

SENSE AND “SENSITIVITY”: EPISTEMIC AND INSTRUMENTAL APPROACHES TO STATISTICAL EVIDENCE

David Enoch* & Talia Fisher**

Statistical evidence is the subject of a heated and ongoing debate. Courts and legal scholars often view statistical evidence with suspicion, treating it as inadmissible even when it is probabilistically equivalent to individualized evidence. But attempts to vindicate the suspicion or to dismantle it altogether have been largely unsuccessful. The aim of this Article is to provide a comprehensive answer to the statistical evidence debate. The Article offers a novel explanation for the suspicion toward statistical evidence, pointing to the epistemic inferiority of statistical evidence due to its lack of “Sensitivity”—namely, the requirement that a belief be counterfactually sensitive to the truth as a necessary condition of “Knowledge.” After exposing the epistemic distinctions between statistical and individualized evidence, the Article turns to examining their implications for the legal arena. It claims that while the epistemic story provides an explanation for the suspicion toward statistical evidence, it does not provide a justification for this suspicion, for Sensitivity (like epistemology more generally) is not significant in the legal arena. Instead, this Article proposes an incentive-based vindication of the reluctance to use statistical evidence in court and points to the interesting interaction between the epistemic and the incentive-based approaches. After laying down the theoretical foundation, this Article demonstrates its descriptive potential. It demonstrates the proposed theory’s capacity to explain the prevailing le-

* Jacob I. Berman Professor of Law and Professor of Philosophy, Hebrew University of Jerusalem.

** Associate Professor of Law, Tel Aviv University. For excellent comments and suggestions, our thanks go to Ori Aronson, Oren Bar-Gill, Hagit Benbaji, Selim Berker, Mitch Berman, Kevin Clermont, Bruce Chapman, Stewart Cohen, Hadar Dancig-Rosenberg, Markus Dubber, Cian Dorr, Kate Elgin, Daniel Epps, Nir Eyal, Yuval Eylon, Jeff Fagan, Johann Frick, Sharon Hannes, Caspar Hare, Amnon Knoll, Netanel Lipshitz, Ofra Magidor, Ofer Malcai, Jennifer Nagel, Christian Piller, Ariel Porat, Jim Pryor, Amit Pundik, Mike Redmayne, Melinda Roberts, Andrea Roth, David Sklansky, Simon Stern, Carol Steiker, Alex Stein, Martin Smith, Sergio Tenenbaum, Omri Yadlin, Ruti Weintraub, Dan Wikler, participants of the 2015 Criminal Justice Roundtable at Harvard Law School, participants of the Critical Legal Analysis Series at University of Toronto Faculty of Law, participants of the Constitutional and Criminal Law Workshop at Bar-Ilan University, participants of the faculty seminar workshop of the Economics Department at Ben-Gurion University, participants of the faculty seminar at Sapir College, participants of the faculty seminar at the Buchmann Faculty of Law at Tel Aviv University, and the editorial board of the *Stanford Law Review*.

gal doctrine and the rules governing the admissibility and sufficiency of statistical evidence across various categories, including DNA evidence and propensity-for-crime evidence as well as incriminating versus exonerating statistical evidence. On the prescriptive front, the Article provides criteria for legal reform and suggests that the admissibility of statistical evidence should be contingent on the type of offense or misconduct alleged against the defendant.

INTRODUCTION.....	558
I. A CRITICAL REVIEW OF PROPOSED SOLUTIONS TO THE STATISTICAL EVIDENCE PUZZLE.....	565
A. <i>Exogenous-Factor Claims</i>	566
B. <i>About-Relation Claims</i>	566
C. <i>The Causal-Connection Argument</i>	567
D. <i>Defendant-Specific Claims</i>	568
E. <i>Justice-in-the-Particular-Case Claims</i>	568
F. <i>Autonomy Considerations</i>	569
G. <i>Social-Acceptability Considerations</i>	570
H. <i>The Guaranteed-False-Conviction Argument</i>	570
II. THE THEORETICAL FRAMEWORK.....	571
A. <i>Justifying Beliefs or Actions?: We Require an Epistemic, Not a Practical, Framework</i>	571
B. <i>Some Epistemology: Introducing “Sensitivity”</i>	573
C. <i>From Epistemology to More Practical Concerns: Should the Law Care About Sensitivity (or Knowledge)?</i>	577
1. <i>The remaining puzzle: why care about knowledge?</i>	578
2. <i>Sanchirico on character evidence</i>	581
3. <i>Solution: the instrumental significance of being sensitive</i>	583
III. SOLVING SOME DOCTRINAL PUZZLES.....	585
A. <i>DNA Evidence</i>	587
B. <i>Propensity-for-Crime Evidence at the Guilt Phase of Trial</i>	592
C. <i>Incriminating Versus Exonerating Statistical Evidence</i>	596
D. <i>Admissibility Versus Sufficiency of Statistical Evidence</i>	602
IV. PRESCRIPTIVE ANALYSIS.....	607
CONCLUSION.....	609

