STATISTICAL EVIDENCE: THE CASE FOR SENSITIVITY

David Enoch and Talia Fisher[*]

INTRODUCTION

"For nearly twenty years, law journals have been the forum for a bitter debate about the use at trial of overtly probabilistic evidence and methods," wrote Jonathan J. Koehler and Daniel N. Shaviro in 1990.[1] More than two decades have passed since then, but these words still hold true. Despite the voluminous body of literature dedicated to the issue of statistical evidence, it continues to generate great controversy in evidence law scholarship. Questions regarding the admissibility and sufficiency of statistical evidence arise in court in ever-growing numbers, with seemingly inconsistent treatment in the case-law.[2] The aim of this article is to dispel some of the confusion surrounding the use of statistical evidence in the legal arena. This will be done by viewing statistical evidence through the prism of Sensitivity—namely, the epistemological requirement that a belief be sensitive to the truth as a necessary condition for knowledge. We will use the epistemological discourse on Sensitivity as well as an instrumental analysis to address the descriptive and prescriptive challenges that statistical evidence poses.

[1] Jonathan J. Koehler & Daniel N. Shaviro, *Veridicial Verdicts: Increasing Verdict Accuracy through the Use of Overtly Probabilistic Evidence and Methods*, 75 CORNELL L. REV. 247, 247 (1990) (see also the sources cited therein).
[2] For further discussion on the topic of statistical evidence, see Ronald J. Allen, *On the Significance of Batting Averages and Strikeout Totals: A Clarification of the "Naked Statistical Evidence" Debate*, 65 TUL. L. REV. 1093 (1991); Craig R. Callen, *Adjudication and the Appearance of Statistical Evidence*, 65 TUL. L. REV. 457 (1991).

PRESENTING THE PROBLEM

The natural starting point for the statistical evidence debate is the Blue Bus hypothetical,[3] which is a variant of *Smith v. Rapid Transit,* a seminal case in modern evidence law.[4] The hypothetical is as follows: A bus causes harm. In the first scenario, an eyewitness recognizes it as belonging to the Blue Bus Company. The witness, however, is imperfectly reliable. To illustrate, let us assume her to be 70% reliable in such circumstances. Most people would have no qualms about accepting her testimony as the basis for a positive finding that the bus belonged to the Blue Bus Company and, perhaps, for finding the Company liable on the basis of that testimony. Now consider a second scenario: A bus causes harm, but there are no eyewitnesses. The court is presented with uncontested data regarding the distribution of buses in the relevant area, showing that the Blue Bus Company owns 70% of the buses. In this second scenario, many will typically object to basing a positive finding of fact—and certainly liability—on this kind of evidence. Such was the case in *Smith v. Rapid Transit*, where the court rejected the statistical market-share evidence as the basis for finding that the defendant's bus was implicated in the accident.[5]

Regardless of the reasons why (and more on this in what follows), an overwhelmingly common and strong intuition among practitioners and scholars alike is that the second scenario should be treated with some suspicion, that in some sense, market-share evidence is inferior to eyewitness testimony. Yet in both the Blue Bus scenarios—eyewitness testimony and market-share evidence—all of the apparently relevant features are held constant. In particularly, the

---

[3] For one discussion of this common hypothetical, see Mike Redmayne, *Exploring the Proof Paradoxes*, 14 LEGAL THEORY 281, 282 (2008).
[4] **S**mith v. Rapid Transit, Inc., 58 N.E.2d 754 (Mass. 1945). In *Smith*, it was agreed that the presence of the defendant's bus line at the time and place of the accident was consistent with the schedule. The trial court ruled for the defendant and the Supreme Judicial Court of Massachusetts upheld the verdict, ruling that liability could not rest solely on the "mathematical chances" that support it. For a more elaborate discussion of *Smith v. Rapid Transit*, see Gary L. Wells, *Naked Statistical Evidence of Liability: Is Subjective Probability Enough?,* 62 J. PERSONALITY & SOC. PSYCH. 739 (1992).
[5] *Id.* at 755.

probabilities are equal. This means that the chances of a finding of liability against the Blue Bus Company being mistaken are equivalent in the two cases. And of course, there is something statistical about the eyewitness testimony as well. She is certainly fallible; out of an arbitrary set of one-hundred buses that she recognizes as belonging to the Blue Bus line in similar circumstances, thirty will belong to other companies.[6] How, then, can we distinguish between the evidence that the relevant bus is a member of a set 70% of buses that belong to the Blue Bus Company (insufficient to substantiate a finding of liability under *Smith*) and the fact that the bus was a member of a set 70% of buses that are accurately identified by the eyewitness (sufficient for a finding of liability)? Can the evidential distinction between these two scenarios be accommodated in a theoretically respectable manner?[7]

Consider the analogous gatecrasher hypothetical,[8] where it is uncontested that of a thousand people attending a stadium event, only ten purchased tickets. If any individual attending the stadium is sued or, even more clearly, is prosecuted, a finding against him or her merely on the strength of the (very strong!) statistical evidence would seem to be inappropriate, even where conviction based on a probabilistically similar piece of direct individual evidence, such as videotape, seems perfectly fine. The problem generalizes, of course, and surfaces in many different scenarios. Thus, it can appear in civil or criminal trials; arise with different levels

---

[6] Nothing in our relevant intuitions depends on the evidence being from a witness. Think about other types of clearly individual evidence, like videotape from a surveillance device or physical evidence such as the accused's gun found on the scene.

[7] One major issue we will not address here is the reference class problem. We will assume, more or less throughout, that the statistical evidence latches on to the relevant frequencies. Such a simplifying assumption cannot be objected to in our context, as it arguably arises both for statistical and for individual, direct evidence (for instance, regarding the 70%-reliability of the eyewitness account).

[8] David Kaye, *The Paradox of the Gatecrasher and Other Stories,* ARIZONA ST. L.J. 101, 104 (1979); Robert J Rhee, *Probability, Policy, and the Problem of Reference Class*, 11 INT'L J. OF EVIDENCE & PROOF 286, 289 (2007).

of probability; relate to past or future events; be the sole piece of evidence or supplement other types of evidence; and so on.[9]

Indeed, the legal doctrine has failed to systematically resolve questions regarding the use of statistical evidence in court. In the first half of the twentieth century, when such evidence first began to appear in court, many judges responded antagonistically, deeming it inadmissible and devoid of any probative value.[10] To this day, courts continue to exhibit a general preference for individuating evidence, rejecting base-rate evidence, despite its potential to promote accuracy in legal fact-finding.[11] The doctrinal picture, however, is not clear-cut: *First,* the legal doctrine regarding the use of statistical evidence for imposing liability is very incoherent, and there is conflicting case-law on this matter.[12] While statistical evidence is often, indeed, considered irrelevant and inadmissible in court,[13] at other times, the courts treat it as admissible when presented as supplementary evidence. In yet other instances, statistical evidence is even treated as sufficient for a finding of liability.[14] For example, in contrast to the ruling in *Smith*, in *Kaminsky v. Hertz*,[15] the appellate court ruled for a presumption of ownership, when the only

---

[9] As will be demonstrated in further parts of this article, for certain purposes, there may be important distinctions among these different scenarios. For a list of examples, and some insistence on the significance of the differences between them, *see*: Redmayne, *supra* note 3 at 281. At this point in the discussion, no further distinctions will be made. The problem of accommodating the distinction between statistical and individual evidence is a general one, and as a first step, a general solution is called for.

[10] *See, e.g.,* Virginia & S.W. Ry. Co. v. Hawk, 160 F. 348, 352 (6th Cir. 1908); Evans v. Ely, 13 F.2d 62, 64 (3d Cir. 1926); Commercial Standard Ins. Co. v. Gordon Transports, Inc., 154 F.2d 390, 396 (1946); Smith v. Rapid Transit, Inc., 58 N.E.2d 754, 755 (Mass. 1945); People v. Collins, 438 P.2d 33, 40-41 (Cal. 1968) (en banc); State v. Carlson, 267 N.W.2d 170, 176 (Minn. 1978). *See also* Alon Harel & Ariel Porat, *Aggregating Probabilities Across Cases: Criminal Responsibility for Unspecified Offenses*, 94 MINN. L. REV. 261, 294 (2009).

[11] Jonathan J. Koehler, *When Do Courts Think Base Rate Statistics Are Relevant?,* 42 JURIMETRICS 373, 373 (2002); Chris Guthrie, Jeffrey J. Rachlinski & Andrew J. Wistrich, *Inside the Judicial Mind*, 86 CORNELL L. REV. 777, 810 (2001).

[12] Koehler, *supra* note 11, at 373.

[13] Koehler, *supra* note 11, at 377.

[14] In *Turner v. United States*, 396 U.S. 398, 417 (1970), the Supreme Court held that base-rate evidence indicating that more than 98% of all heroin consumed in the U.S. is illegally imported can suffice for inferring that its possession amounts to possession of a smuggled drug.

[15] Corp. 288 N.W. 2d 426 (Mich. Ct. App., 1980). But see Richard A. Posner, *An Economic Approach to the Law of Evidence*, 51 STAN. L. REV. 1477, 1508 (1999), contesting the description of *Kaminsky v. Hertz Corp.*, 288 N.W.2d 426 (Mich. App. 1979), as

evidence brought forth against the defendant, Hertz, was the unchallenged testimony that the truck that had caused the accident bore a Hertz logo and that Hertz owned 90% of the yellow trucks bearing a Hertz logo.[16]

Second, not only are there seemingly random inconsistencies in the court rulings on statistical evidence in the general class of cases, but exceptions to the general approach of admissibility also exist in specific categories of cases. Such is the case, for example, with DNA profiling, which is increasingly relied upon by the courts.[17] To date, most American courts admit DNA evidence, despite the apparent similarities to other, inadmissible, types of statistical evidence.[18] In light of the considerable inconsistency in the case-law regarding the admissibility or sufficiency of base-rate statistical evidence, the extrapolation of the legal doctrine on this matter has been deemed "part science and part art."[19]

---

contrary to *Smith v. Rapid Transit*, in light of the fact that the corresponding percentages were 90% and 10% and the nonstatistical evidence pointing to the defendant's ownership of the truck that had caused the accident.

[16] For a similar case, see Kramer v. Weedhopper of Utah, Inc., 490 N.E. 2d 104 (Ill. Ap. Ct. 1986). Kramer was injured when a bolt sheared off the aircraft that he had assembled from a Weedhopper kit. Weedhopper received 90% of its bolts from Lawrence and 10% of its bolts from Hughes. Based on this base rate, an Illinois appellate court reversed a trial court's summary judgment in favor of the primary bolts manufacturer. The appellate court held that the statistical evidence allows the inference that the bolts in Weedhopper's bin and the bolt supplied to Kramer were purchased from Lawrence.

[17] Brian C. Smith et al., *Jurors' Use of Probabilistic Evidence*, 20 LAW & HUM. BEHAV. 49, 49 (1996). For further examples of statistic-based scientific evidence, see People v. Fierro, 821 P.2d 1302, 1318-19 (Cal. 1991); People v. Alzoubi, 479 N.E.2d 1208, 1210 (Ill. Ct. App. 1985).

[18] Kristen Bolden, *Note & Comment: DNA Fabrication, A Wake Up Call: The Need to Reevaluate the Admissibility and Reliability of DNA Evidence*, 27 GA. ST. U.L. REV. 409 (2011) (citing United States v. Jakobetz, 955 F.2d 786, 799 (2d Cir. 1992)). In addition, in some jurisdictions, the reliability of DNA evidence has been statutorily established.

[19] J. Koehler, *When Do Courts Think Base Rate Statistics are Relevant?*, 42 JURIMETRICS 373,401(2002). Some scholars have attempted to deal with the doctrinal puzzles and apparent inconsistencies by unearthing the conditions under which courts are more likely to allow statistical evidence to be used at trial. The most comprehensive and notable attempt has been made by Jonathan Koehler. Koehler examined the use of base-rate statistics in appellate courts, extrapolating from the case-law that appellate courts are more likely to view such statistical evidence as relevant and admissible under the following conditions:

a) when the evidence arises in cases appearing to bear a statistical structure;

b) when the evidence is offered to refute chance occurrence theories

c) when the evidence is computed using reference classes that bear particular features of the focal case; or,

d) when the evidence is offered in cases where individuating evidence is impossible to obtain.

In addition to the inconsistencies in the legal doctrine, the legal scholarship is also fraught with discrepancies on this topic. The decision in *Smith*[20] triggered a heated, still on-going debate in the legal academia around the questions of whether judicial fact-finding ought to be grounded on standard probability logic and whether statistical evidence regarding the base-rate for liability is sufficient grounds for a ruling in criminal or civil trials.[21] Some scholars, such as Daniel N. Shaviro and Jonathan J. Koehler, question the refusal to ground verdicts in favor of plaintiffs or prosecutors on statistical evidence.[22] They claim that the objective of verdict accuracy requires that courts hold in favor of the party whose case seems more likely to be founded [23] and that the only relevant misgivings about imposing liability are doubts related to risk of error, not its overtness.[24] But those evidence law scholars who advocate the use of statistical evidence are outnumbered by academics, who, in turn, tend to be strongly against statistical evidence and object to its submission in trial for reasons of probative or moral deficiency and for policy considerations.[25]

---

[20] Another well- known statistical evidence case that also led to much debate is People v. Collins, 438 P.2d 33 (Cal. 1968).

[21] Some of the extensive discussions of the issue of statistical evidence can be found in the *Harvard Law Review* (1970-1971), *Boston University Law Review* (1986), and *Tulane Law Review* (1991). *See also* Roger C. Park & Michael J. Saks, *Evidence Scholarship Reconsidered: Results of the Interdisciplinary Turn*, 46 B.C. L. REV. 949, 992; ALEX STEIN, FOUNDATIONS OF EVIDENCE LAW (2005); HO HOCK LAI, A PHILOSOPHY OF EVIDENCE LAW (2008); William Meyerson, *Significant Statistics: The Unwitting Policy Making of Mathematically Ignorant Judges*, 37 PEPP. L. REV. 771 (2010); Amit Pundik, *Statistical Evidence and Individual Litigants* (SSRN).

[22] Daniel Shaviro, *Statistical-Probability Evidence and the Appearance of Justice*, 103 HARV. L. REV. 530 (1989); Jonathan J. Koehler & Daniel N. Shaviro, *Veridicial Verdicts: Increasing Verdict Accuracy Through the Use of Overtly Probabilistic Evidence and Methods*, 75 CORNELL L. REV. 247 (1990).

[23] Shaviro, *supra* note 22, at 532.

[24] *Id*. at 543.

[25] Scholars who object to statistical evidence on grounds of probative deficiency include Richard B. Wright, who claims that statistical evidence lacks the appropriate causal connection. Richard B. Wright, *Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof: Pruning the Bramble Bush by Clarifying the Concepts,* 73 IOWA L. REV. 1001 (1988). Alex Stein argues that statistical evidence lacks in case specifity. STEIN, *supra* note 21. Those who object to statistical evidence on moral and policy grounds include Brilmayer and Kornhauser, who claim that statistical evidence is antithetical to the defendant's individuality and violates the moral directive to judge individuals on the basis of their own conduct. Lea Brilmayer & Lewis Kornhauser, *Review: Quantitative Methods and Legal Decisions,* 46 U. CHI. L. REV. 116, 149 (1978).Pundik claims that statistical evidence undermines the individuality and autonomy of the person against whom it is being used, in light of the assumptions decision-makers have to make when inferring information regarding an individual's behavior from statistical evidence. Amit Pundik, *Statistical Evidence and Individual Litigants: A Reconsideration of Wasserman's Argument from Autonomy*, 12 INT'L J. OF EVIDENCE & PROOF 303 (2008). Laurence Tribe asserts a qualitative distinction between the outcome of a wrongful conviction

The apparently inconsistent treatment of statistical evidence in both the legal doctrine and evidence law literature has created a need for an overarching theory. The objective of this article is to provide such a theoretical framework, to solve the doctrinal puzzles and dispel some of the incoherence associated with the use of statistical evidence in trial.

The theoretical framework we offer in this article views statistical evidence through the prism of epistemology: It connects statistical evidence to a broader epistemological discussion of similar phenomena, and uses this epistemological discourse to highlight Sensitivity – the requirement that a belief be counterfactually sensitive to the truth– as a way of epistemically explaining the legal suspicion towards statistical evidence. The second layer of our theory is devoted to the claim, that the epistemological distinction cannot serve as a satisfactory vindication of the reluctance to rely on statistical evidence. Knowledge, Sensitivity, and epistemology– we claim-- carry little, if any, legal value. Instead of the epistemological story, we therefore offer an elaborate incentive-based story, vindicating the suspicion towards statistical evidence. However, as we show in the article, the epistemological story and the incentive-based story are closely knit and interestingly related, Using these theoretical foundations we turn to expose the intuitions underlying the prevailing differential treatment of statistical evidence in the doctrine, as well as explain why some types of statistical evidence are regarded by the courts to be admissible and sufficient for substantiating liability, while other types are rejected as such. In

---

when the trier has been fully convinced of the defendant's guilt and a wrongful conviction when the trier is conscious of potential error. Laurence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329 (1971). Charles Nesson makes a similar claim, arguing that rulings based upon statistical evidence may be illegitimate and unacceptable, despite any potential probative value, because of their adverse effect on the public trust in the adjudication system. Charles Nesson, *The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts*, 98 HARV. L. REV. 1357, 1378 (1985). Richard Posner challenges the sufficiency of statistical evidence for findings of liability due to the expected increase in the costs of trial that such evidence may generate. Richard A. Posner, *An Economic Approach to the Law of Evidence*, 51 STAN. L. REV. 1477, 1508 (1999). Some fail to distinguish between the different strands of objection or launch a dual attack against statistical evidence. *See, e.g.,* Jonathan J. Koehler, *The Probity-Policy Distinction in the Statistical Evidence Debate,* 66 TULANE L. REV. 141 (1991).

addition, the article highlights the prescriptive contribution of our theoretical framework, in providing criteria for legal reform and revising the treatment of statistical evidence in certain contexts.[26]

The article proceeds as follows. Part I reviews some of the existing theoretical endeavors to vindicate the distinction between statistical evidence and individual evidence and points to their shortcomings. Part II presents an alternative theoretical framing of the statistical evidence debate. Part III then applies this theoretical framework to the legal sphere, showing its capacity to solve doctrinal puzzles and guide legal reform. Part IV concludes the discussion.

