

THE SPECIFIC EVIDENCE RULE. REFERENCE CLASSES — INDIVIDUALS — PERSONAL AUTONOMY

Kyriakos N. Kotsoglou

Associate Professor in Law

Northumbria University Newcastle upon Tyne, U.K.

kyriakos.kotsoglou@northumbria.ac.uk

ABSTRACT: This paper grapples with the issue of naked statistical evidence in general and the reference class problem (RCP) in particular. By analysing the reasoning patterns underlying the RCP, I will show, first, that the RCP rests on theoretical presuppositions which we are by no means bound to accept. Such a presupposition is, what I will call, the wholesale approach in decision-making. Secondly, I will show that the very effort to increase the level of precision to a maximum so that a reference class contains a single member only is theoretically inconsistent insofar, as it deprives reference classes of their general (and thus scientific) character. Thereupon, I will argue, thirdly, that the decision to enact a specific evidence rule is a political one and reflects deep moral and jurisprudential values, not scientific propositions. Such a value is personal autonomy, which I go on to illuminate briefly. Whether the trier of fact will treat cases in a wholesale approach or not depends on constitutional arrangements and legal values putting emphasis on the individual and the latter's dignity.

KEYWORDS: reference class problem, individualisation, specific evidence, discretion, personal autonomy, statistical inferences.

SUMMARY: 1. INTRODUCTION: 1.1. Uncertainty about the Value of Probabilistic Evidence. 1.2. Is There a «Specific Evidence Rule»?— 2. THE REFERENCE CLASS PROBLEM: 2.1. The RCP as a Paradox. 2.2. Formal Logic. 2.3. Reference Classes and Individuals. 2.4. Fallibilism.— 3. THE VALUES OF LAW: 3.1. The Values of Criminal Law. 3.2. Personal Autonomy as a Legal Value. 3.3. Discretion. 3.4. *Specific Evidence* and *Direct Evidence*.— 4. OUR CRAVING FOR GENERALITY.— 5. CONCLUSIONS.— BIBLIOGRAPHY.

«This requirement that evidence should focus on the defendant must be taken to be a rule of law relating to proof distinct from the general rule governing the quantum of proof.»

Glanville Williams

1. INTRODUCTION

1.1. Uncertainty about the Value of Probabilistic Evidence

There has been a long discussion on the aptness and usefulness of formal methods in general and numerical methods in particular in criminal adjudication. At its core, the discussion pivots around the requirements, quantitative or qualitative in nature, for sufficient proof of guilt. To what extent does (accurate) statistical evidence yield a specific inference to the individual (defendant) and warrant a criminal verdict?

A series of recent cases concerning the use of the most prominent member of statistical evidence, DNA profiles, exemplify the tension around the evidential rules regarding the sufficiency of a sole item of evidence. Can DNA evidence provide a safe basis for a criminal conviction? For example in England and Wales, the Court of Appeal seems to be oscillating between its original position¹, according to which DNA as a sole item of evidence provides an *insufficient* basis for conviction, and the latter's diametric opposite proposition, according to which «there is no evidential or legal principle which prevents a case solely dependent on the presence of the defendant's DNA profile on an article left at the scene of the crime being considered by a jury»².

Legal uncertainty remains thus as regards the evidential value of DNA profiles and statistically analysed evidence more generally. Can a probabilistic piece of evidence like the grenade firing pin in *Jones*, the scarf in *Ogden* or the balaclava in *Grant*, even when taken at their *Galbraith* highest, provide sufficient evidential support to the *probandum*? Can statistical evidence warrant an inference to the specific individual? It is not clear what the law on that matter is both in England and Wales and in other jurisdictions nor what the solution to the problem should be.

¹ See *R v Lashley*—unreported; CPS Policy Directorate, Guidance on DNA Charging, 2004, implemented in *R v Grant* [2008] EWCA Crim 1890; *R v Ogden* [2013] EWCA Crim 1294; *R v Bryon* [2015] EWCA Crim 997).

² *R v Tsekiri* [2017] EWCA Crim 40 at 21). For the *Tsekiri*-type of cases probative sufficiency of evidence is thus to be determined in relation to a non-exhaustive list of surrounding facts of the case. The open texture of the abovementioned list makes it difficult to determine whether a certain combination of elements should constitute a case to answer and provide a safe basis for conviction, especially in view of the *Tsekiri*-test according to which «each case will depend on its own facts». Recently, however, the same Court of Appeal has signified—although not in an entirely clear way—that we cannot single out the defendant/appellant as the source of the DNA to the exclusion of all others when we lack individualistic evidence (see *R v Jones (William Francis)* [2020] EWCA Crim 1021 (03 Aug 2020). See also Kotsoglou and McCartney (2021, p. 135-140).

1.2. Is there a «Specific Evidence Rule»?

One should not think that the problem of applying naked statistical evidence to the individual case is some idiosyncratic, theoretically cryptic or rather uncommon feature of criminal adjudication. For example, the ENFSI³ in its recent roadmap understands biometrics as a technique which «allows a person to be *individualised* and authenticated, based on a set of recognisable and verifiable data, which are very distinctive»⁴. More fundamentally, as the abovementioned document made clear, «pattern recognition of features of comparison for individualisation and source attributions»—what is widely known as S.A.D. (Source Attribution Determination)—is still to be counted among the «fundamentals in forensic science»⁵. This policy document echoes thus the forensic science's credo, *i.e.*, individualisation—Kirk (1963, p. 236) dubbed individualisation the «essence of forensic science»— and manifests the latter's ubiquitous character. As Paul Roberts (2007) remarked, the reference class problem (hereafter: RCP⁶) is despite its «mathematical connotations [...] pervasive in legal adjudication, and will have been encountered in some form or another by every legal practitioner and scholar of legal procedure» (p. 243).

From Justice Antonin Scalia, who noted that statistical evidence «is worlds away from “significant proof”»⁷, over the German Federal Labour Court, which held that statistical data (in that case: a Monte-Carlo Simulation) is not conclusive for the individual case⁸, to the U.S. Court of Appeals (Second Circuit), which regarded the application of naked statistical evidence to the individual as «surmise» and made clear that the latter «will not of course substitute for specific proof»⁹, several higher courts in Western jurisdictions have continuously and consistently quashed decisions which were based entirely on naked statistical evidence. We seem to be able to detect the outlines of a hitherto not clearly articulated «specific evidence rule».

At the same time, our main question remains unanswered: What *is* the evidential weight of statistical evidence? And how do we resolve the conflict between higher courts (at least in the jurisdictions identified above) requiring «specific evidence», on

³ The European Network of Forensic Science Institutes [ENFSI] comprises more than seventy forensic Institutes from European countries (including the U.K.), whose overarching goal is to «ensure that the quality, development and delivery of forensic science throughout Europe is at the forefront of the world». See ENFSI, Vision of the European Forensic Science Area 2030.

⁴ *Ibid.*, § 1.1. (emphasis added).

⁵ *Ibid.*, § 1.3.

⁶ As I will show the RCP is a theoretical account of the practice of individualisation.

⁷ *Wal-Mart Stores Inc. v. Dukes et al.*, 564 U.S. 338 (2011), Opinion (Scalia), at 14.

⁸ Although there was no mention of the reference class problem *eo nomine* in the decision, the Federal Court raised once again questions of sufficiency of proof by making clear that proof of unlawful behaviour hinges on «statistical data being conclusive for the employer in question», to wit: on specific evidence. Federal Labour Court [2009] — 8 AZR 1012/08, § 68.

⁹ *United States v Shonubi*, 998 F 2d 84 (2d Cir. 1993) [*Shonubi II*], at 16.

the one hand, and the practice of regarding the individualisation of naked statistical evidence as part and parcel of free assessment of evidence, on the other? Surely, it would be a legalistic fallacy to assume that the «specific evidence rule» *is* conceptually and evidentially sound in England let alone elsewhere simply because for example the Court of Appeal (E+W) stresses that probabilistic statements warrant no conclusion «in relation to the individual case»¹⁰. As Judith J. Thomson (1986) put it concisely, «[f]riends of the idea that individualized evidence is required for conviction have not really made it clear why this should be thought true» (p. 206). The *Shonubi* case¹¹ neatly encapsulates the abovementioned tension. In *Shonubi V* the District Judge made clear that he was not ceding the main point in the argument about the evidential value of statistical evidence; he was merely deferring to the higher court's authority. «The specific evidence requirement of *Shonubi* I[II] and IV», he lamented, «is a denigration of the modern evidentiary principles of free admissibility and free evaluation of probative force by the trier»¹², accusing at the same time the U.S. Court of Appeals (Second Circuit) that it «distorts the Federal Rules of Evidence» and that it required a result which is «compassionate», alas lacked legal basis. There is no such basis, Judge Weinstein added, for the specific evidence rule¹³.