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

1. Jonathan J. Koehler & Daniel N. Shaviro, *Veridical Verdicts: Increasing Verdict Accuracy Through the Use of Overtly Probabilistic Evidence and Methods*, 75 CORNELL L. REV. 247, 247 (1990); see also *id.* at 247 n.1 (listing contributions to the debate).

admissibility and sufficiency of statistical evidence arise in court with ever-growing frequency, 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 by connecting the statistical evidence debate to broader epistemic discussions and by highlighting “Sensitivity”—that is, the requirement that a belief be counterfactually sensitive to the truth—as a way of epistemically explaining the suspicion toward statistical evidence. After exposing the epistemic distinctions between statistical and individualized evidence, the Article turns to examining their implications for the legal arena. We will use the epistemic discourse on Sensitivity as well as an instrumental analysis to address the descriptive and prescriptive challenges that statistical evidence poses.

One starting point for the statistical evidence debate is the classic Blue Bus hypothetical,³ which is a variant of *Smith v. Rapid Transit, Inc.*,⁴ a seminal case in modern evidence law. The hypothetical consists of two cases. In both cases, a runaway bus injures the plaintiff, and the case goes to trial against the eponymous bus company. In the first case, the evidence includes eyewitness testimony that one of the Blue Bus Company’s buses caused the injury. The witness, however, is imperfectly reliable. To illustrate, let us assume her to be 70% reliable in such circumstances. In the second case, however, there is no eyewitness to the accident. Instead, the plaintiff seeks to introduce evidence about the Blue Bus Company’s market share in the area where the accident took place. The uncontested market share data show that the Blue Bus Company owns 70% of the buses in the relevant area. This, the plaintiff argues, shows that it is more likely that one of the Blue Bus Company’s vehicles was involved in the accident, because Blue Bus is the largest bus company in the area with the greatest number of buses on the road.

Even though the evidence in both cases may be of equal probative value, our responses to the two cases are very different. Most people (lawyers and laypeople alike) find nothing problematic in basing a finding on the eyewitness evidence in the first case but are very reluctant to ground a finding on the mar-

2. For further discussion on the topic of statistical evidence, see generally Ronald J. Allen, *On the Significance of Batting Averages and Strikeout Totals: A Clarification of the “Naked Statistical Evidence” Debate, the Meaning of “Evidence,” and the Requirement of Proof Beyond a Reasonable Doubt*, 65 TUL. L. REV. 1093 (1991); and Craig R. Callen, *Adjudication and the Appearance of Statistical Evidence*, 65 TUL. L. REV. 457 (1991).

3. For one discussion of this common hypothetical, see Mike Redmayne, *Exploring the Proof Paradoxes*, 14 LEGAL THEORY 281, 281-82 (2008).

4. 58 N.E.2d 754 (Mass. 1945). In *Smith*, there was evidence to show that the presence of the defendant’s bus line at the time and place of the accident was consistent with its schedule. *Id.* at 754-55. The trial court ruled for the defendant, and the Supreme Judicial Court of Massachusetts upheld the directed verdict, ruling that liability could not rest solely on the “mathematical chances” that support it. *Id.* For further discussion of *Smith*, see Gary L. Wells, *Naked Statistical Evidence of Liability: Is Subjective Probability Enough?*, 62 J. PERSONALITY & SOC. PSYCHOL. 739 (1992); and Richard W. Wright, *Causation in Tort Law*, 73 CALIF. L. REV. 1735, 1826 (1985).

ket share evidence in the second case.⁵ And many courts seem to agree.⁶ Indeed, the second case closely resembles the *Smith* case mentioned above, in which the court rejected market share evidence as the basis for finding that the defendant's bus was responsible for the accident.⁷