PART I: A CRITICAL REVIEW OF PROPOSED SOLUTIONS TO THE STATISTICAL EVIDENCE PUZZLE

A good way to appreciate the depth of a problem is to explore the attempts that have been made to tackle it. The literature on statistical evidence is rather extensive, and it offers several avenues for contending with the problem of justifying the distinction between statistical evidence and individual evidence.[27] We will begin by mapping out the most influential suggestions in the literature and briefly noting their shortcomings. This will underscore for the reader the gravity of the problem, as well as enable an appreciation of the distinctive features (and advantages) of the account that we will propose, in our attempt to address it.[28]

---

[26] The article offers a detailed account of the incentive-based story underlying statistical evidence, and addresses the derivative doctrinal issues. The epistemological account, in contrast, is outlined in a preliminary fashion only, as the epistemological intricacies of statistical evidence are of less interest to a legal audience. For a comprehensive and detailed account of the epistemological foundations underlying our theory, see: David Enoch at al., *Statistical Evidence, Sensitivity, and the Legal Value of Knowledge*, PHILOSOPHY AND PUBLIC AFFAIRS (forthcoming, 2013)

[27] For a good survey, see HOCK LAI HO, A PHILOSOPHY OF EVIDENCE LAW 137-39 (2008). For an earlier, much more critical discussion of many of these suggestions, see Ferdinand Schoeman, *Statistical vs. Direct Evidence.* 21 NOÛS 179 (1987).

[28] While this critical survey is too brief to be conclusive, it does offer a rough sketch of the research in the field.

One line of reasoning raised in the literature against statistical evidence relates to exogenous factors: perhaps—as Posner claims—the very resort to statistical evidence is in itself proof that no other evidence could be found; and perhaps that in itself is evidence against the plaintiff (or the prosecution).[29] If this is, indeed, the case, statistical evidence should be accorded less weight, a priori, simply because it tends to be submitted in circumstances in which the case of the party presenting the evidence is weaker. This may be so—though much empirical work would be necessary to substantiate this argument. But we can safely abstract away from all of this by stating—as we did at the outset of this article—that we hold all other things constant between the statistical-evidence scenario and the individual-evidence scenario. Yet**,** even after holding all else equal, the intuition of a distinction between the two types of evidence seems to persist. So explanations of this kind will not suffice.

Another claim made in the literature is that there is an important distinction to be made between evidence that is genuinely about the relevant defendant as opposed to merely statistical evidence, which is considered to be unrelated to the defendant's matter**.** Following this line of reasoning, in the Blue Bus hypothetical, the eyewitness testimony is about the Blue Bus Company, whereas the market-share evidence is not—the latter is in no way relevant to determining what happened in the specific case.[30] This about-relation claim does not, in our opinion, help.[31] In the context of evidence, the only "about" that is relevant is the epistemological "about," the "about" of indication. And with both the individual and the statistical pieces of evidence the relevant evidence indicates that the bus was blue. In this sense, the statistical evidence, too, is "about" the Blue Bus Company. Now, there may be nothing objectionable about

---

[29] Richard A. Posner, *An Economic Approach to the Law of Evidence*, 51 STAN. L. REV. 1477, 1509c (1999).

[30] Richard B. Wright, *Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof: Pruning the Bramble Bush by Clarifying the Concepts*, 73 IOWA L. REV. 1001, 1050 (1988).

[31] For a similar claim, see Ferdinand Schoeman, *Statistical vs. Direct Evidence.* 21 NOûS 183 (1987).

using such about-talk to capture the intuitive distinction between statistical and individual evidence. But doing so without giving considerably more details about the about-relation amounts not to an explanation or vindication of the distinction but, rather, merely to its re-naming. And providing the necessary further detail would require, we suspect, conducting the kind of discussion that this article is engaging in.

Judith Jarvis Thomson suggests that the difference between statistical and individual evidence should be understood causally.[32] Individual evidence, she claims, is causally linked in an appropriate way to the thing for which it is taken as evidence: in the Blue Bus case, it is the fact that the bus that inflicted the harm was blue that caused the eyewitness testimony and (it seems) in an appropriate way. In the case of statistical evidence, however, no similar appropriate causal link is present. That the relevant bus was blue in no way, apparently, caused the market-share evidence. Thomson holds such causal links with evidence to be a necessary condition for knowledge.[33] She also holds that they are necessary for justifiable legal fact-finding,[34] presumably at least partly because she believes that knowledge is a necessary condition for justifiable legal fact-finding (at least in criminal cases).

Yet in our opinion, the causal mechanism does not capture the legal distinction between statistical evidence and individual evidence. For instance, courts may sometimes need to accept evidence (expert witness testimony, for example) regarding certain mathematical truths. It is very hard to see how the causal requirement can be met here, given that mathematical truths are (arguably) causally inert. Also, causal links—even appropriate ones—can be notoriously complicated. And cases can easily be constructed—cases with multiple causes, independent

---

[32] Judith Jarvis Thomson, *Liability and Individualized Evidence, in* RIGHTS, RESTITUTION, AND RISK 225 (1986).
[33] *Id*. at 230.
[34] *Id*. at 244.

causal chains, different facts that suffice causally only together but not each on its own, different facts each of which suffices causally alone, etc.—where it is not clear what follows from a causal theory and, to the extent that it is clear, the implications are intuitively unacceptable. Lastly, consider a certainty case where, for instance, no one at the stadium purchased a ticket. There, intuitively the evidence is still statistical (100% is also a probability, isn't it?), but it nonetheless seems sufficient for conviction. It is not clear, however, how a causal theory can accommodate this result: after all, there is no appropriate causal link between no one having purchased a ticket and, say, John's gatecrashing. Thomson explicitly addresses the certainty case, [35] but instead of showing how her theory can accommodate the desired result there, she introduces an explicit exception for the certainty case. This, of course, is objectionably ad hoc. At the very least, a theory that would account for the desired result in the certainty case as a natural particular instance (rather than as an ad-hoc exception) would be the better for it.

Another argument raised in the literature to defend the difference between the two types of evidence focuses on the specific defendant. The claim made is that the defendant ought not to be punished for being a member of a reference class.[36] True, there is indeed something problematic about convicting a defendant for gatecrashing, based purely on the percentage of gatecrashers among those at the stadium: This, after all, is just a repetition of the intuitive suspicion against statistical evidence. But it is highly misleading to say that in such a case the defendant will have been convicted for his membership in the relevant reference class. If we end up punishing the defendant, it will be for crashing the stadium gates. But since we do not have omniscient knowledge of the facts, we must determine—by relying on the basis of evidence—whether the

---

[35] *Id*. at 248.

[36] *See* Mark Colyvan et al., *Is It a Crime to Belong to a Reference Class?*, 9(2) J. POL. PHILO. 168 (2001); Richard O. Lempert, *The Economic Analysis of Evidence Law: Common Sense on Stilts (Symposium: New Perspectives on Evidence,* 87 VA. L. REV. 1619, 1669 (2001).

defendant did in fact gatecrash. And in making this determination, the statistical evidence seems relevant—and if it is not, this still remains to be shown. This point is underscored by the fact that there is something statistical about individual evidence as well. Indeed, it is precisely in this context that it becomes tempting to insist—as some have[37]—that in actuality, at bottom, all evidence is statistical evidence. But this presumably does not show that in the eyewitness scenario, we are punishing someone for being a member of the class of people who would be recognized by the eyewitness. In both cases of statistical and individual evidence, we punish *for* the offense, *by* relying on evidence.

Relatedly, it is sometimes maintained that since the court's primary duty is to do justice in the specific case before it, justice in that case cannot be compromised in order to achieve the more efficient result in the overall class of cases or the result that is likely to minimize the global risk of error.[38] This argument, however, is also insufficient to validate the distinction between statistical and individual evidence, for the court that is instructed to ignore all global effects and to strive solely to do justice in the case at hand still has to resort to evidence—some evidence—to determine what doing justice in the specific case demands. And thus far, no compelling claim has been made showing statistical evidence to be any less appropriate for this purpose than individual evidence.[39]

---

[37] *See, e.g.,* Jonathan J. Koehler & Daniel N. Shaviro, *Veridicial Verdicts: Increasing Verdict Accuracy through the Use of Overtly Probabilistic Evidence and Methods*, 75 CORNELL L. REV. 247, 263 (1990)

[38] Eric Lillquist, *Recasting Reasonable Doubt: Decision Theory and the Virtues of Variability*, 36 U.C. DAVIS 85, 140 (2002).

[39] For similar reasons, talk of collective punishment, or of the need to address the specific defendant rather than a group, will not help here. For many references, see HOCK LAI HO, A PHILOSOPHY OF EVIDENCE LAW 139 (2008). Ho himself is guilty of similar errors, when he talks about relying on statistical evidence as intentionally taking a gamble at the defendant's expense. In cases of statistical evidence, "we saw an inadequacy in the evidence and we intentionally subjected the defendant to an open risk of injustice: we gamble on the facts at his expense." *Id*. at 142. But of course, there is always this inadequacy with fallible evidence. And criminal procedure always involves intentional subjection of the defendant to a risk of injustice. Individual evidence is in no way better in this regard than statistical evidence.

Yet another argument in the literature is that relying on statistical evidence violates the relevant party's autonomy and individuality, indeed perhaps even her free will and agency.[40] By relying on statistical evidence to convict a gatecrasher, for example, are we not in effect saying that she was bound to crash? Or perhaps we are being disrespectful of her full autonomous individuality, treating her as simply a member of a statistical group and not as a genuine person. If so, are we not—by relying on statistical evidence—in some sense degrading her? Is this not reason to reject such evidence? In our opinion, this line of argument can also be rejected While there need be nothing wrong with excluding degrading evidence, even when it is acknowledged as genuinely probative, this reasoning cannot justify the distinction between individual and statistical evidence. This is firstly because it does not plausibly generalize to all the relevant cases and cannot explain cases like the Blue Bus hypothetical. Secondly and more importantly, this claim confuses epistemology and metaphysics: Statistical evidence is relevant only as evidence. By taking something to give reason to believe or find that the defendant gatecrashed, we are not expressing any belief that he or she has always been bound to do so. Nor are we implying that he or she is anything less than a fully autonomous individual. We are just taking one thing as an indicator of another.

Another attempt at rationalizing the distinction between the two types of evidence is Nesson's well-known claim that verdicts based upon statistical evidence are socially unacceptable.[41] This claim too rejected: First, its empirical basis is unpersuasive, as it is near impossible to delineate the boundaries of what would be acceptable to the public. Second, even assuming the empirical problematic away, there is room to question whether public trust is even

---

[40] David Wasserman, *The Morality of Statistical Proof and the Risk of Mistaken Liability*, 13 CARDOZO L. REV. 935 (1991); Pundik, *supra* note 25.

[41] Charles Nesson, *The Evidence or The Event? On Judicial Proof and the Acceptability of Verdicts*, 98 HARV. L. REV. 1375. 1379 (1985).

a goal that ought to be attained.[42] While it is arguably important that the legal system enjoy some public confidence (though questions may be raised as to the soundness of this as an intrinsic aim, independent of whether the legal system *merits* public trust) and even if perhaps (though this is far thornier) securing that trust can sometimes justify catering to the prejudices of the masses—if there is no other way to justify the traditional skepticism towards statistical evidence, then this feature of public opinion is indeed a prejudice, which renders the call to accommodate it suspicious, at the very least. Furthermore, for our purposes here, we can simply assume the problem away with the premise that the public is going to form an accurate opinion about statistical evidence. In this (perhaps hypothetical) case, then, nothing about public opinion and trust can justify an otherwise unjustified approach to statistical evidence. Of course, (justified) public opinion could supplement any other justification for the traditional approach, but it would be the other justification that provides the primary rationale. Here, too, then, the public opinion argument can be safely dismissed.[43]

One final explanation for differentiating between the two types of evidence, which may initially seem plausible but must ultimately be rejected, goes as follows: To return to the gatecrasher hypothetical, were we to prosecute each and every person who exited the stadium and use the statistical evidence to convict, each and every one of them would be guaranteed to be found guilty, including the ten innocent people who had purchased tickets. In non-statistical cases, though the probability of finding an innocent party guilty might be higher than in each of the gatecrasher trials, there would be no certainty of conviction of the innocent. Since a guaranteed wrongful conviction is something we—as a society—seek to avoid, only non-

---

[42] *See* Talia Fisher & Issachar Rosen-Zvi, *The Confessional Penalty*, 30 CARDOZO L. REV. 871, 908 (2008).

[43] For a more elaborate discussion of similar claims made against Nesson, see Daniel N. Shaviro, *Statistical Probability Evidence and the Appearance of Justice*, 103 HARV. L. REV. 530 (1989).

statistical evidence ought to be accepted, therefore. But this reasoning cannot justify the full extent of the distinction, nor explain it. The following two points demonstrate why: First, in any criminal legal system, innocent defendants are virtually (if not logically) guaranteed to be convicted. The only way to avoid wrongful convictions is to abolish the criminal justice system altogether. The second point can be illustrated with a variant of the gatecrasher hypothesis, where it is possible to indict only one person, because, for instance, all the other attendees fled the stadium before the police arrived. Relying on statistical evidence under such circumstances would not guarantee the conviction of an innocent defendant, but the intuitive reluctance to convict on the basis of statistical evidence would still be fully present.

We should emphasize that this quick critical survey was not intended to be either conclusive or comprehensive in scope. Still, we hope it succeeds in giving a sense of the depth of the problem and that it portrays the theoretical attempts at contending with it, as well as conveying the need for a new resolution.

PART II: THE THEORETICAL FRAMEWORK

1. TWO KINDS OF SOLUTIONS TO THE PROBLEM

Broadly speaking, there are two possible strategies for vindicating the distinction between statistical evidence and individual evidence: an epistemological strategy and a practical strategy. The epistemological strategy aims at demonstrating that statistical evidence is epistemically inferior to individual evidence. Such attempts seek to show, for instance, that statistical evidence never justifies belief; or (more plausibly) that it is harder for statistical evidence than for (probabilistically similar) individual evidence to do so; or that individual evidence can sometimes suffice— and statistical evidence cannot—for being entitled to hold on to a belief or warranted in having a certain degree of confidence; or that statistical evidence cannot—or is

much less likely to—render a belief rational; or that individual evidence can support knowledge, but statistical evidence cannot. All of these are epistemological matters, as is clear when distinguished from more practical matters. Thus, it seems rather indisputable that statistical evidence can render some *actions* rational; it can justify, for instance, avoiding the sandwiches from the deli with the questionable track record.   But justifying actions or rendering them rational is one thing, and justifying beliefs or rendering them rational is arguably another. An epistemological strategy insists on a difference between statistical and individual evidence that is of this latter, epistemic kind. Of course, different manifestations of this strategy may focus on different epistemic concepts (justification, entitlement, warrant, knowledge, rationality, epistemic reasons), and they may vary on other dimensions as well. But all share the insistence on the difference being roughly of this kind, rather than of the practical kind.

Instances of the practical strategy assume—at least for the sake of argument—that there is no epistemic difference between statistical evidence and individual evidence. Roughly speaking, as far as truth or conduciveness to truth is concerned, once we keep the probabilities constant across the two kinds of scenario, the game is over. But epistemology is one thing, and evidence law quite another. And so it is possible that there are practical reasons—such as instrumental reasons having to do with institutional features, with administrative costs, or with differential incentives—why the law should take individual evidence more seriously than it does statistical evidence.

Given this distinction between epistemic and practical strategies (for defending the discrepancy between statistical and individual evidence),[44] we can now make the following

---

[44] See Ferdinand Schoeman, *Statistical vs. Direct Evidence,* 21 Noûs 179, 187 (1987), for a similar distinction. Redmayne also introduces this distinction, but he adds a third strategy, in terms of attacking the inference from the statistical evidence to the

prediction: If the best vindication of the distinction between the two types of evidence is along practical—certainly, instrumental—lines, it is likely to be law-specific. That is, if what justifies the differential legal treatment of statistical and individual evidence is essentially related to the instrumental payoffs of the law so treating it, then it is essentially tied to *the law's* so treating it. It is, after all, quite possible that the instrumental considerations relevant to the law are different from those applicable to other institutions, or perhaps outside any institutional context at all. If, however, there is an epistemic vindication of the distinction, then it is likely to apply much more widely, indeed perhaps as widely as the relevant epistemic notion (justification, perhaps, or knowledge) reaches. Going in the other direction now, if the problem stretches much farther than the law, an epistemic solution, rather than an instrumental one, seems to be called for. We return to this prediction and to the evidence law sphere later on, after introducing the epistemic condition Sensitivity, which will be important in what is to come.