This brings us to our main issue, the meaning of «specific evidence», and the contested validity of the RCP. *Should* statistical data—accurate as it can ever be—motivate action in general and warrant a legal decision in particular? The question at its kernel is whether an epistemic inference *from* a relevant population serving as a basis for calculating and assigning probabilities *to* an individual can ever be valid if the only evidence that we have is information about the reference class in question. Since we deal with the problem of factual generalisations and individualisation, we—rather unwillingly—have to raise fundamental questions about the nature of our reasoning processes. Unsurprisingly, these issues have spawned an extensive debate¹⁴—for very good reasons, since legal adjudication aspires to be rational. However, there is no consensus on what lessons to draw. The discussion between the opposing parties has stalled. It would not be exaggerating to say that we have reached the point «where one would like just to emit an inarticulate sound» (Wittgenstein, 2009, § 261).

This paper will provide a *theoretical diagnosis* of the RCP. I will show that the question of embracing and deploying formalised reasoning patterns as proxy for de-

¹⁰ *R v Jones (William Francis)* [2020] EWCA Crim 1021, at 31.

¹¹ In *US v Shonubi*, the prosecution relied on statistics to prove the amount of drugs Shonubi had smuggled into the USA. However, the appellate court quashed the sentence twice because it was not based on «specific evidence». See also *United States v Shonubi*, 802 F Supp 859 (EDNY 1992) [*Shonubi I*]. *United States v Shonubi*, 998 F 2d 84 (2d Cir. 1993) [*Shonubi II*]. *United States v Shonubi*, 895 F Supp 460 (EDNY 1995) [*Shonubi III*]; *United States v Shonubi*, 103 F 3d 1085 (2d Cir. 1997) [*Shonubi IV*]; *United States v Shonubi*, 962 F Supp 370 (EDNY 1997) [*Shonubi V*].

¹² *United States v Shonubi*, 962 F Supp 370 (EDNY 1997) [*Shonubi V*], at 375.

¹³ *Ibid.*, at 376.

¹⁴ See the special issue edited by Allen and Roberts, (2007), for more discussion and further references. See also Colyvan *et. Al.* (2001, p. 168-181). Tillers (2005, p. 33-49).

cision-making cannot be addressed let alone answered in a normative vacuum, *i.e.*, independently of the procedural architecture of legal systems examining and validating, say, criminal charges. In other terms, the reference class problem is not an analytic one. By examining the reasoning patterns underlying the RCP, I will show, *first*, that the RCP rests on theoretical presuppositions which we are by no means bound to accept in Western, anthropocentric legal orders. Such a presupposition is the wholesale approach in decision-making (part 2). *Secondly*, I will show that the very idea of a group-to-individual inference is anything but inevitable or legitimate; the idea of a reference class containing a single member only is theoretically inconsistent insofar, as it deprives reference classes of their general (and thus valid) character (part 3). Thereupon, I will argue, *thirdly*, that the decision to enact a specific evidence rule is a political one and reflects deep moral and jurisprudential values, not scientific propositions. Such a value is human dignity and moral autonomy, which I go on to illuminate briefly. Whether the trier of fact will treat cases in a wholesale approach or not depends on constitutional arrangements and legal values, which in the case of England and Wales and other similar legal orders put emphasis on the individual and the latter's dignity (part 4). This theoretical diagnosis will show how the RCP dissolves once we look at it from the right angle.

2. THE REFERENCE CLASS PROBLEM

2.1. The RCP as a Paradox

The RCP boils down to the question of whether one is inferentially justified in drawing a group-to-individual inference, *if* all that we know is the latter's membership to the former. The question is thus whether we can apply naked *statistical* evidence to the *individual case* qua unique historical event.

As shown above, the RCP is—despite its philosophical provenance—ubiquitous in adjudicative contexts, indeed it surfaced in litigation early in the twentieth century¹⁵. However, it was not until a ground-breaking monograph by the philosopher J. L. Cohen (1977) that the *RCP eo nomine a)* could be articulated, and *b)* the paradoxical results of applying axiomatised inference patterns especially the axioms of the theory of mathematical probability in litigation, were fleshed out. Cohen's analysis sparked an *academic* interest in the foundations of evidence and proof in adjudication and in the RCP in particular (p. 74-81). To investigate the claim whether the adjudicative process indeed any vernacular decision-making context could ever be axiomatised, Cohen puts us in the setting of a rodeo and informs us that according to fully reliable information, among the 1,000 spectators only 499 paid for admission. At the same time, we learn that no tickets were issued and there can be no (reliable)

¹⁵ See, *e. g.*, the American case *Smith v. Rapid Transit, Inc.* 317 Mass. 469, 58 N.E.2d 754 (1945).

testimony as to whether the person in question—call him S_1 —paid for admission or whether he had climbed over the fence. Therefore, we should decide based on a mathematical probability alone (0.501) whether we have sufficient evidence that S_1 paid for admission and is thus to be found guilty. The abovementioned probability was not chosen randomly, for according to the rather mainstream view in Anglo-American Law of Evidence¹⁶ the standard of proof (hereafter: SoP) in civil adjudication can be expressed as a .501 probability¹⁷.

The logical outcome of that thought process is that, based exclusively on the information that S_1 is one of the randomly chosen spectators, the rodeo organisers would be entitled to recover compensatory damages for the admission money. The reasoning pattern described above is plausible insofar, as it captures the essence of the mainstream view in theory of evidence. Merely by allowing the trier of fact to draw an inference from naked statistics to the individual case (S_1), *i.e.*, to individualise the statistical proposition, we admit implicitly that the same can be done with regards to every other person among the 1000 spectators: S_2 , S_3 , S_4 and so on. However, by juxtaposing all instantiations of the abovementioned reasoning pattern, our set of conclusions about the spectators

$$\{S_1, S_2, S_3, \dots S_{n-1}, S_n, n=1000\}$$

becomes contradictory. Our reasoning pattern warrants a set of conclusions, according to which the rodeo organisers can justifiably raise a claim against all 1.000 spectators, although only 499 among them had paid for admission. Given the fact that scientific models are meant to represent target phenomena and that the only justification of such a model of proof is «solely and precisely that it is expected to work» (Von Neumann, 1955), the reasoning pattern deployed in the rodeo example simply *does not work*. The model is unbearably *over-inclusive*. As Cohen (1977, p.75) remarked, it is «manifestly unjust» that a randomly chosen individual should lose his case although there is a .499 probability that he had paid for admission, which is barely better than tossing a coin.

Legal adjudication aspires to be rational and, most importantly, free of paradoxes¹⁸. But getting a handle on what exactly the RCP is and what to do about it, proves to be a difficult task. This article proceeds from the conviction that the RCP will not give away, if we simply confront it *head on*. To gain clarity, we need to focus on the *logical structure* of the RCP by looking at the facts of a case similar to *Shonubi* or the Gatecrasher Paradox.

¹⁶ I use capital letters to distinguish the academic understanding of the subject matter from the subject matter itself, *i.e.*, the law of evidence as criminal/civil courts have shaped it.

¹⁷ For a critical introduction see Redmayne, 1999, p. 167-195.

¹⁸ The Rationalist Tradition was first described in Twining (2009, p. 211-249).

2.2. Formal Logic?

For reasons of simplicity, we can start our investigation by looking at a *deductive* version of the RCP. Let us assume that a company C operates under a general policy of (blatant) discrimination based on sex so that *all* female employees are not to be promoted just because they are female¹⁹. Following the reasoning pattern of modus ponens [hereafter: MP], we can draw the inference that a randomly chosen female employee in that company, F_1 is subject to discrimination. In more detail:

- | | |
|---|-----|
| Every female employee in company C is subject to discrimination | [1] |
| F_1 is a female employee in company C | [2] |
| —MP, [1] and [2]— | |
| F_1 is subject to discrimination | [3] |

Knowingly, formal logic provides us with reasoning patterns that eliminate discretion insofar, as they unpack the content of general propositions and yield necessarily true conclusions. In a deductively valid argument, the premises necessitate the truth-value of the conclusion insofar, as accepting the premises means that we have—on the pain of irrationality—to accept the conclusion too. The evidence of discrimination against S_1 is a function of a formal calculus (MP) since the truth-grounds of the conclusion are already contained in those of the premises: the conclusion ([3]) necessarily follows from the set of premises ([1, 2])²⁰. Most importantly, the soundness of the conclusion is a consequence of formal properties of the logical machinery underpinning the syllogism, not the result of the fact-finder's justification.

At the same time, the syllogism depicted above captures the architecture of formal reasoning patterns, for *it eliminates the very need for decision-making*. The problem is, of course, that deductive reasoning patterns are nothing but a caricature of complex (procedural) reality (Toulmin, 1976, p. 149). As David Schum (1994) remarks, «the price paid for necessity is vacuity, since the content of the conclusion of a deductive argument is already present in the premises» (p. 454). In other words, we had actually known in advance that F_1 had been subject to discrimination²¹.

Of practical relevance for adjudicative contexts are *inductive* patterns of reasoning. Frequentists, and oftentimes forensic scientists and courts, use the probabilistic counterpart of modus ponens (Prob-MP) to draw a categorical conclusion and indi-

¹⁹ This case-scenario is not fictional. The reasoning pattern discussed here was the *ratio decidendi* in a ruling of the Labour Court in Berlin/Germany. See Kotsoglou, 2017 for more discussion.

