

NATURALIZED EPISTEMOLOGY AND THE LAW OF EVIDENCE REVISITED

Ronald J. Allen*

ABSTRACT: This article revisits «Naturalized Epistemology and the Law of Evidence», published twenty years ago. The evolution of the relative plausibility theory of juridical proof is offered as evidence of the advantage of a naturalized approach to the study of the field and law of evidence. Various alternative explanations of juridical proof from other disciplines are examined and their shortcomings described. These competing explanations are similar in their reductive, a priori approaches that are at odds with an empirically oriented naturalized approach. The shortcomings of the various a priori approaches are driven by their common methodological commitments to employing weird hypotheticals to engage in intuition mining about the American legal systems that in turn ignore important aspects of the actual legal systems and consistently make impossible epistemological demands. As a result, both their descriptions of and prescriptions for American legal systems are implausible, unlike the relative plausibility theory.

KEYWORDS: naturalized epistemology, juridical proof, jurisprudence, probability, relative plausibility

RECOMMENDED CITATION: ALLEN, R. J., 2020: «Naturalized Epistemology and the Law of Evidence Revisited», in *Quaestio facti*, 2: 253-284. Madrid: Marcial Pons Ediciones Jurídicas y Sociales. DOI: http://dx.doi.org/10.33115/udg_bib/qf.i2.22446

* 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. I am also indebted to Marcello Di Bello for helpful comments on a draft of this article, and to the Visiting Scholars and SJD students studying with me this year at Northwestern: from Brazil Vitor de Paula Ramos and from China Zhuhao Wang, Siyun Liao, Kaixin Zheng, and Xiaobiao Xiong. I am particularly indebted to Regina M. Hurley, a Ph.D. student at Northwestern University for very helpful discussions on various aspects of this paper, although all of the views expressed herein are my own.

Twenty years ago, Brian Leiter and I published an article examining the relationship between naturalized epistemology and investigations into the field and law of evidence. In brief, our argument was that the naturalized turn in philosophy inspired by Quine and others contained valuable lessons for scholars interested in the field of evidence and the various rules for trials that emerge from that field. «Naturalized epistemology» can mean different things to different people. The way we used the phrase was, first, to denote the difference between a priori conceptual/normative analysis on the one hand and on the other empirical inquiry into the actual state of some phenomena (in our case, the law of evidence and the nature of litigation), and second, to emphasize that, no matter what else, «should implies can». For conceptual/normative work, and any other kind for that matter, to have value for understanding or theorizing about the objects under inquiry, it must be responsive to what is in fact possible given the human condition. We then employed that methodological approach to demonstrate that an emerging theory of juridical proof—the relative plausibility theory—was systematically superior to the then reigning probabilistic conception¹.

Much has happened since then. On the theoretical side, philosophical disputations over «naturalized epistemology» continue, as has the implications of that debate for law in general and jurisprudence in particular². On the applied side, there is good reason to believe that relative plausibility has supplanted probabilism as the dominant paradigm for understanding juridical proof³. This has been aided by the shoring up of the foundations of relative plausibility that began in the original Allen/Leiter article through incorporation of aspects of inference to the best explanation informed by the reality of the American trial process and the manner in which people reason in real life⁴. Together these developments are evidence of the significance of naturalized epistemology for the field of evidence.

Perhaps in part responsive to these developments, the editors of *Quaestio Facti* suggested that I revisit the implications of naturalized epistemology for the field of evidence in the form of a contribution to the journal's «Conjectures and Refutations» section, and this article is the result. Before turning to substance, a clarification is in order concerning the meaning of «naturalized epistemology», as I would express it in somewhat different terms (but not obviously with significant change of meaning). As I am using the phrase, it refers to 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. More particular issues include how such legal systems structure dispute

¹ ALLEN and LEITER, 2001.

² See the symposium on Brian Leiter's book in *Law and Philosophy*, 2011, 30: 377-516.. Also see LEITER, 2011, 2020.

³ See the symposium on Relative Plausibility and Its Critics, *International Journal of Evidence and Proof*, 2019, 23 (1-2): 1-217.

⁴ See, e.g. PARDO and ALLEN, 2008.

resolution, the relationship between that structure and the field of evidence with its resultant rules, and how all of this is influenced by cognitive capacities.

To be clear, the objective here is not to canvas such a sprawling field but instead provide a relatively brief but hopefully provocative essay to stimulate «refutations», and in doing so to stress two things. First, I will briefly review the successes of the naturalized approach. Second, in an admittedly solipsistic vein, I will lament some of the shortcomings of alternative approaches to theorizing about the field of evidence if the objective is to understand the actual state of the subject of the inquiry—to help us jurists in other words. This last point is critical. Different disciplines do different things, are motivated differently, and have differing objectives. For whatever reasons, a cottage industry focusing on juridical proof has arisen from various disciplines. Using what are taken as examples from the legal arena and subjecting them to scrutiny from the perspective of some other discipline may very well be valuable to that discipline, a matter beyond my competence. However, the utility of such efforts for understanding juridical proof may be another matter, and it often is for three reasons. The first is the weird hypothetical problem. Hypotheticals are posited that are supposed to reflect some aspect of the legal system or the field of evidence, but do not, and then subjected to scrutiny. Because such hypotheticals do not reflect critical aspects of any existing legal system, the scrutiny they are subject to yields little of value for understanding of or prescribing for American legal systems. Second, and relatedly, some theorizing neglects critical aspects of the actual state of affairs in the object being theorized about. Neither of these necessarily violates the «should implies can» canon, but tying analysis to either realistic hypotheticals or the actual state of affairs might be a good idea if one is trying to understand those states of affairs. The third problem more intimately focuses on «should implies can» and that is that much of the theorizing about the field of evidence from the cognate disciplines makes essentially impossible epistemological demands, and to the extent that is true the resulting analysis is of little use for the jurist.

Collectively, these three problems may be driven by a factor that is at least contingently if not necessarily highly correlated with a priori analysis, which is the tendency to attempt to reduce problems to a small set of variables—one if possible—in order to identify «the» solution or key to a problem or answer to a question. At a very fundamental level, this effort to reduce things to various sorts of algorithms misconceives the nature of the inquiry (if again one is trying to explain juridical proof). The American legal systems are not simple and static but instead complex and dynamic, and to be understood must be approached from that perspective⁵. Legal systems, like markets and democratic governments, are complex adaptive systems, and their «parts» (for example, the field and rules of evidence or procedure, conventional trial practices and so on) are part of the construction of that complex adaptive system. To

⁵ See, e.g. ALLEN, 2013. For an argument about the misdirection of «analytic jurisprudence» that is sympathetic to the central argument of this paper, see PRIEL, 2020.

approach the field of evidence as though it were a simple and static matter usefully subjected to a priori, top-down analysis would be like asking what are the simple rules governing the Amazon River's relationship to the rain forest, or asking what the point of either the Amazon River or the rain forest is. Answers to such questions will be uninformative at best. So too will be algorithmic answers to the nature of juridical proof.

That, at least, is the burden of this essay. It is not to give advice to other disciplines how to do their work; rather, it is to explain why what they do often cannot be used by the jurist—pure solipsism in other words, but from a perspective «to highlight the importance of continued dialogue between legal epistemology and formal epistemology»⁶ at least for the jurists. I start with an examination of the success of relative plausibility in supplanting probabilism as the best explanation of juridical proof, which includes a discussion of the quite odd debate over probabilistic/statistical evidence, an odd debate stimulated in large measure by its failure to attend to the lessons of naturalized epistemology. I then turn to some possible causes of that oddness: weird hypotheticals, neglecting the actual state of legal systems, and impossible epistemological demands.

The interrelationship of law and probability is ancient⁷, but sustained scholarly attention was not focused on it until the seminal article by John Kaplan in 1968⁸. Probability theory in general and decision theory as its companion seemed to provide a compelling framework for understanding much of juridical proof. Indeed, it is not too much to say that the idea swept the field that conventional probability theory applied directly to legal decision-making, and that there remained only idiosyncratic problems like naked statistical evidence⁹. Cracks in the façade appeared when L. Jonathan Cohen identified a set of proof paradoxes latent in the probability approach¹⁰. He thought he could resolve them by replacing one formalism—probability theory—with another—a formalized Baconian inductive logic¹¹. Shortly thereafter, the paradoxes were extended and another somewhat formal solution offered to solve the problem—a reconceptualization of trials rather than of probability¹². The problems were not solved, and deeper anomalies were discovered¹³.

The anomalies and the failure to resolve them were driven by the variables noted above: weird hypotheticals that distract rather than facilitate inquiry, failure to attend

⁶ HEDDEN and COLYVAN, 2019. For an example of the general proposition of this article implemented, see ALLEN, 2020.

⁷ See HACKING, 2006. For criticisms of the first edition of Hacking's book, see GARBER & ZABELL, 1979.

⁸ KAPLAN, 1968.

⁹ See e.g., KAYE, 1979a, 1979b, 1982.

¹⁰ COHEN, 1977.

¹¹ *Ibid.*

¹² ALLEN, 1986.

¹³ ALLEN, 1997.

to the actual conditions of juridical proof, and impossible epistemological demands, all driven by the reductive tendency of a priori analysis. Take weird hypotheticals and the related problem of not attending to the actual conditions of juridical proof first. Much discussion has focused on hypotheticals that are designed to elucidate the issues surrounding statistical evidence in particular and probability theory more generally. Two in particular dominate: the Gatecrasher hypothetical of Cohen, and the Blue Bus example of just about everybody. Rather than being enlightening about juridical proof, the attention focused on these (and similar) hypotheticals may have simply impeded progress.

The Gatecrasher involves a rodeo with 1000 attendees but only 499 have paid. The rodeo operator picks one at random to sue for the entry fee, and those are the only facts before the court. Who should win? According to a version of probabilism, the rodeo operator, as the probability is greater than .5 that the randomly selected person did not pay. This, in turn, is said to violate our intuitions that purely statistical evidence is not sufficient for liability, and thus counts against a probability interpretation of juridical proof¹⁴. The Blue Bus hypothetical is similar and is said to be inspired by a real case¹⁵. Imagine an accident where an injured party says she was injured by a bus, that the Blue Bus Company operates 70% of the buses in the city and the Red Bus Company operates 30%, and that is the only evidence. Should the Blue Bus Company be liable? A number of legal academics as well as those from other disciplines apparently think no and that that is the result the legal systems would reach. How does that compare to a witness who is 70% reliable and testifies that a Blue Bus hit her? By comparison to the directly statistical, those same academics often shift to yes, the answer they assert that would also be given by the legal system¹⁶.

¹⁴ Enoch et al. assert that «the powerful, uncompromising intuitions of pretty much all of [evidence] scholars» is that there is a critical difference between statistical and non-statistical evidence and that the law is consistent with that intuition. ENOCH, et. al, 2012. For what it is worth, that is neither my intuition nor more importantly my reasoned judgment and as discussed below it also grievously misstates the «law». Perhaps reflecting a similar impatience with weird hypotheticals, Christian Piller, «criticize[s] safety without participating in the production and discussion of ingenious counterexamples». PILLER, 2019: 307. Jonathan Weinberg, Shaun Nichols, and Stephen Stich provide some evidence that Gettier intuitions are culturally dependent, and from that suggest that intuitions should not be relied upon. WEINBERG et al., 2001. A recent article, GARDINER 2020, argues that many of the intuitions concerning these various hypotheticals are unjustifiable in large measure due to the failure to account for differences in the scenarios. From this, she questions whether a unified explanation of various hypotheticals can be provided. This recognizes the complexity being addressed, and to that extent is consistent with my argument; she does not take into account the implications of adaptability, which makes the matter of complexity all that more complex.