2.   Sensitivity

Think of the following version of the lottery paradox (for knowledge):[45] In the first scenario, you buy a lottery ticket, where the chances of winning are, literally, one in a million. You hold on to it for a day, and now the winning ticket has already been picked, but you receive no indication about the results. Do you know that your ticket has not won? The overwhelmingly plausible answer seems to be "No!" You may know that it's highly unlikely that you hold the

---

relevant finding. *See*: Redmayne, *supra* note 3 at 245. We fail to see how this forms a third kind of strategy here: Either the problem with the inference prevents it from establishing the relevant belief (on which the finding is based), or it does not. If it does, the problem is epistemic; if it does not, the problem seems practical. Either way, the relevant cases fall into one or the other of the two strategies differentiated in the text. Notice that the practical strategy is not necessarily limited to instrumental considerations. Perhaps, for instance, the autonomy-line mentioned above relies on non-instrumental but still practical reasons that differentiate between statistical and individual evidence.

[45] A more extensive presentation of the knowledge-related lottery puzzles can be found in JOHN HAWTHORNE, KNOWLEDGE AND LOTTERIES (2004). This section is survey-ish in nature: we do not purport to be making an original contribution here. The view we are concerned with is one of several accounts that engage with various variants of HENRY E. KYBURG, PROBABILITY AND THE LOGIC OF RATIONAL BELIEF (1961).

winning ticket; you may be justified in gambling against your ticket at rather high odds; you may, of course, believe that it hasn't won. But you do not *know* that it has not won (even if, as things turn out, your ticket has not in fact won).

Now compare the second scenario: You buy a lottery ticket with somewhat better odds—one in a thousand, perhaps. You hold on to it for a day, and now the winning ticket has already been picked, and you find the winning numbers in today's newspaper. They do not match your ticket's numbers. Newspapers are pretty reliable on such matters but not, of course, infallible. Let's suppose that factoring in all the probabilistically relevant information here—the initial odds, the probability that the newspaper made a mistake, and whatever else may be relevant—the probability that your ticket is nonetheless the winning one is one in a million. Do you now know that your ticket did not win? In this second scenario, the overwhelmingly plausible (and common) answer would be "Yes!" Indeed, it is hard to see how this answer is avoidable without deteriorating into a rather global kind of skepticism.

We were careful to hold probabilities constant, yet intuitively—at least when it comes to knowledge—there is an important difference between the two scenarios. In the first scenario (where your evidence that your ticket has not won is just the odds of the lottery), you do not know that the ticket did not win. In the second scenario (where your evidence partly consists of the newspaper item), you do know that the ticket has not won. Given that the probabilities are held constant, what can possibly explain this difference?

One reasonable answer makes use of some relevant counterfactuals:[46] What would you have believed, in both scenarios, had your ticket in fact been the winning ticket? In the first scenario, you would have still believed that the ticket was not going to win. After all, your pessimistic belief was based on the statistical data that is still present, unchanged, even in a case where the pessimistic belief is false (because the ticket actually wins). In the second scenario, though, things are different. In that scenario, your belief is based partly on what was written in the newspaper. And the newspaper—while in no way infallible—is still, so we're assuming, at least reasonably sensitive to the facts. Had your ticket in fact been the winning one, then, in all likelihood this is what would have appeared in the newspaper. And then, in line with the newspaper's information, this is also what you would have believed. It thus seems highly plausible to say that in the newspaper scenario, had the belief (that your ticket didn't win) been false, you would not have held that belief. And so we have a distinction between the two lottery scenarios. There is a kind of counterfactual that differs in truth-value in the two cases: The counterfactual "had the relevant proposition been false, you would have not believed it" ends up being true in the second scenario, where knowledge is present, and false in the first scenario, where knowledge is absent.

This is not merely an interesting curiosity or mere coincidence. For such counterfactuals seem to capture something that is intuitively of tremendous epistemic significance. Without committing ourselves to anything more precise at this point, we can say that when such a counterfactual is false—when, in other words, a true belief of yours is one you would have held on to even had it been false—then your belief (true though it may be) is not adequately sensitive

---

[46] Prominent proponents of this kind of view include ROBERT NOZICK, PHILOSOPHICAL EXPLANATIONS (1981); Keith DeRose, *Solving the Sceptical Puzzle*, 104 PHILOSOPHICAL REV. (1995); Alvin Goldman, *Discrimination and Perceptual Knowledge*, 78 J. OF PHILOSOPHY 771 (1975); and Fred Dretske, *Conclusive Reasons*, 49 AUSTRALASIAN J. OF PHILOSOPHY 1 (1971).

to the truth. Indeed, the fact that your belief is true may be understood as a kind of epistemic accident: after all, you would have held on to it even had it been false. And so it may be thought that there's no genuine epistemic achievement here on your part—you simply lucked out, as it were. But where the counterfactual is true (where, in other words, had the belief been false, you wouldn't have held it any longer), your belief does seem appropriately sensitive to the truth; you do seem entitled to some intellectual credit here. It is not mere accident that you believe truly (after all, had that proposition been false, you would have no longer held this belief).

We can now, then, introduce Sensitivity:

> **Sensitivity**: S's belief that p is sensitive $=_{df}$ Had it not been the case
>
> that p, S would (most probably) [47]not have believed that p.

Reflection on the two lottery scenarios lends initial intuitive support to the thought that Sensitivity is a necessary condition for knowledge; that insensitive beliefs do not constitute knowledge. But even without this contention that Sensitivity is a necessary condition for knowledge, it is sufficient for our purposes merely to assert it as an epistemological desideratum: i.e., that with other things held equal, a sensitive belief is epistemically better than an insensitive one.

At this point, we can return to statistical evidence: The two evidence scenarios in the Blue Bus hypothetical (market-share evidence and eyewitness testimony) are parallel to the two lottery scenarios (the belief is based solely on the odds versus being based also on the newspaper

---

[47] This qualification is rooted in counterfactual semantics. An account of the semantics of such counterfactuals lies outside the scope of this article. For further discussion of this qualification, using the most influential philosophical account, *see*: Enoch at al., *supra,* note 26.

report).[48] The analogy does not stem only from the intuitive similarity. Rather, in both cases, Sensitivity marks a step in the right direction. We have already seen that this is the case in the lottery scenarios. Let's revisit, then, some of the examples of statistical evidence: Suppose, that in both Blue Bus scenarios, we find for the plaintiff and against the bus company. Where we do so based on the individual evidence—the eyewitness testimony—our finding seems to be sensitive. Had it not been a Blue-Bus bus, would we have found the Blue Bus Company liable? Probably not. Our eyewitness is not infallible, of course, but she is fairly reliable. So had it not been a Blue-Bus bus, she would have probably not testified that it was, in which case, we would not have found the Blue Bus Company liable. In this scenario, then, the finding is appropriately sensitive. Things change, however, if we base our finding solely on statistical evidence, as in the second scenario. In that scenario, we find against the Blue Bus Company solely on the basis of the market-share evidence. Now, regardless of whether or not it was one of the Company's buses that caused the harm, the market-share data would not change. Either way, the Blue Bus Company still owns 70% of the buses that run in the relevant area; it's just that the bus that caused the harm ends up not being one of its buses (it's a Red-Bus bus). In this case, we would still have the exact same statistical evidence available to us, and so we would find the Blue Bus Company liable here too. Thus, by relying on statistical evidence, we render our findings insensitive.

---

[48] The evidence law literature on statistical evidence (or on the proof paradoxes) has recently started to address the more general epistemological issues here, but hasn't yet appreciated the full significance of comparison with lottery cases and the like. While Stein mentions a lottery paradox in a related context, he deals with a version thereof that is not relevant to our concerns. ALEX STEIN, FOUNDATIONS OF EVIDENCE LAW 67 (2005). Redmayne discusses our version of the paradox and explicitly draws an analogy between the evidence law cases and the epistemological literature on the (relevant kind of) lottery paradox, but he fails to mention the relevance of something like Sensitivity (instead, he discusses the related, albeit less appropriate here, Safety condition, and even then, only in a very brief way). Mike Redmayne, *Exploring the Proof Paradoxes*, 14 LEGAL THEORY 281 (2008). Ho briefly mentions the similarity but fails to put it to theoretical use, HOCK LAI HO, A PHILOSOPHY OF EVIDENCE LAW 168 (2008). No one, so far as we are aware, discusses in this context Sensitivity in sufficient detail to shed light on why the law should care about this distinction or to show how this way of understanding the distinction can help shed light on some related doctrinal features. In the more philosophical literature on statistical evidence, the parallel was made earlier and more often (*see, e.g.,* Judy Thomson, *Liability and Individualized Evidence*, in RIGHTS, RESTITUTION, AND RISK 236 (1986)).

Consider the gatecrasher hypothetical: Sensitivity can explain the difference between finding against the defendant merely on the strength of the strong statistical evidence and convicting on the strength of a probabilistically similar piece of direct individual evidence. If the percentage of gatecrashers is quite high, convicting a defendant merely on the basis of the statistical evidence makes the conviction insensitive. For even had the defendant not been guilty of gatecrashing—had he been one of the small number of law-abiding people who actually purchased tickets [49]—he would still have been convicted. The same does not hold true for the individual evidence.

When discussing epistemic and practical strategies for defending the distinction between statistical and individual evidence, we argued that if the phenomenon to be explained can be found beyond the evidence law problem, the required solution will tend to be of the epistemic kind. As can be seen from the lottery examples, we are now in a position to claim that the problem with the distinction between statistical and individual evidence indeed arises in far more contexts than the legal one. Indeed, the problem emerges even where there is no clear institutional context and where the instrumental considerations that may apply are few, weak, and different from those relevant to evidence law. Yet the problem is clearly the very same problem, and the reluctance to rely on statistical evidence is clearly the very same reluctance.[50]

---

[49] According to the standard semantics of counterfactuals, weare to evaluate the counterfactual's consequent in the possible world(s) closest to the actual world where the antecedent is true. So we are to change as little as possible. With the counterfactual in the text, then, we ought to imagine a world in which John is innocent but that is otherwise as similar to our own world as is possible. That world is one in which John entered the stadium having lawfully purchased a ticket. As mentioned above, an account of the semantics of such counterfactuals lies outside the scope of this article. For further discussion of the leading philosophical account, *see*: Enoch at al., *supra,* note 26.

[50] Pundik stresses that in some contexts—say, the context of giving a medical diagnosis—we are perfectly willing to rely on statistical evidence. He takes this as reason to believe that the problem is not as general as we claim in the text. *See* Pundik, *supra* note 25. But note that Pundik is interested in the extent to which we are willing to act on statistical evidence in different practical contexts. The point in the text above is that even outside the context of any action at all, we are unhappy to attribute knowledge (and perhaps even justified belief) to a subject who bases her beliefs on statistical evidence alone. And here the reluctance seems to be very general indeed.

This means that instrumental attempts at vindicating the distinction within the law of evidence—even if successful on their own terms—still fail to capture the phenomenon in its entirety and, so, are not as good an explanation as may be hoped for. The broad scope of the phenomenon accordingly mandates an epistemological justification. Focusing attention on Sensitivity provides just such a vindication. By highlighting Sensitivity, we do just that: its epistemic significance explains why there is something suspicious about statistical evidence across the board—in the legal context, in the lottery scenarios, and any other context in which we care about knowledge or beliefs being sensitive to the facts.

Another, closely related attempt at epistemological vindication of the sort needed here is explanatory in nature. Suppose some evidence misleads you; that is, although E was evidence for p, it turned out that not-p. In the case of statistical evidence, such circumstances invite a "you-win-some-you-lose-some" attitude. If 70% of the buses are owned by the Blue Bus Company, and that was our reason for thinking that the relevant bus was a Blue-Bus bus, then we knew from the outset that we were going to be mistaken roughly 30% of the time. But when individual evidence misleads, the outcome is different:[51] If we were to rely on the eyewitness testimony in ruling against the Blue Bus Company, and it turns out that the bus that caused the harm actually belonged to the Red Bus Company, some explanation for the discrepancy would seem to be warranted. Certainly, settling for a "you-win-some-you-lose-some" attitude seems inadequate. To phrase this in Martin Smith's terms, this means that only individual evidence *normically*

---

[51] A fuller discussion of mistakes than we need or can afford to conduct here would include—among other things—discussion of the ways in which the mistakes of many procedures are not themselves entirely random (newspapers printing erroneous lottery results are not equally likely to print all erroneous results, etc.). Note that such details (it seems reasonable to expect) will be closely connected to the requirement to explain the relevant kind of mistake, as emphasized in the text.

supports that which it is evidence for, so that when it misleads, an explanation seems to be in order.[52]

Now, given some plausible hypotheses about the relations between the truth values of counterfactuals and the nature of the relevant kind of explanation, this explanatory story—in terms of normic support—seems closely related to our Sensitivity story. And in a fuller epistemological discussion, we would have to ask which one (if any) is more basic, which one does the ultimate epistemological work here. But for our purposes, it suffices that Sensitivity-like counterfactuals capture—often enough, in sufficiently central cases—an epistemically relevant feature of the distinction between statistical evidence and individual evidence. We do not claim explanatory ultimacy for the relevance of Sensitivity. So even if what does the ultimate explanatory work here is something like normic support, as long as something like Sensitivity is still epistemically relevant, our vindication of the distinction between statistical and individual evidence holds.[53]

---

[52] Martin Smith, *What Else Justification Could Be*, 44 NOÛS 10 (2010). Smith goes much further than merely noting the explanatory point in the text. He thinks of normic support as grounding epistemic justification, indeed as doing so even against probabilities—so that one belief may be justified while the other is not even if the latter is more probable (for the thinker), as long as the former is normically supported by the evidence. Relatedly, Mark Schroeder suggested to us in email correspondence that while statistical evidence can support credences, it cannot support all-out beliefs (which, it follows, do not supervene on credences). Furthermore, two recent works by Timothy Williamson (2009) and (forthcoming) defending a sharp separation between chance and epistemic probability are relevant here. For our purposes, however, there is no need to discuss these interesting suggestions.

[53] Nevertheless, let us make the following points about Smith's interesting suggestion. First, he explicitly addresses Sensitivity (Martin Smith, *What Else Justification Could Be*, 44 NOÛS 10, 23 (2010)), rejecting it—if we understand him correctly—because misleading evidence that normally supports the relevant belief (and thus grounds epistemic justification) is not sensitive. But this, it seems to us, is beside the point: Of course *bad* or *misleading* evidence—individual evidence included—can fail Sensitivity. (If a bad or misleading eyewitness testifies that p, it may be the case that she would have so testified even had it been the case that not-p. Indeed, this is one of the standard ways of discrediting eyewitnesses.) The crucial point for us is that even *good* statistical evidence fails Sensitivity. Second, Smith's suggestion attempts to explain epistemic justification using thoughts about what does and what does not warrant explanation. And we agree, as stated in the text, that he's on to something important here, at least regarding the correlation between good, justification-grounding, and knowledge-grounding evidence and what mistakes call for explanation. Still, given the opacity of *calling for an explanation*—the question "What calls for explanation?" seems to us profound, and we do not know of any eye-opening answers to it—it is hard to see Smith's contribution as explanatory progress. Talking about which mistakes call for explanation (rather than about which evidence supports which beliefs) does not seem to reduce mysteriousness. Perhaps the truth values of Sensitivity-like counterfactuals—or the law-like connections that support them—are better candidates for being the more basic explanatory story.

3.   Should the Law Care about Sensitivity (or Knowledge)?

Let's recap. Using (one version of) the lottery paradox, we introduced and motivated the intuitive requirement that beliefs should be appropriately sensitive to the truth. We then formulated Sensitivity, according to which for A's belief that p to be sensitive it is necessary for it to be the case that had p been false, A would probably not have believed that p. We suggested that Sensitivity is plausibly considered an epistemically relevant condition (even if not quite necessary for knowledge and even if there is some deeper-still epistemological story—perhaps in explanatory terms—as to why Sensitivity is relevant). We then returned to the topic of statistical evidence and presented an epistemological vindication of the distinction between statistical and individual evidence, relying on Sensitivity. We argued that given the lottery paradox and related contexts where the very same phenomenon—the reluctance to rely on merely statistical evidence—is present, outside any legal setting, an epistemological justification (rather than an instrumental one) is precisely what me must look for.

But even if all of this is correct—indeed, even if Sensitivity is necessary for knowledge— why should the law of evidence care about knowledge? Why, in other words, should it make a legal difference whether a certain belief constitutes knowledge? In what follows, we first present the remaining doubts in more detail (section 3.1). We then proceed to present Sanchirico's incentive-based discussion of character evidence (section 3.2). Our solution to the puzzle (in section 3.3) is going to concede that the law should not much care about knowledge or, indeed, about epistemology in general. And we are going to endorse a generalization of Sanchirico's instrumental reasoning as vindicating the difference in the law's treatment of statistical and individual evidence. At the same time, we will point to the congruence between the instrumental reasoning and the epistemological discussion, and we will demonstrate that this is no mere

coincidence. The law should not care about epistemology, but it should care about something that is epistemically relevant—it should care about Sensitivity. Or so we are about to argue.

3.1   The Remaining Puzzle: Why Care about Knowledge?

Accuracy is paramount in legal fact-finding.[54] There may not be complete consensus as to how important it is that courts not err; or which mistakes it is more important to avoid;[55] or whether error avoidance is more or less important than other social goals.[56] But it is unequivocally agreed upon that courts ought to avoid too many, too "big" mistakes. Whatever the functions of the law, whatever goods it can help achieve, its ability to do so depends on fact-finding accuracy.[57] Furthermore, parties seem entitled to courts using procedures that will render mistakes that infringe on their rights sufficiently improbable.[58] Now, statistical evidence can help improve courts' reliability. Indeed, it can serve to minimize error just as much as individual evidence can. In cases of the kind we have been focusing on in this article, the relevant piece of statistical evidence is probabilistically on a par with the relevant piece of individual evidence.