²⁰ See Wittgenstein (1958, proposition 5.121).

²¹ Proof in formal logic, Wittgenstein (1958, p. proposition 6.1262), stressed, is «merely a mechanical expedient to facilitate the recognition of tautologies in complicated cases». Algorithms using computational power take this simply to the extreme. After all, even the most labyrinthine logical argument is just a tautology in disguise, so that the reasoning process and conclusion are equivalent.

vidualise statistical data²². Consider now the following modification of the previous example; according to the new case-scenario, only 96% of a specified population group²³, *i.e.*, of the female employees in company C, are subject to discrimination. According to a rather mainstream approach among frequentists, the abovementioned base-rates do confer a probability on the individual case. In more detail:

96% of female employees in company C is subject to discrimination [4]

F_1 is a female employee in company C [5]

—Prob-MP, [10] and [11]—

There is a 96% probability that F_1 is subject to discrimination [6]

Proposition [6] is a valid conclusion but misses the target. The *probandum* cannot be deduced from the premises; all that can be known is that the *probandum* is more or less likely to have occurred. Instead of deductive entailment, we have inferential support. However, our primary goal is to find out whether F_1 has been subject to discrimination, not to calculate any probability. *If*, therefore, we assign a probability to a single case, then we can use a threshold as a cut-off point to draw a conclusion:

There is a 96% probability that F_1 is subject to discrimination [7]

The SoP lies at 95% [8]

—96% > 95% —

F_1 is subject to discrimination [9]

Relatedly, we can think of the diametric opposite of the previous example, in which, all else being equal, 94% of all female employees in company C are subject to discrimination.

There is a 94% probability that F_1 is subject to discrimination [10]

The SoP lies at 95% [11]

—94% < 95%—

F_1 is not subject to discrimination [12]

Let us take stock. *Firstly*, it is striking that simply by drawing a group-to-individual inference, we eradicate any differences between deductive and inductive methods of reasoning²⁴. For in both inductive examples outlined above the *decision* is only ostensibly one. Similar to the deductive version, the reasoning pattern necessitates, *i.e.*, generates automatically the respective conclusion; see propositions [9] and [12]. What is more, we do not really *decide* that the specific individual in question, F_1 , has or has not been subject to discrimination. The respective conclusion represents a function of the mere comparison between the probability that *any* member of the

²² For more discussion on this, see Sober (2002, p. 65-80).

²³ By that way, the usual problem of fluctuation of the statistical probability depending on the reference class vanishes.

²⁴ Anderson, Schum and Twining (2005, p. 100), have already pointed out that the «inductive form of an inference can be converted to a quasi-deductive form by identifying an articulating the generalization upon which it depends». See also Toulmin (1976, p. 112-13).

reference class was subject to discrimination and the respective (numerical) decisional threshold. (Legal) proof is thus reduced to an empirical question concerning a *class of events*. *Secondly*, we see how the proof paradox articulated by Cohen emerges. The inference in relation to a certain individual, F_1 (see propositions [9] and [12]), can be drawn using the same data set for *every* other individual belonging to the respective reference class:

$$\{F_2, F_3, \dots, F_{n-1}, F_n\}.$$

Note that this is contrary to our information that only 94/96% of the female employees in company C respectively, are subject to discrimination. As I will show further below, these two features identified above, *i.e.*, the eradication of the need to make decisions and the proof paradox itself, are closely related (see part 3).

Ultimately, the RCP boils down to the realisation that within the space of formalised reasoning patterns there is no warranted way of singling out the *individual case*. In other words, we cannot examine whether, *e. g.*, Alice Green (F_1) or Sarah Murray (F_2) *qua* unique individuals have been subject to discrimination or not. Mathematical models of proof allow us to look at an individual only through the lens of the latter's membership to a reference class. These patterns reduce unique historical events or individuals to empirically or otherwise informed statistical traits.

The spectator in the Gate-Crasher example or the female employee in our example above, are thus regarded as nothing but the embodiment of a set of statistical traits, which forces us to willy-nilly treat these individuals *in a wholesale manner*. The fact-finder has no normative toolkit to separate, as it were, the wheat from the chaff. They can attribute a property φ , *e. g.*, being subject to discrimination, to a reference class only. On the flipside, they cannot determine whether or how the said property relates to each of the specific members of the reference class. There is a hint there to be taken. But what would that be? I shall come back to that later.

For now, it is important to note that the *rigidity* of any formalised approach generating the RCP is striking. The mathematical framework described above captures a *class* of female employees, rather than being tailored to a particular individual (F_1) including her actions, omissions, or misfortunes. To be clear: due to its syntactic structure, the inferential model described above cannot even envisage individual cases to adjudicate them separately. The very term *individual* is just not predicated. Humans are viewed as the totality of attributed statistical traits.

Let me also stress that the core idea underlying the deployment of mathematical models of proof is the claim that axiomatised systems can yield conclusive answers as regards individual cases. The idea of reasoning patterns which take some data as input and return some other data as output is neither new nor is it free of theoretical commitments with regards to the decisional domain. Actually, the view that (scientific) knowledge is a product of an *empirically uninterpreted formal calculus* based on for-

mal logic lies at the heart of a now obsolete²⁵ paradigm in philosophy of science, *i.e.*, the hypothetico-deductive method (syntactic view of theories)²⁶. The RCP is thus not free of problematic theoretical presuppositions. On the contrary, it is intertwined historically and conceptually with the idea that human knowledge as such can be axiomatised. By deploying formalised reasoning patterns (of the type underpinning the RCP) in decision-making, we commit ourselves to so much more than a trite observation. The set of logical principles and theoretical commitments described above is a highly sophisticated and contentious conception of theoretical reasoning commonly known as the *syntactic view* of theories—an idea which, as we saw, is necessary to understand the theoretical underpinnings of the RCP and formalised reasoning patterns in general.

2.3. Reference Classes and Individuals

Our main discussion so far revolved around the first part of the «specific evidence rule», *i.e.*, the question of whether compiled statistical data can ever yield a conclusion qualifying as *specific* evidence. In other words, the question is whether we can attribute a trait ϕ to an individual member of a reference class, if all that we know about this person is his or her membership to the said class. This is in effect the problem of the *single case probability*, *i.e.*, the necessary condition for the very possibility of applying naked statistical evidence to individuals. The intelligibility of the single case probability is all the rage not only in legal evidence scholarship but in the theory of probability too. Frequentists—this includes, notably, Hans Reichenbach who was one of the first to use the term «reference class»²⁷—have for a long time tried to

²⁵ The idea of an empirically uninterpreted deductive/inductive system, *i.e.*, the so called Received View in philosophy of science, as well as its central component, the Deductive-Nomological/Inductive-Statistical (D-N/I-S model), have been widely criticised from the very beginning. In the words of A. J. Ayer: «nearly all of it was false» (A. J. Ayer who played a major role in introducing logical positivism to England, made this comment in a televised interview with Professor Bryan Magee). Briefly, Carl Hempel, one of the main protagonists of logical empiricism and father of the D-N model, from which the RCP derives, sang very publicly the model's swan song. During the opening talk at a symposium on the structure of scientific theories, *i.e.*, a subject dominated by the Received View, Hempel conceded that the research program was fundamentally wrong, for the idea of syntactic axiomatisation could not overcome criticism coming wave after wave (see Suppe, 2000).

²⁶ Logical empiricism pivoted around the deductive-nomological model (D-N/I-S model), according to which explanations have the structure of sound and with necessity derivable logical arguments in which a general law of nature occurs as an essential premise (Hempel and Oppenheim, 1948). The basic idea of the D-N model is that one deduces the *explanandum*, which describes the phenomenon to be explained, from an *explanans*, consisting of one or more laws, typically supplemented by true sentences about initial conditions. The model is intended to apply to both the explanation of «general» regularities by other laws and the explanation of particular events, although subsequent developments have largely focused on the latter. The structural similarity with the RCP should already be obvious.

²⁷ Reichenbach (1971, p. 376), regarded the idea of the single case probability insofar as meaningless, as it represents «an elliptical mode of speech». Furthermore, a basic component of Richard

tackle this problem head on. If we want to make sense of the utterance «single case probability», as probability theorists suggest, then we should translate that probability «into a statement about a frequency in a sequence of repeated occurrences» (Reichenbach, 1971, p. 376-377).

Consider the example of a man called James Smith wondering whether he will survive his sixty fifth birthday. It would only make sense to consider the fact that James Smith is, say, Liverpudlian and not just an English man, if the probability to survive his sixty fifth birthday is bigger (or smaller) given that he comes from Liverpool and not just from England. Similarly, it would only make sense to mention that James Smith comes from a leafy area of Liverpool rather than a deprived one, if and only if that would give us additional information about his life expectancy. By compiling *more and more* data, the reference class is becoming *narrower and narrower* until at some point eventually we will be able, frequentists claim, to validate the *narrowest* class possible (Reichenbach, 1971, p. 374). Instead of differentiating *between* the members of a wide reference class, frequentists suggest we partitioned the reference class itself into increasingly smaller sets. For the *narrowest* class we would need a set of predicates

$$\{P^1, P^2, P^3 \dots P^m\}$$

which can sufficiently and exhaustively describe the *narrowest* class possible. In other words, the narrowest reference class could be regarded in legal terms as a «commonality»²⁸ which—if relevant—would allow a reference-class-wide resolution «in one stroke»²⁹. The narrowest-class-method would have thus the capacity to revolutionise the efficiency of the system of legal adjudication especially in times of cuts and austerity. The same method could perhaps solve the RCP. I have my doubts.

