Advisors on Science and Technology</i> (Consejo de Asesores del Presidente en Ciencia y Tecnología).
QAS	<i>Quality Assurance Standards</i> (Estándares de Garantía de la Calidad).

CIENCIA  
PARA EL PROCESO



## (UN-)INTERPRETABILITY IN EXPERT EVIDENCE: AN INQUIRY INTO THE FRONTIERS OF EVIDENTIAL ASSESSMENT

Alex Biedermann\*

Kyriakos N. Kotsoglou\*\*

**ABSTRACT:** Evidence law regimes across several contemporary legal orders provide a host of doctrinal devices designed to probe various sorts and sources of information, especially with respect to their accuracy and reliability. These legal provisions, however, are vulnerable to or even tolerant towards uninterpretable evidence, that is evidence whose probative value cannot be assessed. This article critically examines and discusses the «littering paths» of uninterpretable evidence in legal proceedings. We point out the misinformative character of such evidence as well as the potentially misleading impact on criminal adjudication when using scientific evidence for purposes that, strictly speaking, the evidence cannot help with. We identify common causes and possible remedies and argue that to track uninterpretability, it is necessary to understand what exactly it means to say that a piece of evidence is probative in the procedural space of reasons, and how to draw the line between informative and non-informative items of evidence.

**KEYWORDS:** scientific evidence, interpretability, probative value, propositions, forensic science

---

\* University of Lausanne, Faculty of Law, Criminal Justice and Public Administration, School of Criminal Justice, 1015 Lausanne-Dorigny (Switzerland). The authors gratefully acknowledge the support of the Swiss National Science Foundation through grant BSSGI0\_155809, the Fondation pour l'Université de Lausanne and the Société Académique Vaudoise. – alex.biedermann@unil.ch

\*\* Northumbria University, Faculty of Law & Business, Newcastle Upon Tyne, NE1 8ST (UK). – kyriakos.kotsoglou@northumbria.ac.uk

**SUMMARY:** 1. INTRODUCTION. 1.1. Seemingly probative scientific evidence? 1.2. Aims and scope.—2. THEORISING (UN-)INTERPRETABILITY OF SCIENTIFIC EVIDENCE. 2.1 Delimiting the scope of inquiry: the model-based view of interpretation. 2.2. Principles of scientific interpretation. 2.3. Role, limitations and necessity of empirical measures of scrutiny for forensic science evidence. 2.4. Uninterpretability: conceptualising the frontiers of interpretability. 2.5. Distinguishing (un-)interpretability from probative value and relevance. 2.6. (Un-)interpretability vs. inconclusiveness.—3. PROBING FOR (UN-)INTERPRETABILITY. 3.1. Distinguishing between evaluative, investigative and technical reporting. 3.2. Uninterpreted and partially interpreted vs. uninterpretable evidence. 3.3. The central role of main objects of litigation and propositions in interpretation. 3.4. Exemplifying (un-)interpretability: probabilistic genotyping.—4. CONCLUSIONS

**RECOMMENDED CITATION:** ALEX BIEDERMANN; KYRIAKOS N. KOTSOGLOU, 2021: «(Un-)interpretability in expert evidence: an inquiry into the frontiers of evidential assessment», in *Quaestio facti*, 3: 483-517. Madrid: Marcial Pons Ediciones Jurídicas y Sociales. DOI: [http://dx.doi.org/10.33115/udg\\_bib/qf.i3.22599](http://dx.doi.org/10.33115/udg_bib/qf.i3.22599)

## 1. INTRODUCTION

### 1.1. Seemingly probative scientific evidence?

It is not an exaggeration to say that the field of evidence & proof has primarily focused—at least for the past few decades—on the *reliability* of various forensic science<sup>1</sup> disciplines. From the ground-breaking *Frye*<sup>2</sup> decision which set the initial standards for scientific validity in the 1920s, widely considered even beyond the American context, over the «DNA wars» of the 1990s<sup>3</sup> to the ongoing discussions over how to define criteria of scientific validity,<sup>4</sup> scholars, criminal courts and scientific organisations have tried to lay down the procedural and methodological architecture for the operation of forensic practitioners in the criminal justice system.

These developments raise the question of what else could and, indeed, should have been discussed, *in addition* to reliability. As we will argue below, the main and usually sole requirement for admissibility in criminal proceedings is a core feature of forensic scientists' operations that, so far, is not adequately illuminated. Whereas reliability—based on «validity» in the US<sup>5</sup> and «sufficiently reliable scientific basis»

<sup>1</sup> For the purpose of this paper, forensic science is broadly understood as «the application of scientific or technical practices to the recognition, collection, analysis and interpretation of evidence for criminal and civil law or regulatory issues». See PRESIDENT'S COUNCIL OF ADVISORS ON SCIENCE AND TECHNOLOGY (PCAST), 2016: 1, hereinafter cited as PCAST Report, and PCAST as council.

<sup>2</sup> *Frye v. United States*, 293 F. 1013 (D.C. Cir. 1923).

<sup>3</sup> See ARONSON, 2007.

<sup>4</sup> See the PCAST Report for an example of a report that stirred considerable debate in recent years.

<sup>5</sup> *Daubert v. Merrell Dow Pharm., Inc.* (509 U.S. 579 [1993]). The *Daubert* standard is the test currently used on the federal level and in several jurisdictions on the state level. In the federal court

in England and Wales<sup>6</sup>—invokes the exception to the opinion rule which raises a general ban on opinions from any other person than the triers of fact, relevance is the main requirement for *admissibility*, save for any exclusionary rules.

Relevance too seems to be a straightforward topic insofar as it describes a relationship between two facts. A fact *x* is relevant to a fact *y* when the former has some material bearing on the latter. For example, to ascribe criminal liability to or acquit a defendant, we need evidence which is capable of helping to prove or disprove some element of the respective criminal allegation. This is obviously easier said than done. Admitting relevant evidence at trial and excluding irrelevant evidence is similar to the fictional rugby manager's instruction «To win, you need to score more tries than the other side». An expected response to that can only be: «Sure, but *how?*»<sup>7</sup>

In common law, James Fitzjames Stephen's classical definition seems to offer a helpful approach to criminal courts: «The word "relevant" means that any two facts to which it is applied are so related to each other that according to the common course of events on either taken by itself or in connection with other facts proves or renders probable the past, present or future existence or non-existence of the other».<sup>8</sup> Explicit reference to the term «probability» in the above definition has been replicated in modern legislature, most notably in FRE 401 (US) according to which «Evidence is relevant if (a) it has any tendency to make a fact more or less probable that it would be without the evidence; and (b) the fact is of consequence in determining the action». Similarly, in England and Wales Lord Simon of Glaisdale noted that «Evidence is relevant if it is logically probative or disprobative of some matter which requires proof [...] It is sufficient to say [...] that relevant (i.e. logically probative or disprobative) evidence is evidence which makes the matter which requires proof more or less probable».<sup>9</sup>

But before the critical reader is led to think that judges and jury in criminal trials deploy any sort of probability calculus to determine relevance and weight,<sup>10</sup> it is worth reminding one of the core elements in Stephen's definition. What is necessary, according to the abovementioned determination, is a vast array of background generalisations about the world, contextual information and, often, cultural understanding in order to place the evidence into context. The law predicates admissibility on relevance, but as Thayer put it, «the law furnishes no test for relevance».<sup>11</sup> It is

---

system, it replaced the *Frye* standard, which, nota bene, is still used in some states.

<sup>6</sup> *R v Dlugosz* [2013] EWCA Crim 2.

<sup>7</sup> See WILLIAMS, 1995: 239.

<sup>8</sup> STEPHEN, 1948: art. 1.

<sup>9</sup> *DPP v Kilbourne* [1973] A.C. 729 at 756 HL.

<sup>10</sup> See *R v Adams* [1996] 2 Cr App R 467, 481, 482 (England and Wales), where the Court of Appeal remarked that to «introduce Bayes' Theorem, or any similar method, into a criminal trial plunges the jury into appropriate and unnecessary realms of theory and complexity deflecting them from their proper task».

<sup>11</sup> THAYER, 1898: 265.

important to note here that background knowledge is neither cognitively present nor can it easily be articulated.

We need to distinguish the *laissez-faire* approach to, let us call it, normal evidence, as outlined above, from expert opinion evidence. The latter is admissible in England and Wales if, inter alia, it is «relevant to a matter in issue in the proceedings». <sup>12</sup> Experts, however, need to do much more than simply claiming that fact *x* is relevant to fact *y*. They need to explain *why* that is the case in their expert opinion. This insight helps us fully grasp the structural difference between relevance for non-expert evidence and relevance for expert evidence: in addition to the *qualitative* assessment which is expressed in a binary mode (relevant-irrelevant), experts need to lay out their underlying assumptions and invoke/present the respective scientific model at play, including the inferential steps which allow them to reach a given conclusion, i.e. the *quantitative* measure of how strongly or weakly (if at all) the evidence favours one proposition over the other. An expert, whose duty in England and Wales is «to help the court to achieve its aims», <sup>13</sup> thus needs to explain his or her «methods», including the «inferential steps», and the extent and quality of the data on which the expert's opinion is based. <sup>14</sup>

Although this soft-law instrument provides devices with which court judges can decide on the reliability and validity of the proffered method, it is important to note that factual relevance is a function of reliability, not vice versa. A reliable method has a high degree of specificity and sensitivity. In other words, reliability is a precondition for relevance. A method whose results are only slightly better than chance barely deserves the name scientific. In fact, such a pseudo-method would be reminiscent of the diviner's claim that when he holds the rod, a certain sensation in his hand indicates the existence of water five feet under the ground. This fails to establish any relationship between the «feeling in the hand» (method) and the empirical fact of water being present at certain depth under the ground. <sup>15</sup> Similarly, and in a more realistic context, any polygraphic output cannot indicate the truthfulness or falsehood of a statement. <sup>16</sup> Both methods are blatantly unreliable.

Where these requirements cannot be met, we argue here, expert evidence should not be adduced in the criminal process. However, forensic science practice is far from these theoretical understandings. In many instances, forensic science evidence is used to support various strains of argument despite being *uninterpretable* or, at best, only partially interpretable. Such unwarranted uses of evidence, we cannot stress this enough, are a cause of concern and call for an inquiry into the understanding of what

---

<sup>12</sup> CPD V Evidence 19.A.1

<sup>13</sup> CPD V Evidence 19B.1.

<sup>14</sup> CPD V Evidence 19.A.5(a-c).

<sup>15</sup> See WITTGENSTEIN, 1958: 1-15.

<sup>16</sup> See KOTSOGLU, 2021.

exactly it means to say that a piece of expert evidence is informative and, hence, how to probe for evidence interpretability, i.e. the feasibility of assigning probative value.

To illustrate the problem of interpretability, consider a type of forensic science evidence that can be presented in a way that should arouse the interest of prosecutors: no matter what the analyst observes, the respective finding can be said to be in agreement with the prosecution's case. Such evidence may seem too good to be true, yet it actually exists. A well-known example is the discipline of firearms discharge residue (FDR, also called gunshot residue [GSR]) analysis. Suppose a case in which the analyst reports the finding of particles classified as GSR on the hands of a suspected shooter. Suppose also that those particles have the same qualitative composition and morphological features as those found in the head hair of the victim and in spent cartridges recovered from the crime scene. While there are many subtleties and circumstantial aspects to be taken into account in shooting cases, for the illustrative purposes here it shall suffice to mention that, at least in principle, forensic literature considers the presence of such particles «a positive result».<sup>17</sup> This is so, as is widely thought,<sup>18</sup> because «GSR is not common to the average's person daily environment».<sup>19</sup> The exact meaning of the proposition that a finding is «positive» pertains to a discussion in its own right, ranging from broad views such as understanding «positive» merely in a descriptive sense as the presence of target particles (or as a factual statement thereof), to stronger views according to which «positive» means incriminating, i.e. having probative force in relation to someone's alleged activities, or even criminal liability.

But suppose now that a search has been conducted and that *no* particles were found on the hands of a person of interest. Interestingly, such a finding can also be presented in a way that suits the prosecution's case. It suffices to invoke references to specialised literature in the field. For example, scientists have asserted that «[t]he absence of gunshot residue on a person's hands does not eliminate that individual from having discharged a firearm»<sup>20</sup> and that «... negative results [...] are not necessarily exculpatory».<sup>21</sup>

Thus, both the presence and absence of particles seems to be a useful type of evidence for the prosecution. Yet, this seemingly catchall probative property comes at a high price: the same findings can also be «fitted» to the narrative of the defence. On the one hand, the finding of no particles is what we typically expect to see on a person that is *not* related in any way to the discharge of a firearm.<sup>22</sup> On the other

<sup>17</sup> HEARD, 2008: 255.

<sup>18</sup> We call this a mere thought because, as argued later in §2.2., one cannot directly jump from observations to conclusions. In particular, the mere rarity of a given type of evidence among a certain group of persons by itself does not make an item of evidence probative.

<sup>19</sup> HEARD, 2013: 274.

<sup>20</sup> TRIMPE, 2011: 29.

<sup>21</sup> SCHWOEBLE & EXLINE, 2000: 127.

<sup>22</sup> See also *Pole v. Randolph* 570 F.3d 922 (7th Cir. 2009) for an example of a case in which the appellant argued that the retention of negative GSR findings (i.e. no detected GSR particles) implied

hand, any finding of particles raises the question of contamination and, hence, the problem of false positives.<sup>23</sup> In view of these diverging assertions, laypersons may wonder what, if anything, one is logically entitled to conclude without giving undue advantage to one of the competing parties to the matter. Expert witness testimony and forensic science in general thus has to be antithetical to the approach of Pythia, the Oracle of Delphi. For her reply to those going to war («You will go, you will return not in the war shall you die»<sup>24</sup>) could conveniently validate mutually exclusive outcomes. She allowed those asking to hear what they wanted to hear.

We thus see that even a slight scratch on the surface features of a given type of forensic science evidence readily prompts a host of questions. Namely, the questions of what it means to interpret an item of scientific evidence in the first place, what an admissible interpretation of evidence is and, more fundamentally, what it means to say that an item of evidence is interpretable. Moreover, inquiring about whether an item of evidence is interpretable, and to what extent, cannot be separated from the questions of the frontiers of interpretability and how those frontiers ought to be drawn. In this paper, we aim at investigating these questions in the particular context of scientific evidence.

## 1.2. Aims and scope

The main thesis defended in this paper is that settling questions regarding the nature of (probative) value—a necessary preliminary to the assessment of relevance—and the feasibility of ascribing evidential value, hinges upon our ability to cope with a notion that we will call, for present purposes, *(un-)interpretability*. As a device for drawing the frontiers of the possibility to assign probative value, the notion of interpretability is not understood here as an abstract property of evidence that could be contemplated in the absence of a human reasoner, equipped with knowledge and background information regarding the circumstances of the instant case. Instead, interpretability is understood as a function of a reasoner's assumptions, knowledge and understanding about selected target systems, and the intended purpose for which evidence is being adduced. In a nutshell, we will call evidence *uninterpretable* when

---

that vital evidence was kept from the defence («The exculpatory evidence [...] consisted of two favorable gunshot residue tests. [Defendant] contended that the swabs of his hands and tests of the glove were negative for gunshot residue...», at 930).

<sup>23</sup> See also *Hudson v. Lafler* Civil No. 04-CV-74001-DT (E.D. Mich. Jun. 8, 2006) («the prosecution's expert witness admitted that the presence of gunshot residue on a person did not necessarily indicate that the person had discharged a gun», at 7).

<sup>24</sup> The utterance was successful (and profitable) because its meaning changed depending on whether the comma was put before or after the word «not». If the former was the case (i.e. «you will go, you will return, not in the war shall you die») the person believed that he would survive the battle. If the latter was the case (i.e. «you will go, you will return not, in the war shall you die») the person believed that he would die. See YEROULANOS, 2016: 421.

it is not possible to assert whether—and to what extent—evidence is informative in one way or another with respect to a contested evidential claim. Underlying this definition is the idea of ascertainable meaning with respect to an aspect of *primary* interest in a case. Note that this does not exclude the possibility of evidence being interpretable with respect to some other, but subsidiary or even uncontested aspect of a case.<sup>25</sup> In other words, interpretability is not an intrinsic feature of the evidence, but a relational one. It is predicated on the procedural space of reasons and depends on the parties' specific propositions.

We will investigate the consequences of uninterpretability by pointing out how it can undermine subsequent reasoning processes and lead to unwarranted conclusions. An example for this is evidence presented in support of a particular proposition or claim, but tacitly carried over—in an inferentially flawed way—to support the resolution of a disputed matter about which that evidence, in strict terms, is actually uninformative. This results in a spuriously or a seemingly probative conclusion. Where this happens, evidence is «littering», so to speak, the inferential grounds on which subsequent juridical decision-making needs to be based. Thus, there is more to uninterpretable evidence than it being evidence that doesn't answer the question.

Section 2 starts by delimiting the scope of inquiry to the science-centred view of interpretation. This part provides a working definition of interpretability as applied to scientific evidence and its apprehension in legal proceedings. Section 2 will also explain where the demand of interpretability comes from and why it is not covered by standard measures for evaluating the fitness of forensic science disciplines, such as ground-truth testing and related notions such as error rates, foundational validity and validity as applied.<sup>26</sup> Section 3 focuses on ways to probe and enact interpretability, especially determining its limits. For this purpose, we will argue that the distinction between investigative and evaluative uses of evidence,<sup>27</sup> as well as the notion of hierarchy of propositions<sup>28</sup> is crucial. This section also explains the differences between (un-)interpretability and other commonly encountered notions, such as inconclusiveness and probative value. Though this paper is mainly a theoretical inquiry, Section 3 will include an example of evidence (un-)interpretability in the context of modern DNA profiling, namely probabilistic genotyping. The conclusions, in Section 4, emphasise that uninterpretability is not merely a synonym for evidence that does not answer a question of primary interest, nor that it is an end itself. Instead, we argue, uninterpretability is a means to an end, that is avoiding the use of evidence for conclusions it is not able to support.

---

<sup>25</sup> Uncontested propositions in the adversarial criminal process do not need to be proven.

<sup>26</sup> On foundational validity and validity applied see, e.g. PCAST, 2016: 43.

<sup>27</sup> JACKSON, 2000.

<sup>28</sup> COOK et al., 1998a; JACKSON & BIEDERMANN, 2019.

## 2. THEORISING (UN-)INTERPRETABILITY OF SCIENTIFIC EVIDENCE

### 2.1. Delimiting the scope of inquiry: the model-based view of interpretation

Since the mid-1990s, the term interpretation is «experiencing a steady period of evolution»<sup>29</sup> in forensic science. In a landmark paper, a group of scientists of the now axed Forensic Science Service (FSS) considered interpretation as the «drawing of rational and balanced inferences from observations, test results and measurements».<sup>30</sup> This understanding of interpretation as a cognitive process is the result of an FSS project called Case Assessment and Interpretation (CAI).<sup>31</sup> It focused on scrutinising and structuring forensic science services towards delivering output that meets the needs of the mandating party. These needs, however, are not necessarily defined in a narrow sense, specific to scientific evidence. They are not materially different from the reasons that drive evidence processing in general and inquiries thereinto. As Aitken et al. put it succinctly: «The essential issue is: what does the evidence *mean*?».<sup>32</sup> This question echoes the notion of meaning and the quest for understanding highlighted in Section 1.

These developments have been accompanied by a series of scholarly works centred around the notion of *interpretation* of scientific evidence.<sup>33</sup> They focus on the use of forensic science evidence (e.g. DNA, glass fragments, etc.) to help consumers of expert information discriminate between competing propositions of interest regarding, for example, the source of recovered trace material (i.e. identification), or alleged activities that led to the deposition of trace material, including considerations of phenomena such as transfer and persistence of trace materials. Arguments and principles presented in these works rely on, and invoke, concepts of probability and statistics to deal with uncertainty.<sup>34</sup> The mathematical nature of these works has equally attracted criticism<sup>35</sup> and, from the mid-1980s on, enthusiastic support.<sup>36</sup> It is important to note, however, that these debates over the adequacy of formal methods of rea-

<sup>29</sup> COOK et al., 1998b: 152.

<sup>30</sup> Ibid.

<sup>31</sup> For reviews of CAI see, e.g. JACKSON et al., 2014; JACKSON & JONES, 2009.

<sup>32</sup> AITKEN et al., 2010: §2.2.

<sup>33</sup> E.g. AITKEN & STONEY, 1991; AITKEN, 1995.

<sup>34</sup> Evett and Weir «view the interpretation problem as one of updating uncertainty in the light of new information» (EVETT & WEIR, 1998: 23). See also FIENBERG, 1989, and the two volumes by GASTWIRTH, 1988.

<sup>35</sup> For a widely cited critique, see TRIBE, 1971.

<sup>36</sup> See, e.g. the New Evidence Scholarship movement described by LEMPERT, 1986, in a contribution to Symposium on Probability and Inference in the Law of Evidence held at Boston University in April 1986.

soning mainly concern questions about evidence & proof in legal adjudication as a whole, rather than specific items of scientific evidence.

At this juncture, it is worthwhile to include a critical distinction regarding the term interpretation that is often misunderstood. As mentioned above, when (forensic) scientists talk about interpretation in the sense outlined above, they mean assessing the value of evidence in terms of its capacity to help discriminate between competing propositions of interest. Interpretation, thus, is conditioned on the assumptions made by the scientist as well as the specific information available to the scientist at the time the interpretation is made.<sup>37</sup> Most importantly, expert witnesses are confined—at least in England and Wales—within the questions asked in relation to live procedural issues. Yet, this interpretation is not the end of the matter because «the ultimate value of any particular piece of evidence, scientific or otherwise, must always be assessed contextually, in the light of its contribution to the case as a whole».<sup>38</sup> This latter dimension of interpretation thus amounts to assessing the value of evidence using possibly higher-level propositions—*different* from those with respect to which the scientist assessed the findings—and using all the information and other items available in the case.

The remainder of this paper will focus on the former, science-centred understanding of interpretation because it is a crucial preliminary to the latter, broader sense of interpretation. While the science-centred view of interpretation is well established in literature and, to some extent, in practice,<sup>39</sup> there is yet no structured account of the frontiers of interpretability, especially of the notion of uninterpretability and its implications for advanced stages of the legal process.

## 2.2. Principles of scientific interpretation

An early, if not the earliest account of what today is known as the «three precepts for the forensic scientist»<sup>40</sup> can be found in Evett's overview of his «personal odyssey».<sup>41</sup> In this account, he refers to these precepts as the «the principles of our science».<sup>42</sup>

---

<sup>37</sup> Typically, the scientist does not need to know all the information. This is neither practically possible, nor desirable, because it may bias the interpretation. For example, in the case of glass fragments detected on a jumper, task-relevant information includes aspects such as the mode of breaking and the time elapsed between the collection of the jumper and the breaking of the window. Whether the person of interest confessed, or whether she was identified by an eyewitness is not relevant information for the scientist (though it is so for the recipients of expert information). See also NATIONAL COMMISSION ON FORENSIC SCIENCE, HUMAN FACTORS SUBCOMMITTEE, 2015.

<sup>38</sup> AITKEN et al., 2010: §3.28. On the notion of inferential contextualism, see KOTSOGLOU, 2013.

<sup>39</sup> See Section 2.2 for references.

<sup>40</sup> EVETT & WEIR, 1998: 29.

<sup>41</sup> EVETT, 1991: 19-20.

<sup>42</sup> *Ibid.*, at 19. Here, by «our science», Evett means forensic science. In later writings, he refers to the precepts as principles of «scientific interpretation» (EVETT & WEIR, 1998: 29) and the «scientific

In essence, the three precepts can be summarised as follows:<sup>43</sup>

1) The meaning of an item of evidence cannot be assessed in isolation, i.e. with respect to only one proposition.<sup>44</sup>

2) The relevant questions for forensic scientists are of the following kind: *given* a particular proposition, what is the probability of seeing this evidence?<sup>45</sup>

3) In order to assess the strength of the evidence, the scientist must consider not only the probability of the evidence (see above point 2) given each proposition of interest, but also the framework of circumstances. The scientist should ask: given the available information (i.e. framework of circumstances), is the evidence better conceivable with one proposition (i.e. account of the contested event) rather than with the stated alternative? If so, to what extent is the evidence better conceivable with one proposition rather than with the stated alternative?