¹⁵ Smith v. Rapid Transit, 58 N.E.2d 754 (Mass. 1945).

¹⁶ See, e.g., ENOCH et al., 2019 for a version of the Blue Bus hypothetical that concludes that if there is no other evidence, and even if «the Blue Bus Company controls 70 percent of the market in the relevant area, the law [...] refuses to find the Blue Bus Company liable». This is usually equivalent to the statement that statistics are an inadequate basis for liability. Even though repeated ad nauseam in the literature, the statement is false, as the text elaborates. There is some empirical support for the proposition that people behave consistently with these intuitions. WELLS, 1992.

Employing hypotheticals such as these is a popular strategy in philosophy and it can be a helpful way of abstracting away from unnecessary details of real-life cases that are orthogonal to the questions at hand. However, for these hypotheticals to be helpful (to jurists), they must be crafted without losing sight of the actual object of inquiry, namely, the legal system. Both fail to do so, and thus neither of these hypotheticals is enlightening about juridical proof. In the Gatecrasher hypothetical, the obvious problem is that in the real world there are numerous sources of evidence that would be proffered by the defendant, such as their own testimony, the ticket stub, a receipt, the testimony of someone who observed the person buy the ticket, to name just a few¹⁷. Cohen assumed no such evidence could be forthcoming, but the reason given demonstrates the problem with the hypothetical: «Perhaps the defendant was prevented by death from giving evidence on his own behalf and the relentless management pursued the case against his estate»¹⁸. In that case, in many states the issue would never arise because the dead man statute would prohibit interested parties from testifying to interactions or transactions with the dead person¹⁹. But what if you were not in a dead man state? The death of the defendant would not preclude any of the sources of evidence noted above, and there are many more in addition. A plethora of hearsay exceptions would apply that allow testimony from third parties about the person's intentions, and thus what the person likely did²⁰, and in any event who was the person sitting next to? Did anyone see him or her at the rodeo, and did they discuss the price of tickets? The hypothetical does not abstract away from unnecessary complications in order to distill the essence of some question facing American legal systems; it abstracts away from many aspects of the legal system that connect the hypothetical to reality²¹.

Perhaps focusing attention on limited and attenuated aspects of the formalities of proof and evidence, rather than on the structure of actual legal systems and dispute resolution, led to the neglect that trials pursue many objectives in addition to truth, one very important one being a governance function that creates incentives for ap-

¹⁷ The next few paragraphs are little more than an elaboration on KAYE, 1982, who argued that it was sensible to construct incentives to produce more evidence when it is available, but also to decide cases on the basis of what he called naked statistical evidence when that is all the evidence that there is. That is exactly right, although the structure of those incentive programs can be complicated.

¹⁸ COHEN, 1981a.

¹⁹ See, e.g., *Spencer v. Wayne*, 80 N.E.3d 764 (2017 IL App (2d)); *Matter of Estate of Owen*, No. 35879-8-III, 2019 WL 6876791, at *6 (Wash. Ct. App. Dec. 17, 2019) («[Washington's dead man's statute] provides that in an action against an estate, an interested party may not offer testimony adverse to the estate regarding transactions made with the decedent». (Citing WA ST 5.60.030 (West 1977))).

²⁰ For a famous example, see *Mutual Life Insurance Co. of New York v. Hillmon*, 145 U.S. 285 (1892).

²¹ Responding to a version of the unrealism point in the text, David Enoch and Talia Fisher write: «This is an empirical matter lying outside the scope of this Article». That captures the distance between the opposing approaches to these problems that being addressed. ENOCH and FISHER, 2015: n. 17. The hypothetical also involves strange epistemological demands. Some critical facts are taken as absolutely certain, with no explanation as to how that could be so. For further on this, see n. 57, *supra*.

propriate behavior²². Allowing recovery in cases like the Gatecrasher would create incentives for liability traps—hold events, and start suing people randomly for entry fees. That is absurd, of course, but what is not absurd is for the law to discourage people like rodeo operators from creating anything analogous to liability traps by disallowing recovery when the real problem is the very conditions that the plaintiff created rather than the asserted behavior of others. In more concrete terms, it is perfectly appropriate for the law to conclude that it is the rodeo operator's problem if he or she does not fence off the rodeo adequately, have security or a controlled entrance point, fails to install video cameras, or whatever²³.

When more general incentive considerations did enter the picture, the result was equally strange. For example, one recent work argued that a serious problem with statistical evidence of the Gatecrasher sort—and a potential basis for distinguishing it from testimonial evidence—is that it eliminates an individual's incentive to comply with the law²⁴. If one knows that the mere fact of being selected will result in liability, why bother in the Gatecrasher context to buy a ticket? Just crash the rodeo and take your chances. You cannot be worse off as the worst-case scenario is you get picked randomly and end up having to pay the entry fee, whereas if you actually buy a ticket you might end up having to pay twice. The answer is parallel to the discussion above about incentives for rodeo owners: if you buy the ticket and keep the stub, get a receipt, have a witness observe you, or speak the language of the country the trial is in so that you can testify yourself, you will not suffer the inconvenience of being sued. It seems like an easy trade-off. Only if your selection is random and unknown forces of nature prohibit you from any of the escape hatches will you be liable, but the existence of such a world is fanciful²⁵.

The Blue Bus hypothetical may be somewhat less ridiculous than the Gatecrasher hypothetical, but it, too, contributes little of interest regarding the nature of juridical proof. It is simply an example of lawyerly ineptitude, which seems to be what caused the court in the *Smith v. Rapid Transit* case to do what it did. In the actual

²² ALLEN, 2015. Prof. Nance has extended the governance function to include incentives to produce the best evidence, which may serve the epistemological aspirations of the American legal systems. See NANCE, 2016. His book is the best and most systematic examination of the implications of a general best evidence principle that animates a significant portion of the law of evidence.

²³ Prof. Kaye long ago noted a possible incentive effect concerning the production of evidence. KAYE, 1979c. The difficulty, which made the debate at the time «odd», ALLEN, 1986: 411, is that such an incentive effect operates, or can operate, on all parties to litigation. There is no reason to limit the analysis to plaintiffs or defendants.

²⁴ ENOCH and SPECTRE, 2019. Also, ENOCH et al., 2012.

²⁵ There are other arguments in the literature about incentives that neglect the almost infinite array of choices a person can make to influence their likelihood of liability. See SANCHIRICO, 2001. Character evidence is said to be an example of an incentive killing evidence rule but in fact it just creates different incentives—like don't go to bars where fights occur or jewelry stores where robberies happen. Or if you do go, go with a responsible companion. For a discussion of such matters, see PARK & SAKS, 2006. The incentive arguments are discussed further below.

case, no evidence was adduced as to the prevalence of private buses in the area, of the safety record of the company being sued, of the likelihood of a bus being in the area of the accident at the appropriate time given the bus schedules of the company, of interrogations of the bus drivers that might have been present, of any effort to search for other eyewitnesses, and so on²⁶. If the case stands for anything, it is to vindicate Craig Callen's point that an explanation of juridical evidence should contain an escape hatch for cases that are just woefully insufficient, regardless of any other circumstances²⁷. And as David Kaye argued with respect to the Gatecrasher hypothetical, to give lawyers and parties cost/benefit justified incentives to do their jobs properly²⁸.

The problems with the debates over probability theory extend beyond weird hypotheticals disconnected to reality to the third problem—impossible epistemological demands. Probabilities have to come from somewhere and in a typical case (unlike employment discrimination or DNA cases, or stylized hypotheticals), they would come from the massive amounts of discrete pieces of qualitative evidence adduced at trial. There are numerous debilitating difficulties for probabilistic theories of juridical proof resulting from the manner in which juridical proof actually unfolds²⁹. For example, any effort to deal with a mass of evidence in a probabilistic manner immediately confronts computational intractability³⁰ and without computation there is little point to probabilistic approaches (otherwise people are just making up numbers)³¹.

There is a still deeper problem here. For probability theory to be explanatory of evidentiary processing at trials, people would need to know the necessary probabilistic data to appraise evidence, such as accurate base rates and likelihoods. But to do that, they must already know the outcome of the case being tried. Evidence at trial is obviously contingent in the sense that the implications of any particular piece of evidence is a function of all the other data relevant to the case. Litigated cases are not like pari-mutuel betting in which fully specified statistical rules determine precisely the effect of each new bet. Evidence of the existence of a dead body does not mean that a murder occurred, for example, because subsequently admitted evidence may show the death to have been accidental or in self-defense. The evidence that the de-

²⁶ There was evidence that indicated the accident occurred on the defendant's bus routes at a time within the bus schedule's route, but little else. No effort to show that, given the bus schedule, a bus should or could have been close to the scene of the accident at the asserted time it occurred, and as the court said, «While the defendant had the sole franchise for operating a bus line on Main Street, Winthrop, this did not preclude private or chartered buses from using this street; the bus in question could very well have been one operated by someone other than the defendant». *Smith v. Rapid Transit, Inc.*, 58 N.E.2d 754 (Mass. 1945).

²⁷ First articulated in CALLEN, 1991.

²⁸ KAYE, 1979c.

²⁹ ALLEN, 1997.

³⁰ CALLEN, 1982.

³¹ For an example of the futility of making up numbers, see HEDDEN and COLYVAN, 2019, and ALLEN, 2020.

defendant was in the vicinity can be quite exculpating if there is further evidence that he always visits his sick mother at that particular time. If evidence is not contingent in this sense from a fact finder's perspective, then the fact finder already must know that there was no accident or self-defensive act or that the defendant did commit the crime, but that is exactly the point. To implement a probabilistic theory of evidence such as Bayesian updating requires that the outcome already be known, which of course makes the trial itself superfluous³².

The astute observer will notice, and possibly complain, that analytical points have been met with pragmatic considerations. For example, a primary concern about the Gatecrasher hypothetical is that it is ridiculous on its face and would never arise in an American legal system, and the significant point about the Blue Bus hypothetical is that it is an example of lawyerly ineptitude. Perhaps these responses really are "fighting the hypothetical", but even if so perhaps it is justified. One should not fight the hypothetical when it may actually be a productive object of reflection, but it is a different matter when that is not true. The hypotheticals do not test intuitions about an actual legal system, though they purport to do so. Instead, the assumptions that they rely upon misconceive the object of inquiry, and thus reflecting upon them may obscure rather than enlighten about an actual legal system. Furthermore, as discussed below, when a real case analogous to the hypotheticals actually arises—such as cold hit DNA cases—the results are precisely what they should be if the point is error allocation. And again, there may be other «points» than error allocation justifying differing rules. One must actually look at a legal system to see if this is true.

By comparison to the formalistic probability approaches, what one sees when one actually looks at the American legal systems, and the scope of the project of understanding juridical proof, was recently succinctly explained:

Its scope is [...] about the entire process of proof, including (1) the form, securing, and presentation of evidence, (2) the forms of argumentation employed at trial, (3) the manner in which humans process and deliberate on evidence, (4) the trial structure created by the rules of evidence and procedure, (5) the structure of litigation before and after trial, (6) the manner in which judges and juries, on the one hand, and trial and appellate judges, on the other hand, interact, and (7) to some extent about the meaning and nature of rationality. In other words, all of these features comprise the sprawling entity that we refer to as "juridical proof", and relative plausibility is an attempt to explain that entity in all of its aspects, from beginning to end³³. How any part of the process operates is thus relevant to the project...

³² For a discussion, see ALLEN, 1997. By contrast, statistical evidence plainly can be probative in the context of deciding the relative plausibility of competing explanations of events. This emphasizes why it is important to keep in mind what one is trying to explain. As noted in the text, relative plausibility is attempting to explain the proof process from beginning to end, and I am taking the propounders of the weird hypotheticals discussed herein to be engaged in an analogous effort. It's quite a different matter if one is simply trying to explain idiosyncratic cases or evidentiary proffers.