---

[54] *See*, *e.g*., William Twining, *Evidence and Legal Theory*, 47 MOD. L. REV. 261, 272 (1984) ("Nearly all of the Anglo-American writers from Gilbert to Cross have shared essentially the same basic assumptions about the nature and ends of adjudication and about what is involved in proving facts in this context. There is undoubtedly a dominant underlying theory of evidence in adjudication, in which the central notions are truth, reason, and justice under the law. It can be restated simply in some such terms as these: the primary end of adjudication is rectitude of decision, that is the correct application of rules of substantive law to facts that have been proved to an agreed standard of truth or probability."); Dale A. Nance, *The Best Evidence Principle*, 72 IOWA L. REV. 227, 232-33, 236 (1988) ("The reasonably accurate determination of disputed factual issues is…the pivotal task to be performed at
Trial."). Even if these claims are too strong, and even if one rejects the notion of accuracy as the *exclusive* object of trial,, they do demonstrate the unequivocal and important role accuracy plays at trial.

[55] For instance, the beyond a reasonable doubt standard of proof in criminal proceedings reflects the notion that it is more important to avoid false convictions than it is to avoid false acquittals. *See* George P. Fletcher, *Two Kinds of Legal Rules: A Comparative Study of Burden-of-Persuasion Practices in Criminal Cases*, 77 YALE L.J. 880, 888 (1968).

[56] *See* William Twining, *Evidence and Legal Theory*, 47 MOD. L. REV. 261, 272 (1984) ("The pursuit of truth in adjudication must at times give way to other values and purposes, such as the preservation of state security or of family confidences; disagreements may arise as to what priority to give to rectitude of decision as a social value and to the nature and scope of certain competing values ... . But the end of the enterprise is clear: the establishment of truth.").

[57] For a more elaborate discussion of accuracy in legal fact-finding, see Louis Kaplow, *The Value of Accuracy in Adjudication: An Economic Analysis*, 23 J. LEGAL STUD. 307 (1994).

[58] Of course, other considerations, too, may be relevant to determining the right procedures. Robert G. Bone, *The Role of the Judge in the Twenty-First Century: Securing the Normative Foundations of Litigation Reform*, 86 B.U.L. REV. 1155, 1162 (2008).

Why exclude it, then? Is it really just because statistical evidence cannot ground knowledge or because it would provide a basis for one to believe the relevant proposition even if it were false? Yet why should the law of evidence care about knowledge or about epistemology more generally? It should care, undoubtedly, about truth, accuracy, and the avoidance of error. But why is it important that courts base their findings on *knowledge*? Insisting that the law should, after all, accord significant weight to knowledge or to epistemology in general amounts to a willingness to pay a price in accuracy. Indeed, excluding statistical evidence amounts to excluding what may be genuinely probative evidence. And this means that the legal value of knowledge—if it has legal value and if that value is what grounds the differential treatment of statistical and individual evidence—sometimes outweighs the value of accuracy; that, in other words, in order to make sure that courts base their ruling on knowledge, we are willing to tolerate more mistakes than we otherwise would have to and, in fact, a higher probability of mistake on this or that specific case. This just seems utterly implausible.[59] In sum, regardless of whether knowledge is important in other contexts,[60] it is very hard to assert that it has legal value, indeed enough value to justify tolerating higher rates and probabilities of mistakes.[61] The

---

[59] The problem described runs along parallel lines to something that has recently been receiving much attention in epistemology. For even in epistemology, it is not clear why we should care about knowledge. In this context, too, it seems, we should care about truth. And perhaps we should also care about the justificatory status of certain beliefs or inferences—whether, say, it's rational to have some belief given some evidence or whether we are entitled to infer certain propositions from certain others or some such. But we already know (at least since Gettier) that truth and justification do not suffice—not even taken together—for knowledge. So why should we care about whatever else is needed for knowledge? It makes sense, the thought goes, to aim at truth and, perhaps, also at justification. But why aim at knowledge? This is, to repeat, a controversial question that has recently been the focus of much epistemological attention. *See, e.g.,* D. Pritchard & J. Turri, *The Value of Knowledge. The Stanford Encyclopedia of Philosophy* (2011), http://plato.stanford.edu/entries/knowledge-value.

[60] For example, the epistemological value of knowledge is debated. For further discussion of the knowledge value problem *see:* Duncan Pritchard,& John Turri, *The Value of Knowledge*. THE STANFORD ENCYCLOPEDIA OF PHILOSOPHY, (2011) http://plato.stanford.edu/entries/knowledge-value/

[61] Note that this remains so even if we engage in "knowledge-first" epistemology, perhaps partly because of a (purported) constitutive relation between assertability and knowledge. *See* TIMOTHY WILLIAMSON, KNOWLEDGE AND ITS LIMITS (2000). This is why Ho's way of addressing the proof paradoxes in the law of evidence seems to us unsatisfying. *See* HOCK LAI HO, A PHILOSOPHY OF EVIDENCE LAW 140-43 (2008). Even given his Williamsonian assumptions, why should the law care about, say, assertability—indeed, why should the law care about it enough to tolerate a higher rate and probability of mistakes?

following thought experiment can assist in this regard: Suppose you have to choose the legal system under which your children will live, and you can choose only as between systems A and B. System A is epistemologically superior– perhaps its courts find against the defendant only when they know (or think that they know) of liability, or perhaps they only rule based on sensitive evidence, or perhaps only based on evidence that normically supports the conclusion that the defendant is liable. System B is not as good epistemically as System A. But system B is more accurate, so that the chances of System B erroneously imposing liability are lower than the chances of System A doing so. Which system do you choose for your children – the Epistemologically-Fine-But-Not-That-Reliable System A, or the More-Reliable-But-Not-That-Epistemically-Respectable System B? It seems to us that choosing System A amounts to an objectionable kind of epistemological fetishism, and this regardless of how much more accurate System B is.[62]

The point applies equally to the explanatory suggestion made above: Suppose, then, that statistical evidence cannot ground knowledge or even justification, because mistakes based on it do not call for explanation. Why should the law care in particular about avoiding mistakes that call for explanation? Mistakes that do not call for explanation seem—absent some story telling otherwise, at least—just as socially harmful and just as detrimental to the relevant party as mistakes that do call for explanation.

The point is not that the law—not even evidence law—should "care" only about accuracy. Other considerations (such as the inviolability of certain relationships, security, privacy, or dignity considerations and the opportunity costs of the litigation process) may, at times, trump

---

[62] This doesn't establish that knowledge (and the like) have no legal value whatsoever. Consistently with this judgment, perhaps knowledge does have legal value, but value that is always lexically inferior to that of accuracy. Though we suspect knowledge has no legal value at all, for our purposes here we are happy to settle on the somewhat weaker claim – to the extent that it has value, its value is lexically inferior to that of accuracy.

accuracy.[63] This is true in general,[64] and it may very well be true in our context as well. Perhaps, in other words, there are some cases of statistical evidence where other considerations trump accuracy. This may be the case with respect to certain profiling cases—where human dignity trumps accuracy. Our point is merely that *epistemological* considerations never seem to justifiably defeat considerations of accuracy when it comes to legal policy.

In this way, then, the story of Sensitivity as an epistemically relevant condition may be thought of not as a vindication of the distinction between statistical and individual evidence, but as a diagnosis of the relevant common intuitions and, indeed, perhaps even the beginnings of a debunking of these intuitions. This story helps to see what these intuitions track—something like evidence that can support knowledge. But now that we know that the law of evidence should not care about what these intuitions track, we should perhaps discard those intuitions, at least when it comes to the law. The Sensitivity-based epistemological story perhaps renders the relevant intuitions understandable, but not defensible as a cornerstone of legal policy. A different story is going to have to be told, then, if the distinction between statistical and individual evidence is to be vindicated. But that story, we will argue, is very closely related to the knowledge story. For in this story, though knowledge has no legal value, it will end up being indirectly relevant after all. To see this, we need to first consider Sanchirico's work on character evidence.

---

[63] Chris William Sanchirico, *Character Evidence and the Object of Trial*, 101 COLUM. L. REV. 1227, 1227 (2001) ("Most evidence scholarship takes as given that trial is at its core a search for truth, a sorting out of past events. Although commentators emphasize that truth seeking competes with other considerations, such as the sanctity of certain relationships, the dignity of the parties, and the opportunity costs of process, few would consider these rival claims part of the purpose of trial. They are rather constraints, to be accommodated or compromised. The reason to encroach at all upon these competing principles lies, by most accounts, in the value—inherent or instrumental—
of discovering what really happened."). For further discussion of the tradeoff between accuracy in legal fact-finding and the costs of trial, see Jonathan R. Macey & Geoffrey P. Miller, *Judicial Review of Class Action Settlements,* 1 J. LEGAL ANALYSIS 167, 178 (2009).

[64] See, for instance, Mitch Berman's interesting discussion of how the fact that indisputable or conclusive evidence is the standard used in many sports to allow instant-replay to reverse the initial call is best explained not as a concern for accuracy but by other values. Mitch Berman, *Replay,* 99 CAL. L. REV. 1683 (2011)

3.2   Sanchirico on Character Evidence

Criminal law exhibits an ambivalent attitude towards character evidence. Such evidence is typically admitted at the sentencing phase of trial,[65] but is inadmissible, in most contexts, at the guilt phase.[66] This is in spite of the underlying suspicion that this type of evidence has probative potential to facilitate a more accurate decision at the guilt phase too. If character evidence is deemed inadmissible at one phase of the criminal trial, what is the justification for admitting it at a later stage of the same proceeding? Why ban evidence of such probative value when deciding on the crucial question of guilt? In a fairly recent paper, Chris Sanchirico addresses this question.[67]

His core argument is that the rule prohibiting the use of character evidence for propensity reasons cannot be explained convincingly in terms of enhancing fact-finding accuracy. Rather, character evidence rules can be justified only by the broader scheme underlying evidence law—namely, the creation of incentives for proper out-of-court conduct. While character evidence has predictive (and, therefore, probative) value, claims Sanchirico, it has no incentive value: its presence dampens the incentives to refrain from the proscribed acts. The reason for this is that at the juncture most relevant for incentives—when an agent is deliberating whether and how to break the law—the relevant character evidence is already a given and can be used to his detriment whether or not he chooses to engage in the misconduct. This leads to a decrease in the marginal cost of engagement in the criminal activity ex ante. Ideally, in order to generate the

---

[65] In fact, a defendant's criminal record (alongside offense severity) is the weightiest determinative factor in sentence gravity. *See* Youngjae Lee, *Recidivism as Omission: A Relational Account*, 87 Tex. L. Rev. 571, 571 (2009).

[66] Character evidence is inadmissible if it is submitted for the purpose of showing that a defendant likely acted in conformity with a certain character trait. Fed. R. Evid. 404(a). Exceptions do exist. For example, the "Rule of Mercy" allows the defendant to introduce evidence of good character traits inconsistent with the charged conduct. If the defendant chooses to resort to the Rule of Mercy, the prosecution is then permitted to rebut the evidence with evidence of negative traits. Fed. R. Evid. 404(a)(1).

[67] Sanchirico, *supra*, note 63

efficient incentives, we would want the actor to know that the likelihood of his being (charged and convicted and) punished strongly depends on whether or not he decides to break the law here and now. The weaker this dependence, the weaker the incentive provided by the law to not engage in this specific criminal behavior. Thus, admitting character evidence at the trial stage would be counterproductive in terms of incentives. The prohibition on bad-character evidence promotes deterrence by avoiding a decrease in the marginal cost of engaging in criminal behavior.[68]

Sanchirico's argument underscores an important dimension of the purpose of evidence law: evidence law should be construed as being also (perhaps primarily) about supplying good incentives for primary behavior, behavior of agents outside the courtroom, and outside the legal procedure more generally. Of course, Sanchirico's claim need not be construed as asserting that giving the right incentives to primary behavior is the only normative consideration governing the rules on character evidence. But even if other underlying rationales do apply, Sanchirico has succeeded in drawing attention to another kind of consideration, one that it would be foolish for a legal system to ignore.

Sanchirico's article is devoted to character evidence, not statistical evidence. The similarities and distinctions between the two types of evidence will be further pursued below.[69] At this stage of the discussion, however, the relevant point is that his general strategy can be easily applied to statistical evidence as well. Think, for instance, about John, the potential gatecrasher who is now deliberating, weighing the options of purchasing a ticket or gatecrashing

---

[68] And given some plausible assumptions about the difference between the trial stage and the sentencing stage (such as which is more relevant for deterrence), perhaps this line of thinking can begin to validate the above-mentioned mixed treatment of character evidence.

[69] Character evidence may be thought of as a type of intra-personal statistical evidence. And just as with statistical evidence there is an intuitive feeling that the evidence is not sufficiently directly about the relevant individual, with character evidence, too, there is an intuitive sense that the evidence is not sufficiently directly about the relevant specific action.

or going home and doing something else altogether. We are assuming, of course, that John has no influence on the behavior of the others at and near the stadium. This means that he has almost no influence on the relevant statistical evidence—the percentage of those who enter the stadium without a ticket is only to a miniscule degree impacted by the outcome of John's deliberation. For all intents and purposes, he should think of it as already a given. If so, though, our willingness to rely on statistical evidence almost entirely eliminates whatever incentive the substantive criminal law can give John not to break the law. For if the statistical evidence is strongly against him— say, because 98% of those at the stadium are gatecrashers—John already knows that he will be convicted, regardless of whether or not he buys a ticket. And if the statistical evidence is not strongly against him, he knows that it will constitute strong exonerating evidence, whether or not he is guilty of gatecrashing. Either way, then, he might as well go ahead and gatecrash; whatever he decides will have negligible influence on the likelihood of his being punished.

Similarly, Sanchirico's analysis can also be employed in the Blue Bus context: if statistical evidence regarding the 70% market share of the Blue Bus Company were admissible at trial, deterrence would be undermined. This is due to the fact that the Blue Bus Company's expected cost of engaging in negligent behavior is a function of the difference between the probability that liability will be imposed given negligence and the probability liability will be imposed given engagement in the socially desirable behavior. Admitting the market share statistical evidence would enhance the probability of liability in the latter type of cases. In other words, introducing statistical evidence at trial (*ex post*) would lower the marginal cost of negligent behavior for the Blue Bus Company, thereby dampening its incentives to take the necessary precautions or to engage in the desirable level of activity (*ex ante*). At the same time, the Red Bus Company— holding only 30% of the market share-- will also be disincentivized to adopt the socially optimal

precautions or activity level so as to prevent the occurrence of negligent accidents. The reason for such adverse incentive effects is trivial-- the introduction of the statistical evidence will lower the prospects that it will be held liable for accidents, thereby reducing the expected cost of negligent behavior.[70]

3.3   Solution: The Instrumental Significance of Being Sensitive

At this stage, we find ourselves in the following predicament: The scope of the initial phenomenon to be explained—the resistance to relying on statistical evidence—is much wider than its appearance in the law of evidence, applying even in more purely epistemological settings (where nothing resembling the instrumental considerations relevant to the law is relevant). An epistemological explanation is thus called for, and we tried to formulate one in terms of Sensitivity. But the Sensitivity-based vindication is not germane to the law, certainly not in a way that could justify tolerating a higher rate of inaccuracy. In the legal context, what is needed, it seems, is an instrumental account, which we suggested (following Sanchirico's writing on bad-character evidence). But of course, the instrumental story cannot assist with the lottery paradox or other non-legal cases where talk of incentives seems out of place. Is there no way out then? Furthermore, is it mere coincidence that the epistemological and instrumental considerations align so neatly, at least when it comes to the law? The answer to these questions is no:

Think about incentives as in the case of John, who is deliberating about whether or not to purchase a ticket. He is now thinking in terms of conditionals: "If I gatecrash the stadium, they

---

[70] In addition to these precautionary distortions, the admissibility of statistical evidence may also impair market competition: The Blue Bus Company will absorb higher liability costs, which will lead to a decrease in its market share, and in general to less than fully effective incentives to grow. Another possibility is that each of the companies will attempt to hold less that 51% of the market share, so as not to suffer from the evidentiary disadvantage that a larger market share imposes. Either way, the statistical evidence may lead to a monopoly in the bus industry. *See:* Posner, *supra*, note 15 at 1510.

will punish me. If I don't, they won't." And typically, when at some point in time, some such conditionals are true, at a later point in time, (some of) the very same facts are captured by counterfactuals, or subjunctive conditionals. Suppose that John proceeds to gatecrash. Then his conditional "If I don't gatecrash the stadium, they won't punish me" now corresponds with the counterfactual that can be presented, say, when John is on trial, that "Had he not gatecrashed, we would not have punished him." And this counterfactual should sound familiar: it is the relevant instance of Sensitivity! In other words, though the epistemological story is not itself of legal value and though the instrumental story that is of legal value is not itself epistemologically respectable, both of them nonetheless stem from the same source—Sensitivity-style counterfactuals. Such counterfactuals are necessary both for knowledge (or are in some other closely related way epistemically relevant) as well as for a reasonably efficient incentives-structure. While the epistemological story and the instrumental story do not depend on one another, they are also not totally independent of each other either, for both are contingent on Sensitivity and related counterfactuals.[71]

What we end up with is the following: There is a need for an epistemological story, one that will treat lottery cases and legal cases alike. Sensitivity and its epistemic significance do that. There is also a need for a practical, most probably instrumental story, one that will vindicate the legal significance of the distinction between statistical and individual evidence without resorting to knowledge-fetishism. The generalization of Sanchirico's account does that.[72] But the

---

[71] As is often the case with explanations of coincidences, one may still ask the question whether the explanation itself is a mere coincidence. Is it, in other words, mere coincidence that Sensitivity and related counterfactuals are relevant both practically and epistemically in this way? Or is there perhaps some even-deeper story that can be told here? We do not know, but we can't deny that it would be exceptionally nice if such a deeper story were to exist.