But this now presents a problem. The Blue Bus hypotheticals were carefully designed to keep all other things equal. This means, for instance, that relying on the two kinds of evidence will yield, over the long run, the same rate of mistaken decisions: in the words of the hypothetical, the percent chance that the witness mistook the color of the bus in the first case is equal to the market share of all other bus companies in the second case. And yet despite these similarities, almost everyone—judges and legal scholars, lawyers and laypeople alike—seems to draw a sharp distinction between eyewitness testimony and market share evidence,⁸ even when all other things are held equal. How can such a discrepancy be explained? Why *do* we treat these kinds of evidence so differently? And can this intuitive distinction be vindicated? In other words, *should* we treat market share evidence and eyewitness testimony differently? If so, why?⁹

Consider the analogous gatecrasher hypothetical,¹⁰ in which it is uncontested that of 1000 people attending a stadium event, only 10 purchased tickets. If any individual attending the event is sued for crashing the gate or, even more clearly, is prosecuted for such gatecrashing, a finding against him or her merely on the strength of the statistical evidence seems inappropriate. On the other

5. See Wells, *supra* note 4. In this well-known study, students and experienced trial judges found the defendant liable in fewer than 10% of cases on average when the case was based on statistical evidence. When the case was based on probabilistically identical individualized evidence, however, subjects found the defendant liable in over 65% of cases on average. See also Redmayne, *supra* note 3, at 281-82 (describing this reluctance as an “overwhelming intuition”).

6. For an extensive discussion of the antagonism toward statistical evidence across jurisdictions and over time, see Jonathan J. Koehler, *When Do Courts Think Base Rate Statistics Are Relevant?*, 42 JURIMETRICS J. 373, 380-85 (2002).

7. See *Smith*, 58 N.E.2d at 755.

8. Note that 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 at the scene.

9. One major issue we will not address here relates to the relevant reference class with respect to which the statistical evidence is drawn. We will assume, more or less throughout, that the statistical evidence latches onto the relevant frequencies. Such a simplifying assumption is unobjectionable in our context, as it arguably arises both for statistical evidence and for individual, direct evidence (for instance, regarding the 70% reliability of the eyewitness account).

10. This is a variation of Jonathan Cohen's famous gatecrasher scenario, in which it is uncontested that, among 1000 people attending a stadium event, only 499 unidentified attendees purchased tickets. See L. JONATHAN COHEN, *THE PROBABLE AND THE PROVABLE* 75 (1977); see also David Kaye, *The Paradox of the Gatecrasher and Other Stories*, 1979 ARIZ. ST. L.J. 101, 101; Robert J. Rhee, *Probability, Policy and the Problem of Reference Class*, 11 INT'L J. EVIDENCE & PROOF 286, 287 (2007).

hand, conviction based on a probabilistically similar piece of direct, individualized evidence, such as a videotape, seems perfectly fine.

The puzzle concerning the differential treatment of probabilistically equivalent statistical and individualized evidence, exemplified by these hypotheticals, surfaces in many different scenarios. This puzzle can appear in civil or criminal trials, arise with different levels of probability, relate to past or future events, and so on.¹¹

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 statistical evidence first began to appear in court, many judges responded antagonistically, deeming it inadmissible and devoid of any probative value.¹² To this day, courts continue to exhibit a general preference for individualized evidence and to reject base-rate evidence despite its potential to promote accuracy in legal factfinding.¹³

The doctrinal picture, however, is not clear-cut. First, the legal doctrine regarding the use of statistical evidence for imposing liability is incoherent, and there is conflicting case law on the matter.¹⁴ Although statistical evidence is often considered irrelevant and thus inadmissible in court,¹⁵ at other times, when presented as supplementary evidence, courts sometimes treat it as admissible.¹⁶ In yet other instances, statistical evidence has even been treated as sufficient for a finding of liability.¹⁷ For example, in contrast to the ruling in *Smith*, the appellate court in *Kaminsky v. Hertz Corp.*¹⁸ ruled that the plaintiff had established a rebuttable presumption of ownership when the primary evidence brought forth against the defendant, Hertz, was unchallenged testimony

11. As we demonstrate in Part III of this Article, there may be important distinctions among these different scenarios for certain purposes. For a list of examples, see Redmayne, *supra* note 3, at 282-85. 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.