We saw above that the narrowest reference *class* is (almost by definition) satisfied by at least more than only one individual, thing, or event, *e. g.*, James Smith *and* David Black. Therefore, the reference class is obviously not the *narrowest* possible. For there can be at least one further predicate P_n ($n > m$) which applies to James Smith but not to David Black, so that the set of predicates

$$\{P^1, P^2, P^3 \dots P^m\}$$

is a proper *subset* of the new (extended) set

$$\{P^1, P^2, P^3 \dots P^m, P^n\}^{30}.$$

It becomes thus apparent that in order to increase the level of precision and the grade of «single-case» probability, *i.e.*, in order to reach down to the level of the indi-

von Mises' frequentist theory of probability which was later also adopted by his brother, Ludwig, is the constraint that numerical probabilities can only be calculated for series of events which he called «collectives» (Mises, R. von, 1964, p. 12). See also Nagel (1939).

²⁸ See, *e. g.*, Rule 23(a)2 Federal Rules of Civil Procedure (USA).

²⁹ *Wal-Mart Stores, Inc. v. Dukes U.S.*, 564 U.S. 338 (2011), at 9 (per Justice Scalia).

³⁰ Fetzer (1977, p. 214).

vidual, we would have to keep enlarging our set of predicates until we inevitably end up with an aggregate of features which are tailored to the unique features of a certain individual, thing or event. We would need thus to fractionate our reference class by taking every single detail into account, until we get sufficiently partitioned reference classes that reach the level of the individual. In other words, we would only then be allowed to speak about the narrowest class possible, when our set of predicates

$$\{P^1, P^2, P^3 \dots P^m, P^n\}$$

comprises descriptors designating distinct properties of a *unique* individual or historical event, to wit: when the narrowest reference class description is satisfied by no more than one member. However, this would be an oxymoron insofar, as we would have:

- a) reference classes that contain a single case, and
- b) as many reference classes as cases.

I cannot stress enough that propositions (a) and (b) are precisely what the mainstream approach in forensic science espouses. For example, individualisation has generally been defined as a «special case of identification, where the [...] class is populated by one object only»³¹. Reference classes with one single object are not only inefficient but also theoretically inconsistent. Statistical data which are valid only with regards to a single case contradict the very concept of scientific explanation which needs to be *general* in scope. Purporting to eliminate discretion by deploying reference classes which purportedly contain a single member does not eliminate subjectivity or uncertainty, but merely distracts us from the very problem of decision-making. As Allen and Pardo (2006) remark «the only class that would accurately capture the “objective” value [of probability] would be the event itself, which would have a probability of one or zero respectively» (p. 114). If every individual case can be described as the solitary member of some—let us call it—super-narrow class, then frequentism loses its thrust, indeed its explanatory power. A reference class whose granularity reaches the maximum level so that it contains a single observation can only be perceived as a post-modernist «joke» (Sober, 2008, p. 90)³².

In all, frequentism and the idea of eliminating the need for decision-making under uncertainty—this aspiration lies at the heart of the RCP—end up being stripped of their distinctive and nominal feature, *i.e.*, their ability to describe things in the abstract, which is a contradiction in terms. As W.C. Salmon (1974) put it pithily, «God would be unable to construct an inductive-statistical explanation of any physical event [...] not as a limitation of His power but as a reflection of His omniscience» (p. 165). Why? As outlined above, God would be able to detect the tiniest difference between seemingly similar cases and classify them separately. Proponents of formalised reasoning patterns who would not perceive the RCP as a real *problem*,

³¹ For a critical introduction, see Champod (2009, p. 1508-1511).

³² See also Fetzer (1977, p. 200).

must alas jettison generality to deal with the individual case, which is fatal for any scientific account of the world indeed for any frequentist *theory* of probability. There are identifiable limits to our capacity to apply—in a warranted way—general propositions to individual circumstances, things or events. The complexity of historical events outstrips anything that could be validated by general rules³³.

2.4. Fallibilism?

We saw above that formalised systems of proof *by design* treat reference classes in a rigid way, and as a result fail to capture the *heterogeneity* of the latter. For example, formal systems cannot differentiate between employees who have been subjected to discrimination and those who were spared the agony. This is, one could say, antithetical to the overriding objective of (criminal) adjudication, *i.e.*, justice which includes factual accuracy. Proponents of formalised decision-making processes in legal adjudication, however, disagree even with the way in which the problem is framed. Speaking of «failure», they contend, is merely a way to beg the question. Kaye (1979) makes this point explicit when he stresses that outcomes which have traditionally been called «paradoxical» are «perfectly appropriate in the uncertain and imperfect world of litigation» (p. 38). It is possible, Schauer (2003) remarks, that reliance in generalisations «known from the beginning to be imperfect», *e. g.* non-spurious naked statistical evidence «might still be empirically superior to relying on allegedly direct or individualized assessments» (p. 98). Admittedly, one has to provide an explanation of what precisely is—from an evidential point of view—the benefit of case-specific evidence as dictated by the specific evidence rule. After all, it is a run-of-the-mill claim that «all evidence is probabilistic»³⁴.

Individualised assessments too, proponents of this approach claim, rely ultimately on empirical rules that are statistical in nature. The difference is, so the same line of defence, that proponents of individualistic methods fail to articulate the exact magnitude of the accuracy for their methods, *i.e.*, the «power» of their respective test³⁵. This is an important point and merits clarification. Logical systems of proof generate erroneous outcomes whose extent we can express in a comprehensive quantitative manner. The difference between quantitative and qualitative approaches, one could say, is that the former are equally fallible albeit they are at least transparent about their rate of errors: no matter which path we will follow, we will still end up

³³ This insight significantly weakens the claim that group-to-individual inferences are «inevitable» and «sometimes epistemically legitimate» see Tillers (2005, p. 33). At the same time, Tiller's proposition (at 39) that «it is *not* always inferentially illegitimate to base inferences about the behaviour of an individual on the behaviour of other people» is entirely reasonable, *if* by that is meant that group-to-individual inferences rely partially on statistical data and are not reduced to the same data.

³⁴ This is a point that even Tribe concedes Tribe (1971, p. 1330).

³⁵ See, *e. g.*, Kaye (2009, p. 1177).

with Blackstone-ratios. It is simply, so the same line of possible defence, a matter of the respective legal order and its underlying political and moral values to find the appropriate trade-off between competing values, which is structurally the exact same situation with legal systems employing individualistic evidence. After all, the choice of the SoP is a way to fix the ratio of erroneous outcomes, *i.e.*, a way to adjust those Blackstone-ratios³⁶. The crucial *difference*, proponents of formalised reasoning patterns add, is that opting for individualistic methods leaves us in the dark about the extent of erroneous outcomes. In that sense, paradoxes of proof, one could say, are a blessing in disguise. Why *should* we, one might wonder, refrain from attributing non-spurious class characteristics to members of a reference class?

The question at its core is whether criminal justice systems and evidence law regimes can rely on *mathematical* answers to problems encountered in adjudication, which pertains to the very nature of (criminal) jurisprudence. To answer that question, we need to dig deeper and investigate the structure of (legal) decision theory as well as the notion of discretion.

3. THE VALUES OF LAW

3.1. The Values of Criminal Law

The previous section provided us with important insights into the deep problem of singling out an individual from a reference class using exclusively axiomatized methods. The abovementioned inability is necessitated by *a)* the *wholesale approach* of formalised reasoning patterns and *b)* the fact that reference classes become inconsistent, indeed collapse as soon as our focus shifts from the reference class to the individual. In fact, the very notion of *individual qua* separate unit of analysis belies the frequentist/general character of statistical models. As the science historian James Gleick (2011) put it: «Probability is about ensembles, not individual events. Probability theory treats events statistically» (p. 329). In other words: probability theory does not offer us a warranted way out of the dilemma α or $\neg\alpha$, for any statistical proposition can produce either event. Nor can probability theory answer questions in the form «how likely is a particular event to happen?»: historical events happened, because they happened (one should not conflate the logical triviality of this tautological sentence with its methodological value). As David Lucy (2013) put it with characteristic verve: The idea of «a frequency being attached to an outcome for a single event is ridiculous» (p. 3-4).