The rationale underlying these precepts stems from both formal considerations (i.e. inferential logic) and common sense. From a formal point of view, the precepts can be recognised as elements of a (statistical) inference method.<sup>46</sup> It is important, however, to emphasise that this is merely to say that the precepts are fundamentally rooted in formal-logical considerations. There is no suggestion that legal evidence or proof processes, as a whole, are a statistical exercise.<sup>47</sup> Instead, the focus here is on understanding the common-sense rationale underlying these precepts and their role in avoiding logically flawed reasoning. While not commonly presented in legal textbooks,<sup>48</sup> the rationale is briefly exposed in the next paragraphs.

Precept number one is about ensuring a balanced perspective. It does so by requiring the consideration of a pair of contrastive propositions. Precept number two seeks to ensure that scientists focus on the findings (evidence), given propositions, and not the reverse. Indeed, the duty of the expert is not to opine on propositions, but to concentrate on the evidence. This may sound counterintuitive, but it suffices

---

approach to interpretation» (Ibid., 245).

<sup>43</sup> For a more recent formulation of these principles, with an emphasis on the framework of circumstances (i.e. conditioning information) see EVETT et al., 2017.

<sup>44</sup> A proposition is a statement or assertion about a contested real-world event, that is one version of an event of interest. For example, a pair of what are referred to as source-level propositions is «The semen came from Mr. B» and «The semen came from some other man» (COOK et al., 1998a: 232). Usually, propositions reflect positions held by parties (i.e. prosecution and defense) though there is no procedural requirement for a party to specify a proposition.

<sup>45</sup> This principle has also been referred to as «the single most important lesson for the logic of evaluative forensic science» (EVETT, 2009: 159).

<sup>46</sup> Specifically, in the case here, the precepts conform to the logic of the likelihood ratio, which is a core element of Bayesian statistical inference (EVETT, 1991: 19). On the notion of likelihood ratio, see also FRIEDMAN, 2017: 70.

<sup>47</sup> E.g. EVETT & WEIR, 1998, note: «It is not our claim that Bayesian inference is a panacea for all the problems of the legal process. However, we do maintain that it [...] provides insights that are not otherwise possible» (at 29).

<sup>48</sup> For notable exceptions, see FRIEDMAN, 2017; DENNIS, 2017: Chap. 4.

to consider a reformulation of this precept in the more informal language of causes (propositions) and effects (evidence). Suppose Billy hits a window with a hammer; it is not too difficult, then, to conclude that the probability of a broken window and fragments of glass landing on the floor and, possibly, on Billy's clothing, is large.<sup>49</sup> However, the reverse is not straightforward: given the evidence only (in this case: recovered glass fragments), it is not obvious to infer what the cause was, let alone to extrapolate from this Billy's course of conduct. At least one alternative proposition and more investigation are needed to elucidate the case, a task entrusted to the judiciary.<sup>50</sup> The subtle point here is the difficulty of understanding the structure of conditional probabilities, which is known to cause serious problems.<sup>51</sup> In particular, it is important to stress the difference between, on the one hand, the probability of Billy having smashed the window with a hammer *given* the findings and, on the other hand, the probability of the glass findings *given* that Billy smashed the window.

In turn, precept number three seeks to prevent unwarranted assertions about probative value. For example, the mere fact that the scientist considers the evidence to be well conceivable with a particular proposition (e.g. when saying «if proposition so and so is true, then this evidence is what I would expect to see») does not necessarily mean that the evidence supports that proposition. The reason for this is not only that the evidence may equally well be compatible with the alternative proposition at play, and hence be uninformative.<sup>52</sup> It is also important that the evaluation considers the framework of circumstances of the instant case. Stated otherwise, an item of evidence is not probative in one way or another with respect to a particular proposition *in abstracto*. An item of evidence has value with respect to a proposition A rather than proposition B only insofar as the evidence is better conceivable with proposition A, rather than B, i.e. if the evidence is better accounted for by proposition A rather than proposition B in the light of the task-relevant information (derived from the framework of circumstances).

Though they may seem technical at first sight, these precepts of scientific interpretation can be stated in largely non-technical language,<sup>53</sup> and they are fundamental in the sense that many cases of recorded misuse of scientific evidence can be reconstructed as a violation of one or more of these principles.<sup>54</sup> This is why they

<sup>49</sup> This example is adapted from PEARL & MACKENZIE, 2018: 99.

<sup>50</sup> EVETT et al., 2017: 19.

<sup>51</sup> A typical example is the notorious prosecutor's fallacy (THOMPSON & SCHUMANN, 1987).

<sup>52</sup> Suppose, for example, that in the above example involving a broken window the alternative proposition is that Johnny smashed the window with a hammer, and Billy was a bystander.

<sup>53</sup> The precepts can be stated without resort to either probability or statistics. Thus, while a formal justification for the three precepts is based on statistical principles and methods, it is not necessary—for practical purposes and ease of communication—to label the precepts openly as statistics or statistics based, let alone Bayesian.

<sup>54</sup> A good example is the case *People v. Collins*, 68 Cal. 2d 319 (1968), discussed in FINKELSTEIN & FAIRLEY, 1970. For a more recent example in England and Wales see, e.g. *R v Barry George* [2007] EWCA Crim 2722.

have been incorporated in standard and guidance documents for forensic examiners and legal practitioners,<sup>55</sup> and are regularly reemphasised in the light of controversial cases.<sup>56</sup> But, as well-intentioned and mildly constraining these precepts are, a realistic assessment of the current reporting practice of (forensic) scientists must conclude that, as a whole, forensic scientists continue to follow a variety of different reporting formats and conclusion scales—some of which do not conform to the three principles exposed above.<sup>57</sup>

While the causes of this are operational constraints and adherence to conventional procedures, which are managerial topics in their own right, there are two conceptual considerations that echo deeper challenges. The first is the widely held view that the «problem of scientific evidence» is, first and foremost, an empirical one, i.e. centred around the question of whether forensic scientists can reliably do what they claim to be able to do.<sup>58</sup> Section 2.3. will briefly address this issue. The second consideration has to do with the question of whether the principles of scientific interpretation are sufficient as a safeguard against uninterpretability, let alone able to define limits of interpretability. This aspect will be addressed in Section 2.4.

### 2.3. Role, limitations and necessity of empirical measures of scrutiny for forensic science evidence

The discussion so far might have inadvertently alienated legal scholars and practitioners whose perspectives regarding admissibility of scientific testimony are influenced by the quests for (empirical) validity derived from legal provisions, such as FRE 702. The notion of validity has been prominently brought to the broader judiciary's attention in the PCAST Report with its strong emphasis on what the Council termed foundational validity and validity as applied.<sup>59</sup> In a nutshell, the Council views forensic science as a discipline that renders direct opinions about the truth or otherwise of particular propositions<sup>60</sup> (e.g. the proposition that the person of interest rather than an unknown person is the source of a particular trace) which is why, the Council considers, it is necessary to require empirical performance indicators for this mode of operation. Specifically, the Council notes that «scientific

<sup>55</sup> E.g. ASSOCIATION OF FORENSIC SCIENCE PROVIDERS (AFSP), 2009; EUROPEAN NETWORK OF FORENSIC SCIENCE INSTITUTES (ENFSI), 2015 (hereinafter cited as ENFSI Guideline); AITKEN et al., 2010.

<sup>56</sup> AITKEN et al., 2011.

<sup>57</sup> For a review and discussion see, e.g. THOMPSON, 2019.

<sup>58</sup> PCAST, 2016, is a prominent example for this.

<sup>59</sup> PCAST, 2016: 42-43, explicitly relates foundational validity and validity as applied to FRE 702 (c, d).

<sup>60</sup> Note that this is contrary to principle no. 2, exposed in Section 2.2., according to which the expert should focus on the probability of the findings, given propositions, *not* the reverse (i.e. focus on the propositions given the evidence).

validity and reliability require that a method has been subjected to empirical testing [...] and the error rates are determined».<sup>61</sup>

Prima facie, this sounds entirely reasonable, for who can really afford to ignore hard data regarding the accuracy of a source of information, i.e. the number of times a source of information—using everyday language—«got it right and wrong»? The point is that this perspective falls short of the scope of interpretability, which is why it is worthwhile to properly distinguish these concepts from one another.

To be clear, there is nothing wrong with false positives and false negatives<sup>62</sup> where the concern is process output that conforms to the Council's description of forensic science. Typical examples are cases in which the examiner's task amounts to applying a diagnostic test (e.g. a drug test) and reporting is limited to the actual outcome (i.e. positive or negative). However, even in these cases, a factual report of a test's outcome, accompanied by an error rate, does *not* directly inform us about whether or not a target proposition (e.g. presence or absence of a detectable amount of a drug) is true.<sup>63</sup> Empirical statements—we cannot stress this enough—are probabilistic in nature, and categorical conclusions are unwarranted.

The main reason for this is that a conclusion about the truth-value of a proposition depends on more than a test outcome alone.<sup>64</sup> Moreover, consumers of expert information need to critically inquire about the extent to which the aggregate performance measure, e.g. the underlying error rate data, refers to types of cases (and related case circumstances) that are comparable to the instant case.<sup>65</sup> Again, this is not to say that error-rate data are useless; indeed, they are helpful to inform discourses regarding admissibility of a method or technique *in general*.<sup>66</sup> In other words, error-rate data, base rates and similar statistics are—at best—informative, but not conclusive, about the single event. This insight reminds us about the conceptual and methodological problems in the process of communication between forensic *scientists* on the one hand and *triers of fact* on the other hand.

Therefore, once a type of examination (or testimony) is deemed to be admissible (unless some exclusionary rule applies), we need to assess the probative value of the test outcome in the light of the task-relevant circumstantial information of the case at hand. The factual report of the test outcome, accompanied by aggregate performance indicators, is an *uninterpreted* finding. It is uninterpreted because it does not

<sup>61</sup> PCAST, 2016: 46.

<sup>62</sup> For a discussion of the topic of error rates see, e.g. KOEHLER, 2013.

<sup>63</sup> For example, there is usually non-scientific data in a case that informs what in more technical accounts is called a prior probability (i.e. a probability prior to considering the scientific evidence). See, e.g. PARMIGIANI, 2002: Chap. 1.

<sup>64</sup> *Ibid.*

<sup>65</sup> For a call of a more detailed scrutiny of the parameters of the studies from which data are drawn and used for assessments in instant cases, see IMWINKELRIED, 2020.

<sup>66</sup> *Ibid.*

directly inform the consumers of expert evidence *how* they should revise, in the light of the test outcome, their view on the pair of contrasting propositions of interest.

In order to give the test outcome an interpretation in the sense understood in this paper—i.e. the principles of scientific interpretation (Section 2.2.)—one not only needs a probability assignment for observing the test outcome (e.g. a positive result) *given* the target proposition (e.g. presence of a detectable amount of illicit substance). One also needs the probability of obtaining the test result given the appropriate alternative proposition (absence of illicit contents). Common proxies<sup>67</sup> for these probabilities are the sensitivity for the former and the probability of a false positive (i.e. 1-specificity) for the latter.<sup>68</sup> Combined, these two assessments characterise the value of the evidence: the evidence is probative in one way or another to the extent that the method's sensitivity differs from the complement of the specificity (i.e. the probability of a false positive).

Again, this should conform to laypeople's understanding and common intuition. As an example, consider a situation in which the sensitivity is the same as the probability of a false positive. That is, the probability of obtaining a positive test result is the *same*, regardless of whether or not a detectable quantity of illicit substance is present (i.e. the target proposition is true). Such a test result would have *no* probative value. A positive result has probative value if and only if the probability of the positive result is greater given the first rather than the second proposition, or vice-versa. More specifically, the greater the sensitivity compared to the false positive probability, the more probative the positive test result is with respect to the first rather than the second proposition.<sup>69</sup> Stated otherwise, the smaller the false positive probability, compared to the sensitivity, the greater the diagnosticity of the evidence. To summarise, a good test or method should have a high sensitivity and a high specificity (i.e. low false positive probability). Thus, interpreting a finding amounts to addressing points 1 to 3 exposed in Section 2.2., also known in the context here as «the logic of forensic inference».<sup>70</sup>

One question that these considerations raise is whether the PCAST Report depicts an unsuitable approach to the interpretation of forensic science evidence and forensic science as such. Some commentators think so, calling it a «flawed paradigm for forensic evaluation».<sup>71</sup> In a less severe view, however, one could argue that the PCAST Report focused on attempting a minimal first move ahead because, as mentioned above, aggregate performance indicators (e.g. error rates) are an important preliminary consideration (i.e. with respect to admissibility).<sup>72</sup> Empirical data to

<sup>67</sup> We use the term proxy here because, strictly speaking, the statistics refer to observations across multiple cases, whereas the problem to be dealt with here is assessment in the individual case.

<sup>68</sup> KAYE, 1987: 360.

<sup>69</sup> Ibid.

<sup>70</sup> EVETT et al., 2017: 17.

<sup>71</sup> Ibid, at 18, and IMWINKELRIED, 2020.

<sup>72</sup> See also LANDER, 2017: 367-368.

help assign probabilities for false positive outcomes are also helpful by providing, in certain applications, half of the «equation» stipulated by the three principles of scientific interpretation (Section 2.2.). But, by considering only part of the principles of coherent reasoning, the perspective advocated by the PCAST Report leaves scientific evidence essentially uninterpreted. It provides no guidance regarding what exactly recipients of expert information are logically entitled to infer from a particular test result when merely accompanied by an error rate only.<sup>73</sup> Stated otherwise, the PCAST report did not proceed to address what may be called the micro-level of value of evidence assessment. That is, it stayed at the macro-level of admissibility. In this sense, the perspective advocated by the PCAST report abstains from the case-based assessment of the probative value of scientific evidence.

Despite all scepticism, it is worthwhile to consider a possible argument in favour of the PCAST Report. If one considers that, as mentioned above, black-box-like studies represent an essential, though minimal, component of performance assessment, the PCAST report triggered the hope that the field would agree to conduct more ground truth testing (i.e. the generation of data for empirical performance indicators).<sup>74</sup> Rather disturbingly, however, even that hope set the bar too high for some quarters of forensic science that expressed their discontent with the PCAST Report's insistence on empirical grounding.<sup>75</sup>

#### 2.4. Uninterpretability: conceptualising the frontiers of interpretability

Suppose, for the time being, that we accept the PCAST Report's insistence on empirical foundations as an important preliminary to the consideration and admissibility of expert witness testimony in the legal process. Assume further that, on the micro-level of interpretation in the instant case, it is reasonable to require that the forensic scientist addresses the tripartite question outlined in Section 2.2. These questions seek to determine whether, and to what extent, an item of evidence is probative with respect to particular propositions of interest. According to this view, a probative item of evidence is one that is capable to influence, in one way or another, one's view (or belief) regarding the propositions of interest. In legal scholarship, the

---

<sup>73</sup> This is not a problem specific to the PCAST Report. Academics remain vague on how to use diagnostic performance measures properly too. For example, EDMOND et al., 2014, write «[t]he juror can reason with this information to infer something about the particular case» (at 10), without explaining what exactly this inference should amount to and despite the fact that the answer to this question is largely available in literature.

<sup>74</sup> The authors are grateful to Professor Christophe Champod for this reflection.

<sup>75</sup> For a list of published statements in response to the PCAST Report see [https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast\\_forensics\\_2016\\_public\\_comments.pdf](https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_forensics_2016_public_comments.pdf) (last accessed March 4 2021).

latter property is readily recognised as (logical) relevance,<sup>76</sup> a notion that can also be reconstructed in probabilistic terms.<sup>77</sup>

Broadly speaking, assessing evidence thus means assessing whether the evidence is probative with respect to particular propositions of interest. Central to this understanding is that the value of evidence is a property that depends on the intended purpose, the questions of interest in the case at hand (i.e. the litigated matter),<sup>78</sup> the assumptions made and the information available to the interpreter at the time an interpretation is made.<sup>79</sup>

While this working definition of interpretation purports to answer the question of what interpreting an item of scientific evidence means, it does not answer the question of what an (*un-*)*interpretable* item of evidence is. The former, i.e. interpretation, has to do with the logic of inference whereas the latter, (*un-*)interpretability, is an ascription made to a piece of evidence by an interpreter. Prima facie, it might be understood as a property of an item of evidence, but ultimately, it hinges upon the interpretative capacities of the interpreter. We can now delve deeper into these intricate aspects.

The definition of interpretation of scientific evidence through the three principles exposed in Section 2.2. allows us to derive a *provisional* definition of interpretability as a substantial property of the relationship between an item of evidence and the latter's interpreter. That is, an item of evidence can be said to be *interpretable* if and only if it is possible to enact the principles of interpretation.<sup>80</sup> In turn, an item of evidence is said to be *uninterpretable* when the interpreter cannot enact the principles of interpretation and, hence, no statement can be given as to the probative value of the evidence (i.e. asserting which of the competing propositions, if any, is favoured by the evidence over the respective alternative). This definition may seem a statement of the obvious, though it is crucial to ask—as emphasised in the remaining parts of this paper—whether the mere enactment of the principles of interpretation is sufficient to ensure that an interpretation is actually purposeful.

The argument made here is that interpretability is a *relational property*, in the sense that an item of evidence may be interpretable with respect to a given pair of propositions, but uninterpretable with respect to another pair of propositions.<sup>81</sup> For example a DNA profile is, strictly speaking, interpretable mainly as regards the source of the biological material, but requires additional considerations when discussed in

<sup>76</sup> For a probabilistic understanding of the term relevance, in the forensic sense, see STONEY, 1994.

<sup>77</sup> E.g. LEMPERT, 1977; FRIEDMAN, 2017: 49-84.

<sup>78</sup> AITKEN et al., 2010: §2.2.

<sup>79</sup> JACKSON, 2000: 85.

<sup>80</sup> Note, however, that an interpretable item of evidence is not necessarily probative. See §I.E. regarding the distinction between interpretability and probative value.

<sup>81</sup> As an aside, note also that interpretability may vary across examiners, depending on their level of expertise for the evidence in the instant case.

the context of alleged transfer mechanisms.<sup>82</sup> Again, this may seem obvious, but the problem is that in practice, this ambivalence is not always properly acknowledged, or even understood, which may result in seemingly probative evidence, leaving vital considerations of uninterpretability unnoticed—possibly to the detriment of the parties.

To account for the ambivalent nature of interpretability, it is helpful to consider a distinction between what we call here *weak* (or, limited) and *strong* (or, strict) interpretability:

— *Weak (or limited) interpretability*: evidence is said to have weak (or, limited) interpretability if, at best, it is interpretable with respect to a pair of propositions at a hierarchically low level, such as source level (i.e. a proposition regarding the source from which a trace or mark comes from).<sup>83</sup>

— *Strong (or strict) interpretability*: evidence is said to have strong interpretability if it is interpretable at an advanced propositional level, in particular with respect to propositions regarding alleged activities.

The pertinence of this distinction is illustrated by the fact that scientists may content themselves with limited interpretability, which is problematic because it may deprive recipients of expert information from the fact that the evidence is actually uninterpretable with respect to higher propositional levels that are closer to the ultimate issue of a case. Worse, when left unassisted, recipients of expert information may—in an unwarranted way—apprehend weakly interpretable evidence as strongly interpretable evidence. At the very least, the notion of weak interpretability emphasises the understanding that «[s]ource level propositions [...] may have limited utility for criminal adjudication».<sup>84</sup>

To illustrate the ambivalent nature of interpretability, consider an example involving DNA recovered on a discarded object, such as headgear, a tool or a weapon, found on a scene of investigation:<sup>85</sup> a small quantity of DNA (also sometimes called trace or touch DNA) is found on the object. The DNA profile corresponds to the DNA profile of a person of interest. Depending on the quality of the DNA profile (which may be a partial or complete, mixed or unmixed), the observed correspondence in DNA profiles may be probative with respect to the propositions according to which the person of interest rather than an unknown person is the source of the

<sup>82</sup> For more discussion see e.g. KOTSOGLOU & McCARTNEY, 2021.

<sup>83</sup> Regarding the notion of hierarchy of propositions, see COOK et al., 1998a. Broadly speaking, the higher a proposition in the hierarchy, the closer it approaches the ultimate issues in a case, and hence the more useful an interpretation is for fact-finders. For a recent overview of the hierarchy of propositions, see also GITTELSON et al., 2016.

<sup>84</sup> ATKEN et al., 2010: §3.9. For a recent example of this argument in the context of a case involving small quantities of DNA, see *United States v. Gissantaner*, No. 19-2305 (6th Cir. 2021).

<sup>85</sup> For an example of a real case involving a hat found at a crime scene, see BIEDERMANN et al., 2011. Similarly, *United States v. Gissantaner*, supra note 84, involving a weapon.

DNA.<sup>86</sup> Thus, according to our account, the DNA evidence is interpretable with respect to (sub-)source level propositions. This, however, represents only limited interpretability because the source level propositions are distant from the ultimate issues to be decided by the triers of fact. The latter are primarily interested in questions such as whether or not the person of interest is the individual who wore the headgear at the crime scene (or handled the weapon at the relevant time, etc.). Such propositions are at a higher hierarchical level, and interpretability with respect to such propositions would require *strong* interpretability. However, ensuring strong interpretability in this sense is a highly intricate task because it requires more than scientific knowledge regarding the rarity of the genetic features. The scientist would need to discuss scientific knowledge, if available, regarding phenomena such as transfer, persistence and rates of recovery. This requires data from empirical research. When knowledge and data on these topics are scarce, or even missing, the scientist may be unable to assess the probative value of DNA profiling results with respect to the alleged activities. This inability to enact the principles of interpretation will make the evidence *uninterpretable at this advanced propositional level* (i.e. activity level), despite the interpretability at source level. As mentioned above, evidence interpretability is relational insofar as it refers to a specific proposition, not to some intrinsic property of a piece of evidence.

The critical reader may wonder why this poses a problem and why it is not sufficient that scientists fall back to a less ambitious interpretational level (i.e. source level), and thus simply do what they «know best»—after all, this should lead to more robust findings. There are two problems with this intuitive idea. First, the operations of forensic experts is not an anything-goes activity, for this would run contrary to the «good old way» according to which expert witnesses act simply as «helpers of the court».<sup>87</sup> Secondly, a critical problem is that consumers of expert information may be tempted to carry the evaluation given source level propositions over to a conclusion regarding alleged activities—in an argumentatively unwarranted way.<sup>88</sup> This transition may be unwarranted because, as argued above, such a move requires specialised knowledge (e.g. regarding phenomena such as transfer and persistence) which, in case of evidence uninterpretability, is unavailable to the scientist. And, if such specialised knowledge is unavailable to scientists, it is also unavailable for consumers of expert information.

This problem may remain undetected, and thus be aggravated, when the scientist does *not* alert fact-finders that the evidence is *uninterpretable* with respect to higher-level propositions (i.e. alleged activities). The bottleneck, thus, is lacking spe-

---

<sup>86</sup> Note that, strictly speaking, this would be called sub-source level propositions because the question of source only refers to the DNA, not to the trace as such. The reason for this is that because of the low quantity of recovered material, the type of biological material (e.g. saliva) from which the DNA comes from, cannot be determined. See, e.g. EVETT et al., 2002: 521.

<sup>87</sup> THAYER, 1892: 665.

<sup>88</sup> E.g. GILL, 2014.

cialised knowledge. This creates the danger that laypeople—unaware of the informational problem—ignore this bottleneck and draw conclusions from the evidence *despite* its uninterpretability with respect to higher-level propositions.<sup>89</sup> Where this happens, uninterpretable evidence can litter the paths of reasoning towards primary matters of litigation (ultimate issues). Such evidence, thus, is only seemingly probative.