12. See, e.g., *Commercial Standard Ins. Co. v. Gordon's Transps., Inc.*, 154 F.2d 390, 394 (6th Cir. 1946); *Evans v. Ely*, 13 F.2d 62, 64 (3d Cir. 1926); *Va. & S.W. Ry. v. Hawk*, 160 F. 348, 352 (6th Cir. 1908); *Smith*, 58 N.E.2d at 755. For two other early decisions critical of statistical evidence, see *People v. Collins*, 438 P.2d 33, 40-41 (Cal. 1968) (en banc); and *State v. Carlson*, 267 N.W.2d 170, 176 (Minn. 1978).

13. See Chris Guthrie et al., *Inside the Judicial Mind*, 86 CORNELL L. REV. 777, 810 (2001); see also Richard Goldberg, *Epidemiological Uncertainty, Causation, and Drug Product Liability*, 59 MCGILL L.J. 777, 782 (2014); Koehler, *supra* note 6, at 373-74; *infra* Part III.

14. See Koehler, *supra* note 6, at 377, 401; *infra* Part III.

15. See Koehler, *supra* note 6, at 377.

16. See *id.*

17. In *Turner v. United States*, 396 U.S. 398, 414-17 (1970), for example, the Supreme Court held that base-rate evidence indicating that more than ninety-nine percent of all heroin consumed in the United States is illegally imported sufficed to permit an inference that its possession amounted to possession of a smuggled drug.

18. 288 N.W.2d 426 (Mich. Ct. App. 1979).

that the yellow truck that had caused the accident at issue in the case bore a Hertz logo and that Hertz owned ninety percent of such trucks.¹⁹

Second, not only are there seemingly random inconsistencies in the court rulings on statistical evidence in the general class of cases, there are also exceptions to the general approach of inadmissibility in certain categories of cases. Such is the case, for example, with DNA profiling, which is explicitly statistical in nature. The standard DNA profile is extracted from only a small portion of the donor's entire DNA. So even under the assumption that a unique set of DNA characterizes each and every individual, two or more people can still share an identical DNA *profile*.²⁰ Despite its statistical nature, DNA evidence is increasingly relied on by the courts.²¹ To date, most American courts admit DNA evidence despite its apparent similarities to other, inadmissible types of statistical evidence.²² In light of the considerable inconsistency in the case law regarding the admissibility or sufficiency of statistical evidence, the extrapolation of the legal doctrine on this matter has been deemed "part science and part art."²³

19. *Id.* at 427, 429. *But see* Richard A. Posner, *An Economic Approach to the Law of Evidence*, 51 STAN. L. REV. 1477, 1508 n.68 (1999) (contesting the description of *Kaminsky* as contrary to *Smith* in light of "the fact that [in *Kaminsky*] the corresponding percentages were 90% and 10%" and that "there was also nonstatistical evidence pointing to the defendant's ownership of the truck that had caused the accident").

For a case similar to *Kaminsky*, see *Kramer v. Weedhopper of Utah, Inc.*, 490 N.E.2d 104 (Ill. App. Ct. 1986). The plaintiff, Kramer, was injured when a bolt sheared off the aircraft that he had assembled from a Weedhopper kit. *Id.* at 105-06. Weedhopper received 90% of its bolts from Lawrence and 10% of its bolts from Hughes. *Id.* at 106. Based on this base rate, an Illinois appellate court reversed the trial court's grant of summary judgment in favor of the primary bolts manufacturer. *Id.* at 107-08. The appellate court held that the statistical evidence allowed "the inference that . . . the bolt supplied to Kramer [was] purchased from Lawrence." *Id.* at 108.

20. ROBERTO PUCH-SOLIS ET AL., ASSESSING THE PROBATIVE VALUE OF DNA EVIDENCE: GUIDANCE FOR JUDGES, LAWYERS, FORENSIC SCIENTISTS AND EXPERT WITNESSES 13 (Royal Statistical Soc'y, *Communicating & Interpreting Statistical Evidence in the Admin. of Criminal Justice*, Practitioner Guide No. 2, 2012).

21. 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) (en banc); and *People v. Alzoubi*, 479 N.E.2d 1208, 1210 (Ill. App. Ct. 1985).

22. See Kristen Bolden, Note, *DNA Fabrication, a Wake Up Call: The Need to Reevaluate the Admissibility and Reliability of DNA Evidence*, 27 GA. ST. U. L. REV. 409, 419-20 (2011). For a comprehensive list of state DNA statutes, see Davina Dana Bressler, *Criminal DNA Databank Statutes and Medical Research*, 43 JURIMETRICS J. 51 app. at 68 tbl.1 (2002).