Nevertheless, the normative question remains unanswered: *If* naked statistical data *are* available, then *should* we apply them to the individual? *Should* we attach evidential weight to, say, a very high probability so that we can equate the latter with sufficient

³⁶ *In re Winship* 397 U.S. 358 (1970).

proof and «specific evidence»? If not, as proscribed by the specific evidence rule, by what authority? These questions cannot be answered independently or regardless of the structure and internal *values* of a legal order. Causal relationships which empirical/natural sciences investigate, do not inform us sufficiently about *decisional* issues, *i.e.*, they cannot dictate what *we should do* let alone pre-empt the outcome of a legal process. Note that any choice of action requires the input of (personal, societal or otherwise) values, and will reflect *value judgments*. Notably, legal orders are nothing but sets of values enshrined in statutes, judicial rulings and legal principles. Let me now clarify what exactly I argue for. I do *not* submit that in a rationality-driven world, scientific facts should not *guide* our lives. Ordinarily, scientific facts are generated using a methodologically controlled and thus reliable process. What I *am* saying is that statistical propositions, being *general* in nature, cannot *on their own* fully determine our course of action in the context of the individual case. Statistical propositions aspire to provide a *general* account of a given target system under study³⁷.

On the flipside, the objective of fact-finders in liberal (as opposed to authoritarian) legal orders is not to provide any kind of *general* explanation of a domain under study, but to ascribe liability to an individual's actions or omissions by rendering a verdict. It is the *particular* defendant and the *particular* legal dispute which fact-finders have to resolve. In order to answer *practical questions* in relation to a single event, we need more than the available (statistical) data and more than our degree of belief in a particular hypothesis³⁸. Individual cases can be answered only with recourse to external values or values which are internal vis-à-vis a given legal order (Kuhn, 1996, p. 110). As Tillers (2005) reminded: «Valid inference[s] can be unfair» (p.46). Mathematical results are not self-applying in the way envisaged by logical positivism nor can statistical data automatically motivate action.

(Legal) values are not an appropriate object of inquiry for empirical sciences. Thinking otherwise would amount to conflating empirical issues with normative (moral, political or legal) ones. Any effort to sidestep the thorny question of values by focusing merely on «pure facts» is not only intellectually reckless but also suspicious. Such a move would advance certain ideology-laden values over others without even having to argue for them³⁹. A decision signifies a logical jump, a leap of faith, which resolves a practical issue arising from contested factual claims. This does not mean, of course, that fact-finders are given unbridled discretion in making factual determinations. A criminal verdict is a *normatively* structured decision under uncertainty. Therefore, the central choice of embracing and deploying formalised reasoning patterns as proxy for decision-making cannot be addressed let alone answered in a *normative vacuum*, *i.e.*, independently of the procedural architecture of legal systems examining and validating, say, criminal charges.

³⁷ For more analysis, see Kotsoglou and Biedermann (2022).

³⁸ See also Sober (2008, p. 7).

³⁹ See Hughes (2012).

We can now refine the question asked above: *which* values are enshrined in legal orders? Obviously, unless someone is willing to defend some (weak or strong) version of Natural Law, this question cannot be answered in the abstract. Different legal orders will validate different values. There is, however, a core set of values of *general relevance* for Western legal orders—especially in the wake of the Second World War—putting emphasis on *personal autonomy, individual responsibility, and human dignity*.

3.2. Personal Autonomy as a Legal Value

From Aristotle who noted that «praise and blame attach only to *voluntary* actions» over Kant (1788) who defined autonomy as the capacity to decide for oneself to contemporary philosophers who indefatigably emphasise the fundamental value of autonomy in modern societies⁴⁰, personal autonomy has occupied centre stage in Western philosophy and polity. Arguably, personal autonomy and human dignity can and should be regarded as axiomatic (moral) truths⁴¹ which articulate humans as rational agents with the capacity to make, and act upon, judgments for which they are to be held responsible. As Aristotle expressed this axiomatic truth a few millennia ago: «To distinguish the voluntary and the involuntary is presumably necessary for those who are studying the nature of virtue, and useful also for legislators with a view to the assigning both of honours and of punishments»⁴².

Expectedly, a number of legal scholars have already highlighted that individuality and personal autonomy are central components of Western legal systems and that naked statistical evidence demean the notion of individualised justice⁴³. Personal autonomy and human dignity are the flipside and prerequisite of the very possibility of ascribing criminal liability in a *liberal* justice system. From the conditions of criminal liability in the context of omissions and strict liability over the requirements of consent in the context of sexual offences⁴⁴ to the very essence of discrimination widely regarded as a distinction «based on personal characteristics attributed to an individual solely on the basis of association with a group»⁴⁵, personal autonomy permeates the fabric of our legal orders. It is not irrelevant that Art. 1 of the Universal Declaration of Human rights prescribes that «all human beings are born free and equal in dignity and rights. They are endowed with reason and conscience». In other

⁴⁰ See Feinberg (1989) and Wasserman, (1992, p. 935).

⁴¹ See for example Dahl (1989, p. 100), who among other democratic theorists talks about the «presumption of autonomy». «To accept the idea of personal autonomy among adults, then, is to establish a *presumption* that in making individual or collective choices each adult ought to be treated—for purposes of making decisions—as the proper judge of his or her own interests.»

⁴² Aristotle, *Nicomachean Ethics*, Book III: Moral Virtue, Chapter I (emphasis added).

⁴³ For a detailed analysis, see Zuckerman (1986, p. 487) and Pundik (2008, p. 303-324).

⁴⁴ See the English case *R v Konzani (Feston)* [2005] EWCA Crim 706; see also Weait, 2005, p. 763-772.

⁴⁵ *Andrews v Law Society of British Columbia* [1989] 1 SCR 143 — NB: this is a Canadian case.

words, the same values endow human beings with agency so that the latter can be held responsible for *their* actions—not for the features of their reference class. The possessive adjective *their* does the conceptual heavy lifting at this juncture. In any modern society, a person should be held legally responsible for his or her actions only, not for those of his or her family, tribe, or reference class.

The main point here is structural, not moral, in nature. Reducing a verdict to someone's membership to a reference class is a matter of policy. In other words, societies could—if they choose to—reduce decisions to statistical data, but only if they opted for *collectivist* or other similar values. High-ranking officers of the Soviet regime concisely sum up this idea: «We are not fighting against single individuals», writes Martin Latsis who headed the Ukrainian secret communist police (Cheka),

[w]e are exterminating the bourgeoisie *as a class*. It is not necessary during the interrogation to look for evidence proving that the accused opposed the Soviets by word or action. The first question you should ask him is *what class does he belong to*, what is his origin, his education and his profession. These are the questions that will determine the fate of the accused. Such is the sense and essence of red terror⁴⁶.

What is more, it is not just a linguistic coincidence that the closely related word *fascism* derives from the Italian verb *fascio* which literally means «to bundle»⁴⁷. Western legal orders with their requirement for «specific evidence» choose *not* to bundle citizens but to treat them with dignity on a case-by-case basis so that each case relies on its own facts⁴⁸.

On a more practical level, the Grand Chamber of the Strasbourg Court has held that «the notion of personal autonomy is an important principle underlying the interpretation of the Convention guarantees». This notion needs thus to be seen «as an essential corollary of the individual's freedom of choice»⁴⁹. With personal autonomy and human dignity (note that these two concepts go hand in hand⁵⁰) being important values in Western *anthropocentric* legal orders, it becomes unpalatable to regard a human being merely as the embodiment of a set of statistical traits and, most importantly, to hold him or her responsible not for his or her own actions/omissions but for the actions of a group of people (reference class). Such a reductionist

⁴⁶ In the newspaper *Red Terror* (November 1, 1918), cited by Solzhenitsyn (2018, p. 21) (emphasis added).

⁴⁷ It can be seen as remotely related or as historical coincidence that von Mises, one of the fathers of frequentism, was an admirer of fascism: «It cannot be denied that fascism and similar movements aimed at the establishment of dictatorships are full of the best intentions and that their intervention has for the moment saved European civilization. The merit that fascism has thereby won for itself will on eternally in history» Mises, L von (1927:). See also Tillers (2005, p. 45).

⁴⁸ *R v Tsekiri* [2017] EWCA Crim 40, at 21.

⁴⁹ In that case in the context of Art 11. *Sørensen & Rasmussen v Denmark* (applications nos. 52562/99 and 52620/99, § 54. See also *Pretty v United Kingdom* 2346/02 [2002] ECHR 427 (art 8); Personal autonomy in the context of Art 8 ECHR, See *Bărbulescu v. Romania* [GC] Application no. 61496/08. See also *Vörður Ólafsson v. Iceland* 20161/06. Judgment 27.4.2010.

⁵⁰ See *Avram and Others v. Moldova* (Application no. 41588/05) 5 July 2011, § 36.

approach would violate the very anthropological and political tenets underpinning currently existing human rights regimes according to which the individual is a moral agent capable of making choices, and not a mere avatar of their sex, race, religion or reference class more generally.

The respective evidence law regime by implementing the abovementioned political and constitutional tenets cannot *a)* fall prey to any kind of *statistical essentialism* in which individuals exist merely as Venn diagrams where various statistical traits intersect and *b)* willingly ignore heterogeneity within the group. Obviously, no court judgment can or should capture the slightest detail. However, a deliberate and entirely avoidable lack of sensitivity to context or to the particular characteristics of the defendant is deeply troubling. As Nietzsche (1974) put it: «Seeing things as similar and making things the same is the sign of weak eyes» (§ 128). I submit therefore that the *specific evidence rule* as laid down by courts around the world is an instantiation of that deeper moral, political and legal principle. The specific evidence rule is an instantiation of personal autonomy.