[72] Our focus on Sanchirico's account neither entails nor presupposes that no other considerations can contribute here. But any other account would have to be checked for whether it coincides, as Sanchirico's account does, with the epistemic story of Sensitivity.

incentive-based account derived from Sanchirico's argument also relies on the truth of relevant counterfactuals, indeed the very same counterfactuals the epistemological account relies on. Sensitivity is (a part of) the answer to both the epistemological and practical questions. Note that what is relevant for policy recommendations purposes is the incentive-story, rather than the epistemological one (otherwise, we really would have a case of knowledge-fetishism). If there are cases, then, where the instrumental payoffs the incentive-account relies on are not in place, or if they are in place but are outweighed by other instrumental considerations, then even if relying on the relevant piece of evidence would violate Sensitivity, we do not see a practical reason not to rely on it.[73] In what follows, we will apply this theoretical structure to the legal doctrine, to demonstrate its capacity for solving some doctrinal puzzles, and offer prescriptions for legal reform.

PART III: SOLVING SOME DOCTRINAL PUZZLES

Evidence in the form of statistical assessment has extended far into the legal universe and plays a growing role in court.[74] Statistical inferences based upon genetic profiling, the demographic characteristics of populations, as well as cause-and-effect inferences derived from experimental studies and analyses of sample surveys are all used in the legal arena on a daily basis and unfold in a host of legal settings.[75] Thus, DNA is often used for purposes of

---

[73] The extent of the overlap between the epistemological considerations and the instrumental ones is to a large degree contingent. Perhaps if the overlap is significant enough, there are second-order considerations (having to do with administrative costs or the instrumental value of the simplicity of the relevant legal rules) against relying on (insensitive) statistical evidence even in cases where other instrumental considerations do not so imply. But it is quite possible that instrumental considerations will sometimes just not be there to back up the epistemological ones to the degree necessary to compensate for the loss in accuracy that is always involved in ruling out probabilistically respectable evidence. In those cases, our account will not support taking statistical evidence any less seriously than individual evidence.

[74] For further discussion of the origins and use of statistical evidence in trials, see DAVID W. BARNES & JOHN M. CONLEY, STATISTICAL EVIDENCE IN LITIGATION 3 (1986); Michael I. Meyerson & William Meyerson, *Significant Statistics: The Unwitting Policy Making of Mathematically Ignorant Judges*, 37 PEPP. L. REV. 771 (2010).

[75] Stephen E. Fienberg & Miron L. Straf, *Statistical Assessments as Evidence*, 154 J. R. STATIST. SOC. A. 410, 410 (1982).

identification[76] in criminal trials[77] and family law disputes alike.[78] Statistical assessments are

used in discrimination cases, most notably in the employment context to substantiate disparate

impact claims by pointing to a discrepancy between the proportion of minority group members

hired by the employer and the proportion of minorities among the group of qualified people in

the relevant market.[79] Courts also allow statistical assessments of a similar nature to be admitted

in voting contexts[80] in jury selection cases[81] and in constitutional and human rights cases.[82] In

tort law cases, statistical evaluation is used to prove causation.[83] It is also used for the calculation

of economic damages, i.e., when courts use base rates that identify incomes lost by similarly

situated individuals.[84] Such evidence is likewise used in contexts involving patent violations[85]

and competition law contexts.[86]

But we are not interested in all instances in which statistical assessments are used as

evidence in courts of law. What we wish to highlight, rather, is the use of statistical evidence in

---

[76] DNA random match probabilities ascertain the frequency with which a genetic profile would occur in a reference population and are considered admissible by almost all courts. *See* Jonathan J. Koehler, *When Do Courts Think Base Rate Statistics Are Relevant?*, 42 JURIMETRICS 373, 388(2002).

[77] *See, e.g.,* United States v. Hannigan, 27 F.3d 890, 893 n.3 (3d Cir. 1994).

[78] *See, e.g.,* Everett v. Everett, 201 Cal. Rptr. 351 (1984) (DNA testing in a paternity suit).

[79] *See, e.g.,* Anderson v. Westinghouse Savannah River Co., 406 F.3d 248 (4th Cir. 2005) (statistical evidence to prove employment discrimination on racial grounds); Scott Baer, *Defining "Otherwise Qualified Applicants": Applying an Antitrust Relevant-Market Analysis to Disparate Impact Cases,* 67 U. CHI. L. REV. 725 (2000).

[80] United States v. Blaine County, Montana, 363 F.3d 897 (9th Cir. 2004).

[81] Batson v. Kentucky, 476 U.S. 79 (1986); Michael O. Finkelstein, *The Application of Statistical Decision Theory to the Jury Discrimination Cases*, 80 HARV. L. REV. 338 (1966).

[82] Adarand Constructors, Inc. v. Slater.

[83] Statistical evidence of this nature played a central role in the tobacco litigation, *see, e.g.*, Blue Cross v. Phillip Morris, 113 F. Supp. 2d 345 (EDNY 2000); *see also In re Hanford Nuclear Reservation Litig*., 292 F.3d 1124 (9th Cir. 2002). But this type of evidence was rejected in other cases, *see, e.g., In re Fireboard Corp*., 893 F.2d 706 (5th Cir. 1990). *See also* Amit Pundik, *The Epistemology of Statistical Evidence*, http://ssrn.com/abstract=1134655. In product liability cases (when the market-share liability doctrine is implemented), statistical evidence is used to determine market share. *See* Sindell v. Abbott Laboratories, 26 Cal. 3d 588 (1980); *In re Agent Orange Product Liability Litigation*, 597 F. Supp. 740 (1984).

[84] Jonathan J. Koehler, *When Do Courts Think Base Rate Statistics Are Relevant?*, 42 JURIMETRICS 373, 398 (2002) (citing Raymond S. Strangeways & Michael T. Zugelder, *General Versus the Specific: Forecasting Wage Growth in Injury and Death Cases,* 8 J. LEGAL ECON. 1, 3 (1998); Contemporary Mission, Inc. v. Famous Music Corp., 557 F.2d 979, 981; Wilson v. B.F. Goodrich Co., 627 P.2d 1280 (Or. Ct. App. 1981)).

[85] David H. Kaye & David A. Freedman, *Reference Guide on Statistics, in* FEDERAL JUSTICE CENTER, REFERENCE MANUAL ON SCIENTIFIC EVIDENCE 83, 85 (2d ed. 2000).

[86] United States v. Columbia Pictures, 25 FRD 497 (SD New York 1960).

contexts similar to the Blue Bus and gatecrasher hypotheticals: namely, cases in which the statistical evidence provides a base rate for the defendant's liability[87] (as opposed to a factual characteristic that is isolated from the ultimate legal question of guilt or liability) and where it establishes a conclusion that is itself non-statistical. More specifically, we will focus on situations in which liability or guilt is grounded on inferences as to the defendant's conduct based upon reference to membership in a particular population or reference class.[88]

We begin with the two extreme points: DNA evidence, which courts tend to endorse, and propensity for crime evidence, which courts tend not to admit at the guilt phase of trial. In what follows, we will demonstrate how our theoretical framework can explain the differential legal regulation of these types of statistical evidence under prevailing law.

DNA

DNA evidence is an interesting illustration of the adaptability of our theory to the legal arena. It is interesting, because despite its statistical nature, discernible when it identifies the frequency with which genetic profiles occur in reference populations, courts seem rather happy to rely on it.[89] The theoretical model presented above can shed some light on this notable exception to courts' general resistance to statistical evidence. We start out by describing the legal doctrines governing the admissibility of DNA evidence and then show how our account of the

---

[87] Liability here refers to full liability, as opposed to the statistical liability that underlies the market-share liability doctrine. Under the market-share liability doctrine, liability is apportioned according to the expected harm posed by individual defendants as determined by their share of the relevant market. For further discussion of market-share liability, see Sindell v. Abbott Laboratories, Inc., 607 P.2d 924, 936 (Cal. 1980).

[88] Note, however, that another critical characteristic shared by the Blue Bus and gatecrasher hypotheticals is that in these cases, the statistical base-rate evidence is the single piece of evidence presented at trial—often termed in the literature cases involving "naked statistical evidence." Jonathan J. Koehler & Daniel N. Shaviro, *Veridicial Verdicts: Increasing Verdict Accuracy through the Use of Overtly Probabilistic Evidence and Methods*, 75 CORNELL L. REV. 247, 257 (1990). We, in contrast, will discuss the use of statistical evidence from both the admissibility and sufficiency perspectives.

[89] Andrea Roth, *Safety in Numbers? Deciding when DNA Alone is Enough to Convict* 85 N.Y.U.L. REV. 1130, 1132.(discussing the tendency of courts to rely on DNA evidence) .

distinction between statistical and individual evidence can explain and, to an extent, justify this exception.

DNA evidence first surfaced in American courtrooms in the 1980s and has since emerged as the most important forensic scientific breakthrough of the twentieth century,[90] leading to numerous convictions and hundreds of post-conviction exonerations.[91] DNA testing was depicted by one court as the "single greatest advance in the search for truth ... since the advent of cross-examination"[92] and has been analogized to "the finger of God."[93] Like other courts around the world, American courts—at both the federal and state levels—sweepingly admit DNA evidence,[94] in civil cases (mostly in paternity suits)[95] and in criminal trials. In 1988, the Florida Court of Appeals was the first appellate court in the United States to admit DNA evidence in a criminal proceeding, in *Andrews*.[96] This was followed by the *Jakobetz* trial,[97] which marked the first federal court admission of DNA evidence. By 1990, thirty-eight states had admitted DNA evidence,[98] and by the mid-1990s, most states were allowing DNA test results to be admitted as evidence in criminal trials. Whether it is the *Frye* standard, *Frye-Kelly* standard, or *Daubert*

---

[90] Joseph L. Peterson & Anna S. Leggett, *The Evolution of Forensic Science: Progress Amid the Pitfalls*, 36 STETSON L. REV. 621, 630 (2007).

[91] For further information about DNA exonerations, see http://www.innocenceproject.org/Content/Facts_on_PostConviction_DNA_Exonerations.php (last visited May 19, 2011).

[92] People v. Wesley, 533 N.Y.S.2d 643, 644 (N.Y. Sup. Ct. 1988).

[93] Kristen Bolden, *DNA Fabrication, A Wake Up Call: The Need to Reevaluate the Admissibility and Reliability of DNA Evidence*, 27 GA. ST. U.L. REV. 409, 409 (2011).

[94] Jonathan Kahn, *Race, Genes, and Justice: A Call to Reform the Presentation of Forensic DNA Evidence in Criminal Trials*, 74 BROOKLYN L. REV. 325, 325 (2009).

[95] *See* Carl W. Gilmore, *Challenging DNA in Paternity Cases: Finding Weaknesses in an Evidentiary Goliath*, 90 ILL. BAR J. 472 (2002).

[96] Andrews v. Florida, 533 So.2d 841 (Fla. Dist. Ct. App. (1988).

[97] United States v. Jakobetz, 955 F.2d 786 (2d Cir. 1992), *cert. denied*, 113 S. Ct. 104 (1992).

[98] L. Damon Whitmore, *Note, The Admissibility of DNA Evidence in Criminal Proceedings*, 39 WAYNE L. REV. 1411, 1411 (1993).

standard that is applied, DNA evidence is currently almost universally accepted in both federal and state courts.[99]

Another dimension of the role of DNA evidence in court is the evidentiary weight it is ascribed. In this context, too, courts have shown a general tendency to endorse DNA evidence, viewing this technology as bringing an unprecedented degree of certitude to the courtroom. For reasons which will be specified below,[100] courts are more prone to convict on the basis of DNA when it is corroborated by other types of evidence, but most courts do not rule out the possibility of convicting on DNA alone.[101] Indeed, when the probabilities have been sufficiently high, courts have convicted solely on the basis of DNA evidence.[102] Moreover, some courts have declared DNA evidence alone as a sufficient basis for conviction even in the face of conflicting eyewitness evidence.[103]

The theoretical foundations presented further on can offer some insight into the doctrinal treatment of DNA evidence. Before proceeding, however, the following three preliminary points should be made: First, our discussion will not relate to the scientific foundations of DNA

---

[99] *See:* Frank B. Ulmer, *Using DNA Profiles to Obtain "John Doe" Arrest Warrants and Indictments*, 58 WASH. & LEE L. REV. 1585, 1598 (2001) (claiming that all U.S. jurisdictions allow DNA evidence to be admitted into court). As a practical matter, the analysis of the evidence presented by forensics labs is subjected to close scrutiny by the court. Most states require statistical probability analysis to interpret DNA "match" evidence, as a precondition to admissibility. *See*, *e.g.,* People v. Coy, 620 N.W.2d 888, 897-99 (Mich. Ct. App. 2000).

[100] *See*: ###

[101] Roth, *supra*, note 89, at 1155 (describing the "emerging phenomenon of 'pure cold hit' DNA prosecutions in which the entirety of the government's case against the suspect, aside from his prior conviction, is a DNA profile match or a match accompanied only by general evidence").

[102] *Id. See* also: Brooke G. Malcom, *Convictions Predicated on DNA Evidence Alone: How Reliable Evidence Became Infallible*, 38 CUMB. L. REV. 313, 315 (2008) (listing court decisions ruling that DNA is sufficient for conviction). This is also the case in England. *See:* Adams, [1996] 2 Crim. App. at 468 where the court held: "there is…nothing inherent in the nature of DNA evidence which makes it inadmissible in itself or which justifies a special, unique rule, that evidence falling into such a category cannot found a conviction in the absence of other evidence." *Id.* At 1155

[103] *See* Brooke G. Malcom, *Convictions Predicated on DNA Evidence Alone: How Reliable Evidence Became Infallible*, 38 CUMB. L. REV. 313 (2008) (citing People v. Rush, 630 N.Y.S.2d 631, 634 (Sup. Ct. 1995), *aff'd*, 672 N.Y.S.2d 362 (App. Div. 1998); State v. Toomes, 191 S.W.3d 122, 131 (Tenn. Crim. App. 2005); Roberson v. State, 16 S.W.3d 156, 172 (Tex. App. 2000); Springfield v. State, 860 P.2d 435, 453 (Wyo. 1993)).

evidence. Rather, it will be conducted under the assumption that DNA is highly probative. Indeed, we are going to restrict our attention to just those cases where the probability that the accused is guilty, given that there is a DNA match, is extremely high, though not 1. (In symbols: $P(G|M)=1-\varepsilon$, for a positive but arbitrarily small ε.) Second, the discussion will focus on the hard cases of "cold hit" DNA: namely, where DNA is the only evidence and where it was obtained without some prior suspicion (in other words, DNA evidence was obtained, run against some database, and a match was found, not that a suspect was pinpointed and then tested for DNA). And third, we will be restricting our attention to the use of DNA evidence as evidence for the prosecution (in a criminal case).[104] With these stipulations in place, then, can anything be said in favor of using DNA evidence, especially given the background of suspicion towards statistical evidence in general? How do we solve this doctrinal puzzle?

One obvious feature that distinguishes DNA evidence from most other kinds of statistical evidence is the *extremely* high level of probability underlying it. This suggests one reasonable, albeit unexciting, solution to the doctrinal puzzle, in terms of the relative value of accuracy: Although it may be the case that the same objections to statistical evidence are no less applicable with high probability evidence, the value of accuracy is much weightier in cases with such levels of probability. This very high probability of the evidence is the core of the difference in the rulings in *Kaminsky v. Hertz*[105] and *Smith v. Rapid Transit* in *Kaminsky*, as in DNA cases,

---

[104] At least in the criminal case, the legitimacy of DNA evidence as exonerating evidence is clear enough not to be interesting—it is often non-statistical in nature, and in any case, the relevant high probability of accuracy certainly suffices for reasonable doubt.
[105] Corp. 288 N.W. 2d 426 (Mich. Ct. App., 1980).

considerations related to the value of accuracy outweighed the standard reasons for not relying on statistical evidence.[106]

Another possible account for the preferential treatment of DNA evidence is the incentive story. Recall our generalization of Sanchirico's theory, according to which relying on statistical evidence will create inefficient incentives for, say, the Blue Bus Company as well as its competitor, the Red Bus Company. Sanchirico's reasoning relied on the fact that both companies would be in a position to know that their chances of being found liable are unrelated to their relevant conduct (because liability is determined by their market share). But perhaps in DNA cases—certainly, in *most* DNA cases— the potential offender has no access to such knowledge. Most people possess little knowledge regarding their genome sequence, DNA profile, and its frequency in the relevant population. So, unlike readily available statistical evidence, the incentive story arguably does not apply here, and there is no incentive-based reason to ignore genuinely probative statistical evidence. Despite the plausibility of this explanation, there is room to question whether it captures the full picture. Suppose, for instance, that in addition to DNA, we can also check for DNA* match. DNA* shares with DNA its incentives-relevant properties (things like what knowledge is and is not available ex ante), but is much less effective probabilistically, so that the probability that the accused is guilty given a DNA*-match is, let's say, around 70%. In such a case, too, the incentive story collapses**.** Yet the intuitive reluctance to relying on statistical evidence is still as strongly present.[107]

---

[106] This corresponds with jurors' tendencies to convict on the basis of high probability statistical evidence. *See* Kevin Jon Heller, *The Cognitive Psychology of Circumstantial Evidence*, 105 MICH. L. REV. 241, 301 (2006) ("Although jurors are extremely sensitive to deviations away from certainty… they are generally willing to convict on the basis of probabilistic evidence that ... establishes a 0.995 likelihood of the defendant's guilt.").