It follows from the above that whenever the recipient of expert information cannot be considered to be competent to assess the meaning of evidence with respect to more advanced propositional levels (such as alleged activities), and when these levels reflect the needs in the case, it is not appropriate for the scientist to limit evidence evaluation to only a low propositional level (such as source level).<sup>90</sup> Instead, the scientist should either assess what, if anything, the evidence means with respect to the higher propositional level, provided the necessary specialised knowledge is available, or else alert the fact-finder that the evidence is *uninterpretable* with respect to the advanced propositional level, because of the lack of relevant specialised knowledge.

Notwithstanding these intricacies, the default mode of assessment, especially in forensic genetics, still widely relies on propositions regarding the source of DNA (i.e. so-called sub-source level propositions).<sup>91</sup> This includes applications of modern probabilistic genotyping<sup>92</sup> software that scientists use to evaluate DNA profiling results with respect to propositions according to which the person of interest is (or is not) the (or one of the) contributor(s) to the DNA in the trace.<sup>93</sup>

The sceptical reader may retort that recipients of expert evidence can make disciplined use of evidence and that they do not unduly extend the value of evidence

---

<sup>89</sup> This danger is particularly critical whenever the evidence has a high probative value with respect to source level propositions (i.e. a hierarchically low propositional level), but *not* with respect to higher propositional levels regarding, for example, alleged activities.

<sup>90</sup> This is stipulated, for example, in the ENFSI Guideline: «Propositions are not altered during examination/evaluation unless the key issues in the case and/or the conditioning information have changed. For example, when the issues at hand are at activity level, the absence of data or expert knowledge on transfer, persistence or background level of the trace type under consideration is not a justification to change the set of activity level propositions to a set of source level propositions. In fact, the choice between (sub-) source and activity should not be influenced by the availability of data or expert knowledge but solely from the consideration of factors such as transfer, persistence and background levels that could crucially affect the strength of the findings within the context of the case circumstances» (ENFSI, 2015: 13). Similarly, the Recommendation 4 of the guidelines issued by the DNA Commission of the International Society for Forensic Genetics specifies: «Propositions should be formulated at a level in the hierarchy of propositions, where the forensic scientist brings knowledge that is required, but not readily available to the court. This is crucially important when transfer, persistence and background DNA have a significant impact on the case» (GILL et al., 2020: 5).

<sup>91</sup> On the term sub-source see *supra* 86.

<sup>92</sup> More detailed discussion on probabilistic genotyping is presented later in Section 3.4.

<sup>93</sup> BRIGHT et al., 2016; COBLE & BRIGHT, 2019; BUCKLETON et al., 2019 (noting that «PG [probabilistic genotyping] software and other interpretation and statistics methods evaluate the DNA results at the subsources level», at 395).

to an advanced proposition level where the evidence is actually uninterpretable. So, what could be wrong then with adhering to an evaluation given source level propositions?

The problem is that an evaluation given source level propositions is not a general solution for all cases. Indeed, in some cases, the question of source may not be disputed at all—but then the parties would not need expert evidence to back their claims. To continue with the DNA example introduced above, suppose that the person of interest concedes that the object (e.g. headgear, weapon, tool etc.) is theirs, but then adds that it was lost or stolen, say, sometime prior to the incident.<sup>94</sup> In such a case, it may not be contested that the corresponding DNA is that of the person of interest.<sup>95</sup> Thus, rather than the issue of source, the competing accounts are whether the person of interest is the man who wore the headgear at the crime scene versus an unknown person of interest wore the headgear (and that the person of interest last wore it, e.g. two weeks ago). Clearly, an expression of the value of the evidence with respect to source level propositions is useless here for there is no need to prove something that is not contested. Instead, of interest here is only the extent to which DNA evidence is capable of discriminating between alleged activities; and, if the evidence cannot be assessed with respect to the alleged activities, this means that it is uninterpretable at this propositional level, and thus has no probative value. It would be unwarranted then to shift attention to a lower, but *irrelevant* propositional level (i.e. source level) for which the evidence may be informative.

## 2.5. Distinguishing (un-)interpretability from probative value and relevance

Whether or not meaning can be assigned to an item of evidence is, according to the view defended in this paper, at the core of the notion of (un-)interpretability. It is understood here as a property that characterises the relationship between evidence and specific propositions as articulated by the interpreter. It is important, however, to draw a distinction with respect to the notion of probative value. In particular, it should be emphasised that even though a piece of evidence may be interpretable, i.e. is one can enact the principles of interpretation and assign the value of the evidence, this does not imply that the evidence is probative. Indeed, an interpretable item of evidence may be found to be neutral and, hence, irrelevant. That is, the evidence is not probative in one way or another with respect to the competing propositions of interest to the fact-finders.

---

<sup>94</sup> See, e.g. JACKSON et al., 2014: 50, for an example of this type of case, involving a balaclava discarded by an offender on the scene of a robbery.

<sup>95</sup> Similarly, for example, it is unsurprising to find a DNA profile corresponding to the profile of a person of interest on an object that has been seized in the apartment where that person lives (or in that person's car, etc.). For a practical example see, e.g. *United States v. Gissantaner*, supra note 84.

Consider Table 1 for a summary of the distinction between (un-) interpretability and probative value. Clearly, when the principles of interpretation cannot be enacted, and hence the evidence is uninterpretable, no statement can be made regarding the probative value. In essence, thus, an uninterpretable item of evidence provides no assistance to anyone who seeks to draw an inference about the competing propositions of interest—such evidence is not relevant.<sup>96</sup> But this may also happen with interpretable pieces of evidence. That is, when the principles of interpretation lead the scientist to conclude that whichever proposition is true, the probability of observing the evidence is the same. Such evidence is interpretable, but not probative in one way or another. Evidence is probative in one way or another if and only if the evidence is better conceivable with one proposition rather than the respective alternative proposition.<sup>97</sup>

(Un-)interpretability	Probative value: the evidence ...		
Weak / strong interpretability (Section 2.4)	<i>is probative</i> : it supports the defence proposition over the prosecution's proposition	<i>is not probative</i> : does not support one proposition over the other (i.e. the evidence is inferentially neutral)	<i>is probative</i> : it supports the prosecution's proposition over the defence proposition
Uninterpretability	No probative value can be assigned		

Table 1: Distinction between (un-)interpretability and probative value.

## 2.6. (Un-)interpretability vs. inconclusiveness

It is useful to distinguish the concept of (un-)interpretability from other terms commonly used by forensic examiners. One such term is «inconclusiveness». This term is mainly used by forensic scientists working in disciplines that involve comparative examinations (e.g. fingerprint experts). For example, according to a recent directive by the U.S. Department of Justice (DOJ), called Uniform Language for Testimony and Reports for the Forensic Latent Print Discipline (ULTR), inconclusive is (to be) used when «the examiner is unable to identify or exclude the two impressions as originating from the same source».<sup>98</sup> According to this definition,<sup>99</sup> the term in-

<sup>96</sup> This corresponds to the doctrinal understanding of relevance. See also *supra* note 76.

<sup>97</sup> E.g. FRIEDMAN, 2017: 70.

<sup>98</sup> U.S. DEPARTMENT OF JUSTICE, 2020: 3 (hereinafter cited as ULTR, short for «Uniform Language for Testimony and Reports for the Forensic Latent Print Discipline»).

<sup>99</sup> The term inconclusive is used in a variety of forensic disciplines, and the reasons which lead examiners to use this conclusion are diverse. For example, it may be that there is simply not enough reference material from a known source to conduct meaningful comparative examinations between materials from known and unknown source.

conclusive means that the findings (observations) are such that the examiner cannot assert whether or not the person of interest is the source of a recovered fingerprint. Note, however, that for the logical and procedural reasons explained in Section 2.2., expert witnesses should not opine directly on propositions, i.e. express themselves in terms of whether a person of interest is the source of a mark or trace. When redirecting the focus on the findings, rather than the propositions, inconclusiveness means that the findings are considered not probative in one way or another (i.e. in favour or against the proposition of same versus different source). Inconclusive is, in essence, taken to mean that the findings are neutral. But even this is not particularly helpful because it provides no insight into the deeper reasons for this conclusion.

The notion of (un-)interpretability can help to disentangle the underlying tenets and distinguish itself from inconclusiveness. Table 1 clarifies that there are two situations in which the evidence is uninformative for recipients of expert information. Either the scientist *was able* to assess the probability of the findings given each of two competing propositions, but found that they were the same, and hence the evidence has no value. Note that according to the considerations exposed in Section 2.4., such evidence would be said to be interpretable because the scientist was able to enact the principles of interpretation (defined in Section 2.2.). Or, the scientist cannot enact the principles of interpretation, a situation in which the evidence is uninterpretable and, hence, necessarily uninformative.

More generally, the term inconclusive is internally contradictory because it suggests that one cannot «conclude» anything, yet asserting that the evidence has no probative value is as much a conclusion as asserting that the evidence has (some) probative value. Further, the use of the term inconclusive presupposes that evidence is either categorically conclusive or inconclusive,<sup>100</sup> a view that misconceives the reality that most evidence is only probative to varying degrees. For these reasons, «inconclusive» is a term that should be avoided altogether,<sup>101</sup> not least because it intrigues the consumer of expert information to think that the examiner is presenting an opinion on the propositions, or probabilities thereof, rather than a report on the value of the findings.

### 3. PROBING FOR (UN-)INTERPRETABILITY

Responsible consumers of expert evidence should take an interest in the topic of (un-)interpretability because it can help them gain a better understanding of the limitations of scientific findings, especially in cases where evidence is only weakly interpretable, i.e. at a propositional level that is far away from the questions (ulti-

---

<sup>100</sup> E.g. the ULTR, 2020, states: «The examiner may offer any of the following conclusions: 1. Source identification (i.e. came from the same source) 2. Source exclusion (i.e. came from different sources) 3. Inconclusive» (at 2).

<sup>101</sup> E.g. COLE & BIEDERMANN, 2020.

mate issues) that the fact-finder has to determine—and possibly uninterpretable at more advanced propositional levels, closer to the primary matters that fact-finders are concerned with. Exposing the limitations of interpretability should help parties guard against unwarranted carrying-over of evidence to advanced levels of the justificatory chain that leads to the final verdict. This section presents and discusses a series of considerations that are key to probing for uncertainty as part of a critical review of scientific evidence.

### 3.1. Distinguishing between evaluative, investigative and technical reporting

A useful first step in the review of a report from an expert witness is to understand the type of report insofar as (un-)interpretability—being relational in nature—is not a cause of concern for all report categories. Questions of (un-)interpretability primarily arise in so-called *evaluative* reports. An evaluative report is one that reports on the observations made during comparative examinations conducted on questioned and known materials,<sup>102</sup> and on the probative value the scientist assigns to those observations in the light of the propositions that reflect the views of the parties at various stages in the legal process.<sup>103</sup> Here, the notion of probative value refers to an expression of strength of support for one proposition compared to its alternative, derived by answering the question triad exposed in Section 2.2. It should be emphasised that the focus of a genuine evaluative report is on the observations and findings, and their discriminative capacity with respect to the propositions of interest—the evaluative report must not contain a direct opinion on the propositions, i.e. live issues.

A first urgent remark regarding the requirement outlined above is that the reality of forensic reporting is far removed from this confined methodological space. With the exception of forensic DNA (to some extent),<sup>104</sup> any fair characterisation of current reporting formats reveals that forensic scientists oftentimes opine directly on competing propositions,<sup>105</sup> without being prevented by the parties or the courts. For example, fingerprint analysts commonly state whether or not (they think that) the person of interest is identified as the source of the fingerprint found on the crime scene.<sup>106</sup> The latter is a statement about a live procedural issue rather than about the (probative) value of the similarities and differences observed when comparing the

---

<sup>102</sup> Questioned material is material whose origin or source is unknown. Typical examples are traces (e.g. blood, fibres, etc.) and marks (e.g. toolmarks, fingerprints, etc.) found on a crime scene, victim or a person of interest. Known (or reference) materials are materials whose origin is known (e.g. a named person or object).

<sup>103</sup> E.g. ENFSI, 2015: §1.1 and §2.1.

<sup>104</sup> E.g. GILL et al., 2018.

<sup>105</sup> For a critical analysis and discussion of current reporting language for federal examiners, see COLE and BIEDERMANN, 2020.

<sup>106</sup> E.g. CHAMPOD et al., 2016; THOMPSON, 2018: 775.

questioned fingerprint and the reference print of the person of interest. Although one might infer that the scientist must have considered the findings to be of (at least some) value, and hence interpretable, allowing them to opine directly on propositions is tantamount to (a) skipping the necessary preliminary stage of assessing the probative value of the observations made in the first place and (b) infringing, thus, the fact-finder's decision-making prerogative, which shows disregard for the procedural architecture of modern criminal justice systems, especially—we say this at the cost of oversimplification—the criminal process.<sup>107</sup> The former, (a), obstructs any insight into how the scientist assessed the value of the *observations*. Whether that assessment was thorough thus remains hidden, unarticulated and, ultimately, escapes proper scrutiny.

The above considerations clarify that evaluative reports involve material from a known source, i.e. a person or object (tool), that has been compared with some questioned item of unknown source. The reason for emphasising the need of a careful assessment of probative value, according to the principles exposed in Section 2.2., is that such an assessment can result in evidence for or against a person of interest. This needs to be distinguished from reports that do not involve results of comparative examinations with a potential source—and hence do not tend to directly implicate a person of interest. For example, during early stages of an investigation, when only questioned material from the crime scene is available, but no potential source, examiners may examine traces to provide information useful for investigation. For example, based on blood stains allegedly left by the offender at the crime scene, analysts may provide information regarding the DNA profile that the offender must possess.<sup>108</sup> Similarly, in a hit-and-run case, the examination of paint fragments found on the scene of the incident can reveal information regarding the (type of) vehicle that was involved. In such cases, the examiner's report is confined to describing the features of recovered material only. Such reporting is also sometimes called technical or investigative reporting.<sup>109</sup>

The above distinction between evaluative and investigative reporting is not necessarily neat. For example, it may be argued that a report is evaluative even though it involves no direct comparison with a potential source (person or object). For example, in the case of a seized substance, analyses may reveal the presence of a drug which may result in charges against a person of interest. Thus, strictly speaking, measurements and results should be assessed in the light of two competing propositions, e.g. that the material does (or does not) contain a detectable quantity of illicit substance. While this could be seen as an instance that raises the problem of (un-)interpretability, it is also possible to look at this type of case as an example of classification. This

---

<sup>107</sup> See BIEDERMANN & KOTSOGLOU, 2018.

<sup>108</sup> STONEY, 1994: 18.

<sup>109</sup> It is worth noting, however, that evaluative reporting often involves elements of technical reporting, in particular sections that report on the features of the recovered materials. ENFSI, 2015: §1.1.

is a task that seeks to determine the nature of examined material by assigning it to a particular category (e.g. illicit versus legal substance). The ENSI Guideline, for example, considers this type of reporting (classification) as technical because it is only concerned with ascertaining the nature of recovered material.<sup>110</sup>

The distinction between technical and evaluative reporting is important because it can help guard against the unwarranted use of investigative information for evidence and proof purposes at more advanced stages of the legal process. Suppose, for example, that a technical report asserts that the fragments of paint found on a victim in a hit-and-run case are blue and composed of several distinct layers. While this information may be helpful to direct investigations (i.e. searching a certain type of blue vehicles), it is unsuitable for direct use as evidence against a person of interest who is found to have a blue car. The reasons for this are, first, that detailed comparative examinations would need to be conducted between the recovered fragments and control material from the car of the person of interest. Second, the value of the observed similarities and differences would need to be assessed with respect to a pair of competing propositions (e.g. the recovered fragments come from the car of the person of interest vs. an unknown vehicle). Only such comparative examinations and the assessment of the probative value of their outcomes can lead to a proper evaluative report.

### 3.2. Uninterpreted and partially interpreted vs. uninterpretable evidence

Having clarified that interpretation in evaluative settings (as opposed to investigative settings, Section 3.1.) requires the forensic expert to consider a bipartite question, that is the probability of the observations (findings) *given* a pair of contrasting and competing propositions and the task-relevant information (Section 2.2.), we can now use this framework to delineate different interpretational limitations.

A common objection against the principles of evaluative reporting is that the requirement for at least one alternative proposition could be misunderstood as a practical obligation. Clearly, to ensure helpfulness, the contrasting set of propositions that the forensic scientist must take for granted when assessing the probative value of the findings maps on the actual procedural claims of the parties. But this is an empirical rather than a doctrinal observation, and does not necessarily map neatly on real litigation tactics in each case. For the defendant has the right to remain silent or can put forward some other claim.

Yet, at the time when forensic experts write their reports, they may not have been provided with an alternative proposition because, strategically, the defence may choose to remain silent until they learn what the scientist has observed (e.g. whether

---

<sup>110</sup> ENFSI, 2015: §5.0 («Reporting, technical (factual)').

or not there is a correspondence between the DNA profile of the crime stain and the DNA profile of the defendant, or whether any DNA is found at all). In the absence of an alternative proposition, there are—according to the ENFSI Guideline<sup>111</sup>—three options that the scientist may follow, two of which will leave the findings *uninterpreted*:

- 1) Adopt an alternative proposition that, given the circumstances of the case, most suitably reflects the position of the party.
- 2) Adopt no alternative proposition.
- 3) Present a range of plausible explanations of the findings.

It is important to note that only option 1 can lead to a proper evaluative report, because the evaluation will be based on a pair of competing and mutually exclusive propositions. However, although experts should be prepared to review their evaluation of the findings when they are provided with a proposition other than the one assumed by default, option 1 remains tricky. In fact, it is problematic for a party to leave the choice of an alternative proposition to the scientist because the scientist's choice may result in an evaluation that is less favourable for the party of interest than it could be.<sup>112</sup>

Option 2 implies that the scientist cannot enact the principles of interpretation, thus leaving the findings uninterpreted.<sup>113</sup> When there is only one proposition, the scientist may, at best, assess the extent to which the findings may be expected to occur given the single proposition available. This amounts to working half-way through the principles of interpretation (Section 2.2.), meaning that the findings can be considered only partially interpreted. Most importantly, however well conceivable the findings may be with the single proposition available, *no* statement can be made about the credit (or, support) the findings provide for the single proposition available. For example, when there is a biological stain (evidentiary item) with a DNA profile that corresponds to the DNA profile of a person of interest, this observation—taken on its own—has *no* probative value as long as the scientist has not considered the findings in the light of at least one alternative proposition. In particular, there are no logical grounds to assert that the observed correspondence in DNA profiles somehow supports the proposition according to which the person of interest is the source of the DNA stain, however appealing such a conclusion might appear.

Option 3 is more subtle. It amounts to the scientist exploring a list of explanations for the findings, and perhaps ranking those explanations according to their plausibility. There are two caveats with this option. The first is that explanations,

---

<sup>111</sup> ENFSI, 2015: §4 (Guidance note 1).

<sup>112</sup> It is well known, for example, that in a case where the source of a DNA stain is contested, the default alternative proposition that the DNA comes from an unknown and unrelated person tends to maximize the probative value (BUCKLETON & TRIGGS, 2005).

<sup>113</sup> See ENFSI, 2015: §4 (Guidance note 1): «in the absence of an alternative proposition, it is impossible to evaluate the findings».

as understood by scientists, are not the same as formal propositions used in the standard evaluative setting (Sections 2.2. and 3.1.). Propositions are formal, specified by the parties and strictly anchored in the framework of circumstances of the instant case. In turn, explanations produced by scientists may be open-ended and exploratory (or even speculative) without necessarily being rooted in the case-specific circumstances.<sup>114</sup> Often, such explanations are provided almost by default, as a deflecting shield,<sup>115</sup> without even inquiring about whether they are pertinent to the case at hand.<sup>116</sup> While this may be suitable at the investigative stage, i.e. when the questions of interest focus on what happened and how, contemplating explanations cannot lead to an evaluative report understood in the sense explained in Section 3.1. An evaluative report states the value of the findings for discriminating between propositions of interest that specify an individual or object as the source of a mark or trace, or well-defined activities whereby the mark or trace has been generated.<sup>117</sup> Thus, recipients of expert information should keep in mind that an explorative report in the sense of option 3 does not represent an evaluative report in the sense of the criteria of interpretation exposed in Section 2.2. An explorative report leaves the observations and findings uninterpreted. Explaining the findings and interpreting findings are two fundamentally different tasks.

The above situations illustrate practical and procedural obstacles that can prevent scientists from enacting the principles of interpretation, thus leading to uninterpreted or, at best, only partially interpreted evidence. The proposed report categories can help recipients of expert information understand the nature of experts' reports (or testimony), recognize their limitations and identify instances where the probative value of the evidence remains undetermined. This is the case for situations 2 and 3 as outlined above.

Notwithstanding, it is important to remind that even in a case corresponding to situation 1, i.e. when the scientist evaluates the findings with respect to at least two competing propositions, there is no guarantee that the evidence is actually interpretable. As an example, consider again the hypothetical case presented in Section 2.4., where a small quantity of DNA was detected on an item handled by an offender (e.g. headgear, tool, etc.), and discarded on the crime scene. When the DNA profile of the trace corresponds to the DNA profile of the person of interest, it is not an easy task to assess the value of this correspondence with respect to so-called activity level propositions, e.g. the person of interest wore the balaclava at the relevant time vs. the person of interest wore the item a few weeks ago (before it went missing or

---

<sup>114</sup> EVETT et al., 2000.

<sup>115</sup> The idea of a deflecting shield is the protection of the scientist against accusations of not providing an exhaustive list of all conceivable explanations of the evidence.

<sup>116</sup> For an example see, e.g. the default statements provided in the context of gunshot residue analyses mentioned in Section 1 of this paper.

<sup>117</sup> COOK et al., 1998a.

got lost).<sup>118</sup> The difficulty of assessing the findings given such propositions is due to the fact that data on phenomena such as transfer and persistence of DNA under the circumstances of the instant case are scarce. Specifically, when the scientist cannot assess the probability of observing the DNA findings given each activity level proposition and the task-relevant case information, this means that the principles of interpretation cannot be enacted, and hence that the evidence must be considered uninterpretable at this propositional level, *despite* possible interpretability at source level. It is also important, at this juncture, to recall the practical implication of uninterpretability and to emphasise the critical need for parties to understand this implication: uninterpretable evidence means that it provides no assistance to anyone asked to decide which of the competing propositions is true.

### 3.3 The central role of main objects of litigation and propositions in interpretation

Successfully probing evidence for (un-)interpretability hinges upon one's ability to keep track, as pointed out in Section 2.4, of the relational nature of evidence (un-)interpretability. That is, the feasibility of ascertaining evidential value can vary as a function of the propositions with respect to which the evidence is assessed, the available task-relevant information and any additional assumptions made. But, if the value of evidence depends on the propositions being considered, how should one define propositions suitably? This is a crucial question because expert reports are often unclear about the reasons that led to the definition of the propositions that have been retained. Experts have a natural tendency to shape propositions in such a way that their findings will be interpretable with respect to the chosen propositions. However, as emphasised in Section 2.4., interpretability is not tantamount to, nor does it guarantee or imply helpfulness. Interpretability merely means that the expert *can* enact the principles of interpretation. Whether they *should* enact these principles in a given way is contingent on procedural circumstances: the claims that parties bring forward and the claims that remain disputed. Remember that forensic science serves the purposes of criminal justice, not vice versa. For example, a person of interest may not contest that an item (e.g. headgear, tool, etc.) is theirs, and hence them being the source of trace material (e.g. DNA, fingerprints etc.) found on the item. Instead, there may be disagreement over the activities that led to the deposition of DNA or fingerprints on the item of interest. Clearly, there is no need to adduce evidence for an uncontested issue. Thus, as noted by Aitken et al., it is important to understand what the main object of litigation is:

---

<sup>118</sup> Recall that the question of source (i.e. whether or not the DNA comes from the person of interest) may not be of interest because the person of interest may not contest that the item or object is theirs. Instead, they may contest the way (i.e. activity) that led to the deposition of DNA on the item or object. For further discussion, see Section 3.3.