My analysis so far raised fundamental questions about the extent to which mathematical models of proof infringe upon the autonomy of the individual, by that I mean both the defendant and the fact-finder. Mathematical models of proof result in ascribing liability to individuals on the basis of aggregate characteristics of reference classes. As a result, fact-finders would be bound to ascribe liability adopting a wholesale approach lacking a normative toolkit to separate the «wheat from the chaff». Mathematical models of proof are thus similar to the «rational system of proof» of the Romano-canon inquisition process which specified via clearly established evidentiary standards the quality and quantity of proof and where the trier of fact was «essentially an accountant who totalled the proof fractions» (Shapiro, 1991, p. 3-5). As we saw above any type of wholesale approach to the assessment of evidence would be antithetical to fundamental (constitutional) axioms of modern legal orders. Axiomatised reasoning patterns infringe upon the autonomy of the fact-finder.

The *specific evidence rule* therefore does not appear to be arbitrary, a «denigration of the modern evidentiary principles», or even an unjustified imposition of higher courts' power—as Judge Weinstein perceived it. On the contrary, it is an expression of legal (human dignity, personal autonomy) and moral values. Any type of one-size-fits-all approach to decision-making—whether in the context of policing, adjudication or probation—would raise serious questions about both the legitimacy and lawfulness of the respective decision or verdict⁵¹. Mathematical methods of proof as proxy for decision-making, Professor Tribe (1971) remarked in his seminal paper «Trial by Mathematics», «operate to distort—and in some instances, to destroy—important values which that society means to express or to pursue through the conduct of legal trials» (p. 1330). In Western legal orders, the individual and not the group reigns supreme whereas a wholesale approach to adjudication based on «head-count-

⁵¹ See also Roberts and Zuckerman (2010, p. 10).

ing» calculations would mean that an individual could be punished or treated in a certain way merely for belonging to some reference class, not for *his* or *her* actions or omissions. As Judge Newman put it:

The «specific evidence» we required to prove a relevant-conduct quantity of drugs for purposes of enhancing a sentence must be evidence that points specifically to a drug quantity for which the defendant is responsible. By mentioning «drug records» and «admissions» as examples of specific evidence, we thought it reasonably clear that we were referring to the defendant—*his* admissions and records of *his* drug transactions⁵².

In modern Western legal orders, the system of adjudication is by no means an empirical let alone a purely statistical enterprise. Human dignity, personal autonomy, reasonableness etc. are *normative* features, not (falsifiable) empirical claims. The same features give thrust to what the courts call the «dissimilarities approach»⁵³, which focuses on what *distinguishes* the members of any reference class; not on an alleged shared identity that *unites* them. As the U.S. American Judge Kozinski put it with regards to the perhaps most expansive class action in legal history where roughly one and a half million women alleged gender discrimination in pay and promotion policies and practices in Walmart stores:

[T]he half million members of the majority's approved class held a multitude of jobs, at different levels of [...] hierarchy, for variable lengths of time [...] with a kaleidoscope of supervisors (male and female), subject to a variety of regional policies that all differed [...]. Some thrived while others did poorly. They have little in common but their sex and this lawsuit⁵⁴.

What is more, the idea of reducing a human being to a few statistically articulated traits—an appealing idea for proponents of actuarial justice and forensic scientists—is not just antithetical to core legal principles in Western legal orders; it is in effect the attempt to strip complex historical events to the bone and reduce them to a set of formalised (mathematical) relations. Yet, in the same way that the complexity of physical phenomena outstrips any linear equation, the normative architecture of the individual especially the latter's legally protected dignity forbids any reductionist approach. Heisenberg's wistful dictum that «the equation knows best» (Gleick, 1994, p. 5) might be valid in the context of nuclear physics, but legal adjudication in Western legal orders is not *that* context. Stop-and-searching a citizen solely on the basis of an algorithmically generated similarity score, or even convicting a defendant in a criminal court because of a sufficiently high statistical (match) probability does not fail or succeed from the point of view of *logicality*. Insurance companies consider, and act upon non-individualised statistical scores—with unprecedented financial success. In view of the procedural architecture of Western legal orders, however, epistemic considerations need to be filtered, and validated through a network of constitutional rights, legal and evidential principles and values. The latter are (again:

⁵² *Shonubi IV*, 103 F.3d at 1089-1090.

⁵³ See, e. g., *Dukes v Wal-mart* 564 U.S. (2011), Opinion (per Ginsburg), at 11.

⁵⁴ *Dukes v. Wal-Mart Stores, Inc.*, Case Nos. 04-16688 and 04-16720, 603 F. 3d, at 652 (per Chief Judge Kozinski).

in the case of all modern Western legal orders that I know of) *anthropocentric*, not group-mediated. The requirement of specific/individualised evidence is not yet another component in the algorithmic set-up of some mathematical model of proof. It is essentially the bulwark against automated decision-making processes which can have legal or other significant effects on individuals.

3.3. Discretion

As outlined above, the effort to introduce mathematical models of proof in legal adjudication (and, subsequently, to resolve the emerging paradoxes) is, at its core, generated by the desire to eliminate the need for discretion. In jurisprudence there is a long history with this (unrealistic) aspiration. From voluminous codifications of the law with *numerus clausus* lists of circumstances over the effort of reducing the indeterminacy of legal terms to modern(?) approaches to Law and Technology, judicial discretion has traditionally been regarded as noise in the system. However, the fact that legal systems in an increasingly complex world are unable to anticipate the future and contain rules providing for every possible combination of facts is a historic lesson which we have learnt at least since the (doomed to fail) Prussian Legal Code (1794) with its more than twenty thousand paragraphs. In a constantly evolving world characterised by a radically unpredictable future, any codification or mathematical model—no matter how thorough or voluminous—would need radical revision moments after its enactment in order to catch the multitude of situations that per force occur in real life. Only if we could anticipate all possible combinations of fact, Hart (1961, p. 135) observes, open texture would be an unnecessary feature of rules. Such knowledge is neither possible nor intelligible.

Discretion, *i.e.* a core feature of the specific evidence rule, is to be welcomed rather than deplored (Hacker, 1977, p. 7) since it allows the law to adapt to constantly evolving new challenges. In all fields of experience there is a limit to the guidance general language let alone mathematical models of proof can provide (Hart, 1961, p. 126). In other words, the so-called «rule-following problem»⁵⁵— derived from the insight that rules can *never* dictate their own application—helps us understand that the necessity of judgment, *i.e.*, the need for judicial discretion, is present even if, or: especially when, we have the illusion that we «automatically» apply the law to facts, for non-standard cases will perforce arise (Hacker, 1977, p. 7). Since legal rules—no matter how precise—are intimately bound up with facts, and since facts are infinitely variable, we will unavoidably end up being entangled in the unpredictability of «our own rules (Wittgenstein, 2009, § 125), always in need of a decision (Hart, 1961, p. 122).

⁵⁵ For a comprehensive introduction to the problem of rule-following, see McGinn (1997, p. 73-112).

Unlike formal logic which «takes care of itself» (Wittgenstein, 1958, proposition 5.473), legal rules such as the *specific evidence rule* do not dictate their own application. The discussion concerning the RCP is generated by the commitment to a mathematical approach yielding generalisable results which, so the aspiration, would eliminate the need for discretion. However, systems of criminal adjudication do not provide an *explanation* of the world, which is necessarily general in scope. Criminal courts resolve practical conflicts by making judgments under uncertainty. The fact-finder is not building a scientific (general) model of the world. To ascribe criminal liability, we need an act of will which bridges the inferential gap between statistical data and the individual case. I submit therefore that the function of the *specific evidence rule* is to secure the fact-finder's decision-making prerogative when the latter is confronted with naked statistical evidence.

3.4. *Specific Evidence and Direct Evidence*

We can now move on to a more doctrinal analysis of the term *specific* especially its delineation from the term *direct*. We saw above that courts around the world require «specific evidence», *i.e.*, evidence that points specifically to the defendant and *his* or *her* actions or omissions. As the US Court of Appeals (Second Circuit) held in *Shonubi IV*: the «specific evidence» which is «required to prove a relevant-conduct quantity of drugs for purposes of enhancing a sentence must be evidence that points *specifically* to a drug quantity for which the defendant is responsible»⁵⁶. This requirement was, in Judge Weinstein's opinion, not only an unwarranted restriction of free assessment of evidence but also incompatible with the Federal Rules of Evidence, especially Rule 402. What is interesting at this point is the fact that, as we saw above, Judge Weinstein equates (and as I will show: conflates) specificity with directness. For example, Judge Weinstein equated these two terms when he wrote in *Shonubi III* that the US Court of Appeals had «called for “specific”, *i.e.*, “direct” evidence of drug transactions», expressing allegedly the belief that «only direct evidence suffices in criminal cases» (which Judge Weinstein calls a «shibboleth»), and assuming thus that specific evidence is the opposite of indirect, *i.e.*, circumstantial evidence⁵⁷.