23. Koehler, *supra* note 6, at 401. 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. Koehler has made the most comprehensive and notable attempt. 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

In addition to the inconsistencies in the legal doctrine, the legal scholarship is also fraught with discrepancies on this topic. The decision in *Smith* triggered a heated, ongoing debate in legal academia²⁴ around (1) whether judicial fact-finding ought to be grounded on standard probability logic and (2) whether statistical evidence regarding the base rate for liability is sufficient ground for a ruling in criminal or civil trials.²⁵ Koehler and Shaviro, for instance, question courts' refusal to ground verdicts in favor of plaintiffs or prosecutors on statistical evidence.²⁶ Shaviro claims that the objective of verdict accuracy requires that courts hold in favor of the party whose case seems more likely,²⁷ and that the only relevant misgivings about imposing liability are doubts related to risk of error, not those related to overtness.²⁸ But the evidence law scholars who advocate the use of statistical evidence are outnumbered by other legal academics who tend to strongly oppose statistical evidence and object to its submission in trial for reasons of probative deficiency, reasons of moral deficiency, or other policy reasons.²⁹

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 individualized evidence is difficult or impossible to obtain. *Id.* at 373.

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

25. Some of the extensive discussions of the issue of statistical evidence can be found in Jonathan J. Koehler, *The Probity/Policy Distinction in the Statistical Evidence Debate*, 66 TUL. L. REV. 141 (1991), and Laurence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329, 1332-78 (1971). See also H.L. HO, A PHILOSOPHY OF EVIDENCE LAW: JUSTICE IN THE SEARCH FOR TRUTH 135-43 (2008); ALEX STEIN, FOUNDATIONS OF EVIDENCE LAW 80-106 (2005); Roger C. Park & Michael J. Saks, *Evidence Scholarship Reconsidered: Results of the Interdisciplinary Turn*, 47 B.C. L. REV. 949, 992 & nn.205-10 (2006) (noting that scholars have applied the insights of Bayesian statistics in various contexts, and collecting sources); Amit Pundik, *Statistical Evidence and Individual Litigants* (Mar. 15, 2008) (unpublished manuscript), available at <http://ssrn.com/abstract=987106>.

26. Daniel Shaviro, *Statistical-Probability Evidence and the Appearance of Justice*, 103 HARV. L. REV. 530, 531 (1989). Ron Allen also endorses the view that statistical evidence and other types of evidence are qualitatively similar. In his opinion, the statistical evidence debate is irrelevant, as the problem lies with the very conceptualization of civil trials. See Allen, *supra* note 2, at 1093.

27. Shaviro, *supra* note 26, at 532.

28. *Id.* at 543.

29. Scholars who object to statistical evidence on grounds of probative deficiency include Richard W. Wright, who claims that statistical evidence lacks the appropriate causal connection. Richard W. Wright, *Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof: Pruning the Bramble Bush by Clarifying the Concepts*, 73 IOWA L. REV. 1001, 1054 (1988).

Alex Stein argues that statistical evidence lacks in-case specificity. STEIN, *supra* note 25, at 100. According to Stein, the qualitative distinction between statistical evidence and case-specific evidence is that the latter can be subjected to counterfactual testing and argumentation. See *id.* Under the dictates of "the principle of maximal individualization" (PMI) laid down by Stein, evidence must be case-specific—namely, it must allow opposing parties to present counterfactual arguments (or counter-“[f]act-generating arguments,” as Stein

The inconsistent treatment of statistical evidence in both legal doctrine and the evidence law literature has created a need for an overarching theory. The objective of this Article is to provide such a theoretical framework and to dispel some of the incoherence associated with the treatment 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 epistemic discussion of similar phenomena, and it uses this epistemic discourse to highlight Sensitivity—the requirement that a belief be counterfac-

terms them). *Id.* Stein's argument proceeds as follows: The concept of probability captures, according to Stein, "not only the calculated chances that probability estimates express, but also the degrees of evidential support that probability estimates rest upon." *Id.* at 81. Stein draws a distinction between the calculated chance that a hypothesis is true (which he terms its "probability estimate") and the extent to which the specific facts that underlie the examined hypothesis are confirmed by the evidence (which he terms "weight"). *Id.* at 81-82. Estimated probabilities, claims Stein, differ in the strength of their evidential credentials, and this determines their evidentiary "weight." *Id.* at 81. This introduces the notion of case-specificity to the attachment of estimated probabilities. *Id.* at 82. Only case-specific evidence can provide evidential credentials and support the weight requirements. *See id.* at 82-83. These requirements are formulated in the PMI, which is comprised of two elements: first, decisionmakers must receive all case-specific evidence pertaining to the case in question, and second, any finding against a litigant ought to be susceptible to maximal individualized examination—namely, it must be counterfactualizable. *Id.* at 100. According to Stein, the problem with the Blue Bus scenario, like other instances of statistical evidence and paradoxes of legal proof, is that it is characterized by a lack of case-specific, or weighty, evidence. *Id.* at 85.