³³ At the beginning of litigation, see, for example, the U.S. Supreme Court's insistence that a plaintiff's allegations in a complaint must be «plausible», and that plausibility is assessed by comparing the allegations with other possible alternative explanations. See *Bell Atlantic Corp v. Twombly*, 550

Relative plausibility explains this sprawling entity and its interconnected parts better than conventional probabilism. Relative plausibility explains the behavior of lawyers in preparing for trial, judicial involvement in pre-trial hearings, and the formulation of theories of the case. It explains virtually all of the procedural context at trial from the opening statements through disjointed presentations of evidence to closing arguments to the meaning of relevance (where by contrast probabilism cannot explain the massive overlap of evidence so that it predicts that most of the evidence actually admitted at trial is irrelevant)³⁴. Relative plausibility explains the behavior of trial and appellate judges. It is explained by (the relationships are bilateral)³⁵ the basic requirements of rationality and the potential and limits of cognition. Most of which the conventional probability explanation fails utterly to accommodate. Indeed, to the extent that relative plausibility relies on what decision-makers do, it exploits the necessity of knowing what the alternatives are and how they are examined for truth value in the process of rational decision-making. The primary message of relative plausibility is that from beginning to end the legal system pushes the parties to provide competing explanations, and these explanations structure the decision that is subsequently made (even if the decision is based on an explanation not advanced by the parties)³⁶.

Relative plausibility has received sustained attention in the literature, and need not be belabored here beyond, simply noting that its success as an explanatory theory stands in stark contrast to the consistent pattern of failure of probabilistic explanations. It also bearing noting that the development of relative plausibility was not influenced by weird hypotheticals; it embraced rather than neglected the actual state of the systems under inquiry; and it exploited rather than assumed away human cognitive capacities. I, of course, see causation operating here rather than correlation.

Other reductionist approaches to juridical proof that have followed in the wake of the demise of the probabilistic explanation exhibit similar attributes of invoking weird hypotheticals, not attending to the actual condition of legal systems, and making impossible epistemological demands. I will discuss two examples: the sensitivity

U.S. 544, 567-78 (2007); *Ashcroft v. Iqbal*, 556 U.S. 662, 682 (2009). Toward the end of the process, see, for example, the fact that appellate courts routinely evaluate the sufficiency of evidence to support a verdict by comparing alternative explanations. See *Anderson v. Griffin*, 397 F.3d 515, 521-22 (7th Cir. 2005); *Bammerlin v. Navistar*, 30 F.3d 898, 902 (7th Cir. 1994); *Yeschick v. Mineta*, 675 F.3d 622, 627 (6th Cir. 2012); *O’Laughlin v. O’Brien*, 568 F.3d 287, 304-08 (1st Cir. 2009); *United States v. Beard*, 354 F.3d 691, 692-93 (7th Cir. 2004); *United States v. Newell*, 239 F.2d 917, 920 (7th Cir. 2001). See also *United States v. Sever*, 603 F.3d 747, 753 (9th Cir. 2010) (abuse of discretion in denying defendant’s discovery request for evidence to support defendant’s «alternative explanation»).

³⁴ See *DIAMOND et al.*, 2012; *VIDMAR and HANS*, 2007; *BENNETT and FELDMAN*, 1981; *WAGENAAR et al.*, 1993. On overlapping evidence see, *PARDO* 2013: 576-89.

³⁵ See, e.g., Ribeiro’s elegant presentation of this point: «What I am suggesting is that the testimony is relevant because it makes the hypothesis that the employee stole the necklace a better explanation, not because the evidence is explained by this hypothesis». *RIBEIRO*, 2019: 109.

³⁶ *ALLEN and PARDO*, 2019b: 207-208. The Supreme Court of the United States recently commented on this fundamental aspect of the American legal systems in *United States v. Sineneng-Smith*, *** U.S. *** (decided May 5, 2020): «In our adversarial system of adjudication, we follow the principle of party presentation. As this Court stated in *Greenlaw v. United States*, 554 U.S. 237 (2008), “in both civil and criminal cases, in the first instance and on appeal [...], we rely on the parties to frame the issues for decision and assign to courts the role of neutral arbiter of matters the parties present”. *Id.*, at 243».

and incentive proponents on the one hand and the relatively new idea of «normic» support on the other³⁷.

The sensitivity theorists, perhaps more influenced by the realities of the legal systems, are more cautious in their conclusions than were the probability theorists³⁸, but one sees analogous problems popping up in their discussions of hypotheticals designed to elucidate the nature of juridical proof generally and specifically the difference between statistical and non-statistical. Here is a recent bus color version:

Eyewitness: A bus causes harm, an eyewitness recognizes the bus as belonging to the Blue Bus Company. The witness, however, is imperfectly reliable; let us say that she is roughly 70 percent reliable in matters such as this one. The law has no qualms about accepting the eyewitness testimony as evidence and indeed basing a positive finding that the bus was a Blue Bus bus (and perhaps also that the Blue Bus Company is liable) on the testimony³⁹.

Statistic: A bus causes harm. There is no eyewitness, but we have uncontested data regarding the distribution of buses in the relevant area; in particular, the Blue Bus Company owns roughly 70 percent of the buses there. Here, though, the law typically will not be willing to base a positive finding of fact—and certainly not liability—on just this kind of evidence, sometimes called statistical evidence⁴⁰.

David Enoch, Talia Fisher, and Levi Spectre take this hypothetical as a starting point for analyzing sensitivity. A verdict, v , is «sensitive» if and only if, had v been false, fact finders would (most probably) not have found that v ⁴¹. Verdicts based on the market share statistic are insensitive because, had it not been a blue bus that caused the harm, the statistical evidence nonetheless would lead the jury to falsely find the Blue Bus Company liable for the harm, given that the market share data would remain the same in those cases. The verdict based on the eyewitness testimony, by contrast, is sensitive. Had it not been a blue bus that caused the harm, Enoch et al. assert, the witness would have «probably» testified accordingly and, thereby, the jury probably would have come to correctly find that the Blue Bus Company is not liable for the harm⁴².

The sensitivity theorists are careful to note that their claims about sensitivity are not meant to be definitional of statistical evidence but, rather, are meant to provide a general «instrumental, incentive-based» explanation for why the law generally prefers

³⁷ Also addressed in passing is Sarah Moss' work in *Probabilistic Knowledge*, Moss (2018), which possesses analogous attributes. See n. 68, *infra*.

³⁸ On probabilistic imperialism, see EDWARDS, 1986. See also many of the responses to COHEN, 1981b, in *Behavioral Brain Sciences*, 4(3).

³⁹ ENOCH et al., 2019: 197.

⁴⁰ *Ibid.*

⁴¹ Their formulation of sensitivity is framed in terms of beliefs as follows: «*Sensitivity*: S's belief the p is sensitive =df. Had it not been the case that p , S would (most probably) not have believed that p ». *Ibid.*: 204. I have taken the liberty of changing the epistemic notion of «belief» to the legal notion of «verdict» given that the legal notion is what is of central interest.

⁴² ENOCH et al., 2019: 197, 206, n. 13.

non-statistical to statistical evidence⁴³. They reiterate the argument against statistical evidence that an agent's reasons to comply with the law are «seriously compromised unless the agent knows, *ex ante*, that her action will affect how the law treats her; that is, unless she has very good reason to expect the law's treatment of her to be sensitive to her behavior»⁴⁴ and that «treating statistical evidence as the law treats individual, direct evidence would diminish the incentives supplied by substantive law in sufficiently many central cases»⁴⁵. Although they recognize that sensitivity alone may not be of central concern to the law, they see it to be intimately bound up with their incentive argument.

Regardless of how one gets to their destination, the general point about disincentivizing agents seems particularly peculiar for analogous reasons given above with respect to the Gatecrasher. There are two possible applications of the incentive argument, although Enoch et al. do not sort them out. One is the effect on the most immediately critical agents, the bus drivers, and the other is the effect on their employers, the bus companies. The possible effect on bus drivers is discussed first because it is so obviously wrong, followed by a somewhat more extended discussion of the effect on bus companies.

According to one application of the argument of Enoch et al., the law does not hold the Blue Bus Company liable on the basis of statistical evidence because doing otherwise would disincentivize bus drivers from causing such harm. This would require that bus drivers of both red and blue buses ponder the risks of harm to them or their companies from being held liable on the basis of statistics regarding their company's market share. It would be quite surprising if bus drivers even knew about let alone had these statistics in mind given that there are far more important things to be mindful of in their line of work, and far more incentivizing reasons to avoid causing harm, such as losing their jobs, hurting pedestrians, or being shunned by their colleagues, just to name a few. Moreover, the bus drivers would have to think that they would be in the category where the exclusive evidence is statistical—only if they run over pedestrians where no other probative evidence is available will the statistical evidence be of importance. That seems a bit implausible.

Enoch et al. recognize the limitations of the incentive argument and mention what one might call the larger concerns of the legal system contained within the governance function of evidence law⁴⁶. For the law to be sensible and for the incentive argument to get any traction, the various options and the costs and benefits of all the relevant participants' behavior have to be taken into account, which includes

⁴³ ENOCH et al., 2012: 200; ENOCH and SPECTRE, 2019: 181.

⁴⁴ ENOCH and SPECTRE, 2019: 181, 183.

⁴⁵ *Ibid.*: 182. ENOCH et al. also invoke normic support for the distinction that they are drawing, which is addressed below.

⁴⁶ ENOCH et al., 2019: 219. On the governance function, see *supra.*: n. 22. For the most part, this is just another aspect of socially useful incentives.

the activity of bus companies in addition to their drivers. Brushing aside the larger systemic effects, Enoch et al. conclude that the testimonial evidence is still to be preferred to the statistical «because there is no similar incentive-corrupting effect to relying on individual evidence—even individual evidence that is probabilistically indistinguishable from the relevant piece of statistical evidence»⁴⁷.

This is wrong on both counts. On the one hand, the argument that the testimonial evidence is probabilistically equivalent but somehow qualitatively different from the statistical is another problematic inference from a weird hypothetical. On the other hand, bus companies as well as rodeo operators and attendees have many options in life and the risk of this form of liability is likely to cause them to take precautionary action. I discuss these two points in turn.

The hypothetical demonstrates the subtle fashion in which impossible or outlandish epistemological demands can be smuggled into weird hypotheticals⁴⁸. Recall that according to Enoch et al., what makes the witness testimony sensitive is that «had it not been a Blue Bus, she would have probably not testified that it was»⁴⁹. Well, actually, there is by hypothesis a 30% chance that she would have, but what does that mean? The authors are implicitly asking the reader to assume that the 30% error rate is composed of a random variable over all decisions in the (undefined) reference class, and therefore, the witness would «probably not» identify a red bus as a blue bus in the incident in question (the errors being distributed over all colors of buses). The point based on this implicit distribution is that the probability of any particular kind of error in the class of all bus identifications would be small, which

⁴⁷ ENOCH et al., 2019: 219.

⁴⁸ A sin that seems ubiquitous in this literature. My frequent co-author Michael Pardo falls prey to the weird-hypotheticals-that-make-peculiar-epistemological-demands dynamic in his otherwise quite sensible tracking of the relationship and pitfalls between safety and sensitivity on the one hand and reliability on the other. Here is one example from PARDO, 2018: 66:

Drug Weight: A criminal defendant is tried for possessing a large amount of an illegal substance. According to the criminal statute at issue, the prosecution must prove that the defendant possessed 500 grams or more of the substance. A chemist who sampled and tested the substance will testify as an expert that the amount seized was over a kilogram. The expert is well qualified, employed an acceptable and reliable methodology, and otherwise satisfies the criteria for admitting expert testimony. The chemist's process, however, tends to overestimate weight by a very small amount (say, by one gram or less).