[107] Perhaps, then, we should think of the incentive story as justifying reliance on DNA evidence and debunking our intuitions about DNA*-evidence. Another possible line of thinking would refer to the brief suggestion in our discussion above, that at least

Yet, both the high probability level account and the incentive explanation hold, without a doubt, at least some of the relevant truth about the treatment of DNA evidence. And in fact, we are committed to such non-epistemic stories being what guides legal policy. But the Sensitivity requirement can add an additional dimension to the explanation of the DNA conundrum: Suppose that A is convicted solely on the basis of a DNA match. Had A not been guilty, would we have still convicted him? Well, in fact, had A not been guilty, but the DNA evidence had nonetheless matched his DNA, we *would* have still convicted him. But this is a different counterfactual, one that invites us to travel to a different possible world. The counterfactual that is relevant here is the one we began with, where there is considerable pressure to answer in the negative—had it not been A, we *wouldn't* have found a DNA-sample matching A's DNA at the crime scene. The reason that this is the relevant counterfactual is rooted in possible-world talk According to the dominant view in the semantics of counterfactuals, we are to evaluate the counterfactual's consequent in the possible world or worlds closest to the actual world where the antecedent is true.[108] There's considerable intuitive pressure to think that a possible world in which A is innocent and yet the DNA sample matches his DNA is *farther away* from the actual

---

one of the problematic features of statistical evidence is that systematic reliance on it guarantees some false decisions, indeed, false convictions (think again of the variant of the gatecrasher hypothetical, where we indict all those attending the stadium). This problem—of a guaranteed false decision—doesn't seem to be relevant to DNA evidence, where systematically relying on such evidence does not yield a similar guaranteed result. It does, of course, have the *probabilistic* result that we are *highly likely* to falsely convict. But any system that convicts has this result. This explanation, then, also fails to do all the necessary work here. This is partly because of the doubts mentioned in section 1 (roughly: why think that the difference between a guarantee and a ridiculously high probability that is still smaller than 1 makes all this difference?), but also for the following reason: We can imagine a variant of the gatecrasher hypothetical where the guarantee of a false conviction is absent, say, if many of those attending the stadium escaped before the police arrived. Indeed, suppose (again, as we did in section 1) that only one person was apprehended at the stadium, and only he is brought to trial. In this case, relying on statistical evidence does not have the result of a guaranteed false conviction. But the reluctance to rely on statistical evidence is no less strong. Thus, the guarantee story can't be the full story here. In general, it is an interesting exercise to construct a parallel gatecrasher case for any story about DNA evidence. For instance, in the case of DNA, we typically don't even know (in a specific case) that there is another person whose DNA would match that found at the crime scene. So perhaps we should think about a gatecrasher case where we don't know that some people actually bought tickets; all we have is the probability that some did. Things get complicated. But even in this last version of the gatecrasher case, it seems that the law would not convict solely on the basis of the statistical evidence, and it also seems that this is as it should be. So DNA remains special.

[108] For further discussion of the leading philosophical account regarding possible world talk , see: David Enoch at al., *Supra*, note 26.

world than a world in which A is innocent and no matching DNA sample is found at the crime scene. If this is true, then in the DNA case—unlike other statistical evidence cases—Sensitivity is satisfied. So DNA may be special even epistemically, according to the account in this article.

The explanatory test can further reinforce this point: If we convict someone of gatecrashing solely on the strength of the statistical evidence and later find out that she was a rare ticket-buyer, we do not (nor does it seem that we should) look for a deep explanation—we played the odds, and lost. But in a case where we convict A based purely on DNA evidence and, later on, find out she was innocent, we do look for a deeper explanation, and justifiably so.[109]

In sum, the theoretical account developed here offers an explanation for the exceptional treatment of DNA in contrast to the usual wariness accorded statistical evidence. Unlike other types of statistical evidence, DNA evidence does seem to be sensitive: for had the defendant not committed the offense, we would, in all likelihood, not have found her DNA on the scene, and so we would, in all likelihood, not have convicted her. [110] Moreover, the incentive problem with statistical evidence, described above, does not seem to apply to DNA evidence, because potential perpetrators are very rarely in a position to know whether DNA collected at the crime scene will match theirs.

Propensity-for-Crime Evidence at the Guilt Phase of Trial

---

[109] For a more elaborate discussion of the proximity relation between worlds, see David Lewis's account, referring to the number and size of miracles needed to move from the actual world to the relevant possible world. This fits nicely with the points made in the text: it would seem like a fairly big miracle for A to be innocent and yet for the DNA sample from the crime scene to match A's DNA.

[110] A possible objection relies upon another conceivable counterfactual which can be assumed—had the defendant not committed the offense, *and the evidence would have been present*, would we still have convicted him. This counterfactual needs to be distinguished from the one assumed in the text. Though these two types of counterfactuals sometimes align, this need not always be the case as the DNA example shows.

Propensity-for-crime evidence is a category of base-rate evidence structured in one of two possible forms (leading to different levels of probability): *a certain percentage of the people who commit the crime with which the defendant is charged share a certain demographic or economic trait, which also characterizes the defendant*; or, alternatively,: *a certain percentage of the people who share a particular demographic or economic trait that also characterizes the defendant engage in the criminal activity with which the defendant is charged*. Propensity-for-crime evidence has probative value because it affects the probability of guilt. For example, the statistical evidence regarding the rate at which convicted felons engage in a subsequent offense has probative value in the trial of a convicted felon, because it informs the prior probability of another offense.[111] The same holds true in the context of other demographic traits, such as age, gender, or ethnicity. Statistical evidence pointing to the greater propensity for crime of young males informs the prior probability of guilt in a trial involving a young male defendant.[112] Yet despite the probative value of propensity-for-crime evidence, under prevailing doctrine it is generally disallowed at the guilt phase of trial.[113]

At first glance, propensity-for-crime evidence could be mistakenly conceptualized as the mirror image of evidence that reconstructs past events.[114] But in the context of providing incentives for engaging in certain forms of primary behavior (as well as from the perspective of Sensitivity), the past-future distinction carries great significance. In line with Sanchirico's

---

[111] *Id.*

[112] For further discussion of this hypothetical, see Redmayne's "future violence" example in Redmayne, *supra* note 3 at 281. The difference between predictive evidence and trace evidence can be exemplified as follows: on the question of whether A hammered a nail into the wall, the scratch on the head of the nail serves as trace evidence, whereas the fact that A is a carpenter is predictive evidence. In predictive evidence, the direction of causation goes from the evidenced phenomenon to the conduct in question. In trace evidence, the causal relationship runs in the reverse direction. *See* Sanchirico, *Supra,* note 63

[113] *See* Stephens v. State, 774 P.2d 60, 64 (Wyo. 1989); Hall v. State, 692 S.W.2d 769 (Ark. Ct. App. 1985). According to Koehler, this type of evidence is rejected most frequently in child abuse cases. Koehler, p. 384.

[114] *See, e.g.*, *Note: Prediction Markets and Law: A Skeptical Account,* 122 HARV. L. REV. 1217, 1229 (2009); Denise Meyerson, *Risks, Rights, Statistics and Compulsary Measures,* 31 SYDNEY L. REV. 507, 521 (2009).

argument, incentive-setting is determined by the changes to an individual's payoffs following the commission of particular acts, whereas predictive evidence exists irrespective of the commission of such acts. Predictive evidence, in other words, lacks incentive value. In this particular context, if the predictive statistical evidence relating to the higher propensity for violent crime were to be admitted into the courtroom, as proof of guilt ex post, the marginal cost of engaging in the criminal activity, ex ante, would be lowered. As explained earlier, admitting evidence relating to the propensity for violent crime would enhance the probability of conviction, both for those who commit violent crimes and for those who refrain from such behavior. The disincentive for engaging in crime would thus be weakened accordingly. In the extreme case that such propensity-for-crime evidence could serve as an exclusive basis for conviction, John may face an almost identical payoff in a choice between engaging in the criminal act and not doing so and may, therefore, easily opt to commit the act.[115] Inadmissibility of such evidence therefore promotes deterrence.[116]

As discussed above, Sanchirico applies this basic intuition to the context of bad-character evidence, arguing that admission of bad-character and prior-acts evidence would dampen incentives and impair deterrence by lowering the marginal cost associated with engaging in the criminal behavior. The extension of the incentive-based intuition to the context of bad character, and especially to the sub-category of prior convictions, can, however, be challenged. Sanchirico's analysis focuses on the legal payoff in the period of time following the act suggesting bad character or—in the case of prior convictions—following the involvement in the

---

[115] Sanchirico, *supra,* note 63

[116] There are those who claim that admission of character evidence may actually further deterrence, for individuals with prior convictions can be expected to take special measures in order to avoid situations that might lead to their arrest, in light of the greater probability of conviction once indicted. *See*, *e.g.*, Roger C. Park & Michael J. Saks, *Evidence Scholarship Reconsidered: Results of the Interdisciplinary Turn*, 46 B.C. L. REV. 949, 1016 (2006). This matter lies outside the scope of our discussion.

first offense (as can be deduced from the existence of prior convictions or acts). Indeed, the legal payoff faced by the defendant after the initial act of crime (for which she was already convicted) would be suboptimal if information as to prior conviction were admitted during the guilt phase of any subsequent trial, for the reasons discussed above. But it is not clear why, when devising rules for optimal deterrence, emphasis should be placed exclusively on the incentive structure and on the legal payoff in the period of time following the first offense or initiation of the bad act. Individuals should also be deterred from committing the first act of crime, suggesting that the prism through which the proper incentive structure is viewed and constructed ought to extend to the period of time prior to the involvement in the first offense. Defining the relevant timeframe is crucial in this context, for when focusing on the legal payoff prior to the first crime, a rule that permits information of prior convictions to be submitted to the court could actually further deterrence. It would enhance the expected cost of engaging in the first criminal act (as the expected sanction would now include a greater probability of conviction in any future trial).

This type of problem does not surface in scenarios of propensity-for-crime evidence based on gender, age, ethnic, or demographic characteristics (or other paradigmatic cases in the general category of predictive evidence). Unlike the feature of engagement in criminal activity underlying the prior convictions category of evidence, traits such as age or gender are not a matter of choice for the individual, nor are they problematic from a social-welfare perspective. Exclusion of propensity-for-crime evidence relating to such non-voluntary, welfare-neutral characteristics, at the guilt phase of trial, will more likely result in amplifying deterrence than the bad-character scenario underlying Sanchirico's analysis. Sanchirico's model, in other words, falls prey to a major critique that our use of his model does not fall prey to.

In sum, the inadmissibility of propensity-for-crime evidence relating to features of a non-voluntary nature and/or of a neutral quality from a social-welfare perspective can be justified in light of the role of evidence law in regulating primary activity. The incentive-based analysis does not substantiate, however, a clear-cut case for excluding bad-character evidence at the guilt phase of trial. Although Sanchirico's analysis was intended to explain the inadmissibility of bad-character evidence, it actually offers a better reasoning for more paradigmatic types of propensity-for-crime statistical evidence, such as predictions based upon age, gender, or ethnicity.

The incentive-based analysis can be complemented by the epistemological perspective of our theory: Propensity-for-crime evidence, it has been claimed, poses a fundamental challenge to Judith Jarvis Thomson's analysis, for it satisfies – often enough – the causal connection requirement.[117] The problem, however, is not related to the question of the existence of a causal connection between criminal conduct and age or familial background, but rather stems from the *direction of the causal connection*. In propensity-for-crime evidence, the direction of causation is reversed and runs from the evidenced phenomenon to the conduct in question. Put differently, in propensity-for-crime cases, the evidence provides information as to the conduct in question, but it is not affected by it.[118] Due to this reverse direction of causality (i.e., the fact that the evidence is not a result of the act in question), propensity-for-crime evidence does not satisfy the requirement of Sensitivity. The Sensitivity perspective thereby offers a complementary account of the prevailing doctrine. Although it is the incentive perspective, rather than the epistemological story, that is relevant for policy purposes, the two accounts align here.

---

[117] Redmayne, *supra* note 3 at 281; Wasserman, *supra,* note 40;

[118] Sanchirico, *supra* note 63

Predictive evidence is routinely admitted into court during the sentencing phase of trial, but since our article deals with the use of statistical evidence for purposes of imposing liability, we will only briefly relate to this point. Predictive evidence used at sentencing includes the defendant's age, her rehabilitative potential, and other proxies for her future dangerousness. [119]

Other important evidence relating to the offender's character admissible during sentencing under the U.S. federal sentencing guidelines is her criminal record. [120] Alongside offense severity, prior convictions are the weightiest determinative factor in sentence gravity. State guidelines likewise incorporate prior conviction provisions.

Unlike the guilt phase of trial, admitting predictive evidence at the sentencing phase of trial is expected to further deterrence. In light of the social costs of criminal punishment, imposing a unitary sanction that is grave enough so as to deter all (or almost all) potential offenders entails dead weight loss. Some individuals may be effectively deterred from engaging in the unwanted conduct when exposed to a more lenient, and thereby less costly, sanction. The severe punishment is "wasted" upon them. Tailoring criminal punishments to the "deterrence proneness" of individuals (this can be described as a form "price discrimination") enhances deterrence level per given cost of punishment. Predictive evidence, regarding the age of the defendant, her opportunity costs, and rehabilitative potential, serves as a proxy for such "deterrence proneness." Admitting this evidence at the punishment phase of trial enhances the expected cost of engaging in criminal activity for those high propensity offenders, for those individuals who are less easily deterred. At the same time, it allows for a less costly deterrence for low propensity offenders.

---

[119] 28 U.S.C. § 994(d).

[120] 28 U.S.C. § 994(d)(10)

A similar intuition exists with respect to evidence of prior convictions. Individuals who have been convicted in the past need enhanced penalties to be optimally deterred from re-offending, for by engaging in criminal behavior such individuals have revealed their proclivity for criminal activity. Moreover, when individuals have been subjected to prior criminal conviction and punishment, any subsequent sanction imposed upon them may be effectively eroded: individuals with criminal records have lower opportunity costs; there is a decrease in the marginal cost of imprisonment years; and the additional reputation costs entailed in a greater number of convictions decrease as the number of convictions rises. Holding all things equal, the result of such erosion is a weaker deterrent effect of the criminal sanction for repeat offenders as compared to first-time offenders. In order to achieve the same level of deterrence, repeat offenders should thus be subjected to a higher expected punishment.

Of course, creating incentives for optimal deterrence isn't the only normative consideration policymakers should take into account when formulating rules of evidence. Considerations relating to equality before the law, to the dignity of the defendant, and to the preservation of certain social relationships also play a role. For these reasons, personal characteristics of the defendant based upon race, sex, national origin, creed, or socioeconomic status are not incorporated into the sentencing guidelines. But these are all the exception rather than the rule, for as a general matter, admitting statistical evidence of a predictive nature at the sentencing phase of trial amplifies incentives and furthers deterrence, while admitting these types of evidence at the guilt phase of trial tends to impair deterrence.

Incriminating Statistical Evidence

With the exception of DNA and fingerprint evidence, the use of statistical evidence for conviction purposes is extremely uncommon and very controversial. Note, that not any evidential use of statistics for incriminating purposes is analogous to the blue bus scenario depicted above. Often, statistical pieces of information are used as a part of a general inference to the best explanation. Thus, often – in the law and elsewhere – if there is a phenomenon that calls for explanation, and among competing possible explanation one explanation is better, we have reason to believe the better explanation[121]. And sometimes, probabilities make a difference for how good an explanation is. Thus, if among two suspects one has a motive and the other doesn't, this makes the hypothesis that the first suspect committed the act more plausible than that the second one did. Similarly, if the hypothesis that the butler did it explains all the relevant (non-statistical!) pieces of evidence well, whereas the hypothesis that the gardener did it leaves many details hostages to unlikely (that is, probabilistically low) coincidences, and if we know independently that either the gardener or the butler did it, then we have strong reason to believe that the butler did it, partly based on the low statistics regarding the coincidence "explanation". But this is not what is going on in the Blue Bus or Gatecrasher cases. In those cases, there is no interesting sense in which the relevant hypothesis (that it was a blue bus; that the accused crashed the gates) better explains, compared to alternative hypotheses, some relevant phenomenon. Rather, it's the more pure use of the statistical evidence itself that is at stake. The distinction between these two ways of using statistical information – as a part of a seemingly legitimate inference to the best explanation, or as purely statistical as in the cases we've been discussing – is important, as the discussion that follows shows.

---

[121] There is much discussion, in epistemological contexts, of inference to the best explanation, and some of it is quite critical. For an overview and many references, see Igor Douven (2011) "Abduction", *The Stanford Encyclopedia of Philosophy*, available online here: http://plato.stanford.edu/entries/abduction/
The details and criticisms are not relevant for the limited use to which we put inference to the best explanation here.