Those who object to statistical evidence on moral and policy grounds include Lea Brilmayer and Lewis 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). Amit Pundik claims that statistical evidence undermines the individuality and autonomy of the person against whom it is being used, in light of the assumptions decisionmakers 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. EVIDENCE & PROOF 303 (2008).

Laurence H. Tribe asserts a qualitative distinction between the outcome of a wrongful conviction when the trier has been fully convinced of the defendant's guilt and a wrongful conviction when the trier is conscious of potential error. Tribe, *supra* note 25, at 1382-83. Charles Nesson makes a similar claim, arguing that rulings based on statistical evidence may be illegitimate and unacceptable, despite any potential probative value of the evidence, because of their adverse effect on 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).

Finally, Richard A. 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. Posner, *supra* note 19, at 1512.

Some fail to distinguish between these different strands of objection or launch a dual attack against statistical evidence. *See* Koehler, *supra* note 25, at 141.

tually sensitive to the truth—as a way of epistemically explaining the legal suspicion of statistical evidence.

The theory also claims that the epistemic distinction cannot satisfactorily vindicate the reluctance to rely on statistical evidence. Knowledge, Sensitivity, and epistemology—we claim—carry little, if any, legal value. Instead of the epistemic story, we therefore offer an incentive-based story, vindicating the suspicion toward statistical evidence. However, as we show in this Article, the epistemic story and the incentive-based story are closely knit and interestingly related in light of their similar structures and ramifications in the legal arena. Using these theoretical foundations, we expose the intuitions underlying the prevailing differential treatment of statistical evidence in the doctrine and explain why some types of statistical evidence are regarded by the courts as admissible and sufficient for substantiating liability, while others are not. In addition, the Article highlights the prescriptive contribution of our theoretical framework by providing criteria for legal reform and revising the treatment of statistical evidence in certain contexts.³⁰

The Article proceeds as follows. In Part I, we review some of the existing theoretical endeavors to explain the distinction between statistical evidence and individualized evidence and point out 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 the existing doctrinal puzzles and guide legal reform. Part IV concludes the discussion.

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 extensive, and it contains various attempts to justify the distinction between statistical evidence and individualized evidence.³¹ We will begin by mapping out the most influential suggestions in the literature and highlighting their shortcomings. This will both underscore the gravity of the problem and enable an appre-

30. This Article offers a detailed account of the incentive-based story underlying statistical evidence and addresses the derivative doctrinal issues. The epistemic account, in contrast, is outlined in a preliminary fashion only, as the epistemic intricacies of statistical evidence are of less interest to a legal audience. For a comprehensive and detailed account of the epistemic foundations, see David Enoch et al., *Statistical Evidence, Sensitivity, and the Legal Value of Knowledge*, 40 PHIL. & PUB. AFF. 197 (2012).

31. For a good survey, see HO, *supra* note 25, at 136-40. For an earlier, much more critical discussion of many of these suggestions, see Ferdinand Schoeman, *Statistical vs. Direct Evidence*, 21 NOÛS 179, 187-97 (1987).

ciation of the distinctive features (and advantages) of the account that we propose.³²

A. *Exogenous-Factor Claims*

Richard Posner claims that a resort to statistical evidence is in itself proof that no other evidence could be found, and that this in itself indicates the weakness of the plaintiff's (or the prosecution's) case.³³ If this is, indeed, the case, statistical evidence should be accorded less weight simply because it tends to be submitted in circumstances in which the case of the party presenting the evidence is weaker.

It may be that plaintiffs and prosecutors seek to include statistical evidence more often in weak cases, though much empirical work would be necessary to substantiate the argument. But regardless of the existence of such a correlation, Posner's argument fails to explain why after holding all else equal, including general case strength, the intuition of a distinction between the two types of evidence seems to persist. So explanations of this kind will not suffice.