Pardo uses this to illustrate safety through the use of possible worlds. The testimony is safe when it is nowhere close to whatever the element of the crime is, and otherwise it is not because even though the true amount is somewhat less than the statutory requirement the chemist would testify that it exceeds that amount. This seems to me entirely unnecessary. If one «knows» that the process «overestimate[s] by a» gram, then one «knows» what the true amount is. If the chemist's mistake is within a range, then again you know the range of the true weight. This hypothetical makes sense only if you already know what you need to know, and if you do, then one simply presents it to the fact finder. If one does not «know» the error, then the argument about safety does not run through. Notwithstanding this weird hypothetical, I doubt that his views are much at odds with the general thrust of this article. See Pardo 2005.

⁴⁹ *Ibid.*: 206.

is why they say the witness would «probably» not misidentify the bus. But note the smuggling in of knowledge. How do they know what kind of errors the witness is going to make? How do they know that this witness is making random errors (actually uniformly distributed errors) and simply not able to identify red buses under the conditions present, so that in fact most and maybe all of her error rate is the product of misidentifying red buses as blue?

This implicit smuggling in of an answer to a critical question of «knowledge» that assists the analysis is highlighted by considering the unrealism surrounding the claim that the witness is «70 percent reliable». It is all but impossible to decipher the meaning of that statement. Is the reference class the testimony of all witness under all conditions and the generalization is being applied to the witness? Does it refer to bus accidents at night? In what kind of roadway, weather conditions, and lighting due to the phases of the moon? Did someone do a study of this witness? Under what conditions and relating to what kind of incident? Just like 70% market share does not entail causing 70% of the accidents, 70% reliability does not entail that the witness in this particular case with respect to this particular harm is 70% reliable in reporting the truth about what she saw (again, whatever that might mean).

In any event, presumably Enoch et al. believe that we all live in a causal universe, and whatever causes the witness' errors will cause them every time, just like the statistic will always point to the Blue Bus Company. Summed over all possible cases, these two scenarios are structurally identical which means that they are identical from the point of errors even if not from the point of distribution of those errors. The witness' testimony is not more «sensitive»; rather, one simply does not know what causes her mistakes and thus one does not know how they will be distributed. One form of evidence is being preferred for no other reason than ignorance of its distributional effects⁵⁰. That is peculiar.

In comparing the statistical evidence of bus distributions to testimonial evidence, the bus distribution data (assuming it maps perfectly onto accident rates) is at least as good as the testimonial evidence, as the incentive argument properly understood makes clear. Just like rodeo operators can easily protect their income, bus companies can protect against statistical anomalies based on factors under their control⁵¹. These are complex and adaptive systems, in other words, and not simple

⁵⁰ ENOCH et al. (2012) may agree with this, although the matter is unclear. They suggest at times that perhaps there is a qualitative pertinent difference between sensitive and insensitive evidence, yet they also take the position that the law should not discard evidence even if statistical that contributes to reliable results. Well, how «reliable» is the statistical proffer as compared to the testimony, in their view? If what they mean is that there really is no interesting difference between the two forms of evidence, it would have been helpful, and saved a lot of time, for them to have just said that. For sustained treatment of the lack of pertinent epistemological distinctions between «statistical» and «nonstatistical» evidence, see PUNDIK, 2008, 2011.

⁵¹ Another demonstration (see *supra.*: n. 21) of the distance between some forms of theorizing and the naturalist approach is the response of David Enoch and Talia Fisher to a point analogous to the one

and static. For example, bus companies facing liability because of market share can engage in enhanced safety practices to reduce accidents and thus errors overall, and they can reduce errors at trial through such means as tracking the movement of their buses or installing video cameras. That, in turn, will affect the position and thus incentives of other bus companies, the combined effect of which almost surely will increase social welfare (fewer accidents in society and errors at trial)⁵². The real distinction between the bus and witness evidence has little to do with the incentive to engage in socially useful conduct but, once again, is simply that the statistical form carries its ambiguity on its face whereas the testimonial form does not. If preferring the latter to the former has any impact on overall social welfare, it is probably to decrease it⁵³.

in the text. They say: «Our answer to this critique is that while this effect may take place, as long as there are instances when it is counterbalanced by the chilling effect suggested by Sanchirico, both should be taken into account. And of course, one must also bear in mind that overdeterrence is also a problematic result» ENOCH and FISHER, 2015: 593, n. 130. That is exactly right and emphasizes the lack of utility for the jurist of the formal theorizing. The problem for the jurist is trying to identify in the real world what might maximize total social welfare. See ALLEN, 2015.

⁵² I may be accused here of engaging in my own a priori reasoning, but my defense is that it attempts to take into account real features of the world that bear on the construction of legal systems.

⁵³ To bolster their a priori incentive argument, they summarize Richard Posner's ridiculous a priori assessment of the Blue Bus hypothetical from the perspective of micro-economics. ENOCH and FISHER, 2015: 583, n. 83, say:

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.

Here is the argument from POSNER, 1999: 1510:

There is still another objection to allowing the bus case to go to the jury. If B, though responsible in fact for almost half the accidents, is never held liable and A is always held liable, A will have a big incentive to be careful and B little or no incentive to be careful. As a result, over time, more than half the accidents will be caused by B, increasing the error rate resulting from allowing juries to base decisions on the ratio of the companies' buses on the route in question. Eventually, A, having higher liability costs, will probably withdraw from the route; the rule on burden of proof will have created a monopoly!

The absurdity of this argument because of its neglect of the adaptive nature of the real world was pointed out by ALLEN and LEITER, 2001: 1526:

Our doubts about the value of Posner's form of a priori reasoning for explicating the law of evidence are well captured by this paragraph. Taking the argument on its own terms-terms which we reject, of course-the consequence Posner fears will occur only if A is remarkably stupid. In the world Posner is hypothesizing, all A need do is take at most three buses out of service. If A takes three buses out of service, the ratio of buses now favors A (it changes from 51/49 to 48/49), and B will henceforth be held liable for all accidents. Of course, another economic fear might now arise: B, also not being massively stupid, might respond by taking two of its buses out of service. Perhaps the prediction would now be that «bus reduction» would become recursive, eventually resulting in only one bus company with only one bus driven by a very careful driver, which would surely not serve the needs of the community! Therefore, the logic would go, allowing verdicts based on naked statistics is not economically sensible.

In any of the Blue Bus contexts, that «the law typically will not be willing to base a positive finding of fact—and certainly not liability—on just this kind of evidence, sometimes called statistical evidence»⁵⁴, may be right but for the wrong reason. The reason that the law would not be willing to hold the Blue Bus Company liable on the basis of this statistic has nothing to do with the claim that the statistic is insensitive or incentive crushing. Rather, it is that the statistic is just another example of lousy evidence. The fact that the Blue Bus Company owns 70% of the buses does not entail that there is a 70% chance that the Blue Bus Company caused the harm in question. The Blue Bus Company may own 70% of the buses but have extremely effective safety measures in place; they may have well-seasoned drivers with impeccable driving records; they may not even have buses that frequently visit the location of the harm at the time in question. Actual fact finders could easily regard this statistic as being insufficient to ground liability because these relevant possibilities remain unaccounted for by the statistic about ownership.

Once modified with an element of realism, the hypothetical no longer demonstrates any difficulties with statistical or market share evidence that distinguish it from testimonial evidence. A litigator would know that the relevant statistic would be a risk assessment of how likely it is that a blue bus rather than a red bus caused the harm in question, obviously enough, and construct a case on that basis. That might include market share data but also would include data regarding the records of each company's drivers, statistics about the frequency with which each company's buses visit the location of the harm, statistics about the effectiveness of each one's safety measures, figures about similar harms caused by each company in the relevant location, and so on. The «and so on» is centrally important. If sensitivity theorists are correct, then the number of variables included in the risk assessment should not matter because all statistics will be equally insensitive and incentive eradicating. Including more and more relevant statistical evidence ad infinitum—making this an extremely thorough assessment—should not influence fact finders' judgment about the case, nor evidence theorists' assessment of the evidence.

But does it? Suppose that, instead of the single statistic about market share, a litigator presents an extremely thorough, multi-variable assessment of the risk posed by the Blue Bus Company and Red Bus Company that reveals that there is a robust 70% likelihood that the Blue Bus Company and a 30% chance that the Red Bus

As was also pointed out by ALLEN and LEITER, 2001: 1526, n. 106, taking such a priori thinking seriously and assuming really smart bus companies could lead to the prediction of no liability ever for accidents. The bus companies would simply have to run the same number of buses so that the probability of both would always be .50, which does not meet the preponderance standard, and so the plaintiff would always lose. Such logical outcomes indicate a serious problem with the logic. As Allen and Leiter concluded: «Neither a monopoly nor the essential elimination of the industry would result from allowing probabilistic verdicts. If such verdicts began to accumulate inaccurately against bus companies, they would invest in precautions, many of which are cheaply available».

⁵⁴ ENOCH et al., 2012: 197.

Company caused the harm on the day and at the time in question. Would a sensible legal system still regard the eyewitness evidence as more worthy than the statistical evidence? Of course not. When it comes to the actual trial setting, a complex body of evidence consisting of several pieces of relevant and interrelated statistical information that converges in a way that reveals a 70% likelihood that a blue bus caused the harm is at least as good, if not better, evidence as a body of evidence that includes only the testimony of a single witness who is known to be 70% reliable in identifying the color of the bus that caused the harm. The market share evidence presented is insensitive, while the testimonial evidence is apparently sensitive according to these theorists—but their evidentiary value is either the same or the statistical evidence is better. The relative value of statistical and testimonial evidence does not map on to considerations about sensitivity and deterrence (as the sensitivity theorists conceive it), as the hypothetical is designed to show.

The real difference between statistical evidence and testimony lies in the trial convention of assuming reliability of witnesses until credibility is attacked. This convention obviously does not apply to statistical evidence, as its reliability is immediately apparent (under certain assumptions to make the hypothetical comprehensible). Once a witness' credibility is attacked, a «risk assessment» highly analogous to what might be done of bus companies can occur with the opposing party pointing out the «risks» in the testimony. The risk assessment of the testimony involves identifying reasons why the testimony may be unreliable. In the case at hand, this may mean pointing out that, when the harm occurred, it was dark outside or that the witness was far away from the scene—thus impairing her ability to correctly identify the color of the bus. Or, perhaps, the witness herself points out the risks in her testimony by admitting that she may be misremembering the color of the bus.

It is true that the two forms of evidence come in different frames, and it is up to the parties how to react. The parties are perfectly capable of contextualizing both kinds of evidence. With statistical evidence, it is up to the parties to demonstrate its pertinence to the case at hand while with testimony it is up to the parties to demonstrate its problems. That is the real difference between the types of evidence, and not considerations of sensitivity and incentives. In doing so, they may very well want to argue that certain evidentiary contexts may involve more sensitive conditions, and that may very well be persuasive to a fact finder. Were exegeses of sensitivity directed to advice about the presentation rather than the admissibility or categorical treatments of types of evidence, they would be useful. They are not, however, explanatory of the structure of American legal systems. Once again, weird hypotheticals have led astray.