However, the interpretation of *specific* as outlined by Judge Weinstein among many others is based on a misunderstanding. *Specific* evidence, as opposed to *group-mediated* evidence, is one thing. *Direct* evidence, as opposed to *circumstantial* evidence, is another. Even if the terms *specific* and *direct* seem for several practical purposes to be co-extensive, their contrary terms, *i.e.*, *group-mediated* and *circumstantial* respectively, are not. No evidence scholar or criminal court would in principle object to the possibility of convicting someone based entirely on circumstantial evidence.

⁵⁶ *Shonubi IV*; under section II (emphasis added).

⁵⁷ *Shonubi III* at 478.

For example, in the English case of *McGreevy v DPP*⁵⁸, the defendant was convicted of murder despite the absence of an eyewitness to the killing. The evidence which was deemed sufficient to shoulder a direction to the jury that there is a case to answer comprised information about the time and circumstances of the death, the defendant's opportunity to commit the murder, and the blood stains found on the clothing of the victim and the defendant. In the same case, the House of Lords denied the need for a categorical distinction between direct and circumstantial evidence of facts in issue, accepting that the assessment of evidence relies on a series of inferences from the evidence to the *probandum*, regardless of the direct or circumstantial character of the evidence. Circumstantial evidence *can* thus be sufficient to prove any offence including murder⁵⁹.

We could say therefore in a rather tautological way, that circumstantial evidence is indirect evidence. The important insight, however, is that *indirect* evidence is not the same as *undirected* (group-mediated) evidence. A DNA match as a sole item of evidence will not justify a case to answer for the defendant insofar, as it is undirected (group-mediated). The problem with statistical evidence is not that it's direct or circumstantial—this would depend on what we are using statistics *for*—but that it is not directed at a specific individual. In the same example as before: if the defendant whose DNA profile matched the forensic sample is also the deceased person's stalker, then it is highly unlikely that a court would decide that there is no case to answer. The fact that the defendant is the deceased person's stalker allows the fact-finder to individualise the group-mediated (undirected) statistical evidence. It turns the naked statistical evidence (e.g. the random match probability in the context of DNA evidence) into *specific evidence*. Going back to the *Shonubi* case there is an unbridgeable gap between:

- a) an extrapolation from the four balloons which were actually tested positively for drugs to the 103 balloons which *Shonubi* had swallowed, on the one hand.
- b) an extrapolation from the 8th trip to the US to the prior 7 trips, on the other hand⁶⁰.

The first extrapolation, Professor Finkelstein in his affidavit remarked, involves a tested statistical sample, in which the mechanism for selection was random selection of balloons which *Shonubi had verifiably swallowed*. The second extrapolation involves an inference from the balloons he had swallowed to the ones he assumedly *might* have swallowed during his previous seven first trips⁶¹.

⁵⁸ *McGreevy v DPP* [1973] 1 All ER 503 HL; see also *R v P* [2007] EWCA Cr 3216.

⁵⁹ See, e. g., *R v Athwal* [2009] 1 WLR 2430 [2009] EWCA Crim 789

⁶⁰ *Shonubi* II; under section II (emphasis added).

⁶¹ See *Shonubi* III, 895 F.Supp. at 520, citing Finkelstein's affidavit at 1-3.

4. OUR CRAVING FOR GENERALITY

The answer given in this paper is not universal—nor could it ever be. It applies exclusively to those legal orders which place value on human dignity and personal autonomy. However, in all those legal orders the reference class problem *can be dissolved*. The main problem in this debate has been our preoccupation with *the* method of science. Scientific explanations deal with empirical phenomena and aspire to provide a general account of the respective target system. *Generality* is a central feature of scientific enterprise in empirical domains. However, adopting scientific methods in legal adjudication not simply to educate fact-finders but to replace them, is awry. The fact-finders' objective is not to provide any kind of *general explanation* of a target system but to ascribe liability to an individual by rendering a verdict. It is the particular defendant and legal dispute that courts deal with. Whereas singularities are deeply troublesome for scientific theories, the sheer complexity of cases trumps any significant generalization of the verdict. In the words of L.H. Hoffmann (1975) «the slightest movement of the kaleidoscope of facts creates a new pattern which must be examined afresh» (p. 204).

Wittgenstein provided a similar diagnosis a while ago. He remarked that our «craving for generality» is a synonym of «the contemptuous attitude towards the particular case». Similar to philosophers—*i.e.*, Wittgenstein's target of criticism—, legal scholars too envisage the method of science and feel «irresistibly tempted to ask and answer [questions] in the way science does», *i.e.*, with aspirations to generality. Just that *craving*, Wittgenstein remarks, to wit, this «thirst for numerical kind of analysis is the real source of metaphysics» and leads us «into complete darkness»⁶². More broadly, the British philosopher Stephen Toulmin described in his seminal book *The Uses of Argument* how for the past few centuries the (traditionally synonymous) ideas of «rationality» on the one hand and «reasonableness» on the other were torn apart as a result of the emphasis that seventeenth century philosophers and especially logical empiricism, placed on formal deductive techniques. This move, *i.e.*, mapping rationality (and therefore: reasonableness) onto logicity, Toulmin (2001) adds, «did an injury to our commonsense ways of thought» and led to a substantial loss of legitimacy for established decision-making processes» (p. 216). Mathematical methods of proof directly challenge the validity of any other model of decision-making—*inter alia* the jurisprudential one—lacking a formal logical structure.

In a review article which turned to be a milestone for the study of dynamical systems, Robert May (1976) reminded us that «[t]he elegant body of mathematical theory pertaining to linear systems [...] and its successful application to many fundamental

⁶² See also *People v. Collins* 438 P.2d 33 (Cal. 1968): «[M]athematics, a veritable sorcerer in our computerized society, while assisting the trier of fact in the search for truth, must not [be allowed to] cast a spell over him» (per Sullivan J.).

linear problems in the physical sciences, tends to dominate even moderately advanced University courses in mathematics and theoretical physics». The mathematical intuition so developed, he added, «ill equips the student to confront the bizarre behaviour exhibited by the simplest of discrete nonlinear systems [...] Yet such nonlinear systems are surely the rule, not the exception outside the physical sciences» (p. 467). It has repeatedly been pointed out in theoretically sophisticated approaches to law that legal orders are complex adaptive systems. A typical judicial context involves a «rich, highly complex set of interdependent pieces of evidence» (Allen, 1997, p. 258), and no mathematical model of proof has the capacity to «do the math» by analysing even a modest system of beliefs consisting of one hundred propositions. The process becomes instantly *computationally intractable*, for the cataloguing of the combinatorial possibilities of all those elements would strain even a super-computer. In attempting to eliminate discretion in the assessment of evidence, we are walking blindfold into an even bigger problem.

Legal orders have their own established routines for validating knowledge claims. These routines are neither structurally nor—in terms of content—similar to the methods of natural sciences. The idea that some algorithmically validated (therefore: general) proposition guarantees the (factual and normative) rectitude of a legal verdict (which from a legal theoretical point of view is an individual norm) commits the fallacy of taking the extra inferential step based on assumptions that go far beyond what can be logically warranted by the underlying procedure⁶³. The RCP is a manifestation of the (vain) effort to change the decision-making business from an art to science.

The *specific evidence rule* sets out to make sure that the exercise of discretion will not be sidestepped by numerical methods or mathematical models of proof. We cannot reduce decision-making in law to statistical models as envisaged by proponents of actuarial models. Let's get over it and get on with the difficult task of understanding and structuring complex behaviour such as evidence assessment without resorting to oversimplifying and linear mathematical models which seek to ascribe liability to individuals just because the latter share certain characteristics with others⁶⁴.

5. CONCLUSIONS

In this paper I have offered a theoretical diagnosis⁶⁵ of the reference class problem by showing that it rests on assumptions which we are by no means bound to accept. Such an assumption is that the central choice of deploying formalised reasoning

⁶³ See Biedermann *et al.*, (2008, p. 120-132), for more discussion.

⁶⁴ Similarly, Scott (1993, p. 333): «But what theory of justice denies to individuals a basic right enjoyed by others, essential to the functioning of any society, merely because they share certain arbitrary characteristics with someone else?».

⁶⁵ See also Williams (1995, p. 186-200).

patterns as proxy for decision-making can be addressed in a normative vacuum, *i.e.*, independently of the procedural architecture of legal systems examining and validating criminal charges; in other words: that the answer to questions of justice is mathematical in nature (section 3.1). Through *structural* analysis I explained that the *rigidity* of any formalised approach generating the RCP inexorably deals with a class of individuals, rather than being tailored to a particular individual incl. his or her actions (section 2.3). Unless someone assumes a Natural-Law-based stance, then the rigidity of the abovementioned approach is not as such a problem. It *becomes* a problem in *liberal* legal orders which put emphasis on the value of the individual and personal autonomy. On the flipside, authoritarian legal orders enacting collectivist values are perfectly placed to eliminate discretion by reducing decisions to statistical data—provided that the latter are available.