One cannot assess whether evidence is successful in proving a matter in issue until one knows what the issue is and how the evidence relates to it. This observation might sound banal; but it is not. In fact, nearly all of the reasoning errors [...] are either variations on, or are at least exacerbated by, an elementary failure to identify, with sufficient care and particularity, the question which the evidence is capable of answering.<sup>119</sup>

A corollary for parties thus is that they should insist on the understanding that, first, the meaning of evidence depends on the intended purpose(s) for which it is adduced (i.e. the primary object of litigation), and hence, second, that it is these objects of litigation—not minimal interpretability—that define the propositions with respect to which evidence is to be evaluated. If that may result in evidence being uninterpretable, then this is worthy of being exposed. This is vital, especially for the defence. Indeed, it would be disingenuous to change propositions away from the principal matters of litigation for the sole purpose of making the evidence interpretable (though with respect to an uncontested aspect of the case). Such evidence, if uninterpretable with respect to the principal questions, will remain so; interpretability with another set of non-principal questions is prone to lead to seemingly probative evidence. That is, it is prone to being carried over to a conclusion that the evidence cannot logically support.

### 3.4. Exemplifying (un-)interpretability: probabilistic genotyping

Probabilistic genotyping (PG) is the combined use of biological models, statistics and computational methods to help assess complex DNA profiling results,<sup>120</sup> in particular traces with DNA from more than one person (i.e. DNA mixtures). Various software solutions exist to practically implement PG and they are regularly debated in courts.<sup>121</sup> PG aims at quantifying the probative value of DNA profiling results with respect to propositions according to which a person of interest is, or is not, a contributor to the detected DNA together with, in both propositions, a certain number of other individuals (e.g. claimant and/or unknown person(s)).<sup>122</sup> This type of assessment is fundamentally compatible with the principles of interpretation exposed in Section 2.2.: the assessment informs about the probability of obtaining the DNA profiling results *given* competing propositions regarding the source(s) of the detected DNA. Thus, the output of applying PG in an instant case is an assessment of the probative value of the DNA profiling results with respect to particular propositions. More specifically, these are called sub-source level proposi-

<sup>119</sup> AITKEN et al., 2010: §3.3.

<sup>120</sup> E.g. SCIENTIFIC WORKING GROUP ON DNA ANALYSIS METHODS (SWGDM), 2015; AAFS STANDARDS BOARD, 2020.

<sup>121</sup> E.g. *United States vs. Gissantaner*, 2019 U.S. Dist. LEXIS 178848, 2019 WL 5692183 (W.D. Mich. Oct. 16, 2019) and *United States v. Gissantaner*, supra note 84.

<sup>122</sup> E.g. BUCKLETON et al., 2019: 395-396.

tions<sup>123</sup> which, according to the framework exposed in Section 2.4., lead at best to weak interpretability. The reason for this is that such propositions are distant from so-called activity-level propositions that more closely approach the principal issues of interest to fact-finders. Most importantly, it is crucial to understand that a brute PG result is *uninformative* with respect to activity level propositions. Recall that, as mentioned in Section 2.4., an evaluation of the findings with respect to more advanced propositional levels requires specialised knowledge regarding aspects such as transfer and persistence of DNA in the light of the circumstances of the instant case. But Buckleton et al. insist that this is beyond the scope of PG: «Discussion of transfer and persistence [...] has nothing to do with PG». <sup>124</sup> Consumers of expert information should thus carefully distinguish between a PG result and an assessment of the findings for higher-level propositions. While the former, PG, may lead to impressive numbers, it is critically important to resist temptations to let such assessments unduly influence reasoning paths concerned with the latter, higher propositional levels regarding alleged activities.

For illustration, consider a hypothetical case in which A was physically assaulted by a man, resulting in serious injury.<sup>125</sup> Part of the uncontested case circumstances is that A fought with the offender, grabbing his clothing and his skin vigorously. Shortly after the assault, A's nails were grabbed for subsequent DNA analyses. Suppose that the DNA profiling analyses led to a DNA mixture showing a major and a minor contributor corresponding to, respectively, the DNA profiles of A and B. The defendant, B, denies any involvement. Further, B asserts that he had been scratched by A a few days ago during a minor argument, resulting in A reacting hysterically and disproportionately. Therefore, it might not be surprising —B says— to find a DNA profile that corresponds (partially) to his DNA, as is the case with a mixture between A's and B's DNA. In this case, both parties acknowledge that some sort of physical aggression *has* indeed occurred. This, in turn, can justify the assumption that the source of the DNA is not contested. It is the timing and, maybe, the intensity and duration of the quarrel over which there is disagreement. Hence, conducting PG would not directly help with a primary object of litigation except, maybe, with helping to narrow the circle of individuals to investigate. Recall that PG would only provide a result with respect to propositions regarding persons who contributed to the DNA trace, yet the question of source is not contested in this case.

Instead, what is contested, are activities, as is reflected by propositions of the following kind: «B has been assaulted by A» vs. «an unknown man has assaulted A (but A scratched B a few days prior to the assault)». Stated otherwise, the reasoning is directed towards discriminating between competing versions regarding the way by which, and the timing, DNA was deposited under/on A's nails. Assessing the

---

<sup>123</sup> On the concept of sub-source level propositions, see *supra* note 86.

<sup>124</sup> BUCKLETON et al., 2019: 395.

<sup>125</sup> The authors are grateful to Dr. Tacha Hicks for suggesting this case example.

probability of seeing the DNA profiling results given each of these propositions, and the case-specific circumstances, requires expert knowledge (e.g. structured data from experiments under controlled conditions) regarding the transfer and persistence of DNA under circumstances comparable to the alleged activities in the instant case. Such expertise may be scarce, or fragmentary, leading to only limited, if any, capacity to discriminate between the two propositions of interest, not least because the two alleged activities are similar in nature, differing only in timing. Should the scientist be unable to assess the findings with respect to the proposition regarding the alleged activities and circumstances, then the DNA profiling result would need to be considered *uninterpretable* at this propositional level. Further, the scientist should not be allowed then to retreat to reporting results of PG because the output of such an analysis would refer to other propositions (i.e. the source of DNA), however probative the findings may be with respect to such propositions. Indeed, there is no need to adduce evidence for propositions that are not contested. Doing otherwise is prone to let PG output unduly affect inference about a propositional level regarding which the evidence is actually lacking interpretability.

#### 4. CONCLUSIONS

This paper started with the observation that evidence, as we perceive it, is not necessarily what the mainstream view in forensic science claims to be. Real-world evidence, irrespective of its nature, does not come with a built-in, self-explanatory account of its probative force, i.e. with meta-rules of assessment. At the same time, the terms «interpretation» and «interpretability», though commonly used in scientific research and judicial practice, remain largely undefined in the context of expert evidence.

As a main strain of argument, this paper advocated a rule-governed, balanced and logically defensible definition of interpretability, based on principles derived from formal methods of reasoning, called here the principles of interpretation (Section 2.2.). We have equated (un-)interpretability with the forensic expert's (in-)ability to enact minimal principles of logical reasoning applied to expert evidence. Central to this view is that, fundamentally, interpretability is not an inherent or, as it were, an axiomatic property of evidence, but a function of the purpose for which it is adduced, and the specialised knowledge and expertise available to the scientist, necessary to relate evidence meaningfully to the circumstances of the instant case. Being contrastive by design, evidence analysis focuses on providing an assessment of the extent—which may well be nil (remember: interpretability does not necessitate relevance)—to which the evidence is able to discriminate between competing propositions of interest, but abstaining from providing a direct opinion on the propositions and litigated matters that are in the fact-finder's area of competence.

Though it can be considered a substantive property of the relationship between evidence and its analyst, interpretability is actually underspecified and underrecog-

nised in practice, thus rendering evidential processes vulnerable to unwarranted inferential reasoning, possibly detrimental to litigants. Specifically, where interpretability with regard to low-level case aspects, such as the source of evidential material, is carelessly extended to higher-level questions, such as alleged activities, *uninterpretability* may undermine or—as we argue in this paper—«litter» the fact-finders' inferential paths. Stated otherwise, where informational gaps due to uninterpretability are misconceived, actively ignored or concealed, evidence is prone to become seemingly probative for issues for which it is actually uninformative.

In all, there is more to «uninterpretability» than merely referring to evidence that does not answer a question of primary interest, i.e. an ultimate issue. Probing for uninterpretability is not the ultimate goal, but a means to an end. It is about ensuring the proper elicitation of the meaning of evidence, i.e. the rules for its use, and the avoidance of conclusions that go beyond the realm of what is logically warranted. Sceptical readers might argue that the insistence on the notion of (un-)interpretability is redundant because triers of fact may be able to proceed responsibly without it. This paper has argued that this would be short-sighted insofar as it presupposes that controlling for (un-)interpretability could be achieved in the absence of any inferential framework. Moreover, as we have shown, legal provisions are vulnerable to, or even tolerant towards, evidence whose probative value cannot be assessed.

## BIBLIOGRAPHY

- AMERICAN ACADEMY OF FORENSIC SCIENCES (AAFS) STANDARDS BOARD, 2020: *Standard for Validation of Probabilistic Genotyping Systems, ASB Standard 018* (1st ed.), <https://www.asbstandardsboard.org> (last accessed March 7 2021).
- AITKEN, C., 1995: *Statistics and the Evaluation of Evidence for Forensic Scientists*, Hoboken (New Jersey): John Wiley & Sons.
- AITKEN, C., BERGER, C., BUCKLETON, J., CHAMPOD, C., CURRAN, J., DAWID, A., EVETT, I., GILL, P., GONZÁLEZ-RODRÍGUEZ, J., JACKSON, G., KLOOSTERMAN, A., LOVELOCK, T., LUCY, D., MARGOT, P., MCKENNA, L., MEUWLY, D., NEUMANN, C., DAÉID, N., NORDGAARD, A., PUCH-SOLIS, R., RASMUSSEN, B., REDMAYNE, M., ROBERTS, P., ROBERTSON, B., ROUX, C., SJERPS, M., TARONI, F., K TJIN-A-TSOI, T., VIGNAUX, G., WILLIS, S. M. & ZADORA G., 2011: «Expressing evaluative opinions: a position statement», in *Science & Justice*, 51: 1-2.
- AITKEN, C., ROBERTS, P. & JACKSON, G., 2010: *Fundamental of Probability and Statistical Evidence in Criminal Proceedings*, London: Royal Statistical Society's Working Group on Statistics and the Law.
- AITKEN, C. & STONEY, D. A. (eds.), 1991: *The Use of Statistics in Forensic Science*, New York, London: Ellis Horwood.
- ARONSON, J. D., 2007: *Genetic Witness: Science, Law, and Controversy in the Making of DNA Profiling*, New Brunswick (New Jersey), London: Rutgers University Press.
- ASSOCIATION OF FORENSIC SCIENCE PROVIDERS (AFSP), 2009: «Standards for the formulation of forensic science expert opinion», in *Science & Justice*, 49: 161-164.
- BIEDERMANN, A., KOTSOGLOU, K. N., 2018: «Decisional dimensions in expert witness testimony—A structural analysis», in *Frontiers in Psychology (Cognition)*, 9: 2073.
- BIEDERMANN, A., TARONI, F. & THOMPSON, W. C., 2011: «Using graphical probability analysis (Bayes nets) to evaluate a conditional DNA inclusion», in *Law, Probability and Risk*, 10: 89-121.

- BRIGHT, J.-A., TAYLOR, D., MCGOVERN, C., COOPER, S., RUSSELL, L., ABARNO, D. & BUCKLETON, J., 2016: «Developmental validation of STRmix™, expert software for the interpretation of forensic DNA profiles», in *Forensic Science International: Genetics*, 23: 226-239.
- BUCKLETON, J. S., BRIGHT, J.A., GITTELSON, S., MORETTI, T. R., ONORATO, A. J., BIEBER, F. R., BUDOWLE, B. & TAYLOR, D., 2019: «The probabilistic genotyping software STRmix: utility and evidence for its validity», in *Journal of Forensic Sciences*, 64: 393-405.
- BUCKLETON, J. S. & TRIGGS, C. M., 2005: «Relatedness and DNA: are we taking it seriously enough?», in *Forensic Science International*, 152: 115-119.
- CHAMPOD, C., LENNARD, C., MARGOT, P. & STOILOVIC, M., 2016: *Fingerprints and other Ridge Skin Impressions* (2nd. ed.), Boca Raton, London, New York: CRC Press.
- COBLE, M. D. & BRIGHT, J.-A., 2019: «Probabilistic genotyping software: an overview», in *Forensic Science International: Genetics*, 38: 219-224.
- COLE, S. A. & BIEDERMANN, A., 2020: «How can a forensic result be a “decision”? A critical analysis of ongoing reforms of forensic reporting formats for federal examiners», in *Houston Law Review*, 57: 551-592.
- COOK, R., EVETT, I. W., JACKSON, G., JONES, P. J. & LAMBERT, J. A., 1998a: «A hierarchy of propositions: deciding which level to address in casework», in *Science & Justice*, 38: 231-239.
- , 1998b: «A model for case assessment and interpretation», in *Science & Justice*, 38: 151-156.
- DENNIS, I., 2017: *The Law of Evidence* (6th ed.), London: Sweet & Maxwell Ltd.
- EDMOND, G., THOMPSON, M. B. & TANGEN, J. M., 2014: «A guide to interpreting forensic testimony: scientific approaches to fingerprint evidence», in *Law, Probability and Risk*, 13: 1-25.
- EUROPEAN NETWORK OF FORENSIC SCIENCE INSTITUTES (ENFSI), 2015: *ENFSI Guideline for Evaluative Reporting in Forensic Science*, [http://enfsi.eu/wp-content/uploads/2016/09/m1\\_guideline.pdf](http://enfsi.eu/wp-content/uploads/2016/09/m1_guideline.pdf) (last accessed July 6 2021).
- EVETT, I. W., 1991: «Interpretation: a personal odyssey», in Aitken, C., Stoney, D. A. (eds.), *The Use of Statistics in Forensic Science*, New York, London: Ellis Horwood, 9-22.
- , 2009: «Evaluation and professionalism», in *Science & Justice*, 49: 159-160.
- EVETT, I. W., BERGER, C. E. H., BUCKLETON, J. S., CHAMPOD, C. & JACKSON, G., 2017: «Finding the way forward for forensic science in the US—A commentary on the PCAST Report», in *Forensic Science International*, 278: 16-23.
- EVETT, I. W., GILL, P. D., JACKSON, G., WHITAKER, J. & CHAMPOD, C., 2002: «Interpreting small quantities of DNA: the hierarchy of propositions and the use of Bayesian networks», in *Journal of Forensic Sciences*, 47: 520-530.
- EVETT, I. W., JACKSON, G. & LAMBERT, J. A., 2000: «More on the hierarchy of propositions: exploring the distinction between explanations and propositions», in *Science & Justice*, 40: 3-10.
- EVETT, I. W. & WEIR, B. S., 1998: *Interpreting DNA Evidence*, Sunderland (Massachusetts): Sinauer Associates Inc.
- FIENBERG, S. E. (ed.), 1989: *The Evolving Role of Statistical Assessments as Evidence in Courts*, New York: Springer.
- FINKELSTEIN, M. O. & FAIRLEY, W. B., 1970: «A Bayesian approach to identification evidence», in *Harvard Law Review*, 83: 489-517.
- FRIEDMAN, R. D., 2017: *The Elements of Evidence*, St. Paul (Minnesota): West Academic Publishing.
- GASTWIRTH, J. L., 1988: *Statistical Reasoning in Law and Public Policy*, Cambridge (Massachusetts): Academic Press.
- GILL, P., 2014: *Misleading DNA Evidence: Reasons for Miscarriages of Justice*, Cambridge (Massachusetts): Academic Press.
- GILL, P., HICKS, T., BUTLER, J. M., CONNOLLY, E., GUSMÃO, L., KOKSHOORN, B., MORLING, N., VAN OORSCHOT, R. A. H., PARSON, W., PRINZ, M., SCHNEIDER, P. M., SIJENJ, T. & TAYLOR, D., 2018: «DNA Commission of the International Society for Forensic Genetics: assessing the value of forensic biology evidence—Guidelines highlighting the importance of propositions, Part I: evaluation of DNA profiling comparisons given (sub-)source propositions», in *Forensic Science International: Genetics*, 36: 189-202.

- , 2020: «DNA Commission of the International Society for Forensic Genetics: assessing the value of forensic biological evidence—Guidelines highlighting the importance of propositions. Part II: evaluation of biological traces considering activity level propositions», in *Forensic Science International: Genetics*, 44: 102186.
- GITTELSON, S., KALAFUT, T., MYERS, S., TAYLOR, D., HICKS, T., TARONI, F., EVETT, I. W., BRIGHT, J.-A. & BUCKLETON, J., 2016: «A practical guide for the formulation of propositions in the Bayesian approach to DNA evidence interpretation in the adversarial environment», in *Journal of Forensic Sciences*, 61: 186-195.
- HEARD, B. J., 2008: *Handbook of Firearms and Ballistics* (2nd ed.), Hoboken (New Jersey): John Wiley & Sons.
- , 2013: *Forensic Ballistics in Court, Interpretation and Presentation of Firearms Evidence*, Hoboken (New Jersey): John Wiley & Sons.
- IMWINKELRIED, E. J., 2020: «The admissibility of scientific evidence: exploring the significance of the distinction between foundational validity and validity as applied», in *Syracuse Law Review*, 70: 817-849.
- JACKSON, G., 2000: «The scientist and the scales of justice», in *Science & Justice*, 40: 81-85.
- JACKSON, G., AITKEN, C. & ROBERTS, P., 2014: *Case Assessment and Interpretation of Expert Evidence*, London: Royal Statistical Society's Working Group on Statistics and the Law.
- JACKSON, G. & BIEDERMANN, A., 2019: «“Source” or “Activity”. What is the level of issue in a criminal trial?», in *Significance*, 16: 36-39.
- JACKSON, G. & JONES, P. J., 2009: «Case assessment and interpretation», in Jamieson, A. & Moenssens, A. (eds.), *Wiley Encyclopedia of Forensic Science*, Hoboken (New Jersey): John Wiley & Sons, 1-15.
- KAYE, D. H., 1987: «The validity of tests: caveat omnes», in *Jurimetrics Journal*, 27: 349-361.
- KOEHLER, J. J., 2013: «Proficiency tests to estimate error rates in the forensic sciences», in *Law, Probability and Risk*, 12: 89-98.
- KOTSOGLOU, K. N., 2013: «How to become an epistemic engineer: what shifts when we change the standard of proof?», in *Law, Probability and Risk*, 12: 275-298.
- , 2021: «Zombie forensics. The use of polygraph and the integrity of the criminal justice system», in *International Journal of Evidence & Proof*, 25: 16-35
- KOTSOGLOU, K. N. & McCARTNEY, C. 2021: «To the exclusion of all others? DNA profile and transfer mechanics—*R v Jones (William Francis)* [2020] EWCA Crim 1021 (03 Aug 2020)», in *The International Journal of Evidence & Proof*, 25: 135-140.
- LANDER, E. S., 2017: «Response to the ANZFS council statement on the President's Council of Advisors on Science and Technology Report», in *Australian Journal of Forensic Sciences*, 49: 366-368.
- LEMPERT, R. O., 1977: «Modeling relevance», in *Michigan Law Review*, 75: 1021-1057.
- , 1986: «The new evidence scholarship: analyzing the process of proof», in *Boston University Law Review*, 66: 439-477.
- NATIONAL COMMISSION ON FORENSIC SCIENCE, HUMAN FACTORS SUBCOMMITTEE, 2015: *Views of the Commission Ensuring That Forensic Analysis is Based Upon Task-Relevant Information*, Approved 12/08/2015, <https://www.justice.gov/archives/ncfs/file/818196/download> (last accessed March 12, 2021).
- PARMIANI, G., 2002: *Modeling in Medical Decision Making: A Bayesian Approach*, Hoboken (New Jersey): John Wiley & Sons.
- PEARL, J. & MACKENZIE, D., 2018: *The Book of Why, The New Science of Cause and Effect*, London: Penguin Books.
- PRESIDENT'S COUNCIL OF ADVISORS ON SCIENCE AND TECHNOLOGY (PCAST), 2016: *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods*, Washington, D.C.: Executive Office of the President.
- SCIENTIFIC WORKING GROUP ON DNA ANALYSIS METHODS (SWGDM), 2015: *Guidelines for the Validation of Probabilistic Genotyping Systems*, <https://www.swgdam.org/publications> (last accessed March 7 2021).

- SCHWOEBLE, A. J. & EXLINE, D. L., 2000: *Current Methods in Forensic Gunshot Residue Analysis*, Boca Raton, London, New York: CRC Press.
- STEPHEN, J. F., 1948: *Digest of the Law of Evidence* (12th ed.), London: Macmillan.
- STONE, D. A., 1994: «Relaxation of the assumption of relevance and an application to one-trace and two-trace problems», in *Journal of the Forensic Science Society*, 34: 17-21.
- THAYER, J. B., 1892: *Select Cases on Evidence at the Common Law*, Cambridge: C. W. Sever.
- , 1898: *A Preliminary Treatise on Evidence at the Common Law*, Boston: Little Brown.
- THOMPSON, W. C., 2018: «How should forensic scientists present source conclusions?», in *Seton Hall Law Review*, 48: 773-813.
- THOMPSON, W. C. & SCHUMANN, E. L., 1987: «Interpretation of statistical evidence in criminal trials: the prosecutor's fallacy and the defense attorney's fallacy», in *Law and Human Behavior*, 11: 167-187.
- TRIBE, L. H., 1971: «Trial by mathematics: precision and ritual in the legal process», in *Harvard Law Review*, 84: 1329-1393.
- TRIMPE, M. A., 2011: «The current status of GSR examinations», in *FBI Law Enforcement Bulletin*, May: 24-32.
- U.S. DEPARTMENT OF JUSTICE, 2020: *Uniform Language for Testimony and Reports for the Forensic Latent Print Discipline* (ULTR, vers. 08.15.20), <https://www.justice.gov/olp/page/file/1284786/download> (last accessed 05 March 2021).
- WILLIAMS, M., 1995: *Problems of Knowledge*, Oxford: Oxford University Press.
- WITTGENSTEIN, L., 1958: *Preliminary Studies for the "Philosophical Investigations". Generally known as The Blue and Brown Books*, London: Blackwell Publishers Ltd.
- YEROULANOS, M. (ed.), 2016: *A Dictionary of Classical Greek Quotations*, London: Bloomsbury.



IURIS PRUDENTIA



## LA PRUEBA DE LOS MOTIVOS DISCRIMINATORIOS SEGÚN LA CORTE SUPREMA ARGENTINA: ¿CÓMO MEJORAR SU REGULACIÓN?

Alejo Joaquín Giles\*

Olivia Minatta\*\*

**RESUMEN:** En los litigios por discriminación, probar el motivo discriminatorio puede ser problemático. La Corte Suprema argentina, inspirada en su par estadounidense, estableció una regla que intenta facilitar esa tarea a los litigantes. Sin embargo, entre los procesos civiles de ambos países existen diferencias relevantes relativas a las facultades de las partes para incorporar evidencia al proceso y ajustar sus posiciones en base a la información nueva que surja en su desarrollo. En el contexto argentino, estas diferencias procesales le quitan a la regla potencial para lograr su propio cometido: allanar el camino probatorio de quien aduce discriminación. Luego de desarrollar el problema, en este artículo sugerimos dos cursos de acción posibles para corregirlo. Uno, establecer de forma más precisa las cargas probatorias asignadas a cada una de las partes. Otro, reevaluar algunos aspectos centrales del sistema procesal.