I will give one last example of unrealistic and weird hypotheticals leading astray. In a recent essay, Martin Smith extends his analysis of «normalcy» from the episte-

mological realm⁵⁵ to legal systems⁵⁶. He, like others, is trying to defend the intuition that statistical evidence is different from non-statistical evidence and doing so by a search for the key distinction⁵⁷. His key is «normic support». The central idea is that «a body of evidence E *normically supports* a proposition P just in case the circumstance in which E is true and P is false would be less normal, in the sense of requiring more explanation, than the circumstances in which E and P are both true...»⁵⁸. Smith is quite cautious in his ultimate conclusions, conceding for example that the DNA cases present a «clash with the normic standard» that «could be seen as a reason for resisting the standard and seeking an alternative solution to the legal puzzle of statistical evidence»⁵⁹. And he concludes in a catholic vein that «If we are open, from the outset, to a range of different ways of managing uncertainty, the legal puzzle of statistical evidence need not strike us as a “puzzle” at all»⁶⁰. Perhaps this means that he accepts that normic support is just one of many tools to use in thinking about evidence, which would not be controversial. All of these different efforts to find the solution or key to the problem of legal proof fail as general explanations but are enlightening perspectives on proof that could very easily play a role in a particular case and are clearly encompassed by the field of evidence. An advocate would find studying all these different approaches useful indeed.

Although the central idea of normic support may be a useful heuristic, what is not terribly useful is the way Smith goes about explicating it, and once again there

⁵⁵ SMITH, 2010.

⁵⁶ SMITH, 2017.

⁵⁷ He worries about the distinction between «purely statistical evidence» and some other form, see *ibid.*: 1195 n. 3 and 1213 n. 19, but that is an unnecessary distraction. It is not clear what «purely statistical evidence» could possibly be, or at least what the alternative is that he worries about, and he does not clarify the matter. He raises a question whether a cold hit DNA case is «purely statistical in nature». *Ibid.*: 1213 n. 19. If it is not, then it is doubtful whether the set of «purely statistical cases» has any members, in which case the analysis of that set is rather pointless. On the other side of the coin, there will always be testimony about an evidentiary proffer that contextualizes it in some way. This is another example of the distracting quality of weird hypotheticals. They are offered as if a certain set of facts were stipulated, and thus incontestable, but that would never be the case in reality. There would have to be evidence, contestable evidence, of the operations of the rodeo for example in the gatecrasher scenario and it would have to be explained why that evidence allowed the inference to certainty that there were exactly 1,000 people in the stands of whom only 499 (or whatever) paid, an explanation that would never carry the day. This applies to DNA as well. See n. 74, *infra*. This, it should be noted, is an example of the general problem of impossible epistemological demands that these hypotheticals make. Certain things are just taken as true rather than as a matter of fallible evidence and inference, yet that is a world that does not exist. In the gatecrasher hypothetical, it would require that the defendant stipulate to all the facts so that they are not then a matter of proof, but then in addition somehow being unable to testify, or offer any other form evidence, that she actually bought a ticket. This combines weird hypotheticals and impossible epistemological demands into a truly remarkable brew the only implication of which is as a testament to the creativity of the human mind.

⁵⁸ *Ibid.*: 1208.

⁵⁹ *Ibid.*: 1214.

⁶⁰ *Ibid.*: 1216.

are the problems of weird hypotheticals, not attending to how trials actually function, and in this case not just the demand for too much knowledge but also subtle shifts in the epistemological demands that bake in the answer he is proposing.

Here is the hypothetical he uses to get the analysis off the ground:

Suppose an electronics store is struck by looters during a riot. 100 people walk out of the store carrying televisions, while the transaction record at the cash register indicates that only one television was paid for, though no receipt was issued. Suppose Joe is apprehended carrying a television from the store, but we have no other information about him. Should Joe be prosecuted for theft? Should he be convicted? Should he be punished? If our only evidence against Joe is that he carried a television from the store and that 99 out of the 100 televisions carried from the store were stolen, while one was legitimately purchased, then, by launching into such actions against Joe, we would do him a serious wrong⁶¹.

Problems begin at the beginning. What could it possibly mean that «the cash register indicates that only one television was paid for, though no receipt was issued»? Who would buy an expensive television set and not get a receipt? What television store would allow such a transaction? And the recurring problem of impossible epistemological demands—without the presentation of fallible evidence, how does anyone know how many people walked out with a stolen television set? Did the cash register that indicated only one television was paid for also keep track of how many were stolen? How do we know it is never wrong? Perhaps all this is fighting the hypothetical, but a scenario involving a riot and looting but with a clerk calmly ringing up one customer for their television purchase while chaos ensues around them is a far cry from the scenarios the legal system ever encounters. As is the requirement of infallible knowledge of the surrounding facts.

Nevertheless, once again let's take it at face value. It is false that the only evidence is what is presented. If one television set was purchased during the riot, someone had to be at the register unless we make the even weirder assumption that stores allow self-checkout with television sets but also don't video the encounter (the video would substitute for identification by a witness). Perhaps now we have to add the additional assumptions that the video wasn't working; the customer was honest enough that even though people were streaming out of the store with television sets, they decided to scan it at the self-checkout line; they were unconcerned with the chaos around them; and they did not care about a receipt, having apparently no concerns about proof of ownership in case something was wrong with the television.

Even if a series of adjustments is made to the hypothetical that make it weirder still but preserve the supposedly purely statistical nature of the case, there still remains further evidence, as there always will be in criminal cases—the accused themselves. The accused can testify to whatever they want, and thus no «serious wrong» would be done by charging them with a crime they almost surely committed and seeing how they defend. And if they did not make a satisfactory response, testify, or

⁶¹ *Ibid.*: 1196.

offer any other defense, a jury or judge would, as they should, convict and no appellate court would reverse it. Regardless of hallowed conventions that say that defendants do not have to defend, in fact they do. Everyone in the courtroom knows that at least one person present knows the truth, and that is the defendant. The facts of this hypothetical do more than provide a plausible story of guilt; it is an overwhelming case. The failure to testify in a hypothetical like this would confirm the truth of the prosecution's case, no matter what instructions are given about drawing no inferences from the exercise of Fifth Amendment rights.

The problems go deeper. How would any evidence of the facts of the hypothetical except that there had been rioting, looting, and Joe was caught carrying a TV set, get before the fact finder? Someone would have to testify to the sale of the TV set, a person who could then be asked if they recognized Joe as the person who purchased it. Ok, maybe this witness would just authenticate records kept by the cash register, or self-checkout machine, or whatever. Still, Joe was arrested, which would have been preceded by an on-the-street encounter between the police and Joe that would have entailed on-the-scene questioning. A refusal to answer would be admissible, as would any obviously concocted story Joe told, all confirming guilt. If Joe asserted he had paid cash (any other form of payment would create a record whether a receipt was issued or not) the police would ask where the money had come from, and other questions seeking to confirm or deny the truth—such as how much the television cost, a question a guilty person could probably not answer. To the jurist (at least to me), it is entirely unclear what to make of these matters precisely because they are so divorced from reality (although I will say that charging the hapless Joe with a crime is about as far from a «serious wrong» as is imaginable).

Like just about everybody, Smith relies on the Blue Bus hypothetical but there is a gap in its construction that illustrates the curious epistemological demands being made that generate the underlying intuition that in his view creates the need for some explanation like the notion of normic support on offer. He hypothesizes «90% of the buses operating in the area on the day in question are Blue-Bus buses»⁶² and a witness whose probability of testifying accurately is unstated⁶³. Usually the hypothetical assumes the same level of fallibility as the proportion of buses. Not asserting an error rate changes the hypothetical from prior uses of it in a subtle way. His point is that «it could just so happen that the bus involved was not a Blue-Bus»⁶⁴ but it could not «just so happen» that «the testimony was wrong» (which is his point in the television set hypothetical as well). There would need to be an explanation of what went wrong with the witness, and thus the witness, unlike the statistic, offers normic support for the verdict.

⁶² *Ibid.*: 1202.

⁶³ *Ibid.*: 1196.

⁶⁴ *Ibid.*: 1196.

By not asserting an error rate, Smith is exploiting the natural inclination and trial conventions discussed earlier, namely, that witnesses are to be believed unless a reason is given not to believe them. In a state of ignorance about the witness's testimonial capacity, a reasonable person has no reason to think that a witness is going to make the rather grievous mistake of identifying a red bus as a blue bus. He is thus probably right that «intuitions» would reflect that there is a difference between the two forms of evidence. Until someone introduces a concern about error, the comparison is between a known rate of error and a nonexistent rate of error, and of course the two look different. That outcome is baked into the hypothetical from the beginning. This is not instructive on the difference between statistical evidence and something else but simply the obvious result of trial conventions (don't litigate what does not need to be litigated).

Smith might object that he has not suppressed the possibility of error because part of his hypothetical involves a general recognition of witness fallibility and he goes so far as to say: «If we were forced to come up with some numerical estimate of how likely it is that the bus really was a Blue-Bus bus, given the witness testimony, it's doubtful that we would go quite as high as 90%—that would seem overly trusting. But 90% is, of course, precisely how likely it is that the bus involved was a Blue-Bus bus, given the statistical evidence about which we were so apprehensive»⁶⁵. But fact finders are not «forced» to come up with a number; nor is someone reading the hypothetical. As the hypothetical stands, it is simply exploiting the unstated assumption of witness credibility, which stands until controverted.

Smith, to the contrary, might say that is exactly his point; one case requires an explanation that another does not, but there is a looming problem here. Once the possibility of error is introduced, the two cases become essentially identical; the possibility of error will have a statistical distribution, even if unknown, structurally identical to the statistical evidence. There could be multiple explanations for the 10% inaccuracy rate, or whatever it is, but for purposes of comparing these cases it does not matter what they are. Suppose there is evidence that a witness lacks credibility and it is asserted that the witness has been known to lie. Still the question remains whether the witness is testifying truthfully in this case, and that is the same type of question as to what color bus hit the plaintiff. Unless we live in an indeterminate world, there will be reasons why the witness is untruthful in different sets of cases. Perhaps at all times like the one in question, after a night out with friends and distracted about thinking of her aging mother, the witness's identification is accurate in 90% of the cases. Only if one can pursue the witness's psychology adequately to get to certainty about what happened—a reference class of one, in other words—would this not be

⁶⁵ *Ibid.*: 1196. Why he thinks that this is overly trusting is a mystery. If a single unimpeached (on any ground) eyewitness testified to the color of a bus, that would not only suffice for a verdict absent evidence indicating that it was virtually impossible for that color of a bus to be in the vicinity, but if asked fact finders would probably indicate that they are close to certainty about what happened.

true, and reference classes of one at trial are rare to say the least. Thus, it «could just happen», indeed it does just happen, that the testimony is wrong 10% of the time. On the assumption that 90% of the accidents are caused by blue buses, the two are identical, which is partially explanatory as to why the «intuition» that these two cases are different does not describe American legal systems (see below).

Curiously, Smith seems to admit as much⁶⁶, but now there is an even larger looming problem. It is a rare case in which the possibility of error is not introduced. If there were no questions of credibility and no conflicting evidence, there would be no trial. Thus, the standard, virtually universal, condition at trial is that witness testimony does *not* provide normic support in his terms—in his view it would be identical to statistical evidence in the relevant ways. What Smith is proposing as a general distinction between types of evidence would seem to apply to virtually no evidence at trial⁶⁷. It begins to sound like a broken record, but the real distinction is not between evidence providing or not providing normic support but instead between evidence that carries its ambiguities on its face and that which does not. It is not that witness testimony provides «normic support»; rather, it is that we pretend that it does until someone wants to argue about it, and in virtually all cases the parties do want to argue about it. That is why they are expending their resources on a trial.

In fact, the similarities between the two types of evidence press even deeper and run in both directions. The hypothetical is designed to have the reader compare an immutable and inscrutable statistic—90% Blue Buses in the area—to a malleable and quite scrutable assertion about witness credibility. But just like the witness credibility evidence might call for further explanation, so too does the statistic—as laid out in detail above. Just like one might probe further into whether or why the witness made a mistake, one could probe further into the variables that predict bus accidents. Neither bus accidents nor witness errors in fact «just happen» in a causal world; there are reasons for both that sometimes can be explored further and sometimes cannot. If a particular witness, or witnesses generally, or . . . (note the frustrating ambiguity in specifying the reference class) give accurate testimony 90% of the time, it is no more abnormal for a witness to be mistaken than it is for a Red Bus to cause one of its 10% share of accidents (if that is what the hypothetical is supposed to be)⁶⁸. What deprives

⁶⁶ *Ibid.*: n. 14.