Furthermore, I showed that group-to-individual inferences violate the *specific evidence requirement* which is an instantiation of central tenets of liberal legal orders such as human dignity and personal autonomy. The specific evidence rule is at the same time closely connected to the fundamental concept of judicial discretion which is an intrinsic feature of law (section 3.3.). Thereupon, I clarified the relationship between direct evidence and specific evidence by showing that circumstantial evidence is not the same as undirected (group-mediated) evidence (3.4). Finally, I explained that actuarial models of adjudication have become appealing due to our craving for generality and the epistemic imperialism on behalf of natural sciences which encroach upon legal methods of adjudication dealing with individual cases (part 4). Once we illuminate those assumptions, the reference class problem becomes a harmless observation about mathematical objects of inquiry rather than a thorn in the side of legal adjudication; it dissolves.

ACKNOWLEDGEMENT

The author would like to thank the two anonymous reviewers for their insightful feedback and useful suggestions. I have also benefitted greatly from the constructive criticism of Paul Roberts and Alex Biedermann.

BIBLIOGRAPHY

- Allen, R. J. (1997). Rationality, Algorithms and Juridical Proof: A Preliminary Inquiry. *International Journal of Evidence & Proof*, 1, 255-275.
- Allen, R. J. and Pardo, M. S. (2007). The Problematic Value of Mathematical Models of Evidence. *Journal of Legal Studies*, 36 (1), 107-140.
- Anderson, T., Schum, D. and Twining, W. (2005). *Analysis of Evidence* (2nd ed.). Cambridge University Press.
- Biedermann, A., Bozza, S. and Taroni, F. (2008). Decision Theoretic Properties of Forensic Identification: Underlying Logic and Argumentative Implications. *Forensic Science International*, 177, 120-132.

- Champod, C. (2009). Identification and Individualization. In A. Moenssens and A. Jamieson (eds.), *Encyclopedia of Forensic Sciences* (p. 1508-1511). Wiley.
- Cohen, L. J. (1977). *The Probable and the Provable*, Clarendon Press.
- Colyvan, M., Regan, H. M. and Ferson, S. (2001). Is It a Crime to Belong to a Reference Class?. *Journal of Political Philosophy*, 9, 168-181.
- Dahl, R. A. (1989). *Democracy and Its Critics*. Yale University Press.
- Feinberg, J. (1989). *The Moral Limits of the Criminal Law* (Vol. 3). Oxford University Press.
- Fetzer, J. H. (1977). Reichenbach, Reference Classes, and Single Case "Probabilities". *Synthese*, 34, 185-217.
- Gleick, J. (1994). *Genius: Richard Feynman and Modern Physics*. Abacus.
- Gleick, J. (2011). *The Information: A History, a Theory, a Flood*. Vintage.
- Hacker, P. M. S. (1977). Hart's Philosophy of Law. In P. M. S. Hacker and J. Raz (eds.), *Law, Morality and Society*. Clarendon Press.
- Hempel, C. G. and Oppenheim, P. (1948). Studies in the Logic of Explanation. *Philosophy of Science*, 15, 135-175.
- Hart, H. L. A. (1961). *The Concept of Law*. Clarendon Press.
- Hoffmann, L. H. (1975). Similar Facts After Boardman. *Law Quarterly Review*, 91, 193-206.
- Hughes, A. L. (2012). The Folly of Scientism. *The New Atlantis*, 37, 32-50.
- Kant, I. (1788). *Kritik der praktischen Vernunft*.
- Kaye, D. (1979). The Laws of Probability and the Law of the Land. *The University of Chicago Law Review*, 47, 34-56.
- Kaye, D. (2009). Probability, Individualization, and Uniqueness in Forensic Science Evidence: Listening to the Academies. *Brooklyn Law Review*, 75, 1163-1185.
- Kirk, P. L. (1963). The Ontogeny of Criminalistics. *Journal of Criminal Law, Criminology and Police Science*, 54, 235-238.
- Kotsoglou, K. N. (2017). Commentary: Federal Labour Court [2009] — 8 AZR 1012/08. *Frontiers in Sociology*, 2 (6), available online: <https://www.frontiersin.org/articles/10.3389/fsoc.2017.00006/full>.
- Kotsoglou, K. N. and McCartney, C. (2021). To the Exclusion of All Others? DNA Profile and Transfer Mechanics. *The International Journal of Evidence & Proof*, 25, 135-140.
- Kotsoglou, K. N. and Biedermann, A. (2022). Inroads into the Ultimate Issue Rule? Structural Elements of Communication between Experts and Fact-Finders. *The Journal of Criminal Law*, 86, 223- 240.
- Kuhn, T. S. (1996). *The Structure of Scientific Revolutions* (3rd ed.). University of Chicago Press.
- Lucy, D. (2013). *Introduction to Statistics for Forensic Scientists*. Wiley.
- May, R. M. (1976). Simple Mathematical Models with Very Complicated Dynamics. *Nature*, 261, 459-467.
- McGinn, M. (1997). *Routledge Philosophy Guidebook to Wittgenstein and the Philosophical Investigations*. Routledge.
- Mises, L. von, (1927). *Liberalism*. Ludwig von Mises Institute.
- Mises, R. von, (1964). *Mathematical Theory of Probability and Statistics*. Academic Press.
- Nagel, E. (1939). *Principles of the Theory of Probability*. University of Chicago Press.
- Neumann, J. von, (1955). Method in the Physical Sciences. In L. Leary (ed.), *The Unity of Knowledge* (p. 157-164). Doubleday.
- Nietzsche, F. (1974). *The Gay Science* (transl. by Walter Kaufmann). Vintage.
- Pundik, A. (2008). Statistical Evidence and Individual Litigants: A Reconsideration of Wasserman's Argument from Autonomy. *International Journal of Evidence & Proof*, 12, 303-324.
- Redmayne, M. (1999). Standards of Proof in Civil Litigation. *The Modern Law Review*, 62, 167- 195.
- Reichenbach, H. (1971). *The Theory of Probability* (2nd ed., transl. by Ernest H. Hütten and Maria Reichenbach). University of California Press.
- Roberts, P. (2007). From Theory into Practice: Introducing the Reference Class Problem. *International Journal of Evidence and Proof*, 11 (4) Special Issue on the Reference Class Problem, 243-317.
- Roberts, P. and Zuckerman, A. (2010). *Criminal Evidence* (2nd ed.). Oxford University Press.

- Salmon, W. C. (1974). Comments on “Hempel’s Ambiguity” by J. Alberto Coffa. *Synthese*, 28, 165-169.
- Schauer, F. (2003). *Profiles, Probabilities and Stereotypes*. Harvard University Press.
- Schum, D. A. (1994). *The Evidential Foundations of Probabilistic Reasoning*. Wiley.
- Scott, R. E. (1993). Chaos Theory and the Justice Paradox. *Wm. & Mary L. Rev.*, 35, 329-351.
- Shapiro, B. J. (1991). “Beyond Reasonable Doubt” and “Probable Cause”: Historical Perspectives on the Anglo-American Law of Evidence. University of California Press.
- Sober, E. (2002). Intelligent Design and Probability Reasoning. *International Journal of Philosophy of Religion*, 52, 65-80.
- Sober, E. (2008). *Evidence and Evolution: the Logic behind the Science*. Cambridge University Press.
- Solzhenitsyn, A. (2018). *The Gulag Archipelago*. Vintage.
- Suppe, F. (2000). Understanding Scientific Theories: An Assessment of Developments. *Philosophy of Science*, 67 (S3) Supplement. Proceedings of the 1998 Biennial Meetings of the Philosophy of Science Association, S102-S115.
- Thomson, J. J. (1986). Liability and Individualized Evidence. *Law and Contemporary Problems*, 49 (3), 199-219.
- Tillers, P. (2005). If Wishes Were Horses. *Law, Probability and Risk*, 4 (1-2), 33-49.
- Toulmin, S. E. (1976). *The uses of argument*. Cambridge University Press.
- Toulmin, S. E. (2001). *Return to Reason*. Harvard University Press.
- Tribe, L. H. (1971). Trial by Mathematics: Precision and Ritual in the Legal Process. *Harvard Law Review*, 84, 1329-1393.
- Twining, W. (1982). The Rationalist Tradition of Legal Scholarship. In E. Campbell and L. Waller (eds.), *Well and Truly Tried: Essays on Evidence in Honour of Sir Richard Eggleston* (p. 211-249). Law Books.
- Wasserman, D. T. (1992). The Morality of Statistical Proof and the Risk of Mistaken Liability. *Cardozo Law Review*, 13, 935-976.
- Weait, M. (2005). Knowledge, Autonomy and Consent: R. v Konzani. *Criminal Law Review*, October, 763-772.
- Williams, M. (1995). *Problems of Knowledge*. Oxford University Press.
- Wittgenstein, L. (1958). *Tractatus Logico-Philosophicus* (transl. by D. F. Pears and B. F. McGuinness). Routledge and Kegan Paul.
- Wittgenstein, L. (2009). *Philosophical Investigations* (4th ed., transl. by G. E. M. Anscombe, P. M. S. Hacker and J. Schulte). Wiley-Blackwell.
- Zuckerman, A. A. S. (1986). Law, Fact or Justice?. *Boston University Law Review*, 66, 487-508.