**PALABRAS CLAVE:** discriminación; presunciones; carga de la prueba; estándares de prueba; *discovery*.

---

\* Abogado graduado en la Universidad Nacional de La Plata. Doctorando en Filosofía del Derecho (U. de Génova y U. de Girona). Máster en Razonamiento Probatorio (U. de Girona y U. de Génova) y en Argumentación Jurídica (U. de Alicante). Contacto: alejojgiles@gmail.com.

\*\* Abogada graduada en la Universidad Nacional de La Plata. Magíster en Derecho (Universidad de Columbia de Nueva York) y Magíster en Derecho Administrativo (U. Austral de Buenos Aires). Contacto: oliviaminatta@gmail.com.

**THE PROOF OF INTENTIONAL DISCRIMINATION  
IN ARGENTINA'S SUPREME COURT JURISPRUDENCE.  
HOW TO IMPROVE ITS REGULATION?**

**ABSTRACT:** In discrimination cases, proving the discriminatory intent can be a difficult task. Inspired by the Supreme Court of the United States, the Supreme Court of Argentina created a rule that tries to make the task easier for plaintiffs. However, there are relevant differences in the way civil procedure is regulated in both countries. Particularly, the opportunities to introduce evidence and readjust the pleadings are more limited in the Argentinian context. This diminishes the rule's potential to fulfill its goal, namely to help plaintiffs to prove their case. After delving into the specifics of the problem, this article aims at providing two alternative solutions. First, to better define the burdens of proof allocated to each of the parties. Second, to reassess some aspects of Argentinian civil procedure.

**KEYWORDS:** discrimination; presumption; burden of proof; standards of proof; discovery.

**SUMARIO:** 1. INTRODUCCIÓN.—2. La prueba de los motivos discriminatorios según la Corte Suprema argentina: 2.1. Reglas para la prueba de los motivos discriminatorios. 2.2. No una sino (al menos) dos clases de reglas. 2.3. Las fuentes de las reglas (y un impulso inicial).—3. El «caso discriminatorio prima facie» en la jurisprudencia de la Suprema Corte estadounidense y su encaje procesal: 3.1. «McDonnell Douglas» y su evolución. 3.2. El encaje procesal de las cargas de «McDonnell Douglas».—4. ¿Cómo mejorar la regla para el contexto argentino? Los caminos que la experiencia estadounidense nos invita a pensar: 4.1. Precisar las cargas. 4.2. Considerar el encaje procesal.—CONCLUSIONES.—BIBLIOGRAFÍA.

**RECOMMENDED CITATION:** ALEJO JOAQUÍN GILES; OLIVIA MINATTA, 2022: «La prueba de los motivos discriminatorios según la Corte Suprema argentina: ¿cómo mejorar su regulación?», in *Quaestio facti*, 3: 521-549. Madrid: Marcial Pons Ediciones Jurídicas y Sociales. DOI: [http://dx.doi.org/10.33115/udg\\_bib/qf.i3.22699](http://dx.doi.org/10.33115/udg_bib/qf.i3.22699)

## 1. INTRODUCCIÓN

Uno de los problemas principales en la prueba de la discriminación fáctica es la acreditación del *motivo* por el cual una persona ha tratado a otra de un modo comparativamente perjudicial. Se dará un caso de discriminación fáctica, por definición, cuando ese estado mental esté determinado por la posesión por parte del sujeto perjudicado de uno o varios atributos especialmente protegidos por el derecho, como la raza, el género, la religión, entre otros. Es lo que sucede, por ejemplo, cuando se prohíbe el ingreso a un establecimiento a una persona por el hecho de ser mujer, es decir, por su género. El problema radica —básicamente— en que es difícil conocer el contenido de los estados mentales de las personas al actuar, sumado a que quienes llevan adelante acciones discriminatorias no suelen poner de manifiesto el motivo que las define como tales (en el ejemplo, el género de quien pretendía ingresar al lugar).

La Corte Suprema de la Nación argentina (CSJN), a través de una serie de sentencias, ha establecido una regla con la que intenta aliviar las consecuencias del mencionado problema respecto de las personas que alegan ante los tribunales haber sido discriminadas. La pregunta que orienta este trabajo es la de cómo es posible mejorar la regulación de dicha cuestión, una arista aún incipiente en nuestro contexto. Las respuestas que sugeriremos tienen que ver con dos cursos de acción: el de precisar la formulación de las reglas probatorias involucradas y el de adecuar algunos aspectos del sistema procesal que inciden sobre las posibilidades de que las reglas de ese tipo cumplan con su cometido, es decir, que repercuten sobre el potencial de estas reglas para alivianar dificultades probatorias.

El recorrido argumentativo que seguiremos será el siguiente. Comenzaremos estudiando la jurisprudencia de la Corte Suprema argentina referida a la prueba de los motivos discriminatorios (2). Luego pasaremos a hacer lo propio con la jurisprudencia de la Suprema Corte estadounidense (su fuente de inspiración) y con algunos rasgos del sistema procesal de ese país que influyen sobre la posibilidad de cumplir con las cargas probatorias atribuidas a las partes en los casos de discriminación fáctica (3). Por último, presentaremos dos vías de acción que —según nuestro entender— este cruce sugiere para mejorar el modo en que se aborda la cuestión en la jurisprudencia del máximo tribunal argentino (4).

Intuimos que las observaciones que presentaremos en el último apartado pueden ser extendidas a otros países que, como Argentina, adoptan sistemas procesales de raíz continental. Quienes leen podrán confirmarlo o refutarlo respecto de sus propios países.

## 2. LA PRUEBA DE LOS MOTIVOS DISCRIMINATORIOS SEGÚN LA CORTE SUPREMA ARGENTINA

### 2.1. Reglas para la prueba de los motivos discriminatorios

Las sentencias más relevantes de la Corte Suprema argentina (CSJN de aquí en adelante) referidas al problema de la prueba de los motivos discriminatorios versan sobre casos donde se alega discriminación laboral al despedir o al contratar, por el desarrollo de actividades gremiales o por el género de la persona perjudicada. Ellas son «Pellicori» (2011), «Sisnero» (2014), «Varela» (2018) y «Fontana» (2021).<sup>1</sup> En todas, la CSJN ha establecido —o reiterado— que para la discriminación fáctica rige una regla probatoria según la cual, si la parte actora acredita indicios suficientes para inferir que el trato perjudicial que ha recibido se debe a un motivo discriminatorio,

---

<sup>1</sup> Su cita completa se encuentra en el apartado *Jurisprudencia* al final de este texto. «Pellicori» (2011), «Sisnero» (2014) y «Varela» (2018) fueron adoptadas por unanimidad. «Fontana» (2021) por mayoría.

la parte demandada tendrá la carga de refutar dicha conclusión.<sup>2</sup> Con sus propias palabras:

resultará suficiente, para la parte que afirma dicho motivo, con la acreditación de hechos que, *prima facie* evaluados, resulten idóneos para inducir su existencia, caso en el cual corresponderá al demandado a quien se reprocha la comisión del trato impugnado, la prueba de que este tuvo como causa un motivo objetivo y razonable ajeno a toda discriminación. («Pellicori», considerando 11.)<sup>3</sup>

En síntesis, si el reclamante puede acreditar la existencia de hechos de los que pueda presumirse su carácter discriminatorio, corresponderá al demandado la prueba de su inexistencia. («Sisnero», considerando 5.)

En «Varela» el tribunal avanza un paso más respecto de lo que venía sosteniendo en «Pellicori» y «Sisnero». Luego de recordar el criterio ya enunciado, precisa parcialmente aquello que las partes han de probar para cumplir con su respectiva carga. El caso se trataba de un despido por motivos sindicales. En esa clase de disputas —se sostuvo en el voto principal— para cumplir con su carga quien acciona «debe mostrar *prima facie* o verosímelmente que estaba llevando a cabo una actividad protegida por las normas que invoca» (considerando 9), es decir, una que califique como ejercicio regular de la libertad sindical o como la emisión de opiniones gremiales.<sup>4</sup> Satisfecho eso, la parte demandada —el empleador— tiene la carga de refutar acreditando que el despido se debió a cualquier otro motivo no discriminatorio.<sup>5</sup> Al decidir el caso que tenía ante sus estrados, la Corte consideró que la sentencia recurrida —que rechazaba la pretensión del trabajador accionante— había omitido aplicar dichas reglas al establecer los hechos probados. En especial, no había tenido en cuenta la documentación y los testimonios que avalaban la realización por parte del actor de actividades calificables como sindicales (reclamos de diversa índole) ni que ello era conocido por la empleadora. Además, el tribunal había omitido analizar si la demandada había satisfecho su carga de acreditar las circunstancias invocadas como justificación del despido con causa (impuntualidades al ingresar al trabajo).

<sup>2</sup> En «Pellicori» la Corte se refiere a esta regla como aquella «que en materia probatoria han de regir en los procesos civiles relativos a la ley 23.592, en los que se controvierte la existencia de un motivo discriminatorio en el acto en juego» (considerando 11). En las sentencias posteriores la denomina, sintéticamente, «estándar de prueba». Sin embargo, para evitar confundirla con las reglas que establecen el grado de corroboración suficiente para que una proposición sea tenida por probada (el sentido que suele darse a la expresión en la literatura probatoria), evitaremos usar dicha denominación.

<sup>3</sup> Con una formulación similar: «cuando se discute si [una] medida obedece a un motivo discriminatorio, la existencia de dicho motivo se considerará probada si el interesado acredita de modo verosímil que la medida fue dispuesta por esa razón y, en ese caso, el demandado no prueba que responde a un móvil ajeno a toda discriminación.» («Varela», 2018, considerando 9.)

<sup>4</sup> Esto dado que se invocaban la ley 23.551, que protege el ejercicio regular de la libertad sindical, y la ley 23.592, que prohíbe la discriminación por opiniones gremiales (entre otros motivos).

<sup>5</sup> En la legislación laboral argentina hay dos tipos de despidos, uno donde el empleador justifica su causa y otro donde no lo hace. Ambos tienen reglas y consecuencias parcialmente distintas. Para el despido incausado, dice la Corte, basta con acreditar cualquier otro motivo no discriminatorio. Para el despido por justa causa, en cambio, es preciso demostrar que se ha configurado la causa invocada como justificante de la decisión (y que esta no sea en sí misma discriminatoria, claro).

La sentencia dictada en «Fontana», por su parte, permite observar bajo qué circunstancias más detalladas la Corte podría considerar cumplida la carga colocada sobre la empleadora en este tipo de casos. Se trataba, nuevamente, de un despido por motivos sindicales. El caso llegó a sus estrados con una sentencia que le había dado la razón a la actora, por considerar acreditado, a partir de un «haz de indicios suficientes», que había sido víctima de discriminación por su actividad gremial. La Corte, con un voto de mayoría, revocó esa sentencia al considerar que se sustentaba en una valoración arbitraria de la prueba producida en el proceso. El vicio estaba —sostuvo el voto— en que se había omitido analizar evidencia favorable a la hipótesis según la cual la decisión de despedir a la actora había obedecido a graves problemas económicos de la empresa involucrada (y no a la filiación sindical de la trabajadora). El mercado automotriz donde comercializaba la empresa estaba en caída y junto con la actora habían sido desvinculadas treinta y cuatro personas más por falta o disminución de trabajo. Ello «configura *prima facie* una causa seria y objetiva, ajena a toda discriminación» (considerando 6).

Dejando de lado aquel desenlace, el estudio de este caso da pie para advertir otra cuestión. Como señaló el voto disidente, del otro lado de la balanza se encontraba acreditado que la actora era una asistente activa a las asambleas gremiales, que transmitía información entre los delegados del sindicato y los trabajadores y que el despido cuestionado era el segundo que había sufrido (el primero había ocurrido el año anterior, al culminar otro conflicto gremial). Esto sugiere que, tomando en cuenta todo el acervo probatorio, es posible identificar en el caso dos hipótesis en disputa, que compiten por explicar los motivos reales del despido de la actora. ¿Se debió a su actividad gremial, a la crisis económica de la empresa, o a ambas cosas en cierta medida? ¿Cuál es la hipótesis más probable a la luz de la evidencia? Los interrogantes son típicos de todo razonamiento probatorio. Y en los casos de discriminación se presentan especialmente cuando ambas partes han desplegado actividad probatoria relevante.<sup>6</sup>

---

<sup>6</sup> Mientras terminábamos de escribir este trabajo la Corte dictó sentencia en el caso «Romero» (2021), donde se alegaba que la no renovación de un contrato a tiempo determinado con el Ministerio de Educación de la Nación se había debido a las actividades gremiales del trabajador. Entre otros argumentos, en el voto mayoritario se sostuvo que para tener por probada la motivación discriminatoria de la desvinculación no bastaba con acreditar el mero activismo sindical del actor, especialmente cuando la demandada se había defendido aduciendo que esta obedecía al cumplimiento del plazo convenido en el contrato (considerando 8). Según surge de la reconstrucción de los antecedentes que hace el tribunal, el actor venía prestando servicios desde el año 2014 para una universidad nacional en el marco de un programa del ministerio; y desde septiembre de 2016 para el ministerio mismo, con un contrato por tiempo determinado que culminó en diciembre de ese año y fue renovado por un mes, luego de lo cual se cortó la cadena de renovaciones. En un caso como este podría alegarse que lo que obedeció a un motivo discriminatorio no fue el cese del contrato que unía a las partes sino el corte de la cadena de renovaciones. Para lo cual sería relevante responder a preguntas como estas: ¿Era usual en el área o actividad donde se desempeñaba el actor entablar relaciones laborales con esa modalidad, mediante contratos a tiempo determinado renovados de modo sucesivo e ininterrumpido? ¿Había otros sujetos en las mismas circunstancias que el actor a los que sí se les siguió renovando el vínculo contractual? ¿Se

## 2.2. No una sino (al menos) dos clases de reglas

Los modos en que se enuncia la regla en «Pellicori» y «Sisnero», por un lado, y en «Varela», por el otro, insinúan que hay al menos dos formas distintas de articularla, con relevantes diferencias entre sí pero con el mismo objetivo: aliviar la carga probatoria de quienes alegan haber sido discriminados. Estas posibles formulaciones dan lugar a dos clases de reglas distintas, a las que denominaremos «R1» y «R2». Veamos.

Una formulación, cercana a la propuesta en «Pellicori» y «Sisnero», podría reconstruirse como sigue:

R1: Si hay *indicios suficientes* a favor de una proposición (la existencia de motivos discriminatorios en una cierta acción), entonces se debe aceptarla como verdadera.

Así como está enunciada la regla, da lugar a un razonamiento que se identifica con las llamadas «presunciones *hominis*», inferencias por las cuales, a partir de una serie de premisas tomadas por ciertas (los indicios), se llega a una conclusión sobre hechos desconocidos por el sujeto decisor. Las presunciones *hominis*, según se ha hecho notar, tienen la misma estructura que toda inferencia probatoria no deductiva, sin que haya una diferencia categorial entre ambas.<sup>7</sup> Si se acepta esta última tesis, cabe preguntarse qué hace que una regla como R1 no sea trivial, es decir, que no sea interpretada como indicando lo obvio para todo razonamiento fáctico. Una respuesta plausible, teniendo en cuenta el objetivo perseguido al establecerla (aliviar la carga probatoria de quienes alegan haber sido discriminados), podría ser que su particularidad radica en la idea de «suficiencia». Cuando la pauta se aplique, el grado de suficiencia requerido para trazar el puente inferencial entre los indicios y una conclusión habrá de ser inferior al exigido ordinariamente para tener por probada cualquier proposición.<sup>8</sup> Habiendo indicios —en este sentido— suficientes, la conclusión deberá ser aceptada como verdadera hasta que la parte demandada incorpore

---

dieron acciones de las que pudiera inferirse una animadversión de la empleadora hacia el actor y otras personas que realizaran actividades gremiales similares a las que este realizaba? Si las respuestas a las tres preguntas fueran positivas, y no se incorporara otra información adicional en sentido contrario, la hipótesis del motivo discriminatorio se vería fortalecida. Y lo contrario si las respuestas fueran negativas.

<sup>7</sup> Por la similitud entre las presunciones *hominis* y las inferencias probatorias, y por la diferencia entre las primeras y las presunciones *iuris* (incluyendo a la variante *iuris tantum* que mencionaremos a continuación), puede consultarse, entre otros: GASCÓN ABELLÁN (2010: 123-140) y AGUILÓ-REGLA (2018). Para una síntesis de esas cuestiones, previo a estudiar el problema de la calificación (como presunciones o no) de las denominadas presunciones *iuris et de iure*, véase DEI VECCHI (2019: 3-5).

<sup>8</sup> Este sentido parece tener en cuenta la propia Corte cuando, respecto del *corpus iuris* elaborado por los comités internacionales de derechos humanos, afirma que este se proyecta sobre el régimen probatorio de la discriminación fáctica de dos maneras: «Primeramente, *al reducir el grado de convicción* que, respecto de la existencia del motivo discriminatorio, debe generar la prueba que recae sobre quien invoca ser víctima de dicho acto. Y, en segundo lugar, al modular, a partir de lo anterior, la distribución de la carga de la prueba y la medida en que esta pesa sobre el demandado al que se imputa la responsabilidad por el mencionado acto.» («Pellicori», considerando 5, la cursiva es nuestra.)

información suficiente para refutarla demostrando que es falsa o que su verdad no puede ser sostenida.<sup>9</sup> Así planteada, la regla actúa como un incentivo hacia la demandada para que realice actividad probatoria y evite la simple negación de todo lo afirmado por la actora.

Puede suceder, como vimos respecto de «Fontana», que habiendo desplegado su actividad probatoria ambas partes, del acervo probatorio emerjan con algún aval dos o más hipótesis que le asignen a una acción determinada un motivo distinto, uno discriminatorio y otro no. ¿Cuál cabe declarar como probada? La pregunta, que no podemos responder aquí, remite al problema más amplio de la valoración de la prueba y la decisión sobre los hechos probados.<sup>10</sup>

Una segunda formulación, distinta a la anterior y más cercana a la esbozada incipientemente en «Varela», podría reconstruirse como sigue:

R2: Si se encuentran probadas instancias de las circunstancias A (un trabajador fue despedido), B (el trabajador ejercía actividades sindicales) y C (el empleador sabía que la persona ejercía actividades sindicales), entonces se debe aceptar como verdadera la conclusión D (el despido estuvo motivado en las actividades sindicales ejercidas por el trabajador), salvo que se pruebe la falsedad de D en el caso concreto (acreditando por caso que el motivo fue otro, como la necesidad generada por una crisis económica).

Esta formulación tiene una forma lógica idéntica a la propia de las denominadas «presunciones *iuris tantum*» (que pueden establecerse legal o jurisprudencialmente).<sup>11</sup> Acreditados los antecedentes de la regla (que sucedieron instancias de A, B y C, según el enunciado que funge de modelo), esta manda a aceptar como verdadera una proposición (según la cual ha sucedido una instancia de D) de modo interino y hasta que sea refutada, lo cual se evalúa del mismo modo en que se evalúa la refutación bajo la

<sup>9</sup> Sobre la suficiencia de las razones para refutar una presunción, véase ULLMAN-MARGALIT (1983: 152-154), donde se sostiene que es la fuerza de la presunción lo que determina el peso de las razones requeridas para refutar sus conclusiones. Se proponen las siguientes variantes: si de un lado de la balanza hay una presunción débil a favor de una conclusión, esta última es refutada contando con alguna razón en su contra; si hay una presunción de fuerza intermedia, su conclusión es refutada contando con buenas razones en su contra; y si hay una presunción fuerte, su conclusión es refutada contando con razones concluyentes en su contra.

<sup>10</sup> Por un panorama general al respecto véase FERRER BELTRÁN (2007). En «Pellicori» la Corte hace una referencia magistral a esta cuestión en relación con los casos de discriminación, pero sin aportar mayores orientaciones: afirma que el cumplimiento de la carga que recae sobre cada parte habrá de evaluarse «de conformidad con las reglas de la sana crítica» (considerando 11).

<sup>11</sup> La forma es análoga, por ejemplo, a aquella con la que puede reconstruirse la norma expresada por art. 178 de la Ley de Contrato de Trabajo argentina (n. 20744), relacionada con la prohibición del despido por causa de embarazo: «Se presume, salvo prueba en contrario, que el despido de la mujer trabajadora obedece a razones de maternidad o embarazo cuando fuese dispuesto dentro del plazo de siete y medio (7 y 1/2) meses anteriores o posteriores a la fecha del parto, siempre y cuando la mujer haya cumplido con su obligación de notificar y acreditar en forma el hecho del embarazo así, en su caso, el del nacimiento.» Esto se puede reconstruir así: si se da una instancia de A (el despido de una trabajadora embarazada en el plazo indicado), se debe aceptar como verdadera una instancia de B (que el despido fue motivado en el embarazo de la trabajadora), salvo que se pruebe su falsedad en el caso concreto.

formulación R1. Acreditados los antecedentes de la regla, entonces, la parte demandada pasa a tener un importante incentivo para desplegar actividad probatoria.

Ambas variantes tienen una diferencia fundamental. En R1 se le instruye al sujeto decisor a recurrir a *razones epistémicas* para establecer si es aceptable como verdadera una proposición según la cual un acto tuvo motivos discriminatorios; y a tomar la decisión, a falta de refutaciones, aplicando un estándar de suficiencia flexible o liviano (y así planteado, altamente indeterminado). En R2, en cambio, las razones para aceptar como verdadera una proposición de esa clase son, en un principio, *razones normativas*. Si el juez da por verificadas ciertas premisas, la regla le obliga a aceptar como verdadero, interinamente, que un acto determinado tuvo motivos discriminatorios, con independencia de la fuerza con que las premisas avalen dicha conclusión desde la perspectiva epistemológica. Las razones epistémicas pasan a ser relevantes recién al momento de evaluar si la parte demandada ha logrado refutar esa conclusión.

Como puede advertirse, las variantes mencionadas regulan de modo diverso las circunstancias bajo las cuales quien alega haber sido discriminado satisface su carga probatoria inicial. La carga de la parte demandada, por su parte, consiste en lo mismo bajo ambas formulaciones: satisfecha la de quien alega discriminación, le toca refutar la existencia de los motivos discriminatorios que *prima facie* se infieren o presumen. Las circunstancias del caso «Fontana» son una clara ilustración de cómo podría lograrse esa refutación. Cuando esta carga resulta cumplida, puede decirse que la presunción (sea *hominis* o *iuris*) ha sido *destruida* (MENDONCA, 1998: 97) o *derrotada*.

Al descansar sobre la discrecionalidad del decisor, las reglas del tipo R1 brindan flexibilidad para adaptarse a los matices de cada tipo de fenómeno discriminatorio, al tiempo que —en el otro platillo de la balanza— disminuyen la previsibilidad en las decisiones. Por el contrario, al limitar relativamente esa discrecionalidad, las reglas del tipo R2 resultan inflexibles a los matices de cada caso pero hacen más predecibles a sus actos de aplicación.

### 2.3. Las fuentes de las reglas (y un impulso inicial)

Al justificar sus decisiones —especialmente en «Pellicori» y «Sisnero»— la Corte tuvo la precaución de mostrar que no estaba siendo original al establecer reglas probatorias especiales para la prueba de la motivación discriminatoria. Una de sus principales referencias es el derecho de la Unión Europea y de sus países miembros.<sup>12</sup>

<sup>12</sup> Para un panorama general de la regulación europea puede consultarse el *Manual de legislación europea contra la discriminación* (2018, pp. 264-272) publicado por la Agencia de los Derechos Fundamentales de la Unión Europea y el Consejo de Europa. La Corte invocó, además del derecho de la Unión Europea, diversos informes que sugieren adecuaciones de las reglas probatorias para la discriminación fáctica (en el sentido de aliviar la carga de quien alega discriminación) emanados por órganos internacionales de derechos humanos, como el Comité contra la Discriminación Racial, el Comité de

Tomemos como ilustración lo que establece la Directiva 2006/54/CE del Parlamento Europeo y del Consejo de Europa al respecto:

Los Estados miembros adoptarán con arreglo a sus sistemas judiciales nacionales las medidas necesarias para que, cuando una persona que se considere perjudicada por la no aplicación, en lo que a ella se refiere, del principio de igualdad de trato *presente*, ante un órgano jurisdiccional u otro órgano competente, *hechos que permitan presumir* la existencia de discriminación directa o indirecta, corresponda a la parte demandada demostrar que no ha habido vulneración del principio de igualdad de trato. (Art. 19.1)<sup>13</sup>.