⁶⁷ For an analogous argument to that in the text, see DI BELLO, 2019. For a critique of Smith's arguments that is at least sympathetic to some of the arguments here, see MICHAEL BLOME-TILLMANN (2020).

⁶⁸ Another recent effort to explain juridical proof among other topics is Sara Moss's *Probabilistic Knowledge*. Her weird hypothetical, another twist on gatecrashers, is:

For instance, suppose you are the factfinder in Gatecrasher. The plaintiff proves that an arbitrary person at the rodeo is more likely than not to have climbed over the fence. Meanwhile, the defendant claims that he is an exception to this statistical generalization. He denies that it is more likely than not that he climbed over the fence, insisting that he did not climb over the fence at all. The statistical evidence of the case may justify your having .501 credence that the defendant is liable. But on the basis of this statistical evidence, your credence does not constitute knowledge. You do not know that the defendant is exactly as likely to have climbed over the fence as

an arbitrary person at the rodeo, because you cannot rule out a certain possibility that is inconsistent with this content—namely, that the defendant is an exception to the statistical generalization presented by the plaintiff. As a result, you do not know that it is more likely than not that the defendant is liable for trespass. According to my account of legal proof, it follows that the plaintiff has failed to prove that the defendant is liable by a preponderance of the evidence. Moss, 2018: 210-211.

Weirdness number one lies in the suggestion that an innovative theory is needed to explain this hypothetical. The defendant has first-hand knowledge and testifies that he bought a ticket (presumably that is what «claims» means). Unless quite seriously impeached (of which there is no evidence in the hypothetical), no one would find the defendant liable on these facts. There is nothing to explain.

Weirdness number two is that this argument implies that any time a defendant testifies to a defense, one cannot have whatever Moss thinks is necessary for a verdict. This does not just apply to cases involving statistical proffers of liability. As laid out in the text, a competent empirical study of blue bus accident causing likelihoods can be just as good or better justification for a verdict than an error prone eyewitness. If the equivalent or stronger evidentiary basis is defeated by a single defensive claim, so too in a rational world should be the weaker. In the nonstatistical case, perhaps evidence undermining the claim could be presented, but then it would have to be explained why equivalent evidence is not available in the statistical case. Once in a while that may be true, but categorically it is plainly false. Impeaching evidence is not limited to witnesses providing substantive evidence.

Weirdness number three lies in the implicit assumption that «knowledge», whether probabilistic or not, is the goal of a trial. It is not, at least not in any sophisticated philosophical sense. Instead, the goal is a common sense, naïve realist, folk psychological best effort to discern what happened under conditions of uncertainty. In an effort to contradict the previous sentence, I searched the pattern jury instructions of all 50 American states for references to knowledge as a criterion for a verdict. No state civil or criminal pattern jury instruction available online asks a jury to come to «knowledge». Instead, they call for the party with the burden to persuade to the appropriate standard. See e.g. RAJI (CIVIL) 6th Standard 2 («On any claim, the party who has the burden of proof must persuade you, by the evidence, that the claim is more probably true than not true.»). Alternatively, they call for the jury to weigh the evidence and decide which proposition is more likely true than the other. See e.g. IDJI 1.20.1 («When I[...]use the expression “if you find” or “if you decide”, I mean you must be persuaded that the proposition is more probably true than not true»). Jury instructions in criminal cases follow the same common sense approach. See e.g. 1 IN Pattern Jury Instructions Criminal Instruction No. 1.15 (2020) («The burden is upon the State to prove beyond a reasonable doubt that the Defendant is guilty of the crime(s) charged. It is a strict and heavy burden. The evidence must overcome any reasonable doubt concerning the Defendant’s guilt. But it does not mean that a Defendant’s guilt must be proved beyond all possible doubt.»). The only reference to knowledge found reminds jurors that «There are very few things in this world that we know with absolute certainty, and in criminal cases the law does not require proof that overcomes every possible doubt». Utah Pattern Jury Instruction.

Moss reiterates her knowledge argument in Moss, S. (2021), where she defends her knowledge account from what many if not all legal observers think is ineluctable uncertainty (including, I would say, uncertainty of the probability of possessing knowledge) by claiming that in criminal cases so long as jurors put aside «unreasonable doubts» then they can have «knowledge» that a defendant is guilty as charged. She operationalizes her account with, of course, another weird hypothetical involving a «wacky juror» who is under the influence of apparently equally wacky conspiracy theories, and other «far-fetched possibilities that jurors ought to set aside, such as possibilities involving evil mind-controlling demons». This is decidedly unhelpful. Surely there is universal agreement, quite independent of any particular epistemological commitments, that it is not a good idea to have wacky people and those who believe in evil mind-controlling demons serve as jurors or judges. More importantly, no mechanism to sort out wacky from not-wacky people is provided, neither is either the extension nor intension of the set with mind controlling demons as a member identified. How this argument advances any particular epistemological approach thus is opaque; rather, it seems to be an ad hoc move designed to insulate

the bus statistics of normic support deprives all evidence of it as well once one looks beneath the surface. It thus is not much of an explanation of the distinction between statistical and non-statistical evidence, nor more generally of juridical proof.

Put weird hypotheticals aside. What if it were true that the only evidence reasonably available, and it was reliable (a key point), is that there is a slightly better than .5 probability that the defendant is liable? The answer in the United States is that it has long been true that if there are no other sources of information, the statistical evidence meets certain conditions of reliability, and deciding the case one way or the other would not create cost/benefit justified incentives for socially useful behavior that outweigh the value of more accurate verdicts, verdict would likely, even if not always, be given for the plaintiff⁶⁹. As has been pointed out for some time⁷⁰, the Smith case is not in fact a case about statistical evidence but instead about lousy evidence⁷¹. Pace this «intuition» that courts are skittish about statistical evidence, based in part on a 75 year old case, in general the courts are receptive to statistical evidence and allow verdicts on that basis. In 2000, the Federal Judicial Center published a manual on the use of statistical evidence. The authors sum up the judicial experience with statistical evidence by saying that «Statistical reasoning can be crucial to the interpretation of psychological tests, toxicological and epidemiological studies, disparate treatment of employees, and DNA fingerprinting», and «this list could easily be extended»⁷². For example:

—Employment discrimination cases are often heavily if not exclusively statistical, as are antitrust cases⁷³.

an epistemological argument from its inherent impossible epistemological demands. Her argument is also inconsistent with the legal system she is trying to explicate. She appears to be arguing that the standard for a criminal conviction requires that the fact finder reach knowledge by putting aside unreasonable doubts and finding that all other possibilities of innocence have been ruled out. This is precisely what the standard was in the United States for convictions based on circumstantial evidence, but not otherwise: «where the Government's evidence is circumstantial it must be such as to exclude every reasonable hypothesis other than that of guilt». The Supreme Court of the United States rejected that standard as too stringent half a century ago in *Holland v. U.S.*, 348 U.S. 121 (1954), and virtually every state has now followed suit. The most recent court of last resort to deal with this issue is the Florida Supreme Court in *Bush v. State*, No. SC18-227, decided May 14, 2020. For a better explanation of reasonable doubt, see ALLEN and PARDO, 2019a.

Recently a number of efforts focusing on knowledge in addition to Moss' have attempted to identify what justifies the «intuition» that something is wrong with naked statistical evidence as exemplified by weird hypotheticals, yet there is no agreement on what that justification is. See, e.g., BLOME-TILLMANN, M., 2017; LITTLEJOHN, C., 2017; LEVANON, L., 2019.

⁶⁹ See, e.g., *Kramer v. Weedhopper of Utah, Inc.*, 490 N.E. 2d 104 (Ill. Ap. Ct. 1986). It would not be a surprise if some courts would want more than a > .5 burden to be met, but others would not. As I say, it is a large, sprawling system.

⁷⁰ ALLEN, 1987: 429, n. 67. See also ALLEN, 1991.

⁷¹ See, e.g., LEMPERT, 1986.

⁷² NATIONAL RESEARCH COUNCIL, 2000: 85.

⁷³ *E.E.O.C. v. O & G Spring & Wire Forms Specialty Co.*, 38 F.3d 872, 876 (7th Cir. 1994) («Reliance on statistical evidence by no means diminishes the plaintiff's obligation to prove discriminatory

—In criminal cases, there are cold hit DNA convictions⁷⁴.

—In *Daubert v. Merrill Dow Pharmaceutical, Inc.*⁷⁵, the Supreme Court included control studies with known error rates (statistical evidence) as part of the foundation for expert testimony, and in the follow-up to the Supreme Court case, the 9th Circuit held that liability could be based on a relative risk of 2/1—a direct statistical measure of liability⁷⁶.

—Class actions often depend on statistical evidence⁷⁷.

—Commercial protection cases⁷⁸.

—Intellectual property cases⁷⁹.

—Real estate cases involving property valuations⁸⁰.

—Tax cases⁸¹.

intent—but in some cases, statistical disparities alone may prove intent.») In re Sulfuric Acid Antitrust Litig., 446 F. Supp. 2d 910, 922 (N.D. Ill. 2006) (Rejecting defendant's motion to reconsider allowing plaintiff's expert testimony based on statistical sulfuric acid price data in an antitrust class action).

⁷⁴ For a discussion see ROTH, 2010. Enoch and Fisher attempt to explain away DNA evidence and cold hit cases on the basis of the enormously high probabilities that can be reached. Standing alone, those are the wrong probabilities. They need to be qualified by among other things the possibility of an innocent explanation of a suspect's DNA being found on the scene, police planting evidence, lab error, and the risk of adulteration of the sample. ENOCH and FISHER, 2015.

⁷⁵ 509 U.S. 579 (1993). See JONES and HAGTVEDT, 2002.

⁷⁶ *Daubert v. Merrell Dow Pharm., Inc.*, 43 F.3d 1311, 1320–21 (9th Cir. 1995) (Daubert II) (Because the background rate of limb reduction defects is one per thousand births, plaintiffs must show that among children of mothers who took Bendectin the incidence of such defects was more than two per thousand). In *Cooper v. Takeda Pharm. Am., Inc.*, 239 Cal. App. 4th 555, 593–94, 191 Cal. Rptr. 3d 67, 98 (2015), the Court made clear the statistical implications of the Daubert II holding:

When the relative risk is 2.0, the alleged cause is responsible for an equal number of cases of the disease as all other background causes present in the control group. Thus, a relative risk of 2.0 implies a 50% probability that the agent at issue was responsible for a particular individual's disease. This means that a relative risk that is greater than 2.0 permits the conclusion that the agent was more likely than not responsible for a particular individual's disease. [Reference Manual on Scientific Evidence (Federal Judicial Center 2d ed. 2000) (Ref. Manual),] Ref. Manual at 384, n. 140 (citing Daubert II).

⁷⁷ See, e.g., *Kurtz v. Kimberly-Clark Corp.*, 414 F. Supp. 3d 317 (E.D.N.Y. 2019) (Allowed hedonic regression analysis to show market wide price premium with respect to allegations of false labeling of products.)

⁷⁸ *United States v. Dish Network LLC*, 75 F. Supp. 3d 916, 939 (C.D. Ill. 2014) (Allowing statistical analysis of random sampling of call records to prove claim that satellite television provider violated the TCPA.)

⁷⁹ *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 423–24 (1984) (Relying on statistical evidence to hold that the Betamax is capable of non-infringing uses and thus acceptable for in-home use.)