B. *About-Relation Claims*

Another claim stated in the literature is that there is an important distinction to be made between evidence that is genuinely about the relevant defendant and merely statistical evidence, which is considered to be unrelated to the defendant's matter.³⁴ 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. This about-relation claim does not help.³⁵ In the context of evidence, the only "about" that is relevant is the epistemic "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 using such about-talk to capture the intuitive distinction between statistical and individualized evidence. But doing so without giving considerably more details regarding the about-relation amounts not to an explanation or a vindication of the distinction but merely to its renaming.

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

33. Posner, *supra* note 19, at 1509.

34. See Wright, *supra* note 29, at 1050.

35. For a similar claim, see Schoeman, *supra* note 31, at 183.

C. The Causal-Connection Argument

Judith Jarvis Thomson suggests that the difference between statistical and individualized evidence should be understood causally.³⁶ Individualized evidence, she claims, is causally linked in an appropriate way to the thing for which it is taken as evidence.³⁷ In the Blue Bus case, it is the fact that the bus that inflicted the harm was blue that resulted in the eyewitness testimony, and (so the argument goes) that fact brought about the testimony 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.³⁸ She also holds that they are necessary for justifiable legal factfinding,³⁹ in part because she believes that knowledge is a necessary condition for justifiable legal factfinding (at least in criminal cases).

Yet the causal mechanism does not capture the legal distinction between statistical evidence and individualized 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. Cases can easily be constructed—cases with multiple causes, independent causal chains, different facts that suffice causally only together, different facts each of which suffices causally alone, etc.—where it is not clear what follows from a causal theory and where, to the extent that it is clear, the implications are intuitively unacceptable. For instance, in some versions of the gatecrasher case, the fact (if it is a fact) that the relevant person crashed the gate is partly causally relevant to the precise nature of the statistical evidence. (Had he purchased a ticket, the statistics would have been slightly different.) Still, if the only relevant evidence against him is the statistical evidence, presumably we do not know that he crashed the gate. In order to do justice to the causal theory's underlying intuition, some further restrictions on the nature of the causal relation are needed.

Lastly, consider a certainty case where, for instance, no one at the stadium in the gatecrasher hypothetical purchased a ticket. There, the evidence intuitively 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's having purchased a ticket and, say, John's gatecrashing. Thomson addresses the certainty case, but instead of showing how her theory can ac-

36. JUDITH JARVIS THOMSON, *Liability and Individualized Evidence*, in *RIGHTS, RESTITUTION, AND RISK: ESSAYS IN MORAL THEORY* 225, 230 (William Parent ed., 1986).

37. *Id.*

38. *See id.* at 234-42.

39. *Id.* at 242-45.

commodate its desired result, she “bypass[es]” it.⁴⁰ This is objectionably ad hoc. At the very least, a theory that could 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.

D. *Defendant-Specific Claims*

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 be punished for being a member of a reference class.⁴¹ 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 toward 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 evidence, whether the defendant did in fact gatecrash. In making this determination, the statistical evidence seems relevant—and, if it is not, the fact that it is not still remains to be shown. This point is underscored by the fact that there is something statistical about individualized evidence as well.⁴² Indeed, it is precisely in this context that it becomes tempting to insist—as some have⁴³—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 cases of both statistical and individualized evidence, we punish *for* the offense *by* relying on evidence.

E. *Justice-in-the-Particular-Case Claims*

Relatedly, it is sometimes maintained that since a 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 a more efficient result in the overall class of cases or the result that is likely to minimize the global risk of error.⁴⁴ This argument, however, is also insufficient to validate the distinction between statistical and

40. *Id.* at 248.

41. See Mark Colyvan et al., *Is It a Crime to Belong to a Reference Class?*, 9 J. POL. PHIL. 168, 171-72 (2001); Richard Lempert, *The Economic Analysis of Evidence Law: Common Sense on Stilts*, 87 VA. L. REV. 1619, 1669 (2001).

42. See Michael J. Saks & Jonathan J. Koehler, *The Individualization Fallacy in Forensic Science Evidence*, 61 VAND. L. REV. 199, 205 (2008).

43. See, e.g., Koehler & Shavero, *supra* note 1, at 263 (“All evidence is probabilistic . . . Base rate evidence is different only in that it makes these uncertainties explicit.”).

44. See Erik Lillquist, *Recasting Reasonable Doubt: Decision Theory and the Virtues of Variability*, 36 U.C. DAVIS L. REV. 85, 140 (2002).

individualized 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 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 individualized evidence.⁴⁵

F. *Autonomy Considerations*

Yet another argument in the literature is that relying on statistical evidence violates the relevant party's autonomy and individuality, and perhaps even her free will and agency.⁴⁶ 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 cause to reject such evidence?