La fórmula, como se advierte, es similar a la variante enunciada antes como R1 y a lo esbozado en «Pellicori» y «Sisnero».<sup>14</sup> Pero según el máximo tribunal argentino, el origen del criterio adoptado se ubica en la jurisprudencia de la Suprema Corte de los Estados Unidos. Lo dice así:

Este principio de reparto de la carga de la prueba en materia de discriminación tuvo su origen en la jurisprudencia norteamericana, en el conocido caso “Mc Donnell Douglas Corp. vs. Green” (fallo del año 1973, publicado en 411 US 792). Esa doctrina continúa siendo aplicada por la Corte Suprema de los Estados Unidos hasta la actualidad, aunque con ciertos matices (ver 431 US 324, año 1977, 509 US 502, año 1993 y 530 US 133, año 2000). («Sisnero», considerando 5.)

En lo que sigue analizaremos la evolución de esta regla probatoria en la jurisprudencia de la Suprema Corte estadounidense y el contexto procesal en el que se aplica. La tarea será útil para responder a la pregunta que abre este trabajo.

### 3. EL «CASO DISCRIMINATORIO PRIMA FACIE» EN LA JURISPRUDENCIA DE LA SUPREMA CORTE ESTADOUNIDENSE Y SU ENCAJE PROCESAL

#### 3.1. «McDonnell Douglas» y su evolución

¿En qué consiste y cómo ha evolucionado la regla probatoria instaurada por la Suprema Corte estadounidense en «McDonnell Douglas» de 1973? En las páginas que siguen presentaremos una breve aproximación a la cuestión.<sup>15</sup>

---

Derechos Económicos, Sociales y Culturales, el Comité de Derechos Humanos y el Comité contra la Discriminación de la Mujer. También de la Organización Internacional del Trabajo.

<sup>13</sup> En el mismo sentido regulan otras directivas, como la 2000/43/CE (art. 8.1) la 2000/78/CE (art. 10.1) y la 2004/113/CE (art. 9.1).

<sup>14</sup> Si vamos al plano jurisprudencial, sin embargo, se pueden encontrar sentencias de importantes tribunales que establecen pautas cercanas a la formulación R2. Así lo ha hecho, por ejemplo, el Tribunal de Justicia de la Unión Europea en el caso «Brunnhofen» (2001), sobre diferencia de salario en virtud del sexo. En ese precedente se sostiene que, para los casos como el analizado, la carga inicial de la persona que alega discriminación es satisfecha si logra probar «que percibe una retribución inferior a la que el banco paga a su compañero de trabajo de referencia y que en realidad realiza el mismo trabajo o un trabajo de igual valor, comparable al que efectúa este» (párrafo 58).

<sup>15</sup> Nos detendremos en algunos de los casos más relevantes, de una línea jurisprudencial conformada por una elevada cantidad de decisiones.

Uno de los logros más importantes del movimiento por los derechos civiles que tomó fuerza en Estados Unidos en el siglo xx, fue la sanción de la Ley de Derechos Civiles de 1964, a partir de la cual se prohibió la segregación racial en espacios públicos y la discriminación en el trabajo por origen racial, sexo, religión o nacionalidad.<sup>16</sup> La Ley y sus posteriores ampliaciones dieron lugar al desarrollo de una copiosa jurisprudencia que a lo largo de los años dejó entrever vaivenes y matices respecto de las reglas probatorias aplicables en los casos por discriminación.<sup>17</sup> Nos enfocaremos en la línea jurisprudencial elaborada por la Suprema Corte de Estados Unidos a partir del caso «McDonnell Douglas Corp. v. Green» de 1973. Se trata de una pauta (o «estructura» o «marco» probatorio, como es denominada por el tribunal y la doctrina especializada de su país) aplicable a los casos individuales donde se intenta probar la motivación discriminatoria de uno o varios actos.<sup>18</sup>

¿Qué resolvió la Suprema Corte en «McDonnell Douglas»? En el caso, Green (un trabajador mecánico) demandó a su ex empleadora, McDonnell Douglas Corp. (una fábrica de aeronaves), por haber rechazado su recontratación pese a su calificación para el puesto.<sup>19</sup> Entre sus argumentos, Green aducía que los motivos de la no contratación obedecían a su origen racial y a su activa participación en manifestaciones en defensa de los derechos civiles de los trabajadores afroamericanos. La empresa, de su lado, sostenía que la decisión se había debido a que Green había realizado actos ilegales en su contra.<sup>20</sup> La Suprema Corte resolvió por unanimidad en favor de Green, confirmando que debía revocarse la decisión que no le había permitido llevar su caso

---

<sup>16</sup> Finalizada la Guerra Civil y durante el periodo de Reconstrucción, la Constitución de Estados Unidos prohibió la esclavitud, reconoció a todas las personas allí nacidas o naturalizadas la ciudadanía y estableció el derecho a voto y la igualdad ante la ley independientemente de la raza. Sin embargo, las reformas constitucionales no se tradujeron necesariamente en políticas, prácticas y leyes que reconocieran la igualdad racial, y la segregación siguió existiendo en distintas formas a lo largo y ancho de Estados Unidos. Las leyes de Jim Crow y la famosa doctrina de «separados pero iguales» que elaboró la Suprema Corte norteamericana en el caso «Plessy v. Ferguson» (1896) son claros ejemplos de ello.

<sup>17</sup> En el derecho antidiscriminatorio estadounidense suelen distinguirse casos de discriminación intencional y no intencional, casos de «motivo mixto» y de «patrón y práctica»; los casos de «motivo mixto» obedecen a una reforma de la Ley de Derechos Civiles operada en 1991 (MALAMUD, 1995).

<sup>18</sup> La pauta es aplicada, se afirma, a la gran mayoría de reclamos discriminatorios que se inician actualmente en el campo del derecho laboral. Sin perjuicio de ello, también se ha extendido a otros reclamos, como aquellos por discriminación en el acceso a la vivienda o discriminación en espacios públicos (EYER, 2019).

<sup>19</sup> Green había sido previamente despedido, luego de trabajar ocho años en la empresa, en el marco de una reducción general de la fuerza de trabajo. Esto dio lugar a una serie de protestas contra la política general de contratación de la empresa porque, según se afirmaba, estaba racialmente sesgada.

<sup>20</sup> Según lo relatado en la sentencia comentada, uno de esos actos consistió en el bloqueo, junto con otras personas, del acceso a la planta de McDonnell Douglas Corp. durante un cambio de turno. A raíz de ello Green fue arrestado y acusado de obstrucción del tráfico, por lo que fue condenado a pagar una multa luego de haberse declarado culpable. La empresa también le sindicaba el haber participado en un «lock-in», esto es, el bloqueo (con cadenas y candados) de la salida de un edificio donde funcionaba parte de sus oficinas.

a juicio alegando discriminación racial.<sup>21</sup> Además, aprovechó la oportunidad para expedirse acerca de lo que, entendió, versaba sobre «el orden propio y la naturaleza de la prueba»<sup>22</sup> (411 U.S. 792, p. 793) de los casos del Título VII de la Ley de Derechos Civiles. Sin una base legal al respecto, el máximo tribunal federal estadounidense estableció un esquema de tres pasos para argumentar sobre la prueba de la motivación discriminatoria.<sup>23</sup> Los presentaremos como «C1», «C2» y «C3».<sup>24</sup>

C1: *Caso prima facie*. La actora tiene la carga *inicial* de probar un «caso discriminatorio *prima facie*».<sup>25</sup>

En supuestos como el estudiado, dicha carga se puede colmar, sostuvo el tribunal, con la acreditación de las siguientes premisas: (i) que el actor pertenece a una minoría racial; (ii) que aplicó y estaba calificado para un puesto de trabajo cuya convocatoria estaba abierta; (iii) que, sin embargo, fue rechazado; (iv) y que el puesto de trabajo siguió vacante luego de su rechazo, en búsqueda de un candidato con sus mismas calificaciones. Al hacerlo, se estarán descartando «las dos razones legítimas más comunes en las que un empleador puede basarse para rechazar a un solicitante de empleo: una absoluta o relativa falta de calificación o la ausencia de una vacante en el empleo buscado» (431 U.S. 324, p. 358).<sup>26</sup>

<sup>21</sup> La cuestión principal para la decisión del caso radicaba en lo siguiente. Se había desechado la pretensión de Green de tramitar su caso como uno de discriminación racial (con apoyo en el §703(a) de la Ley de Derechos Civiles) puesto que esta había sido rechazada en la instancia administrativa previa por carecer de sustento. Intervino allí la Comisión para la Igualdad de Oportunidades en el Empleo (EEOC por sus siglas en inglés). La Suprema Corte entendió que esto último no podía ser tenido como un impedimento para la procedencia de una acción judicial. Este fue, en resumen, el argumento principal que resolvió el pleito, sin perjuicio de las pautas incorporadas en la sentencia acerca de cómo debería, en un nuevo juicio, ordenarse la prueba. Los efectos prácticos de la sentencia fueron reenviar a primera instancia para que se llevara adelante un nuevo juicio. En el nuevo juicio Green perdió nuevamente el caso.

<sup>22</sup> Esta cita y las que siguen son todas traducciones propias del texto original de las respectivas sentencias.

<sup>23</sup> Si bien suelen presentarse como pasos sucesivos, en la práctica no siempre se dan en ese orden. Además, su cumplimiento o no suele resolverse de modo acumulado, en una única decisión.

<sup>24</sup> El tribunal establece estos pasos entre las páginas 802 y 805 de la sentencia (411 U.S. 792). Su semilla fue colocada en la decisión de la Cámara de Apelaciones que entendió en el caso.

<sup>25</sup> En una sentencia posterior, «Texas Department of Community Affairs v. Burdine» (1981), esta carga inicial es sintetizada así: «el actor tiene la carga de probar, *mediante la preponderancia de la evidencia*, un caso *prima facie* de discriminación» (450 U.S. 248, pp. 252-253). De otro pasaje resulta claro que lo que se tiene que probar con ese grado son los indicios que permitan presumir, descartando los motivos no discriminatorios más comunes, que un cierto acto tiene motivos discriminatorios («La carga de establecer un caso *prima facie* de trato desigual no es onerosa. El actor debe probar mediante una preponderancia de la evidencia que ha aplicado para un puesto disponible para el que estaba calificado, pero fue rechazado bajo circunstancias que dan lugar a una inferencia de discriminación ilegítima», *id. ant.*, p. 253).

<sup>26</sup> Observa la Suprema Corte que esto «es suficiente, ante la falta de otra explicación, para inferir que la decisión fue discriminatoria» (431 U.S. 324, p. 358, nota 44; 438 U.S. 567, p. 577). Pero también aclara que en los casos del Título VII los hechos pueden variar y que las premisas especificadas no son necesariamente aplicables en los diferentes supuestos (411 U.S. 792, p. 802, nota 13). Eso hace que

C2: *Refutación*. Por ello, la carga probatoria recae en la demandada, quien para desvirtuar la presunción de discriminación ha de articular un motivo legítimo, no discriminatorio, de rechazo.<sup>27</sup>

C3: *Contestación*. Luego, la actora tiene la carga (*final*) de desacreditar el motivo aducido por la demandada, mostrando que no es el verdadero, sino que se trata de un pretexto.<sup>28</sup>

Al detentar esta última carga, la parte actora «*debe ser dotada de la oportunidad de presentar evidencias acerca de que la justificación ofrecida [por la demandada] es meramente un pretexto para la discriminación*» (438 U.S. 567, p. 578).

En el caso concreto, la Suprema Corte entendió que Green había cumplido con su carga inicial (C1) y la empresa con la suya (C2).<sup>29</sup> Pero que, luego de eso, Green no había tenido la oportunidad de dar el tercer paso (C3), es decir, no había podido alegar sobre si los motivos invocados por la empresa eran o no un mero pretexto.<sup>30</sup>

Las pautas establecidas en el precedente al que se viene refiriendo dieron lugar a muchos interrogantes que se plantearían en casos futuros. ¿Cuáles son los alcances de cada una de las cargas? ¿Qué efectos concretos tiene su cumplimiento?

Unos años más tarde la Suprema Corte resolvió el caso «Texas Department of Community Affairs v. Burdine» (1981), en el cual la actora, Burdine, alegaba haber sido perjudicada laboralmente por ser mujer, al no ser promovida a un cargo que se encontraba vacante y al ser despedida junto con otros colegas de la oficina en la que trabajaba (para posteriormente ser recontratada en otra oficina de la misma repartición).<sup>31</sup> La defensa de la demandada, el Departamento de Asuntos Comunitarios de Texas, se basaba en que el despido había sido necesario para reorganizar la oficina y lograr resultados más eficientes en los equipos de trabajo. Para acreditar el

---

el criterio no se convierta en «una formulación inflexible» (431 U.S. 324, p. 358, nota 44; 438 U.S. 567, p. 575).

<sup>27</sup> En lo inmediato veremos distintas alternativas para interpretar los alcances de esta carga.

<sup>28</sup> El estándar para dar por cumplida C3, se aclara en «Burdine», es el de la preponderancia de la evidencia (450 U.S. 248, p. 253).

<sup>29</sup> En este caso, a diferencia de lo que había resuelto para el caso «Griggs v. Duke Power Co.» (1971), la Suprema Corte consideró aceptable como justificación la invocación de motivos no necesariamente relacionados con la calificación para el trabajo, como lo era la conducta previa de Green.

<sup>30</sup> Para superar esa carga, sostuvo el tribunal, sería relevante la evidencia que muestre: (i) que otros empleados blancos involucrados en actividades similares a las de Green fueron retenidos o recontratados; (ii) el trato dado a la actora tiempo antes del episodio; (iii) la reacción de la empleadora ante las manifestaciones legítimas a favor de los derechos civiles; (iv) y la política general de la empresa respecto del empleo de minorías raciales, lo que puede acreditarse mediante estadísticas (411 U.S. 792, pp. 804-805).

<sup>31</sup> Según lo que se destaca en la sentencia, Burdine trabajaba en una oficina del Departamento de Asuntos Comunitarios de Texas dedicada a la promoción del acceso al empleo. Liberado un cargo inmediato superior al que tenía —el de Director de proyecto— la actora se postuló pero no fue promovida. El cargo se mantuvo vacante durante seis meses hasta que fue contratada otra persona, un varón. Sucesivamente, fueron despedidas tres personas en el marco de una reducción de personal, entre ellas la actora. Burdine fue posteriormente recontratada en otra oficina de la misma administración pública, donde recibía el mismo salario y tenía responsabilidades similares —según tiene por acreditado el tribunal— a las que hubiera tenido en el cargo al que no accedió.

mal funcionamiento que había llevado a esos despidos, Texas se apoyaba en el testimonio del funcionario responsable de tomar esas decisiones y en una «evaluación no discriminatoria» de las calificaciones de las personas involucradas. En primera instancia se consideró que la explicación era racional y suficiente para rebatir el caso *prima facie* presentado por Burdine. Sin embargo, la Cámara de Apelaciones resolvió en sentido parcialmente contrario, afirmando que el testimonio del funcionario no resultaba suficiente para refutar la presunción de que el despido de Burdine había estado motivado por su género. La Suprema Corte estadounidense revocó la decisión de la Cámara.

Para resolver el caso, se discutió especialmente bajo qué circunstancias se satisface la carga de refutación (C2) de un caso *prima facie* de discriminación. Una alternativa, adoptada por la Cámara de Apelaciones, era establecer que la demandada debía probar por preponderancia de la evidencia que el acto cuestionado estuvo guiado por un motivo no discriminatorio.<sup>32</sup> Otra alternativa, adoptada por la Suprema Corte, era establecer una exigencia menor según la cual la demandada debía «explicar claramente las razones no discriminatorias de sus acciones», valiéndose de la «introducción de evidencia admisible» al respecto (450 U.S. 248, pp. 255, 260). En opinión de la Suprema Corte, lo primero no estaba configurado en el caso, pero sí lo segundo.

Aquella elección por parte del máximo tribunal estadounidense deriva de la siguiente idea. El cumplimiento de C1 crea una presunción que es suficiente, ante el silencio de la demandada, para que los jueces tengan por verificado el carácter discriminatorio del acto impugnado.<sup>33</sup> Pero eso no significa, aclara el tribunal, trasladar la carga «de persuasión» hacia la demandada, sino simplemente asignarle una carga «de producción» de evidencia tendiente a desacreditar el motivo discriminatorio alegado por la actora. Al incorporar evidencia en ese sentido la demandada estará cumpliendo con C2, el caso *prima facie* habrá sido rebatido y «la investigación empírica [avanzará] hacia un nuevo nivel de especificidad» (450 U.S. 248, p. 255). En tal escenario, la actora seguirá teniendo su carga *final*, la de persuadir acerca de que su hipótesis de

---

<sup>32</sup> A lo que agregaba el deber de «probar mediante evidencia objetiva que quienes fueron contratados o promovidos estaban mejor calificados que la parte actora» (450 U.S. 248, p. 252).

<sup>33</sup> «El establecimiento del caso *prima facie* crea, en efecto, una presunción de que el empleador discriminó ilegalmente al empleado. Si el juez cree en la evidencia de la actora y el empleador mantiene silencio ante la presunción, el juez debe fallar a favor de la actora porque no queda ninguna controversia fáctica que resolver.» (450 U.S. 248, p. 254.) Los intercambios de notas entre el Juez Powell y el Juez Stevens en el caso Burdine dejaron entrever distintas posturas sobre los alcances del caso *prima facie*. Mientras que para Powell lo central del caso *prima facie* era descartar los motivos más comunes del rechazo (como la falta de calificación) para Stevens era necesario que la actora diera lugar a una inferencia de discriminación. Se ha hecho notar que la no taxatividad de las condiciones para cumplir con C1 ha dado lugar a que las cortes inferiores elaboren una serie de «reglas técnicas» que muchas veces sobrecargan el caso *prima facie* y agregan elementos nuevos para su verificación. Muchos de estos elementos, se sostiene, hacen confluír los «tres pasos» de «McDonnell Douglas», porque adicionan aspectos que hacen al análisis del motivo brindado del empleador. Véase EYER (2019).

la motivación discriminatoria es la más probable a la luz de la evidencia.<sup>34</sup> Darle ese alcance a C2 tiene una utilidad, afirma el tribunal: sirve «para enmarcar la cuestión fáctica con suficiente claridad para que el actor tenga una oportunidad plena y justa» de cumplir con C3 (450 U.S. 248, pp. 255-256). Actúa, podría decirse, como un incentivo para que los demandados eviten el silencio e incorporen al proceso hipótesis y evidencias que permitan ampliar la indagación fáctica.

Post «Burdine» los tribunales inferiores estadounidenses tomaron distintas posturas sobre cuál debería ser la consecuencia de que el actor cumpliera con C3, la carga de desacreditar el motivo invocado por el demandado. Algunos consideraron que se generaba automáticamente una cuestión de puro derecho para resolver el caso a su favor sin ir a juicio oral. Otros consideraron que eso habilitaba al juzgador a declarar la causa de puro derecho, pero no era un paso automático. Finalmente, otros tribunales consideraron que esa evidencia era insuficiente para juzgar en favor del actor.<sup>35</sup>

La Suprema Corte de los Estados Unidos se pronunció al respecto en el caso «St. Mary's Honor Center v. Hicks» de 1993. Hicks era un afroamericano que trabajaba como comandante de turno en una prisión de baja seguridad. Luego de unos años de buena performance en el trabajo y a partir de un cambio de supervisor, comenzó a registrar sanciones disciplinarias que derivaron en su despido. Hicks demandó a St. Mary's aduciendo despido discriminatorio. La empleadora contestó alegando que el motivo del despido había sido la comisión de una serie de graves infracciones disciplinarias. En primera instancia, la corte de distrito falló en contra del actor por considerar que pese a que el motivo alegado por la demandada había sido desacreditado, del acervo probatorio surgía un motivo alternativo no discriminatorio para el despido.<sup>36</sup> La Cámara de Apelaciones revocó esa decisión, al entender que una vez probado el caso *prima facie* y desacreditada la razón legítima brindada por el demandado, correspondía al juez fallar en favor de la actora. El juez no estaba habilitado a evaluar motivos no discriminatorios distintos a los alegados por la demandada, sostuvo.

---

<sup>34</sup> «La carga final de persuadir al juez de que el demandado discriminó intencionalmente al actor permanece en todo momento en el actor [...] La carga que recae sobre el demandado, por lo tanto, es la de refutar la presunción de discriminación produciendo evidencias de que el actor fue rechazado, o alguien más fue preferido, por una razón legítima y no discriminatoria. El demandado no necesita persuadir a la corte de que realmente estuvo motivado por las razones indicadas.» (450 U.S. 248, pp. 253, 254; en el mismo sentido: 438 U.S. 567, 439 U.S. 24.)

<sup>35</sup> Véase MALAMUD (1995: 2234, y nota al pie 23).

<sup>36</sup> Al analizar todo el conjunto de evidencias, la corte de distrito entendió que los motivos alegados por la demandada no eran los que verdaderamente habían motivado su decisión (por mencionar un dato, Hicks había sido, entre varios candidatos, el único supervisor sancionado por violaciones cometidas por sus subordinados). En cambio, entendió que era evidente que existía una animosidad personal del supervisor contra Hicks, pero que no estaba comprobado que esa animosidad obedeciera a razones raciales. La Cámara, por su parte, puso en duda que la animosidad personal pudiera considerarse un motivo legítimo.

En un fallo dividido 5 a 4, la Suprema Corte revocó la decisión de la Cámara.<sup>37</sup> El voto de la mayoría se basó fundamentalmente en la siguiente idea acerca de cómo se configuran las cargas que venimos estudiando. Una vez que el demandado alega motivos no discriminatorios y produce evidencias a su respecto, «sean o no persuasivos en última instancia», satisface su carga de producción (C2) y refuta el caso *prima facie*, colocándose en «una mejor posición que si hubiese guardado silencio». Cuando ello se da, «el marco *McDonnell Douglas* —con sus presunciones y sus cargas— no es más relevante» y la cuestión fáctica se decide como cualquier otra, manteniéndose la carga de persuasión en cabeza de la actora (como sucede con cualquier otra presunción regulada por las Reglas Federales de la Evidencia).<sup>38</sup>

Con base en tal idea, la mayoría sostuvo que el hecho de que la actora desacreditase la razón legítima brindada por el empleador (cumpliendo con C3) no obligaba al juez a fallar a su favor sin enviar la causa a juicio oral, aunque podría resultar un elemento a considerar para inferir la discriminación.<sup>39</sup> Según este lineamiento, los jueces deben resolver de puro derecho a favor de la actora solo si no quedan cuestiones fácticas por resolver, lo que en estos casos se da cuando:

(1) cualquier persona racional sería capaz de identificar hechos que constituyen un caso *prima facie*, y (2) la demandada falla en su carga de producción —es decir, falla en introducir evidencia que, *de ser tomada como cierta, permitiría* llegar a la conclusión de que hubo un motivo no discriminatorio para el rechazo. (509 U.S. 502, p. 509.)<sup>40</sup>

Al contrario, los jueces no pueden resolver la cuestión de puro derecho y deben enviar el caso a juicio «si la demandada ha fallado en cumplir con su carga pero mentes racionales podrían *diferir* respecto de que existe evidencia preponderante que estable-

<sup>37</sup> El voto en disidencia redactado por el Juez Souter, acusó a la mayoría de apartarse y dejar sin efecto la estructura probatoria de «*McDonnell Douglas*». La mayoría, por su parte, entendió que simplemente estaba resolviendo por la correcta aplicación de aquella estructura.