⁸⁰ *CSX Transp., Inc. v. Bd. of Pub. Works of State of W.Va.*, 312 F. Supp. 2d 839, 851 (S.D.W. Va. 2004) (relying on statistical Chi-square analysis to determine true commercial and industrial property market value in West Virginia).

⁸¹ *Kikalos v. United States*, 308 F. Supp. 2d 902 (N.D. Ind. 2003) (Allowed government witness opinion, based on percentage mark-up method and government statistics, in liquor store owner's suit to recover overpayment of taxes).

—Market share liability⁸².

—Lost chance litigation⁸³.

The list above is disparate in various ways, but each example involves deciding a necessary element in a case to the extent possible based on statistics⁸⁴.

What some see as difficulties with American legal systems or puzzles that need resolution that they are trying to highlight with weird hypotheticals and solve with formal theories and algorithms do not seem very disturbing viewed through a naturalistic eye. One sees those systems from beginning to end channeling the parties in both civil and criminal cases to offer competing explanations for what happened, how it happened, and once in a while why it happened. The parties provide the evidence that triggers deliberation and suffer the consequences if the evidence is inadequate in the view of the fact finders. The fact finder is left to ponder over those competing explanations by reference to their own knowledge and experience, and perhaps to construct its own explanation of the underlying event. Fact finders employ their natural reasoning processes (including where appropriate probability), and a conclusion is reached as to the verdict⁸⁵. This is all facilitated by liberal rules of admissibility that make virtually any plausibly relevant evidence admissible; the absence of formal rules of sufficiency, which is left to the final judgment of the fact finder almost always; and little concern over what might be labeled «statistical» evidence. There are no algorithms operating here, nor any impossible computational demands made⁸⁶. The system as a whole and fact finders' deliberations are organic

⁸² *Sindell v. Abbott Laboratories*, 26 Cal. 3d 588 (1980).

⁸³ See FISCHER, 2001.

⁸⁴ The objection might be made that in many of these cases there is other evidence in addition to the statistical evidence, and thus that I may misunderstand the argument that focuses on naked statistical evidence. To the contrary, this highlights the unrealism of the discussion. There is no trial involving only «statistical evidence». See n. 57, *infra*. Someone has to present the evidence, authenticate whatever sources are being relied upon, face cross-examination and rebuttal, and so on. If there is a problem here it lies in statistical evidence being the necessary basis of decision. American legal systems have no difficulty with that phenomenon.

⁸⁵ See ALLEN and PARDO, 2019a. Given the actual state of American trials, it could not be otherwise. Evidence comes in serially, fact finders are instructed not to form conclusions or deliberate on the evidence until all the evidence has been heard, and the number of discrete pieces of evidence would overwhelm any computational capacity humans possess. See ALLEN, 1997. These points are true regardless of whether one is talking about objective or subjective Bayesianism. FRANKLIN, 2011. Moreover, there are no virtually no «objective» data points in a typical trial, and no one cares about the truly «subjective» beliefs of fact finders—their job is to determine the facts in a disinterested fashion. These debilitating problems with standard probability theory are some of the reasons behind the paradigm shift that is occurring from the conceptualization of juridical proof as probabilistic to explanatory.

⁸⁶ Including that trials are a search for knowledge, if I may refer to that as an algorithm; rather, there is an effort to construct a reliable decision making process subject to policy constraints. See *supra.*: n. 68. The two primary uses of «knowledge» (other than as elements of offenses and causes of action) are FRE 602's reference to «personal knowledge» and FRE 702's reference to expert knowledge. These references have quite a complicated relationship to «knowledge», as they authorize witnesses to testify in

and fluid rather than shoehorned into «probability», «evidential probability», «probabilistic knowledge», «belief states», «belief functions», «fuzzy logic», or whatever the theory of the day is. If there is shoe horning, it occurs as a result of party choice, not ontological necessity, and even then the fact finder is free to disregard the party's efforts and to conclude that some other explanation is the most plausible on offer.

The empirical description may be apt, but its moral foundation lacking, and intuition mining of weird hypotheticals that ignore the actual operation of legal systems and make impossible epistemological demands may so demonstrate. Perhaps, but there is one more reason for doubt, which ends this article on the same solipsistic note that it began. This literature has an annoyingly imperialistic capacity to refer not just to the various authors' intuitions, but «our» and «ordinary»⁸⁷ intuitions. Many of these do not coincide with mine⁸⁸, and it is interesting where the ones that do not are located. They are usually found in a priori analysis that involves searching for the single key or algorithm that resolves problems rather than engaging in empirical work⁸⁹. Testing intuitions against a naturalistic view of the process informed by an empiricist's slant, there is no conflict at all⁹⁰. American legal systems are not designed to facilitate intuition mining, guarantee that no wrongful decision favoring either party will be made, nor involve a search for «knowledge» in any but the most casual

the same trial to inconsistent propositions, both of which cannot be true, and therefore both of which cannot be knowledge under any conception of knowledge that requires truth.

⁸⁷ See, e.g., Moss, 2018: 217.

⁸⁸ For example, Moss seems to embrace the «intuition» that «one should avoid convicting innocent people at any cost». Moss, 2018: 215. If she does, I demur. Indeed, that «intuition» is a perfect example of the distance between formal reasoning and empirically informed inquiries. And common sense for that matter. There are on average over a million felony convictions each year in the United States, Bureau of Justice Statistics at <https://www.bjs.gov/index.cfm?ty=tp&tid=233>, and probably over 10 million misdemeanors convictions. See NATAPOFF, 2018; NATAPOFF, 2020: 159-162. Having studied the criminal justice process and the field of evidence for over forty years, my conclusion, for what it is worth, is that the only way to stop the convictions of innocents is to stop convictions of everybody, which would lead to social disaster. Such an intuition is an example of the failure to account for the deadly dilemmas of governing. ALLEN and SHAVELL, 2005; LAUDAN and ALLEN, 2008.

⁸⁹ See, e.g., the list of such views in Moss, 2018: 208-210. This also explains why criticisms that some of the explanations of relative plausibility do not meet what might be called philosophical standards of conceptual analysis miss the point. See, e.g. PUNDIK 2008: 36-40. As I have been arguing throughout this article, the methodology of conceptual analysis is not terribly helpful for explaining juridical proof, which requires instead a realistic embracing of complexity and adaptability. This is not to say that conceptual analysis may not serve other purposes in explicating aspects of juridical proof. As I have noted, for example, some of what are offered as general explanations could in fact be enlightening about intellectual tools to be employed along with others in structuring and deciding a case.

⁹⁰ That, of course, may appear to some as circular. I would draw a difference inference—that many of the a priori theorists might do well to spend more time focusing on whether their and what they think are «ordinary» untutored intuitions are justified rather than on how to justify them. Obviously, much of this article is compatible with much of social epistemology, as the original ALLEN and LEITER, 2001, article noted. On social epistemology, see FULLER, 2002, and GOLDMAN, 1999.

sense⁹¹. American legal systems give parties a civilized forum to get a definitive resolution of their dispute that could emerge from any corner of the sprawling human domain by a hopefully disinterested and unbiased fact finder that operates by reference to minimal, not maximal, standards of rationality, precisely because a priori one has no idea what a dispute may entail. It is up to the parties to construct the framework for the disputation, which is done of necessity idiosyncratically. Holding trials to the standards of most formal theories of evidence, proof, or knowledge would make them unworkable, and the costs of that outcome would greatly exceed its benefits, generating the very harms the «normative» analysis is attempting to minimize.

A search for «the» singular key to whatever problem is under investigation probably explains the use of reductive hypotheticals like the ones discussed. The reductionist efforts miss that American legal systems are complex adaptive processes, which is also why individual data points or the intuitions of particular individuals do not count for much. In a closed deductive and static system, inconsistencies are a problem. In complex dynamic ones, they are the order of the day. The fascination with *Smith v. Rapid Transit* is a good example. That is a single, old case, subject to various interpretations. Even if when decided it stood for the proposition that «statistical» evidence is insufficient for a verdict, so what? Other cases plainly say the opposite. In the Amazon rain forests, species come and go, just like in the American legal systems. There are not uniform intuitions that are the uncontested starting points for analysis. There is complexity everywhere. Even cases that may seem alike when presented in the context of weird hypotheticals may, in a real-world context, have quite different implications. The only way to know is to go look⁹².

In any event, the academics' intuitions concerning juridical proof are certainly not borne out by the real world. In fact, the resilience of this «intuition» is especially surprising⁹³ given the plentiful and directly contradicting evidence that statistical ev-

⁹¹ See n. 68, *supra*. Others have suggested this as well. ENOCH et al., 2019: 35.

Here is David Lewis, taking for granted the point we argue for in the text here (and elsewhere):

What if some far-fetched possibility is called to our attention not by a sceptical philosopher, but by counsel for the defence? We of the jury may wish to ignore it, and wish it had not been mentioned. If we ignored it now, we would bend the rules of cooperative conversation; but we may have good reason to do exactly that. (After all, what matters most to us as jurors is not whether we can truly be said to know; what really matters is what we should believe to what degree, and whether or not we should vote to convict). LEWIS, 1996: 560.

⁹² For an early example in the legal literature of discontent with intuition mining as the primary epistemological method, see ALLEN, R.J. and B. LEITER, 2001. See also ALLEN and LEITER (2003): 892: «The naturalistic turn in epistemology came about in no small measure because of the failure of formal theories (elaborated intuitions) to achieve the goal of justifying assertions of knowledge. Justification, in other words, seemed not to be a matter of just consulting one's intuitions, and the effort to justify assertions of knowledge in that manner came to be perceived as the problem rather than the solution. In order to gain knowledge, one needs not just to think but to investigate».

⁹³ With apologies to Mark Twain, that the death of statistical evidence has been greatly exaggerated has been known for over 35 years, yet the myth seems to persist in some circles. See, e.g., ALLEN, 1986: 429, n. 67.

idence, naked or otherwise, faces any serious problems in American jurisprudence⁹⁴. Of course, it is also possible that those who express this intuition are indifferent to what the real juridical world is and the game being played serves the purposes of other disciplines, in which case this article can undergo a timely and well deserved demise—its most basic point is that this form of intuition mining is not explanatory of juridical proof whatever other value it may have, and that what is required is a naturalistic approach to the problem. Or it could be that the description of the American legal systems is itself either ill-informed or, even if accurate, woefully misguided. I would be delighted if this analysis played a role in generating knowledge by stimulating a response justifying either conclusion. The job is not likely to be done with weird hypotheticals and their handmaidens.

How does any of this apply to the European or Spanish world systems? I hesitate to venture far into unknown terrain, but I will observe the discordance between the general European and Spanish scholars' heavy emphasis on pursuing «truth» in civil dispute resolution and the inefficiencies of many of those systems which, in the pursuit of truth, avoid making mistakes by making it difficult to decide cases—indeed in some cases essentially impossible. But not deciding cases is to decide them, as the status quo favors one party or the other; thus, interminable delay in civil cases perpetuates the very same mistakes that the pursuit of truth is supposed to suppress⁹⁵.

- ALLEN, R. J., 1986: «A reconceptualization of civil trials», in *Boston University Law Review*, 66(3): 401-437.
- , 1991: «On the significance of batting averages and strikeout totals: a clarification of the “naked statistical evidence” debate, the meaning of “evidence,” and the requirement of proof beyond a reasonable doubt», in *Tulane Law Review*, 65(5): 1093-1110.
- , 1997: «Rationality, algorithms and juridical proof: A preliminary inquiry», in *The International Journal of Evidence & Proof*, 1(special): 254-275.
- , 2013: «Taming complexity: rationality, the law of evidence and the nature of the legal system», in *Law, Probability and Risk*, 12(2): 99-113.