The first of the criminal law cases in which statistics was used for conviction purposes was the notorious 1899 trial of Alfred Dreyfus, a Jewish captain in the French Army. Amongst the evidence used to convict Dreyfus were letters he had written, which, according to the prosecution, were cipher messages. The prosecution attempted to substantiate its claim with statistical evidence, by showing a disproportionate frequency to certain letters of the alphabet relative to the standard pattern of French prose.[122] These data were found by the court to be sufficient basis to convict Dreyfus. It should be noted that the Dreyfus case is not analogous to Blue Bus or the Gatecrasher cases. Rather, it is a case of inference to the best explanation premised upon statistical data: Thus. within the context of the case, the court seeks what best explains the data, one hypothesis being that Dreyfus is a spy, and the other hypothesis being that it is mere coincidence. And then the (supposed) low probability of the coincidence serves to show that the second explanation is better. was subsequently challenged for the inaccuracy of the data (as opposed to its actual use).[123]

Another well-known and very problematic use of statistics in criminal trial was *People v. Collins*.[124] An eyewitness account of the robbery of an elderly woman described a blonde woman and bearded African-American man fleeing the scene in a yellow car. The defendant and his wife, who generally fit the eyewitness description, were arrested and brought to trial. The prosecution offered statistical evidence of the overall likelihood of a couple meeting these criteria to be 1 in 12,000,000. The prosecution argued that given this low probability, the defendant must be guilty. This, in other words, is another case of inference to the best

---

[122] David H. Kaye, *The Admissibility of "Probability Evidence" in Criminal Trials,* 26 JURIMETRICS J. 342 (1986).
[123] *Id.*
[124] *See*, *e.g.,* Laurence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329 (1971); Michael O. Finkelstein & William B. Fairley, *A Bayesian Approach to Identification Evidence,* 83 HARV. L. REV. 489 (1970).

explanation analogous to the Dreyfus trial The jury returned a guilty verdict, but the conviction was reversed on appeal by the California Supreme Court, partly due to its inadequate statistical basis—namely, due to the fact that the evidence, in itself, did not meet standards of *statistics* (as opposed to a rejection of base-rate evidence *per se*).[125]

The 1991 *Shonubi* case[126] is an example of incriminating statistical evidence of the blue bus type where statistical evidence is being used at a criminal trial, albeit for sentencing purposes. Shonubi, a Nigerian citizen residing in the U.S., was arrested at JFK Airport for smuggling heroin by way of "balloon swallowing." When taken into custody, he was found to be carrying 427.4 grams of heroin. Shonubi was tried in a federal district court and convicted of possession and importing of heroin.[127] At his trial, it was determined that he had made seven other heroin-importing trips, and this finding was not challenged in subsequent proceedings. According to the federal sentencing guidelines, the severity of Shonubi's sentence should depend not only on the quantity of heroin imported on his last trip, for which he was convicted, but also on the aggregate amount of heroin that was imported during the seven known earlier heroin-importing trips. In order to determine this aggregate amount, the sentencing judge simply multiplied the figure of 427.7 grams by eight, resulting in an aggregate amount in excess of the 3000 gram threshold for sentence enhancement (corresponding to a 34 base offense level). Shonubi appealed his sentence, and the Second Circuit Court of Appeals vacated the sentence and remanded the case to the trial court for resentencing.[128] At the resentencing, the state presented data collected by the Customs Service relating to the amounts of heroin found in possession of 117 other Nigerian drug smugglers who had transported heroin via the same

---

[125] 438 P.2d 33 (Cal. 1968).
[126] 802 F. Supp. 859 (E.D.N.Y. 1992) (often referred to as "*Shonubi I*").
[127] *Id.*
[128] 998 F.2d 84 (2d Cir. 1993) (often referred to as the "*Shonubi II*" case).

"balloon swallowing method" in the time-period between Shonubi's first and last smuggling trips. According to this data, there was a 99% probability that on his seven prior trips, Shonubi had smuggled at least 2090.2 grams of heroin, which, added to the 427.4 grams smuggled on his eighth and final trip, amounted to approximately 2500 grams. Based on this data, the resentencing judge concluded that Shonubi had smuggled between 1000 and 3000 grams of the drug and sentenced him accordingly.[129] Shonubi appealed yet again, and the sentence was re-vacated by the Second Circuit. The Second Circuit ruled that the statistical data did not constitute "specific evidence" of the amount of heroin that Shonubi had smuggled during his seven previous trips. [130] Consequently, back in the District Court, Shonubi was sentenced only for the 427.4 grams of heroin found on him upon arrest. [131] Here, again, the rejection of the statistical evidence was focused on the application of the data in the specific circumstances of the case.

Another well-known use of statistics at criminal trial—albeit of the Dreyfus type case-- is the 1999 English case of Sally Clark . On November 9, 1999, Clark, a solicitor, was convicted of murdering her first two children. The death of her firstborn son, who had died three years earlier at less than three months of age, was originally diagnosed as a case of Sudden Infant Death Syndrome. After the death of her second baby, however, who died a year later at the age of two months, she was charged with murdering both infants. During the trial, an expert pediatrician testified for the prosecution that the chances of two SIDS deaths in a single family are 1 in 73 million. This calculation was later found to be statistically flawed. Regardless, Clark was

---

[129] 895 F. Supp. 460 (*Shonubi III*).
[130] 103 F.3d 1085 (2d Cir. 1997) (*Shonubi IV*)
[131] 962 F. Supp. 370, 375 (E.D.N.Y. 1997) (*Shonubi V*). For an elaborate discussion and analysis of the *Shonubi* trials, see Peter Tillers, *Introduction: Three Original Contributions to Three Important Problems in the Law of Evidence*, 18 CARDOZO L. REV. 1875 (1997).

convicted,[132] and her conviction was later upheld by the Court of Appeal.[133] A second appeal was allowed due to the recovery of exonerating evidence, after which Clark was set free (after more than three years in prison).[134] However, in this case too, it was the controversial statistical quality of the data that was found to be lacking and not an objection principle to the use of base-rate evidence for incriminating purposes.

The rarity of cases like those described above underscores the general reluctance courts show to ground convictions on statistical evidence, the exceptions being DNA and fingerprint evidence. The majority of the difficulties and controversy in the cases where statistical evidence was used at trial courts related to their faulty statistical basis and to questions about their validity. Yet it seems that a more principled objection to the very use of such base-rate statistics for conviction can be inferred from the general trend of resistance to this type of evidence and perhaps from these unique cases too.[135]

Accuracy considerations can help explain the suspicion towards statistical evidence for conviction purposes. As argued earlier on, the rules of evidence are designed with the primary purpose of promoting fact-finding accuracy.[136] Statistical evidence seems, at first glance, to do just that: minimize the overall risk of error. But the notion of *error avoidance* is only one component of accuracy in the criminal trial setting. In order to reduce the overall cost of error in the criminal sphere, the reduction of the overall number of errors should be the focal point. For,

---

[132] *R v Clark,* Crown Court, 9.11.1999.

[133] *R v Clark* (No 2) [2000] EWCA Crim. 54 [238].

[134] *R v Clark* (No 2) [2003] EWCA Crim. 1020.

[135] *See* Redmayne, *supra,* note 3, at 282 .

[136] *See* above at ###. *See also* Louis Kaplow, *The Value of Accuracy in Adjudication: An Economic Analysis*, 23 J. LEGAL STUD. 307, 307-08 (1994) ("Accuracy is a central concern with regard to a wide range of legal rules. One might go so far as to say that a large portion of the rules of civil, criminal, and administrative procedure and rules of evidence involve an effort to strike a balance between accuracy and legal costs.").

the social costs of the two types of errors that occur in the framework of criminal proceedings—wrongful convictions and wrongful acquittals—are not commensurate. The social costs of wrongful conviction are considered significantly weightier than those associated with false acquittal. Minimizing the aggregate social costs of error in criminal proceeding thus entails lowering the incidence of false convictions, even by way of increasing the prevalence of false acquittals.[137] In other words, since court decisions entail an inherent uncertainty and errors can never be completely eliminated, another component of accuracy that must be taken into consideration is *error allocation*.[138] Under this calculus, the rules of evidence and criminal procedure allocate the risk of error between the defense and prosecution in a way that promotes errors in favor of the defendant (considered less costly) at the expense of errors in favor of the prosecution (which entail more substantial costs).[139]

This gives rise to a distinction between statistical evidence that is incriminating and exonerating statistical evidence in terms of admissibility at criminal trials. Allowing exonerating statistical evidence to be submitted in court, while rejecting incriminating evidence, aligns with other rules of evidence and criminal procedure aimed at reducing the likelihood of false convictions (including by compromising the certainty of the innocence of the acquitted). [140]

---

[137] Alex Stein, *The Refoundation of Evidence Law*, 9 CAN. J.L. & JURIS. 279 (1996) (the risks of error must be allocated between the defense and prosecution so as to reflect the disutility ratio between wrongful conviction and wrongful acquittal). For further discussion of the cost-minimization approach to evidence law, see Richard A. Posner, *An Economic Approach to Legal Procedure and Judicial Administration*, 2 J. LEGAL STUD. 399, 408-17 (1973) (discussing the cost-minimization model and the search model); Frederick Schauer & Richard Zeckhauser, *On the Degree of Confidence for Adverse Decisions*, 25 J. LEGAL STUD. 27, 34 (1996).

[138] Tom Stacy, *The Search for the Truth in Constitutional Criminal Procedure*, 91 COLUM. L. REV. 1369, 1406–07 (1991) ("[B]ecause no set of procedures can eliminate all erroneous outcomes, any conception of accuracy must also address how errors should be allocated as between erroneous convictions and acquittals.").

[139] *See* Richard A. Posner, *An Economic Approach to Legal Procedure and Judicial Administration*, 2 J. LEGAL STUD. 399, 410–15 (1973); Frederick Schauer & Richard Zeckhauser, *On the Degree of Confidence for Adverse Decisions*, 25 J. LEGAL STUD. 27, 34 (1996).

[140] Let us briefly note that "the-social-cost-of-error consideration" can also explain why there is no room to draw a distinction between pro-plaintiff and pro-defense statistical evidence in the civil trial arena. In the civil context as well, the rules of evidence

Yet in contrast to the accuracy perspective, the incentive-based approach cannot accommodate such a distinction between incriminating and exonerating statistical evidence. The introduction of either type of evidence at trial would dampen incentives and reduce the (ex-ante) marginal cost of engaging in criminal behavior. True, the prevailing character evidence rule, explored in Sanchirico's analysis, does manifest such a distinction: a central exception to the rule against submitting character evidence is the "Rule of Mercy," whereby the defendant may submit (good) character evidence as a defense. The Mercy Rule allows the defendant to bring witnesses and evidence of pertinent character traits for the purpose of establishing reasonable doubt, despite the prohibition on the prosecution (as a general matter) from introducing bad character evidence. But this distinction between good and bad character evidence cannot be accommodated from an incentive-based perspective. Indeed, as claimed by Sanchirico, incentive considerations dictate abolishing the Rule of Mercy.[141]

As for the epistemological story-- at first sight, it may seem that Sensitivity considerations cannot deliver a similar asymmetry between incriminating and exonerating statistical evidence. After all, both the belief that the butler did it *and* the belief that the butler didn't do it must be sensitive if they are to amount to knowledge or to be epistemically respectable in some closely related way, But this is an illusion. True, Sensitivity considerations do not distinguish between a belief and its negation. But this is not the relevant contrast here. For

---

and procedure affect the comparative frequency of each type of error (that is, errors in favor of the plaintiff and errors in favor of the defendant) and reflect the system's assessment of the social costs associated with each type of error. Unlike in the criminal context, however, the underlying assumption of civil procedure is that the two types of error entail equal costs. Undeserved losses are "equally regrettable," whether incurred by the plaintiff or by the defendant. *See* Daniel L. Rubinfeld, *Econometrics in the Courtroom*, 85 COLUM. L. REV. 1048, 1052 (1985) (citing *In re Winship*, 397 U.S. 358, 371 (1970) (Harlan, J., concurring)). *See also* ALEX STEIN, FOUNDATIONS OF EVIDENCE LAW 219 (2005). This is what justifies and even necessitates that the risk of error be allocated between plaintiff and defendant in a roughly equal manner and that pro-plaintiff and pro-defendant statistical evidence be treated symmetrically. *See* Alex Stein, *The Refoundation of Evidence Law*, 9 CAN. J.L. & JURISPRUDENCE 279, 333-35 (1996). The slight tilt in favor of the defendant can be attributed to the fact that "'taking' is perceivable as being generally more harmful than 'not giving.'" *Id*. at 335.

[141] Chris William Sanchirico, *Character Evidence and the Object of Trial*, 101 COLUM. L. REV. 1227, 1302 (2001).

arguably, while conviction requires the belief that the defendant committed the relevant offense, acquittal in no way requires the belief that the defendant did not commit the offense. At most, acquittal requires the *absence* of the belief that the defendant *did* commit the offense. Even this is too strong − acquittal may be called for even when the finders of fact do believe that the defendant committed the offense, so long as they are not sufficiently confident in that belief, or perhaps so long as that belief is not epistemically respectable in some way, or some such[142]. But the point crucial here is just that the belief that the defendant did not commit the offense is in no way required for justified acquittal. This means that while applying a Sensitivity requirement to incriminating evidence makes sense (for conviction requires belief, and Sensitivity is an epistemic-respectability property of beliefs), applying Sensitivity to exonerating evidence does not make sense (for no similar belief is needed for acquittal to be called for, certainly not the belief that the defendant did not commit the offense). In this way, then, Sensitivity considerations too distinguish rather strongly between incriminating and exonerating statistical evidence.

Admissibility versus Sufficiency of Statistical Evidence: Availability of Individual Evidence

Under Koehler's account of base-rate evidence usage, appellate courts show a tendency towards rejecting such evidence and viewing it as irrelevant and inadmissible in cases in which alternative individual evidence could have been obtained.[143] And even where individual evidence

---

[142] And, of course, sometimes acquittal is called for even in the fact of *full* conviction – indeed, *knowledge* – that the defendant committed the offense, as in cases of highly reliable but inadmissible evidence. For simplicity, we ignore such complications in the text.

[143] J. Koehler, *When Do Courts Think Base Rate Statistics Are Relevant?,* 42 JURIMETRICS 373, 401(2002).

is not readily accessible, there is a general reluctance to base judgments on "naked statistical evidence," as opposed to statistical evidence corroborated by individual evidence.[144]

The incentive story may offer an explanation for this apparent preference of individual evidence for its presumably better incentive-generating effect. Take, for instance, the gatecrasher hypothetical: Consider the possibility that John goes home and then has an alibi; or that he purchases a ticket and keeps it as proof; or that he gatecrashes and is videotaped climbing the fence. There is a positive ex-ante effect to allowing these pieces of evidence to be admitted in court. As argued, at least one important normative consideration against relying on statistical evidence is that this will render the law's primary-behavior incentives less effective and less accurate than they would otherwise be. Since there is no parallel incentive-corrupting effect to relying on individual evidence—even individual evidence that is probabilistically indistinguishable from the relevant piece of statistical evidence—it is clear why courts encourage prospective parties to seek the latter type of evidence.

We must further qualify this point: Arguably, the *magnitude* of the incentive-corrupting effect of statistical evidence varies from case to case and is a function of the evidentiary foundations of the particular case (i.e., of the relative contribution of the statistical evidence to the litigation payoff). When statistical evidence is uncorroborated by individual evidence, the incentive-corrupting effect is its most intense. To explain briefly: As a general matter, there is a diminishing marginal utility to evidence presentation. As more and more evidence is accumulated and presented in court, the effect of additional evidence on the outcome of the case

---

[144] *See* Rebecca Haw, *Prediction Markets and Law: A Skeptical Account*, 122 HARV. L. REV. 1217, 1229 (2009). (discussing the prohibition against the use of "naked statistical evidence") *See also* Dale A. Nance, *The Best Evidence Principle,* 73 IOWA L. REV. 227 n.300 (1988) and FREDERICK SCHAUER, PROFILES, PROBABILITIES, AND STEREOTYPES 81 (2003).

tends to diminish.[145] The result is that the greater the likelihood that the decision will be based solely upon statistical evidence, the greater the impact of statistical evidence on the legal outcome—and the greater its incentive-corrupting potential. And vice versa: in the alternative scenario, where the statistical evidence is only supplementary, its potential chilling effect on incentives will be less significant. Restricting the admissibility of statistical evidence to cases in which it is supplementary would induce at least one of the parties to search for individual evidence and thereby reduce the erosive effect of statistical evidence on incentives.

Yet accuracy considerations seem to point to the opposite conclusion. The preference of individual evidence increases the social costs of reaching a particular level of accuracy in legal fact-finding. Due to the discounting of the probative weight of statistical evidence, parties are pushed to search for and submit individual evidence that—under certain circumstances, at least—could be more costly, despite offering no accuracy advantage. The higher-cost assumption must hold true for at least some of the relevant cases, for otherwise creating external incentives by way of suppressing the statistical evidence would be unnecessary.

Posner has attempted to formulate an accuracy-based vindication of the rules against the use of "naked statistical evidence."[146] Posner's argument in the Blue Bus context is that the problem is not the mathematical probability *per se*, but, rather, the attempt to equate the mathematical probability with the probative weight of the evidence: "If the statistic is the plaintiff's only evidence, the inference to be drawn is not that there is a fifty-one percent probability that it was a bus owned by A that hit the plaintiff but that the plaintiff either investigated and discovered that it was actually a bus owned by B (and let us say that B is

---

[145] Richard Posner, *An Economic Approach to the Law of Evidence,* 51 STAN. L. REV. 1477, 1482 (1999). For a similar claim, see Alon Harel & Ariel Porat, *Criminal Liability for Unspecified Offenses*, 94 MINN. L. REV. 261, 291 (2009) ("It is typically much harder—and more costly—to collect the tenth item of evidence than the ninth item, the eighth item, and so on.").

[146] *See infra* Discussion at ###.

judgment-proof and so not worth suing), or that he has simply not bothered to conduct an investigation."[147]

Put differently, the probative weight of evidence is a function not only of the evidence that exists in a given case, but also of the evidence one would expect to find in that case. The very lack of individual evidence, maintains Posner, weakens the probative value of the market-share evidence. And of course, here there is room to differentiate between cases according to the availability and cost of such alternative individual evidence.[148] The less costly and more readily available the alternative individual evidence, the stronger what is signaled by its exclusion from trial. Alternatively, the easier it is to obtain individual evidence, the stronger the assumption that the relevant party failed to search for that evidence or that her search yielded individual evidence that did not support her case to the same extent as the statistical evidence submitted. These differential signaling effects can explain the phenomenon identified by Koehler—namely, the reluctance of appellate courts to accept uncorroborated base-rate evidence when it is offered in cases in which alternative individual evidence could have been obtained.