<sup>38</sup> Lo entrecomillado es tomado de las páginas 509 y 510 de la sentencia. Esta postura ya había sido adelantada en una sentencia de 1983, en el caso «*United States Postal Service Board of Governors v. Aikens*», donde se sostuvo que «cuando el demandado no logra persuadir al tribunal de distrito de que desestime la acción por falta de un caso *prima facie*, pero responde a la prueba del actor ofreciendo evidencia acerca de los motivos de su rechazo [...], la presunción queda por fuera del caso»; asimismo, cuando «el demandado ha hecho todo lo que le sería requerido si el actor hubiera establecido adecuadamente un caso *prima facie*, entonces deja de ser relevante si el actor realmente lo hizo» (460 U.S. 711, pp. 714-715).

<sup>39</sup> La disidencia toma una postura muy distinta: «bajo *McDonnell Douglas* y *Burdine*, la prueba del caso *prima facie* no solo hace surgir una inferencia de discriminación; en ausencia de mayor evidencia, también crea una presunción obligatoria en favor de la actora» (509 U.S. 502, p. 528). En una de sus críticas a la postura mayoritaria, observó que así planteada beneficia al demandado que miente por sobre el que omite. La lectura de «*McDonnell Douglas*» que hace la mayoría de la Suprema Corte en «*Hicks*» ha sido entendida como un intento de insular los casos individuales de discriminación intencional de los potenciales efectos que podría tener sobre los empleadores la estructura desarrollada previamente en «*Griggs v. Duke Power Co.*» (1971). Véase BAUMANN et. al. (1991), MALAMUD (1995).

<sup>40</sup> Dada la levedad de la carga C2 tal como la entiende la mayoría en «*Hicks*», en los hechos podría ser sumamente difícil que la presunción tenga estos efectos.

ce un caso *prima facie*» (509 U.S. 502, p. 509-510), dado que en ese supuesto quedarán cuestiones fácticas a resolver por parte de los jueces de los hechos (*trier of fact*).<sup>41</sup>

El fallo «Hicks» deja entrever dos visiones muy distintas de la naturaleza y funciones de la estructura «McDonnell Douglas». Para la mayoría, esta última es simplemente una forma de organizar la presentación de la prueba y de compeler a la demandada a introducir evidencia, pero no de constreñir la plataforma fáctica sobre la que decidirá el juez (o jurado). Articulado el motivo legítimo e introducida la prueba de su falsedad, el caso se resuelve como un caso civil más en el que la prueba de falsedad puede ser (o no) un indicio relevante en la inferencia de la discriminación. Para la disidencia, en cambio, el mayor propósito del esquema radica en delimitar el marco fáctico sobre el cual el juez debe resolver. Si bien articulado el motivo legítimo los hechos «adquieren un nuevo nivel de especificidad» (y ya no es relevante analizar si la actora acreditó el caso *prima facie*), la indagación en ese nivel se limita al análisis de la prueba del motivo dado por la demandada y a la prueba que intenta desvirtuarla.

En «Hicks» la Suprema Corte no fue clara sobre en qué supuestos la prueba de falsedad de los motivos alegados al cumplir C2 podría ser un indicio determinante. El voto mayoritario expresó que el jurado «podría» entender que el cumplimiento de C3 es suficiente prueba de discriminación, pero también dijo que no basta con «no creerle al demandado», sino que «es necesario creerle a la actora» y que, además de la falsedad del motivo legítimo, corresponde probar el motivo discriminatorio. Posteriormente, en «Reeves v. Sanderson Plumbing Products» (2000), la Corte se ocupó de aclarar, entre otros puntos, que siempre está permitido inferir la motivación discriminatoria a partir de la prueba de la falsedad de los motivos alegados al cumplirse con C2.<sup>42</sup> Salvo los casos en los que, del análisis de toda la prueba, surge de manera conclusiva algún otro motivo no discriminatorio<sup>43</sup>.

---

<sup>41</sup> En estos casos la Suprema Corte distingue entre juez y juez de los hechos porque está analizando decisiones en las que los jueces pueden fallar sobre el mérito del caso o bien dejar que decida el jurado. Pero incluso cuando no hay jurado, el juez puede resolver «de puro derecho» o no cuando debe resolver controversias fácticas. Si bien «McDonnell Douglas», «Burdine» y «Hicks» fueron casos juzgados por jueces de banco (*bench trials*), a partir de una reforma legal operada en 1991, en los casos discriminatorios el juez de los hechos lo conforma el jurado.

<sup>42</sup> En este caso la Cámara de Apelaciones había resuelto que la moción para obtener un *judgment as a matter of law* había sido mal denegada, por cuanto la parte actora no había probado el intento discriminatorio, pese a haber demostrado la falsedad del motivo legítimo articulado por la demandada. La Suprema Corte de Estados Unidos revocó esta decisión, bajo el entendimiento de que, en esos supuestos, el juez podría inferir el intento discriminatorio a partir de la demostración de que el motivo brindado por el empleador es falso. Y por tanto, el juez no podía suplantar el veredicto del jurado. Sin embargo, la Suprema Corte dejó abierta la posibilidad de otorgar esa moción en los casos en que hubiera «prueba conclusiva» de un motivo no discriminatorio. Algunos académicos entienden que «Reeves» contiene la articulación moderna más clara de la estructura McDonnell. Véase SPERINO (2018, Capítulo 3, p. 12).

<sup>43</sup> Esto tiene especial relevancia en el sistema estadounidense, pues puede determinar hasta qué punto el juez tiene la atribución de evitar el veredicto del jurado.

### 3.2. El encaje procesal de las cargas de «McDonnell Douglas»

Luego de brindar un panorama general sobre la evolución de la regla probatoria de «McDonnell Douglas», nos detendremos en un aspecto complementario: el encaje procesal de las cargas que se desprenden de aquel criterio. ¿De qué mecanismos procesales disponen las partes para cumplir con ellas?

El sistema procesal civil de los Estados Unidos se organiza, a grandes rasgos, en tres etapas: (i) la etapa previa al juicio (*pre-trial*); (ii) el juicio (*trial*); y (iii) la etapa posterior al juicio (*post-trial*).<sup>44</sup>

Durante la etapa previa, la actora presenta su demanda (*complaint*) y la demandada sus defensas formales (*motions to dismiss*) y su contestación (*answer*). Su objetivo central es dar una oportunidad a las partes para que preparen sus casos para el juicio oral, descartar los reclamos que son legalmente insostenibles y determinar la plataforma fáctica sobre la que el juez o jurado deberá resolver. Para los fines de este trabajo nos interesa destacar las siguientes características de esta etapa:

(a) Flexibilidad en la fijación del *thema decidendum* (y del *thema probandum*). Las partes concretan sus alegaciones de fondo a través de la demanda y su contestación, inicialmente. Pero también pueden modificarlas posteriormente a través de enmiendas, de acuerdo con lo que se vaya descubriendo acerca de las posturas de sus rivales y a partir de la evidencia recolectada.<sup>45</sup> Esto es destacado por HAZARD y TARUFFO en los siguientes términos:

Bajo el proceso estadounidense moderno, las enmiendas están permitidas libremente, en particular cuando reflejan nueva información revelada en el *discovery* previo al juicio. Por lo tanto, los alegatos presentan las cuestiones de hecho para su determinación en el juicio ajustadas a la luz de la información producida mediante el *discovery*. (HAZARD y TARUFFO, 1993: 107, traducción propia).

(b) Acceso amplio a las evidencias. El mecanismo del *discovery* (descubrimiento) faculta a las partes, antes de ir a juicio, a recolectar las evidencias que se encuentren en poder o bajo el control de sus rivales y que podrían dar apoyo a sus respectivas posiciones.<sup>46</sup> Les habilita a investigar haciendo solicitudes directas a sus contrapartes (sin necesidad de aprobación judicial) y, sobre la base de las evidencias recolectadas y de la información que de ellas surge, ajustar sus pretensiones y las premisas fácticas que les dan sustento.<sup>47</sup> El alcance de lo que se puede obtener en ejercicio de esta

<sup>44</sup> Para un panorama general al respecto, véase JAMES y HAZARD (1985) y HAZARD y TARUFFO (1993). Nos referiremos al proceso civil *federal*, aunque las características que destacaremos suelen estar presentes también en los sistemas procesales de los estados que componen la unión.

<sup>45</sup> Cfr. Regla 15 de las Reglas Federales de Procedimiento Civil (FRCP, por sus siglas en inglés).

<sup>46</sup> Regulado en el Título V de las FRCP, reglas 26 a 37.

<sup>47</sup> También permite apoyar mociones solicitando que se resuelva el mérito del caso antes del juicio: «el *discovery* no solo puede ser usado para prepararse para el juicio, sino también para obtener evidencias que lo impidan» (HAZARD y TARUFFO, 1993: 114, traducción propia).

facultad es amplio: las evidencias referidas a cualquier asunto no privilegiado que sea relevante para las posturas de las partes y proporcional a las necesidades del caso.<sup>48</sup> Esto incluye la obtención de testimonios, documentación, información electrónicamente almacenada y otros objetos tangibles.

(c) Instancias de decisión sin juicio. Se prevén ciertas instancias en las cuales se puede decidir acerca del mérito del caso sin atravesar un juicio oral o sin que el jurado se expida. Una es el *summary judgment* (juicio sumario), que puede ser solicitado hasta 30 días después del cierre del *discovery*.<sup>49</sup> Otra es el *judgment as a matter of law* (juicio o decisión como una cuestión de puro derecho), que puede ser solicitado en cualquier momento del juicio hasta que el caso sea sometido a deliberación del jurado.<sup>50</sup> Ambos le permiten al juez decidir acerca del mérito del caso si no hay hechos relevantes controvertidos, o bien si considera que no hay evidencia suficiente para que un jurado racional le otorgue la razón a la contraparte de quien presentó la moción.<sup>51</sup>

Retomemos el foco en los casos por discriminación fáctica. ¿De qué mecanismos disponen las partes para cumplir con las cargas de «McDonnell Douglas»? Mínimamente disponen, según acabamos de ver, de reglas flexibles para formular sus respectivas alegaciones y del *discovery* como mecanismo para acceder a la evidencia. Así, cuentan con un amplio acceso a la evidencia y con la posibilidad de *reaccionar* a la información que surja en el transcurso de la etapa previa, reformulando sus alegaciones y/o solicitando la producción de evidencias adicionales si lo creyesen necesario.

Imaginemos que un grupo de personas presentan una demanda colectiva donde alegan haber sido despedidas de la empresa donde trabajaban debido al credo religioso que profesan.<sup>52</sup> Acreditan inicialmente que los despidos se dieron luego de que la demandada conociera su religión, que la demandada tenía una manifiesta animadversión hacia la misma y que en los despidos acaecidos ese año se dio la siguiente desproporción: si bien el total de empleados que profesaban el credo en cuestión era de 1/3, el total que adhería al mismo entre las personas despedidas fue de 2/3. Esto podría ser suficiente para dar por cumplida la carga C1. Imaginemos también que, una vez emplazada, la empleadora contesta la demanda y pretende cumplir con C2

<sup>48</sup> Cfr. Regla 26(b)(1), FRCP.

<sup>49</sup> Cfr. Regla 56 de las FRCP. Según esta disposición, el *summary judgment* cabe cuando «no existe una disputa genuina relativa a ningún hecho material» (traducción propia).

<sup>50</sup> Cfr. Regla 50 de las FRCP. El juez puede hacer lugar a una moción de *judgment as a matter of law* en cualquier momento del juicio, «tan pronto como sea evidente que cualquiera de las partes no puede soportar la carga de la prueba que es esencial para su caso» (SUBRIN et al., 2021: 150, traducción propia). El pedido, de ser rechazado, puede ser renovado luego del veredicto del jurado.

<sup>51</sup> Las instancias están relacionadas de modo tal que «el estándar para otorgar un juicio sumario ‘refleja’ el estándar para la decisión como una cuestión de puro derecho, de modo que ‘la investigación bajo cada uno es la misma’» (530 U.S. 133, p. 150, citas omitidas).

<sup>52</sup> El ejemplo es una libre reformulación de un caso similar, «Coates v. Johnson & Johnson» (756 F.2d 524, 7th Cir., 1985), citado en FINKELSTEIN (2009: 149).

alegando que, en realidad, todos los despidos estuvieron motivados en las sanciones disciplinarias recibidas por los empleados durante los doce meses previos al despido, lo que respalda con documentación pertinente. Eso parece refutar la hipótesis de la motivación discriminatoria. Ahora bien, enterada del motivo alegado por la demandada, y de ser cierto lo de las sanciones, la parte actora podría reaccionar durante el *pre-trial* reorientando su estrategia argumentativa. Podría ampliar su investigación, requerir los registros donde constan las sanciones y el tipo de faltas cometidas e indagar y buscar evidencias sobre una hipótesis alternativa que explique tanto las sanciones como los despidos desde el prisma de la discriminación. Esta podría consistir, por caso, en que el criterio para identificar faltas e imponer sanciones estuvo sesgado según la religión de los trabajadores. No es que cometieran más faltas sino que eran los que proporcionalmente más sanciones recibían por ellas. Si esta hipótesis fuera cierta y pudiera acreditarse, la que alega la demandada se vería debilitada y la que afirma un motivo discriminatorio, fortalecida.

A partir de este ejemplo se puede dimensionar cómo, en sistemas procesales como el estudiado, la parte actora se encuentra en condiciones de terminar la preparación de su caso trabajando sobre una serie de hipótesis que, de otro modo, quedarían libradas a su imaginación, por no poder acceder a la información que llevase a plantearlas o por no poder conocer la defensa de su contraparte en tiempo útil. Una vez que el caso *prima facie* resulta rebatido, «la investigación empírica [avanza] hacia un nuevo nivel de especificidad» («Burdine», 1981, 450 U.S. 248, p. 255). En sistemas como el estudiado, la parte actora, así como su contraparte, se encuentran en condiciones de *reaccionar* a la nueva información que desconocían al momento de entablar demanda o defenderse.

El ejemplo también ayuda a ilustrar un punto que es importante para este trabajo: la forma con la que está regulada la etapa previa de los procesos civiles estadounidenses, en especial por su flexibilidad y por el *discovery*, influye positivamente —y muchas veces resulta decisiva— sobre la posibilidad de cumplir con cargas como C1 y especialmente como C3 (en el ejemplo, investigar ampliamente sobre el sesgo en las sanciones), a la vez que contribuye a ampliar la información total de la que dispondrán los jueces para decidir. Esto mismo ha sido destacado por la Suprema Corte de los Estados Unidos. La existencia del *discovery* fue uno de los argumentos que esgrimió en «Burdine» para sostener que el alcance que decidió darle a la carga C2 (un nivel de exigencia leve) no perjudicaba indebidamente a quienes accionan por discriminación.<sup>53</sup> «Antes de que el discovery haya desenterrado hechos y pruebas

<sup>53</sup> En el caso la Suprema Corte sostuvo que no estaba de acuerdo en que «limitar la obligación probatoria de la parte demandada a una carga de producción obstaculice indebidamente a la parte actora» (450 U.S. 248, p. 258). Ello por tres razones: porque se exige que la explicación de la demandada sea razonablemente específica; porque, aunque no tenga una carga de persuasión, la demandada de todos modos conserva un incentivo para persuadir al juez de los hechos de que su decisión acerca del empleo fue legítima; y porque existen reglas de *discovery* liberales que además son complementadas por el acceso a las carpetas de investigación de la EEOC correspondientes a su caso. «Dados estos factores, no estamos

relevantes», sostuvo en otro caso, «puede ser difícil definir la formulación precisa del caso *prima facie* requerido en un caso particular» («Swierkiewicz», 2002, 534 U.S. 506, p. 512).

¿En qué instancias se decide acerca del cumplimiento de las cargas de «McDonnell Douglas»? Los tres pasos esbozados en ese precedente no se corresponden con etapas específicas dentro del proceso. Las partes canalizan sus cargas, a través de los lentes de aquel esquema, en la demanda y su contestación, en las defensas previas, en las mociones para obtener un *summary judgment*, en la presentación de sus casos durante el juicio, etcétera.<sup>54</sup> Se trata de una estructura que ordena el modo en que las partes alegan los hechos y la evidencia en sus distintas peticiones, mociones y contestaciones; y que brinda a los jueces un esquema y pautas para resolver esas peticiones. El cumplimiento de las respectivas cargas, así, se decide de modo acumulado al rechazar o acoger una moción por *summary judgment* o una por *judgment as a matter of law*. Lo mismo ocurre en el juicio oral al expedirse —jueces profesionales o jurados— sobre los hechos probados en el caso.

Distintas cuestiones relativas al marco «McDonnell Douglas» pueden surgir en diversos momentos procesales y depender de la estrategia de las partes. Por ejemplo, es posible que la actora inicie su reclamo indicando los elementos del caso *prima facie* y que se anticipe a los motivos que posiblemente alegará la demandada, o bien puede esperar escucharla y a atravesar el *discovery* para presentar evidencia que contradiga su hipótesis. Del mismo modo, la demandada puede defenderse en su contestación atacando los elementos del caso *prima facie*, o bien aduciendo un motivo legítimo de su actuar (o ambas cosas a la vez). Este mismo esquema se repetirá luego en instancia del juicio, aunque es usual que la actora se adelante a rebatir el motivo al presentar su caso, pues ya habrá advertido en la etapa de *pre-trial* cuáles son las cartas de la demandada.

Es muy frecuente que «McDonnell Douglas» penetre en el modo que se estructuran las decisiones judiciales que se adoptan en instancia del *summary judgment* y ante las mociones de *judgment as a matter of law*. Actualmente, una gran parte de los casos discriminatorios se resuelven en instancias de *summary judgment*.<sup>55</sup> En general, el

---

convencidos de que el actor encuentre particularmente difícil probar que una explicación ofrecida sin una base fáctica es un pretexto» (ídem anterior).

<sup>54</sup> El cumplimiento de estas cargas fue interpretado por algunos tribunales como un elemento de la acción, cuya ausencia configura un vicio formal. Pero esta postura fue descartada por la Suprema Corte en el caso «Swierkiewicz v. Sorema N. A.» (2002), donde se sostuvo que, para ser formalmente admisible, una demanda por discriminación en el empleo no necesita contener alegaciones sobre hechos específicos que configuren un caso *prima facie*. Por ende, no corresponde desestimar un caso discriminatorio a través de una *motion to dismiss* (defensa formal de la etapa previa). Este criterio, sostuvo la Corte en tal oportunidad, se basa en que el sistema dispone de reglas liberales de *discovery* (para terminar de apuntalar las respectivas posturas) y mociones de *summary judgment* (para deshacerse de los reclamos sin mérito).

<sup>55</sup> Véase EYER (2019: 977, nota al pie 55), describiendo una muestra tomada sobre la base de las decisiones adoptadas por las Cámaras de Apelaciones de Estados Unidos durante un periodo de 3 meses: 60 de 63 decisiones analizadas se resolvieron en instancias de *summary judgment*.

empleador demandado interpone una moción por *summary judgment* alegando que la actora no tiene evidencia suficiente para probar el caso *prima facie* y/o articulando un motivo legítimo, según la evidencia introducida en el *discovery*. La actora tiene la carga de responder demostrando que, según la evidencia introducida, hay hechos controvertidos suficientes de modo tal que un jurado racional podría entender que existe un caso *prima facie*, que el motivo alegado por la demandada es falso y que existe un motivo discriminatorio.<sup>56</sup>

#### 4. ¿CÓMO MEJORAR LA REGLA PARA EL CONTEXTO ARGENTINO? LOS CAMINOS QUE LA EXPERIENCIA ESTADOUNIDENSE NOS INVITA A PENSAR

Volvamos sobre la pregunta con la que abrimos este trabajo. ¿Cómo se puede mejorar la regla fijada por la Corte Suprema de la Nación argentina respecto de la prueba de los motivos discriminatorios? La experiencia estadounidense, representada aquí mediante algunos hitos jurisprudenciales de su Suprema Corte, nos invita a pensar en al menos dos caminos a desarrollar, sin que esto suponga seguir las soluciones que ha venido adoptando. Agruparemos estas reflexiones en dos, una sobre la precisión en la formulación de las cargas asignadas a las partes (4.1.) y otra sobre el encaje procesal que posibilita cumplirlas (4.2.).

##### 4.1. Precisar las cargas

Las sentencias estudiadas sugieren que hay distintas alternativas para precisar la formulación de las cargas probatorias de las partes en los casos de discriminación fáctica. También que la falta de precisión puede generar confusiones acerca de cómo actuar en el proceso (lo que afecta a las partes) y acerca de cómo decidir (lo que afecta a los jueces). ¿Cómo pueden ser precisadas las cargas probatorias en los casos de discriminación fáctica? Presentaremos algunas propuestas. Para ello, seguiremos con las notaciones «C1», «C2» y «C3» para referirnos a las cargas de acreditar un caso

---

<sup>56</sup> Se ha señalado que la falta de claridad de «Hicks» respecto de cuál es el criterio para dar por probada la existencia de un motivo discriminatorio ha llevado al abuso de los institutos del *summary judgment* y el *judgment as a matter of law* por parte de empleadores y jueces. En ese sentido, «Reeves» es entendido como un intento de la Suprema Corte para frenar este abuso, al indicar que, si la actora introduce evidencia de falsedad del motivo alegado, la regla es que el caso puede resolverse como «cuestión de puro derecho» solo cuando existe prueba conclusiva de un motivo no discriminatorio. Luego de «Hicks», sería razonable pensar que, una vez cumplida C2, las cortes deberían dejar de lado «McDonnell Douglas» y enfocarse en establecer cuál de las presentadas por las partes es la hipótesis más probable. No obstante, la práctica demuestra que siguen enmarcando sus análisis en aquel esquema, aunque muchas veces hagan confluir el análisis de las tres etapas en uno. En general, si bien declaran estar aplicando la estructura de «McDonnell Douglas», en muchos casos el análisis hace confluir sus etapas en una evaluación combinada. La diferencia es sutil, pero puede tener relevancia práctica. Véase MALAMUD (1995).

*prima facie* de discriminación, refutarlo y contestar a la refutación, respectivamente. Y recuperaremos «R1» y «R2» para dar cuenta de dos formulaciones distintas para la carga C1, una similar a una presunción *hominis* (y por ende a toda inferencia probatoria no deductiva)<sup>57</sup> y otra similar a una presunción *iuris tantum*,<sup>58</sup> respectivamente.

(a) *Caso prima facie (C1)*

La carga C1 puede ser formulada como R1 (presunción *hominis*) o como R2 (presunción *iuris tantum*). O bien como R2 con una regla subsidiaria que remita a R1 para todo caso no regulado por una norma con forma R2 (es decir, siempre que no fuera aplicable una presunción *iuris tantum*). Esta última ha sido la opción adoptada por la Suprema Corte estadounidense y podría ser la que esté dispuesta a incorporar la Corte Suprema argentina.

Formulada a partir del antecedente de R2, C1 tiene un contenido como este:

C1-R2: La parte actora tiene la carga inicial de probar que se han dado instancias de las circunstancias A, B y C (en cuyo caso se debe aceptar como verdadera la conclusión D, salvo que se pruebe la falsedad de D en el caso concreto).

Formulada a partir del antecedente de R1, C1 adopta un contenido como este:

C1-R1: La parte actora tiene la carga inicial de probar *indicios suficientes* a favor de la existencia de motivos discriminatorios en una cierta acción (en cuyo caso se debe aceptarla como verdadera).