⁹⁴ Given that the intuition is inconsistent with reality, it is also not surprising that the American legal system is pretty much oblivious to the complicated efforts to justify the intuition. The only case that I know of that cites to the various efforts to domesticate the statistical evidence «problem» is an Israeli case authored by Alex Stein, a former American academic fluent in these debates who was recently appointed to the Israeli Supreme Court. See *The State of Israel v. Aviv Bezalel*, The Supreme Court of Israel sitting as the Final Court of Criminal Appeals, Criminal Appeal 2921/18 and Criminal Appeal 2961/18. For example, I could find no American opinion in Westlaw's case database that relied on or discussed epistemic safety, epistemic sensitivity, or evidential probability. Search terms included, among others: «evidential probability», «epistemic probability», «“epistemic” & “sensitivity”», «epistemological safety», «epistemological /10 safety», «“epistemological” & “safety”», «epistemological sensitivity», «epistemological /10 sensitivity», «“epistemological” & “sensitivity”», and «epistem! & (safety /30 sensitivity)». The entire set of search strings is available from the author.

⁹⁵ See ALLEN, 2015.

- , 2015: «The Perils of Comparative Law Research», published in Spanish as «Los Peligros De Investigación En Derecho Comparado», in Jordi Ferrer Beltran & Carmen Vazquez (eds.), *Debatiendo con Taruffo*. Marcial Pons, 2016: 21-40.
- , 2020: «Debate: Legal Probabilism—A Qualified Rejection: A Response to Hedden and Colyvan», in *Journal of Political Philosophy*, 28(1): 117-128.
- ALLEN, R.J. and B. LEITER, 2001: «Naturalized epistemology and the law of evidence», in *Virginia Law Review*, 87(8): 1491-1550.
- , 2003: «Naturalized epistemology and the law of evidence: Reply to Redmayne», in *The Michigan State Law Review* 2003(4): 885-892.
- ALLEN, R.J. and M.S. PARDO, 2019a: «Relative plausibility and its critics», in *The International Journal of Evidence & Proof*, 23(1-2): 5-58
- , 2019b: «Clarifying relative plausibility: A rejoinder», in *The International Journal of Evidence & Proof*, 23(1-2): 205-217.
- ALLEN, R. J. and A. SHAVELL, 2005: «Further Reflections on the Guillotine», in *The Journal of Criminal Law and Criminology*, 95(2): 625-636.
- BENNETT, W. and M. FELDMAN, 1981: *Reconstructing Reality in the Courtroom*. London: Tavistock Feldman.
- BLOME-TILLMANN, M. 2002: «Statistical Evidence, Normalcy, and the Gatecrasher Paradox», *Mind* 123(514): 563-577.
- , 2017: «More Likely Than Not' Knowledge First and the Role of Bare Statistical Evidence in Courts of Law», in Carter, A., Gordon, E. & Jarvis, B (eds.), *Knowledge First - Approaches in Epistemology and Mind*, Oxford, UK: Oxford University Press: 278-292.
- CALLEN, C. R., 1982: «Notes on a Grand Illusion: Some Limits on the Use of Bayesian Theory in Evidence Law», *Indiana LJ*, 57(1): 1-44.
- , 1991: «Kicking in rocks with Dr. Johnson: a comment on Professor Allen's theory», in *Cardozo Law Review*, 13(2, 3): 423-444.
- COHEN, L. J., 1977: *The Probable and the Provable*, Oxford University Press.
- , 1981a: «Subjective probability and the paradox of the gatecrasher», in *Arizona State Law Journal*, 1981(2): 627-634.
- , 1981b: «Can human irrationality be experimentally demonstrated?», in *Behavioral and Brain Sciences*, 4(3): 317-331.
- DI BELLO, M, 2019: «Proof paradoxes and normic support: socializing or relativizing?», in *Mind*, 128(512): 1045-1084.
- DIAMOND, S., B. MURPHY, M. ROSE, 2012: «The kettleful of law in real jury deliberations: Successes, failures, and next steps», in *Northwestern University Law Review*, 106(4): 1537-1608.
- EDWARDS, W., 1986: «Summing up: The society of Bayesian trial lawyers», in *Boston University Law Review*, 66(3): 937-41.
- ENOCH, D., SPECTRE, L, FISHER, L., 2012: «Statistical evidence, sensitivity, and the legal value of knowledge», in *Philosophy & Public Affairs*, 40(3): 197-224.
- ENOCH, D. and T. FISHER, 2015: «Sense and sensitivity: Epistemic and instrumental approaches to statistical evidence», in *Stanford Law Review*, 67(3): 557-611.
- ENOCH, D. and L. SPECTRE, 2019: «Sensitivity, Safety, and the Law: A Reply to Pardo», in *Legal Theory*, 25(3): 178-199.
- FISCHER, D. A., 2001: «Tort recovery for loss of a chance», in *Wake Forest Law Review*, 36(3): 605-655.
- FRANKLIN, J., 2011: «The objective Bayesian conceptualisation of proof and reference class problems», in *Sydney Law Review*, 33(3): 545-561.
- FULLER, S., 2002: *Social Epistemology*, Indiana University Press.
- GARBER, D. and S. ZABELL, 1979: «On the emergence of probability», in *Archive for History of Exact Sciences*, 21(1): 33-53.
- GARDINER, G, 2020: «Profiling and proof: are statistics safe? », in *Philosophy*, 95(2): 161-183.
- GOLDMAN, A., 1999: *Knowledge in a Social World*, Oxford University Press.
- HACKING, I., 2006: *The Emergence of Probability (2nd ed)*, Cambridge: Cambridge University Press.

- Hedden, B and Colyvan, 2019: M, «Legal probabilism: a qualified defence», in *Journal of Political Philosophy*, 27(4): 448-468.
- JONES, G. T. and R. HAGTVEDT, 2001: «Sample data as evidence: Meeting the requirements of Daubert and the recently amended federal rules of evidence», in *Georgia State University Law Review*, 18(3): 721-748.
- KAPLAN, J., 1967: «Decision theory and the factfinding process», in *Stanford Law Review*, 20(6): 1065-1092.
- KAYE, D., 1979a: «Naked statistical evidence», in *Yale Law Review*, 89(3): 601-611.
- , 1979b: «The law of probability and the law of the land», in *University of Chicago Law Review*, 47(1): 34-56.
- , 1979c: «The paradox of the gatecrasher and other stories. », in *Arizona State Law Journal*, 1979 (1): 101-110.
- , 1982: «The limits of the preponderance of the evidence standard: Justifiably naked statistical evidence and multiple causation», in *Law & Social Inquiry*, 7(2): 487-516.
- LAUDAN, L. and R. J. ALLEN, 2008: «Deadly dilemmas», in *Texas Tech Law Review*, 41(1): 65-92.
- LEITER, B., 2011: «Naturalized jurisprudence and American legal realism revisited», in *Law and Philosophy* 30(4): 499-516.
- , 2020: «What is a realist theory of law? », Available at SSRN: <https://poseidon01.ssrn.com/delivery.php?ID=658002065071097112110111067081073098030040014018086061111066121075071096018068090030030053038123014014001023069117079127108106046002025038074087123121067123085125067034019009003020006094099026107110025007072096093073113087068112113094127121127113084&EXT=pdf>.
- LEMPERT, R., 1986: «The new evidence scholarship: Analyzing the process of proof», in *Boston University Law Review*, 66(3): 439-477.
- LEVANON, L., (2019): «Statistical Evidence, Assertions and Responsibility», in *Modern Law Review* 82(2): 269-292
- LEWIS, D., 1996: «Elusive knowledge», in *Australasian journal of Philosophy*, 74(4): 549-567.
- LITTLEJOHN, C. ,2017: «Truth, knowledge, and the standard of proof in criminal law», *Synthese* published online at <https://link.springer.com/content/pdf/10.1007%2Fs11229-017-1608-4.pdf>.
- MOSS, S., 2018: *Probabilistic Knowledge*, Oxford: Oxford University Press.
- , 2021: «Knowledge and Legal Proof», forthcoming in *Oxford Studies in Epistemology* vol. 7, Oxford: Oxford University Press.
- NANCE, D A., 2016: *The Burdens of Proof: Discriminatory Power, Weight of Evidence, and Tenacity of Belief*, Cambridge University Press.
- NATAPOFF, A., 2018: *Punishment without Crime: How our Massive Misdemeanor System Traps the Innocent and Makes America More Unequal*, Basic Books.
- , 2020: «Atwater and the misdemeanor carceral state», *Harvard Law Review Forum* 133(6): 147-178.
- NATIONAL RESEARCH COUNCIL, 2000: «Reference Guide on Statistics», in *Reference Manual on Scientific Evidence* (2nd ed), National Academies Press.
- PARDO, M.S., 2005: «The field of evidence and the field of knowledge», in *Law and Philosophy* 24(4): 322-392.
- , 2013: «The nature and purpose of evidence theory», in *Vanderbilt Law Review* 66(2): 547-613.
- , 2018: «Safety vs. sensitivity: Possible worlds and the law of evidence», in *Legal Theory*, 24(1): 50-75.
- PARDO, M. S. and R. J. ALLEN, 2008» «Juridical proof and the best explanation», in *Law and Philosophy* 27(3): 223-268.
- PARK, R. C. and M. J. SAKS, 2005: «Evidence scholarship reconsidered: results of the interdisciplinary turn», in *Boston College Law Review*, 47(5): 949-1031.
- PILLER, C., 2019: «Beware of safety», in *Analytic Philosophy*, 60(4): 307-335.
- POSNER, R. A., 1999: «An economic approach to the law of evidence», in *Stanford Law Review*, 51(6): 1477.
- PRIEL, D., (2020), «Evidence-based jurisprudence: An essay for Oxford», *Analisi e Diritto* 2020, available at <https://poseidon01.ssrn.com/delivery.php?ID=59411307001311511901112408812003107800>

- 206906708700605009306701503000306910311411909309803101111900102904410807712
411608608903105505702803201300401711211212309306601405802806506411309806911
3021099123101092097111121099113027007007008079100015025081065&EXT=pdf.
- PUNDIK, A., 2008: «What is Wrong with Statistical Evidence? The Attempts to Establish an Epistemic Deficiency», Available at SSRN: <https://ssrn.com/abstract=1134656> or <http://dx.doi.org/10.2139/ssrn.1134656>.
- , 2011: «The epistemology of statistical evidence», in *The International Journal of Evidence & Proof*, 15(2): 117-143.
- RIBEIRO, G., 2019: «Relevance, probative value, and explanatory considerations», in *The International Journal of Evidence & Proof*, 23(1-2): 107-113.
- ROTH, A., 2010: «Safety in numbers? Deciding when DNA alone is enough to convict», in *New York University Law Review*, 85(4): 1130-1185.
- SANCHIRICO, C. W., 2001: «Character Evidence and the Object of Trial» in *Columbia Law Review*, 101(6): 1227-1311.
- SMITH, M., 2010: «What else justification could be? », in *Nous*, 44(1): 10-31.
- , 2018: «When does evidence suffice for conviction? », in *Mind*, 127(508): 1193-1218.
- VIDMAR, N. and V. P. HANS, 2007: *American Juries : The Verdict*. Amherst, N.Y.: Prometheus Books.
- WAGENAAR, W. A., et al., 1993: *Anchored Narratives: The Psychology of Criminal Evidence*, St Martin's Press.
- WEINBERG, J. M., NICHOLS, S., STICH, S., 2001: «Normativity and epistemic intuitions», in *Philosophical Topics* 29(1/2): 429-460.
- WELLS, G. L., 1992: «Naked statistical evidence of liability: Is subjective probability enough?» in *Journal of Personality and Social Psychology*, 62(5): 739-752.