Posner's accuracy-based argument, however, can be challenged due to its circular structure. Indeed, against the background of an evidentiary toll levied on the very resort to statistical evidence, the lack of individual evidence may signal to the court that the relevant party failed to uncover such evidence in support of its case and, therefore, that the statistical evidence should be assigned a lower probative weight. But were courts to accord equal evidentiary weight to statistical and individual evidence, the incentives to the parties in the evidence-seeking process

---

[147] Posner, *supra* note 22, at

[148] Some writers have claimed that individual evidence of liability is almost always available. *See, e.g.,* Ron Allen, *A Reconceptualization of Civil Trials*, 66 B.U.L. REV. 401, 409 (1986); David Kaye, *The Paradox of the Gatecrasher and Other Stories*, 1979 ARIZ. ST. L.J. 101, 104; Richard Lempert, *The New Evidence Scholarship*, 66 B.U.L. REV. 439, 454 (1986). This is an empirical matter lying outside the scope of this article.

would be different: there would be no reason for them to prefer finding individual, rather than statistical, evidence. On this alternative legal background, the signaling effects would also shift: the reliance *per se* on statistical evidence would not, in and of itself, signal the groundlessness of the case of the party submitting this evidence in court. Thus, the signaling effect is a *result* of the distinction between the evidentiary payoffs of statistical and individual evidence, rather than the *reason* or justification for that distinction. On the other hand, this dismisses the case for ab nihilo implementation, but under the current regime whereby courts accord unequal evidentiary weight to statistical and individual evidence, this consideration may count against equalization. Another objection which can be leveled relates to the hidden assumption underlying the discussion— namely, that statistical evidence is cheaper and more accessible than individual evidence .While often true, this is not always the case: DNA evidence, for instance, is relatively expensive and generally inaccessible to the public at large. Moreover, we can safely abstract away from all this by insisting (as we have done throughout this article) that we hold all other things equal in the statistical-evidence scenario and the individual-evidence scenario. This includes an a-priori equal level of accuracy, whether it relates to the mathematical probability per se or the mathematical probability updated by the prior odds relating to the nakedness of the statistical evidence.

From the epistemological perspective-- it is interesting to comment on the possible interaction between statistical and individual evidence where both are available. Think, then, of cases where the evidence for the prosecution includes *both* cold-hot DNA evidence *and* some paradigmatically individual evidence, say, the testimony of an eye witness[149]. There are several possible cases: One possibility is that the two pieces of evidence (the DNA evidence and the eye-testimony) were obtained independently – say, by two independent, non-communicating police

---

[149] Of course, if theirs is *only* cold-hot DNA evidence, then the discussion above applies.

departments. In such a case, if the eye-testimony is sufficiently strong to justify conviction regardless of the DNA evidence, the case is easy and immaterial from a statistical-evidence perspective. If – now going in the other extreme – the statistical evidence is very strong, and the individual evidence rather weak, then the discussion in terms of Sensitivity nicely applies: for a conviction based on both pieces of evidence will not be sensitive (as if the defendant did not commit the offense, we would have still had the strong statistical evidence against him, which would in this case suffice for conviction). A more challenging case is one where no piece of evidence suffices on its own for conviction, but only their combined weight (assuming we are considering relying here on the statistical evidence as well) suffices for conviction. If statistical evidence is to be ruled out entirely, then presumably in such a case acquittal is called for (for the individual evidence is ex hypothesi not sufficient for conviction). But a discussion in terms of Sensitivity seems to yield the opposite result. For in such a case, a conviction that is based on both pieces of evidence *is* sensitive: Had the defendant not committed the offense, the statistical evidence would still have been available, but the individual evidence (the eye-witness testimony) would presumably not have been available. And then, we would not have convicted. Thus, if what we are concerned with is that the conviction be sensitive[150], in mixed cases of the kind just described statistical evidence should be admitted and relied upon. And indeed, in such a case, relying on the cold-hit DNA evidence to enhance the evidential weight of the eye-witness testimony does not seem as intuitively objectionable as relying on statistical evidence alone. The abovementioned disinclination of courts and the legal doctrine to base judgments on "naked

---

[150] Still, if what we are concerned with is that the relevant *evidence* be sensitive, then statistical evidence here is as problematic as everywhere else.

statistical evidence," as opposed to statistical evidence corroborated by individual evidence,[151] echoes this intuitive difference, which talk of Sensitivity nicely explains. [152]

But all the cases discussed in the previous paragraph are cases where the statistical evidence (cold-hot DNA evidence) and the individual evidence (eye-witness testimony) were independently obtained. Another possibility is that this is not the case. What if obtaining one of the relevant pieces of evidence depended on the other? Two directions are possible. Suppose that first we get the eye-witness testimony, and then we check the person the eye-witness identified for DNA ("confirmatory DNA"). In such a confirmatory DNA case the conviction can be appropriately sensitive: for presumably, had the suspect not committed the offense, the eye-witness would not have identified him, and we would not have the DNA evidence available either, and so we would not have convicted. Indeed, in such a case the use of the DNA evidence seems to be in line with the inference-to-the-best-explanation strategy explained above: What best explains why we found a DNA match with the person the eye-witness identified is that he committed the offense. We no longer have a problematic case of use of statistical evidence.

The second direction of dependence between the statistical and the individual evidence is more interesting. Suppose, then, that we first obtain the cold-hot DNA evidence. Then, we place the suspect in a conforming line-up, and the eye-witness identifies him. If we convict in such a case, is our conviction sensitive? The details of the answer (but not its bottom line) depends on an issue bracketed above − whether the relevant counterfactual scenario is one where the defendant did not commit the crime, or whether it's one where the defendant did not commit the crime *and the evidence is as it actually is*. Had the defendant not committed the crime, in all

---

[151] Haw, *supra,* note 144, at 1229.
[152] The exception in this regard is DNA evidence, where courts tend to convict on the basis of cold-hit DNA that is not corroborated by other evidence. *See* Roth, *supra*, note 89, at 1155

likelihood we would not have had available the DNA evidence against him, nor would we (consequently) have the eye-witness testimony against him, and so we would not have convicted. If this is the relevant counterfactual, then, the conviction is sensitive. What would have happened had the defendant not committed the crime but we would have still had the DNA evidence against him? Well, in such a case, we would have still held the line-up, but presumably the eye-witness wouldn't have identified him (for it is a reliable eye-witness, and the defendant in this counterfactual scenario had not committed the crime), and so we wouldn't have convicted. Here too, then, the conviction may be sensitive.

This line of thought has another important payoff. Both intuitively, and as a matter of legal doctrine, we distinguish between the use of statistical evidence at trial, and the use of similar evidence as an investigative tool. Relying on statistical evidence as an investigative tool seems much less problematic, if problematic at all.[153] And the previous paragraph nicely explains why. So long as the statistical evidence itself is not admitted at trial, its use as an investigative tool does not compromise sensitivity: Consider again the case of the cold-hit DNA evidence which leads to a conforming lineup. If the eye-witness is sufficiently reliable to ground conviction, then the fact that the lineup wouldn't have occurred if it weren't for the statistical evidence is neither here nor there: True, that evidence is not sensitive, so that we would have had it even had the defendant not committed the crime. But all that this means is that had the defendant not committed the crime, we would have still held the lineup. It most certainly does not entail that had the defendant not committed the offense, we would have still convicted. In

---

[153] COLIN AITKEN ET. AL, COMMUNICATING AND INTERPRETING STATISTICAL EVIDENCE IN THE ADMINISTRATION OF CRIMINAL JUSTICE 13-27 (discussing the central and growing role of statistical evidence and probabilistic reasoning in criminal investigations).

fact, we wouldn't have convicted, for in such a case, we would have held the lineup, and our reliable eye-witness wouldn't have identified the defendant.

The analysis in terms of Sensitivity thus has impressive payoffs to the understanding of the interaction between statistical and individual evidence, (consequently) to the distinction between the admissibility and sufficiency of statistical evidence, and also to the distinction between the appropriate role of statistical evidence in trial and as an investigative tool.

After analyzing the rules guiding the use of statistical evidence that can be extrapolated from prevailing legal doctrine and case-law and showing how the theoretical foundations we presented can support these rules, we can now move from the descriptive to the prescriptive part of the discussion. In what follows, we will briefly illustrate the potential for legal reform that our theoretical model offers and demonstrate with two examples how implementing our model would open the door to categorical distinctions currently not made by the law. We showed that the formal and informal legal distinctions presently governing the use of statistical evidence in courts refer to *the type of statistical evidence* (DNA versus propensity-for-crime evidence), *its probability level* (DNA evidence as an example of uniquely high probability levels), *the type of legal proceeding* (with the criminal trial being a uniquely problematic arena), and *the availability of alternative individual evidence* (the admissibility versus sufficiency issue). However, the proposed theoretical model supports distinguishing among the different uses of statistical evidence based upon additional criteria that are currently unrecognized by the legal doctrine, such as the *type of misconduct* with which the defendant is charged and *party characteristics.*

Admissibility of Statistical Evidence According to Type of Misconduct

Under prevailing legal doctrine, the use of base-rate evidence is not conditioned upon the type of misconduct alleged against the defendant (other than what is prescribed by the rough divide between civil and criminal proceedings discussed above). The theoretical model we presented, however, allows for a distinction to be drawn between acts that occur within the context of a personal relationship (such as spousal abuse) and acts that occur in other contexts. The reason for the distinction is that with acts in the context of a personal relationship, the deliberating perpetrator typically knows that if he chooses not to commit the particular crime or take the injury-causing course of action, no such act will be inflicted on the potential victim. Suppose, for instance, that according to the relevant statistics, an exceptionally high percentage of the spouses of academics interested in epistemology, who are the victims of a violent death, are murdered by their (epistemologist) spouse. In light of the personal dimension of the act of spousal homicide, John, an epistemology-loving scholar, knows that if he chooses not to murder his wife, Sara, she will most likely not be murdered. John's actions significantly impact his chances of conviction— for without his murdering his wife, most chances are that she will not be the victim of a violent death at all. Therefore, the chilling effect on deterrence, discussed throughout the article, does not occur, and there is no incentive-based reason to exclude the statistical evidence regarding the percentage of epistemologists who murder their spouses.[154] This, of course, differs from the gatecrasher case, where the would-be perpetrator knows that even if she doesn't gatecrash the stadium, others will still do so. For this reason, statistical evidence regarding the percentage of gatecrashers and other relevant base-rate statistics in this category of cases threaten the primary-behavior incentives the law provides and should, therefore, be excluded from trial.

---

[154] We would like to thank Mitch Berman for this point.

The incentive story thus has different implications across the two categories of cases: those in which the act would likely be performed by others regardless of whether the would-be perpetrator decides not to engage in it, as opposed to personal context cases, where the act will not likely be carried out by anyone else. In the latter type of case, the statistical evidence against the defendant ought to be admissible at trial.

This result is at least somewhat counterintuitive. Sensitivity can (we think) explain why. Once again, though, we need to be more careful about the relevant counterfactual. If we base a conviction (in such a case) on the relevant statistical evidence, is our conviction sensitive? Well, had John not killed Sara, would we have convicted him? The answer to this counterfactual is plainly "no". Had John not killed Sara, no one would have, and so no one would have been convicted in murdering her[155]. But perhaps the relevant counterfactual is different: *Had Sara been murdered, but not by John,* would we have convicted him? If *this* is the relevant counterfactual, then the answer may very well be "yes", and our conviction then fails Sensitivity. It is not entirely clear to us which is the relevant counterfactual, and indeed what considerations determine the answer to this question. We hope to discuss this question – and the possible relations between it and the incentive story in the previous paragraphs – in future work.

PART IV: CONCLUSION

We set out to dispel the confusion underlying the "bitter debate"-- in Koehler's and Shaviro's words— regarding the use of statistical evidence at trial. This concludes our vindication of the

---

[155] A quick reminder about how counterfactuals work: A counterfactual is true, somewhat roughly, if and only if its consequent holds in the closest possible world in which its antecedent is true. So the fact that it's *possible* for Sara to be killed by someone else, or that it's *possible* for us to convict someone of murdering Sara even if she is still alive (and hiding somewhere) is no threat to what's in the text here. All that these scenarios show is that there are *some, very distant* possible worlds in which these scenarios play out. But in the *closest* possible worlds in which John doesn't murder Sara, no one else does, nor do we convict anyone of murder.

suspicion with which the law views statistical evidence. Perhaps the intuitive and unreflective suspicion as to the admissibility of statistical evidence, shared by evidence law theorists and practitioners alike, is motivated by the epistemological concerns that are also exemplified by the lottery cases. But these concerns, we have argued, do not survive reflection about what considerations should guide the law. Law is (at least partly) a mechanism for regulating behavior. It operates (at least partly) by creating reasons for action-- by providing incentives. The considerations which should guide the law are dictated (at least partly) by its incentive-creating function. The incentive-based story, however, also supports the initial suspicion towards statistical evidence at trial: Statistical evidence may weaken incentives not to engage in undesirable social activity. And the fact that the incentive based practical considerations and the purely theoretical epistemic case against statistical evidence point in the same direction is not a mere coincidence: Both are premised upon Sensitivity-like counterfactuals.

The proposed theoretical framework-- in epistemological terms and, more importantly, in incentive-based terms-- corresponds with the rules that can be extrapolated from prevailing legal doctrine and case-law, regarding the use of statistical evidence at trial: Thus, our theory can explain the unique treatment of DNA evidence as opposed to the usual caution with which statistical evidence is generally treated. Unlike other types of statistical evidence, DNA evidence can be claimed to meet the Sensitivity requirement: for, had the defendant not committed the offense, we would, in all likelihood, not have found her DNA on the scene, and so we would, in all likelihood, not have convicted her. Moreover, the incentive-based case against standard statistical evidence does not seem to apply to DNA evidence, as potential perpetrators are very rarely in a position to know whether DNA collected at the crime scene will match theirs. So, the

epistemological and incentive-based accounts align in the DNA context, pointing both towards admissibility of such statistical evidence (although, under our theory, it is the incentive-based perspective-- rather than the epistemological case-- that is substantial for policy purposes).

The proposed theoretical framework can also vindicate the inadmissibility of propensity-for-crime evidence at the conviction phase of trial: Introducing predictive information, premised upon defendant characteristics of a non-voluntary nature (and/or of a neutral quality from a social-welfare perspective) may lower the marginal cost of engaging in undesirable social conduct ex ante.   This incentive-based case against propensity-for-crime evidence is complemented by the epistemological   perspective.   Propensity-for-crime evidence provides information as to the conduct in question, but it is not affected by it. Due to this reverse direction of causality propensity-for-crime evidence does not satisfy the Sensitivity requirement.

 The descriptive capacity of the proposed framework also refers to the discrepancy between using statistical evidence for acquittal versus conviction purposes: While incentive-based considerations do not support drawing such a distinction (except insofar as incentives depend on accuracy), accuracy considerations can help explain the general reluctance towards using statistical evidence for purpose of conviction. The social costs of wrongful convictions are considered significantly higher than those associated with false acquittals. So, lowering the aggregate social costs of error in criminal proceedings entails reducing the incidence of false convictions, even by way of increasing the prevalence of false acquittals. The Sensitivity perspective unfolds in a more complicated but parallel manner in this context: Because Sensitivity applies to beliefs, it is relevant only for the incriminating usage of statistical evidence: A conviction plausibly involves a belief that the defendant committed the offense,

whereas an acquittal does not (necessarily) involve a belief that the defendant did not commit the offense- merely (at most) the absence of belief that he did.

Finally, our theory can explain why courts exhibit a greater reluctance to base judgments on "naked statistical evidence" as opposed to statistical evidence corroborated by individual evidence. From an incentive-based perspective, which is the perspective pertinent for policy purpose, there is room to claim that the extent of the incentive-corrupting effect of statistical evidence is a function of the relative contribution of the statistical evidence to the litigation payoff. When statistical evidence is uncorroborated by individual evidence, the incentive-corrupting effect is at its peak. The epistemological perspective requires finer distinctions about the precise nature of the interaction between the statistical and the individual evidence, distinctions that – as shown in detail above – shed interesting light both directly on the differential attitude towards statistical evidence when accompanied by individual evidence, and also on related issues, such as the distinction between the appropriate role of statistical evidence in trial and as an investigative tool.

We concluded by displaying our theory's prescriptive potential. Under prevailing legal doctrine, the use of statistical evidence is not conditioned upon the type of misconduct alleged against the defendant. Our model tentatively calls for a distinction between acts that occur within the context of a personal relationship (where the deliberating perpetrator typically knows that if he chooses not to commit the particular course of action, no such harm will be inflicted on the potential victim) and acts that occur outside such context. The reasons for the distinction are rooted in the differential incentive implications across these two categories of cases.

Further distinctions may emerge from the theoretical framework presented in this article. For instance, there may be room to base the admissibility of statistical evidence upon party characteristics, as the incentive structure may differ across corporations and individual litigants. Similar distinctions may apply with respect to high probability versus low probability levels, or to statistical evidence relating to past versus future events. The solution we suggest to the statistical evidence puzzle – in terms of a partly epistemological explanation of the relevant intuitions, and an incentive-based account of the relevant policy considerations – can facilitate further discussions, and guide these and similar issues, which we leave for future research.