La pregunta sobre cuándo hay indicios «suficientes» para activar el consecuente de C1-R1 es traducible como cuál es el *estándar de prueba* para dar por cumplida esa carga. Podría convenirse en que cabe aplicar un estándar más liviano que el comúnmente requerido.<sup>59</sup> ¿Pero cómo expresarlo con más detalle? Un modo de hacerlo es recordar el argumento de la Suprema Corte estadounidense según el cual lo que se exige inicialmente es descartar como hipótesis los motivos no discriminatorios *más comunes* que podrían presentarse en cada caso, según el tipo de caso de que se trate.<sup>60</sup> Mientras que las reglas de forma R2 cristalizan estos motivos en premisas fijas (*v.gr.* que el postulante no esté capacitado, que no haya plazas disponibles para el empleo, etcétera), las reglas de forma R1 los dejan librados a la discrecionalidad judicial. De este modo, quien accione tendrá la carga de despertar la sospecha de que los motivos de la acción cuestionada son discriminatorios, acreditando que los motivos no discrimi-

<sup>57</sup> «R1», recordemos, tenía esta forma: «Si hay indicios suficientes a favor de una proposición (la existencia de motivos discriminatorios en una cierta acción), entonces se debe aceptarla como verdadera».

<sup>58</sup> «R2», recordemos, tenía una forma como esta: «Si se encuentran probadas instancias de las circunstancias A, B y C, entonces se debe aceptar como verdadera la conclusión D, salvo que se pruebe la falsedad de D en el caso concreto».

<sup>59</sup> Esto es, después de todo, lo que le da sentido a la regla.

<sup>60</sup> Brindó esta explicación, recordemos, al desarrollar por qué entendía que la prueba de ciertas premisas permitía dar por cumplida C1 (por ejemplo, en 431 U.S. 324, p. 358).

minatorios más comunes que suelen guiar esa clase de acciones no se presentan en el caso. Basados en esta idea, podría reformularse C1-R1 como sigue:

C1-R1': La parte actora tiene la carga inicial de probar que los motivos no discriminatorios *más comunes* que suelen guiar a las acciones del tipo al que pertenece la acción cuestionada no se presentan en el caso.

(b) *Refutación (C2)*

La carga C2 también puede ser formulada con distintos niveles de exigencia. Una opción, más leve (L), es exigir simplemente una carga de argumentar y de producir evidencias. De modo que:

C2-L: Cumplida C1, la parte demandada tiene la carga de brindar una explicación y producir evidencias que apunten a que el acto cuestionado estuvo guiado por un motivo no discriminatorio.<sup>61</sup>

Otra opción, de exigencia más elevada (E), es requerir que la demanda pruebe que el motivo para realizar el acto cuestionado fue uno no discriminatorio. De modo que:

C2-E: La parte demandada tiene la carga de probar, por preponderancia de la evidencia, que el motivo para realizar el acto cuestionado fue uno no discriminatorio.

(c) *Contestación (C3)*

La carga C3, finalmente, también es susceptible de distintas formulaciones que se traducen en la exigencia de niveles de suficiencia —estándares de prueba— menores o mayores para que se tenga por probada la hipótesis que alega la parte actora. Unas exigen menos esfuerzos epistémicos que otras. Una formulación con un nivel de suficiencia leve (L) consiste en exigirle a la actora que desacredite únicamente la hipótesis alegada por la demandada. De modo que:

C3-L: Cumplida C2, la parte actora tiene la carga de probar, por preponderancia de la evidencia, que es falso que el acto cuestionado se haya guiado por el motivo alegado por la demandada.

Una formulación con un nivel de suficiencia más elevado (E), en cambio, le exige a la actora que pruebe que su hipótesis es más probable que su negación y que todas las demás hipótesis que plausiblemente expliquen el acto impugnado mediante motivos no discriminatorios (hayan sido alegadas o no por la demandada). De modo que:

C3-E. Cumplida C2, la parte actora tiene la carga de probar, por preponderancia de la evidencia, que su afirmación acerca de que el acto impugnado obedeció a motivos discriminatorios es más probable que su negación y que todas las demás hipótesis plausibles según las cuales dicho acto estuvo guiado por motivos no discriminatorios.

<sup>61</sup> Elegir esta opción requeriría ajustar C1-R2 de modo que la presunción sea derrotada no cuando «se pruebe la falsedad de D» sino cuando se alegue y produzca evidencias sobre un motivo no discriminatorio.

Las dos formulaciones de C3 suponen establecer distintos estándares de prueba para que se tenga por cumplida dicha carga. Hay muchas maneras de fijar estos umbrales de suficiencia, de modos menos o más exigentes.<sup>62</sup> Por razones de simplicidad hemos contrapuesto dos alternativas que consideran solo uno de los varios elementos que podrían tenerse en cuenta para graduar la corroboración de una hipótesis. Esta estrategia, sin embargo, permite demostrar nuestro punto: que la regla formulada por la CSJN puede precisarse y que esta operación puede darse (entre varias opciones) en el sentido expuesto.

La elección por una u otra variante para cada carga supone establecer equilibrios distintos en los procesos por discriminación entre los intereses de sus partes.

#### 4.2. Considerar el encaje procesal

La experiencia estadounidense también nos lleva a pensar en el encaje procesal de las cargas probatorias que venimos estudiando. Como se ha advertido, un sistema procesal flexible y con herramientas para acceder ampliamente a la evidencia influye positivamente —y muchas veces resulta decisivo— sobre la posibilidad de cumplir con las cargas probatorias atribuidas a las partes en los casos de discriminación fáctica. La Suprema Corte estadounidense ha sido enfática en que la parte actora debe contar con las oportunidades que un sistema así brinda, en especial para cumplir con C3.<sup>63</sup> Después de todo, ¿cómo podría rebatir la hipótesis alegada por su contraparte antes de que esta la alegue?

Ahora cabe preguntarnos en qué situación se encuentran, a este respecto, los sistemas procesales civiles vigentes en Argentina (que son similares a los vigentes en muchos otros países de la tradición jurídica continental). Se encuentran, pues, en una situación bastante problemática en relación con los dos aspectos relevantes que hemos destacado antes (la flexibilidad y el *discovery*).

Los códigos de procedimiento civil vigentes no suelen prever para los casos por discriminación (tampoco para los demás) ni herramientas que faculten a las partes a acceder *ampliamente* a la evidencia en poder o bajo el control de su contraparte, ni *flexibilidad* en la determinación de las alegaciones y pretensiones que defenderán en juicio (en general se prohíbe la *duplica*).<sup>64</sup> Tomando como ejemplo las previsiones del Código Procesal Civil y Comercial de la Nación, vemos que este limita la produc-

<sup>62</sup> En FERRER BELTRÁN (2021: 208-210) se proponen como ejemplos hasta siete formulaciones distintas.

<sup>63</sup> Ha destacado, por ejemplo, que la parte actora «debe ser dotada de la oportunidad de presentar evidencias acerca de que la justificación ofrecida [por la demandada] es meramente un pretexto para la discriminación» (438 U.S. 567, p. 578).

<sup>64</sup> TARUFFO (2008: 117-118) explica que el *discovery* —y en general la separación entre la fase procesal preliminar y la vista— es uno de los rasgos distintivos más importantes de los sistemas del *common law* en comparación con los del *civil law*. En el mismo sentido véase HAZARD y TARUFFO (1993: 125).

ción de evidencia antes del juicio a supuestos excepcionales, en los que haya alguna urgencia<sup>65</sup>, y no prevé la posibilidad de exigirle a la contraparte que descubra todas las evidencias que posee o están en su poder.<sup>66</sup> El *thema decidendum*, por su parte, se establece por la relación entre la demanda que da inicio al proceso y su contestación. Los hechos a probar serán aquellos que resulten controvertidos en esa dialéctica y las evidencias a producirse las que se hubieran ofrecido en ambas piezas. Excepcionalmente, cuando al responder la demanda «se alegaren hechos no invocados en la demanda o contrademanda, los demandantes o reconvinientes podrán ofrecer prueba y agregar la documental referente a esos hechos, dentro de los cinco días» (art. 334)<sup>67</sup>. En algunas jurisdicciones el alcance de esta facultad se reduce, notablemente, a la incorporación de documentos.<sup>68</sup> Los casos por discriminación, por su parte, se suelen tramitar mediante juicios de amparo y otras variantes de juicios sumarísimos,<sup>69</sup> que son las vías procesales más acotadas en actos procesales y que restringen el diálogo entre las partes al mínimo, eliminando, por ejemplo, la presentación de alegatos.<sup>70</sup>

En sistemas así, las reglas probatorias analizadas previamente podrán incentivar a la parte demandada a desplegar cierta actividad probatoria (de distinta intensidad según su formulación), pero no alivianan la dificultad de la parte actora para cumplir con sus cargas. En especial con una carga de refutación como C3, puesto que conocerá la hipótesis que le toca refutar en un momento donde no podrá adecuar

<sup>65</sup> El código prevé la producción anticipada de evidencias pero solo para el supuesto donde las partes «tuvieren motivos justificados para temer que la producción de sus pruebas pudiera resultar imposible o muy dificultosa en el período de prueba» (art. 326). En ese caso se habilita: tomar la declaración de algún testigo de avanzada edad, gravemente enfermo o próximo a ausentarse del país; el reconocimiento judicial o dictamen pericial sobre la existencia de documentos o el estado, calidad o condición de cosas o de lugares; el pedido de informes; y/o la exhibición, resguardo o secuestro de documentos, etc.

<sup>66</sup> Se da una excepción respecto de los *documentos*: una parte está facultada a solicitar la incorporación al proceso de aquellos *que se encontrasen en poder de su contraparte*; y la negativa de esta última a presentarlos «constituirá una presunción en su contra» si «resultare manifiestamente verosímil su existencia y contenido» (art. 388). Esta facultad se extiende además a los documentos en poder de terceros (art. 396). Otra excepción tiene que ver con la facultad de solicitar informes a «oficinas públicas, escribanos con registro y entidades privadas» sobre hechos concretos controvertidos en el proceso y «que resulten de la documentación, archivo o registros contables del informante» (art. 396).

<sup>67</sup> Estos hechos son los que la doctrina llama *nuevos hechos*. También se prevén otras ventanas: la alegación de *hechos nuevos* que, relacionados con la cuestión ventilada, ocurren o llegan a conocimiento de las partes entre el momento de la traba de la litis (cuando la demanda es contestada) y hasta los cinco días posteriores a la audiencia preliminar (art. 365); y la facultad del juez de tomar en consideración hechos que, aunque no sean alegados por las partes, resulten probados (art. 163, inciso 6).

<sup>68</sup> Así lo establece, por ejemplo, el Código Procesal Civil y Comercial de la Provincia de Buenos Aires (art. 333).

<sup>69</sup> Todos los casos de la CSJN estudiados en #2, «Pellicori» (2011), «Sisnero» (2014), «Varela» (2018) y «Fontana» (2021), fueron tramitados por la vía del amparo o de juicio sumarísimo (lo mismo «Romero», 2021).

<sup>70</sup> Véase el art. 498, inciso 5 del Código y el art. 11 de la Ley 16.986. Cabe destacar que el art. 334 del Código está previsto concretamente para los procesos ordinarios. La extensión de su alcance a los procesos sumarísimos ha sido obra interpretativa, doctrinal y jurisprudencial (véase: PALACIO y ALVARADO VELLOSO, 1994: 74; HITTERS, 2008).

su escrito postulatorio y, sobre todo, se verá impedida de realizar una investigación profunda acerca de las evidencias disponibles relevantes para dicho propósito, más allá de la limitada facultad de ofrecer, en un plazo de cinco días, la producción de nuevas evidencias de cuya existencia supiera, referidas a lo que la demandada hubiera expresamente alegado. La actora, además, carecerá de la posibilidad de *reaccionar* al resultado de la evidencia ofrecida por su contraparte, cuyo contenido conocerá recién durante el juicio.

En el caso que pusimos de ejemplo unas páginas atrás, sobre los despidos que se sindicaban motivados por la religión de los trabajadores desafectados, la parte actora tendría relevantes dificultades para cumplir con C3. La demandada alegaría su motivo no discriminatorio, las sanciones disciplinarias, al contestar la demanda. Por ende, la actora solo podrá solicitar la producción de alguna evidencia acerca de ello de cuya existencia supiera. Una vez recibida esta última, carecerá de oportunidad de requerir la producción de otra evidencia adicional que, a tenor suyo, se demostrara relevante. Además, no podrá reformular su demanda a la luz de la nueva información agregada.

Un problema similar podría darse en el caso «Fontana» resuelto por la Corte Suprema de Justicia argentina. Recordemos: la actora alegaba haber sido despedida por motivos sindicales y la demandada se defendió aduciendo que la decisión de despedirla había obedecido a graves problemas económicos: el mercado automotriz estaba en caída y junto con la actora habían sido desvinculadas treinta y cuatro personas más por falta o disminución de trabajo. Ello configuraría una causa seria y objetiva, ajena a toda discriminación (como sostuvo el voto mayoritario de la Corte), pero siempre y cuando fuera cierto (agregamos). Sin embargo, la actora contaría con una oportunidad muy limitada para cuestionarlo, mostrando que la crisis no era tal, que los despedidos no eran tantos, que de todos los empleados los únicos despedidos fueron quienes ejercían actividades sindicales, entre muchas otras alternativas que —de darse— le quitarían peso a la hipótesis de la demandada.

Pues bien, pese a que la CSJN se inspiró en la jurisprudencia de su par estadounidense, la regla que incorporó no tiene las mismas condiciones de funcionamiento que las que detenta en su jurisdicción original. Sistemas procesales como el argentino colocan en un mayor nivel de dificultad a la parte actora en comparación con la posición en la que estaría en el sistema estadounidense. «McDonnell Douglas» y el modo en que se aplica en los Estados Unidos nos invita, entonces, a repensar esta arista.

Este problema se podría resolver de al menos dos maneras. Una, la más evidente, consiste en modificar las reglas procesales (al menos las aplicables a los casos de discriminación) flexibilizando la constitución de la litis y brindando herramientas para acceder ampliamente a la evidencia en poder o bajo el control de la contraparte. La otra manera está vinculada con el modo de formular las cargas procesales en cabeza de cada parte, partiendo de la premisa según la cual establecer una carga leve para C2, tal y como pareciera exigir la jurisprudencia de la Suprema Corte de los Estados Unidos, no tiene los mismos efectos para las partes en un sistema procesal como el argentino. Como señala Ronald ALLEN (2014: 203), en los sistemas que no tienen

*discovery* la carga de la prueba tiene mucha más relevancia que en los sistemas que sí lo tienen. En un sistema en que las partes tienen acceso amplio a la información y poseen la oportunidad de aportar evidencia que refute los hechos y la prueba aportados por la contraparte, la determinación de quién lleva la carga de la prueba solo tendría relevancia cuando no existiera evidencia suficiente para probar un hecho controvertido. En cambio, en un sistema en que las oportunidades de alegar hechos y pruebas son limitadas, las reglas de la carga de la prueba pueden funcionar como verdaderos *incentivos* para que las partes que se encuentran en una mejor situación para producir evidencia, lo hagan. De este modo, parece que la ausencia de flexibilidad y de *discovery* en el sistema procesal es una buena razón para adoptar una formulación más exigente para C2, la que denominamos C2-E («La parte demandada tiene la carga de probar, por preponderancia de la evidencia, que el motivo para realizar el acto cuestionado fue uno no discriminatorio») o alguna variante, intentando generar los incentivos suficientes para que la parte demandada aporte toda la evidencia de que disponga y no la retace especulando con que la actora no podrá refutar las hipótesis planteadas como defensas.

## 5. CONCLUSIONES

En el texto hemos dado cuenta del modo en que la Corte Suprema argentina formuló una regla para la prueba de los motivos discriminatorios y de cómo fue formulada una regla con similar objeto en la jurisprudencia de su par estadounidense, la cual ha sido su declarada fuente de inspiración. El recorrido y el cotejo nos ha servido para:

1. Distinguir entre dos maneras de formular la carga probatoria inicial de quien acciona, traducible en dos clases de reglas: una similar a una presunción *hominis* (y a toda inferencia probatoria no deductiva) y otra similar a una presunción *iuris tantum*.

2. Notar que uno de los cursos de acción para mejorar la regla tal y como es presentada por la CSJN es precisar su formulación. A ese efecto identificamos tres cargas distintas repartibles entre las partes y dos alternativas de formulación para cada una. La opción por una u otra alternativa incide en el balance de los intereses de las partes en el proceso.

3. Advertir, por último, que las cargas probatorias de las partes encajan de modo distinto en los sistemas procesales estadounidense y argentino. Dada la falta de flexibilidad en la constitución de la litis y la ausencia de mecanismos para acceder ampliamente a la evidencia que caracterizan al sistema procesal civil argentino, formular las cargas como lo hace la Suprema Corte de los Estados Unidos coloca sobre la parte actora una exigencia superior a la que tendría en el contexto estadounidense. Esto se puede solucionar al menos de dos maneras: cambiando el sistema procesal o cambiando la formulación de las cargas, en el sentido que acaba de mencionarse unos párrafos atrás.

## BIBLIOGRAFÍA

- AGENCIA DE LOS DERECHOS FUNDAMENTALES DE LA UNIÓN EUROPEA y CONSEJO DE EUROPA. (2018). *Manual de legislación europea contra la discriminación*, Oficina de Publicaciones de la Unión Europea, [https://www.echr.coe.int/documents/handbook\\_non\\_discr\\_iaw\\_spa.pdf](https://www.echr.coe.int/documents/handbook_non_discr_iaw_spa.pdf)
- AGUILÓ-REGLA, J. (2018), «Las presunciones en el Derecho», *Anuario de filosofía del derecho*, 34, 201-227, <http://hdl.handle.net/10045/86475>
- ALLEN, R. J. (2014), «Burdens of proof», *Law, Probability and Risk*, 13(3-4), 195-219, <https://doi.org/10.1093/lpr/mgu005>
- BAUMANN, P. T. B., BROWN, J. O. y SUBRIN, S. N. (1991), «Substance in the Shadow of Procedure: The Integration of Substantive and Procedural Law in Title VII Cases». *Boston College Law Review*, 33, 211-303, <https://lawdigitalcommons.bc.edu/bclr/vol33/iss2/1/>
- DEI VECCHI, D. (2019), «El carácter presuntivo de las presunciones absolutas», *Revus - Journal for Constitutional Theory and Philosophy of Law*, 38, <https://doi.org/10.4000/revus.5333>
- EYER, K. (2019), «The Return of the Technical McDonnell Douglas Paradigm», *Washington Law Review*, 94, 967-1017, <https://digitalcommons.law.uw.edu/wlr/vol94/iss3/2/>
- FERRER BELTRÁN, J. (2007), *La valoración racional de la prueba*, Marcial Pons.
- (2021), *Prueba sin convicción. Estándares de prueba y debido proceso*, Marcial Pons.
- FINKELSTEIN, M. O. (2009), *Basic Concepts of Probability and Statistics in the Law*, Springer, <http://dx.doi.org/10.1007/b105519>
- GASCÓN ABELLÁN, M. (2010), *Los hechos en el derecho: bases argumentales de la prueba* (3.a ed.), Marcial Pons.
- HAZARD, G. C. y TARUFFO, M. (1993), *American Civil Procedure: An introduction*, Yale University Press.
- HITTERS, J. M. (2008), «Hechos nuevos, hechos sobrevinientes, nuevos hechos y nuevos documentos», *La Ley*, 998.
- JAMES, F. y HAZARD, G. C. (1985), *Civil Procedure* (3rd Edition), Little, Brown and Company.
- MALAMUD, D. C. (1995), «The Last Minuet: Disparate Treatment after Hicks», *Michigan Law Review*, 93(8), 2229-2324, <https://doi.org/10.2307/1289935>
- MENDONÇA, D. (1998), «Presunciones». *Doxa. Cuadernos de Filosofía del Derecho*, 21-I, <https://doi.org/10.14198/DOXA1998.21.1.05>
- PALACIO, L. E. y ALVARADO VELLOSO, A. (1994), *Código Procesal Civil y Comercial de la Nación. Explicado y anotado jurisprudencial y bibliográficamente: Vol. 80*, Rubinzal Culzoni Editores.
- SUPERINO, S. F. (2018), *McDonnell Douglas: The Most Important Case in Employment Discrimination Law*, Bloomberg Law.
- SUBRIN, S. N., MINOW, M. L., BRODIN, M. S., MAIN, T. O. y LAHAV, A. (2021), *Federal Rules of Civil Procedure*, Wolters Kluwer.
- TARUFFO, M. (2008), *La prueba*, L. Manríquez y J. Ferrer Beltrán (trads.), Marcial Pons.
- ULLMAN-MARGALIT, E. (1983), «On Presumption», *The Journal of Philosophy*, 80(3), 143-163, <https://doi.org/10.2307/2026132>

## Jurisprudencia

*Corte Suprema de la Nación Argentina*<sup>71</sup>

*Recurso de hecho deducido por el Estado Nacional —Ministerio de Educación, Cultura, Ciencia y Tecnología— en la causa Romero, Jonathan Iván c/ Ministerio de Educación y Deportes de la Nación s/ juicio sumarísimo*. Fallos 344:3057. 21 de octubre de 2021.

<sup>71</sup> Se pueden consultar las sentencias en: [sj.csn.gov.ar](http://sj.csn.gov.ar).

*Recurso de hecho deducido por la parte demandada en la causa Fontana, Edith Fabiana c/ Cibie Argentina S.A. s/ juicio sumarisimo.* Fallos: 344:527. 8 de abril de 2021.

*Recurso de hecho deducido por José Gilberto Varela en la causa Varela, José Gilberto c/ Disco S.A. s/ amparo sindical.* Fallos: 341:1106. 4 de septiembre de 2018.

*Recurso de hecho deducido por la actora en la causa Sisnero, Mirtha Graciela y otros c/ Taldelva SRL y otros s/ amparo.* Fallos: 337:611. 20 de mayo de 2014.

*Pellicori, Liliana Silvia c/ Colegio Público de Abogados de la Capital Federal s/ amparo.* Fallos: 334:1387. 15 de noviembre de 2011.

### *Tribunal de Justicia de la Unión Europea (TJUE)*<sup>72</sup>

*Susanna Brunnhofer c. Bank der österreichischen Postsparkasse AG.* Asunto C-381/99. 26 de junio de 2001. Sala sexta.

### *Corte Suprema de los Estados Unidos*<sup>73</sup>

*Swierkiewicz v. Sorema N. A.* 534 U.S. 506. 26 de febrero de 2002.

*Reeves v. Sanderson Plumbing Products.* 530 U.S. 133. 12 de junio del 2000.

*St. Mary's Honor Center v. Hicks.* 509 U.S. 502. 25 de junio de 1993.

*United States Postal Service Board of Governors v. Aikens.* 460 U.S. 711. 4 de abril de 1983.

*Texas Dept. of Commun. Affairs v. Burdine.* 450 U.S. 248. 4 de marzo de 1981.

*Board of Trustees, Keene State Coll. v. Sweeney.* 439 U.S. 24. 12 de noviembre de 1978.

*Furnco Constr. Corp. v. Waters.* 438 U.S. 567. 29 de junio de 1978.

*Teamsters v. United States.* 431 U.S. 324. 31 de mayo de 1977.

*Mc Donnell Douglas Corp. v. Green.* 411 U.S. 792. 14 de mayo de 1973.

*Griggs v. Duke Power Co.* 401 U.S. 424. 8 de marzo de 1971.

*Plessy v. Ferguson.* 163 U.S. 537. 18 de mayo de 1896.

<sup>72</sup> Se pueden consultar las sentencias en: [curia.europa.eu/juris/recherche.jsf?language=es](http://curia.europa.eu/juris/recherche.jsf?language=es).

<sup>73</sup> Se pueden consultar las sentencias en: [supreme.justia.com](http://supreme.justia.com).

Qf



Marcial  
Pons



Càtedra  
de Cultura  
Jurídica



Universitat  
de Girona



Diputació de Girona

[www.quaestiofacti.com](http://www.quaestiofacti.com)