This line of argument can also be rejected. While there is 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, for instance, cases like the Blue Bus hypothetical. The Blue Bus Company, after all, does not possess dignity in the same sense real persons do, and so its treatment is not restricted by autonomy considerations in the same way the treatment of real people is. Secondly, and more importantly, this claim confuses epistemology and metaphysics: Statistical evidence is relevant only as evidence. By using a statistic to give reason to believe 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.

45. 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 HO, *supra* note 25, at 138-40. 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. He states that, in cases relying on 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.

46. See David T. Wasserman, *The Morality of Statistical Proof and the Risk of Mistaken Liability*, 13 CARDOZO L. REV. 935, 942-43 (1991); see also Pundik, *supra* note 25, at 1-2.

G. *Social-Acceptability Considerations*

Another attempt at rationalizing the distinction between the two types of evidence is Charles Nesson's well-known claim that verdicts based on statistical evidence are socially unacceptable.⁴⁷ Nesson maintains that the court's legitimacy is contingent on the public's perception of the verdict as a statement about the actual event.⁴⁸ Statistical evidence transforms the message conveyed by the court from one of certainty to one of risk calculation. In doing so, it expressly states the risk of error underlying the judicial verdict and may thereby weaken the system's legitimacy in the eyes of the public.⁴⁹

This claim can also be rejected. First, its empirical basis is unpersuasive, as it is nearly impossible to delineate the boundaries of what would be acceptable to the public. Second, even assuming the empirical problem away, there is room to question whether public trust is even a goal that ought to be attained. True, 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. We are even willing to assume for the sake of argument (because whether this is so is very unclear) that securing public trust can sometimes justify catering to the prejudices of the masses. Still, if there is no other way to justify the traditional skepticism toward statistical evidence, then this feature of public opinion is indeed a prejudice, which renders the call to accommodate it suspicious. Furthermore, for our purposes here, we can simply assume away the problem with the premise that the public is going to form an accurate opinion about statistical evidence. In this (perhaps hypothetical) case, 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—not the fact that it is an opinion generally held by the public. Here, too, then, the public opinion argument can be safely dismissed.⁵⁰

H. *The Guaranteed-False-Conviction Argument*

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

47. See Nesson, *supra* note 29, at 1357, 1359, 1379.

48. *Id.* at 1358.

49. *Id.* at 1378 (“Cases of naked statistical proof present the most provocative example of probable verdicts that are unacceptable. In these cases, the evidence suggests a sufficiently high numerical probability of liability, but the absence of deference-inducing mechanisms in the judicial process is such that the public is unable to view a verdict against the defendant as a statement about what actually happened.”).

50. For a more elaborate discussion of similar claims made against Nesson, see Shaviro, *supra* note 26, at 534-35.

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 nonstatistical cases, although the probability of finding an innocent party guilty might be higher than in each of the gatecrasher trials, there would be no certainty that an innocent person will be convicted. Since a guaranteed wrongful conviction is something we as a society seek to avoid, the conclusion is that only nonstatistical evidence ought to be accepted. 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, some innocent defendants are virtually guaranteed to be convicted. Against the background of inherent uncertainty, the only way to avoid wrongful convictions is to never convict, thereby abolishing the criminal justice system altogether. The second point can be illustrated with a variant of the gatecrasher hypothetical in which 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 present.

* * *

We should emphasize that this quick critical survey is not intended to be either conclusive or comprehensive in scope. It does not cover all attempts that have been made to vindicate the distinction.⁵¹ Still, we hope it succeeds in giving a sense of the depth of the problem, portraying the theoretical attempts to contend with it, and conveying the need for a new resolution.

II. THE THEORETICAL FRAMEWORK⁵²

A. *Justifying Beliefs or Actions?: We Require an Epistemic, Not a Practical, Framework*

In this Part, we propose our explanation for the suspicion of statistical evidence. But before delving into this issue, another preliminary issue must be addressed: In comparing statistical and individualized evidence, we must begin

51. Additional important contributions to the statistical evidence debate include the argument formulated by Richard H. Field, Benjamin Kaplan, and Kevin M. Clermont linking naked statistical evidence to various instances of failure to satisfy the burden of production. See RICHARD H. FIELD ET AL., *MATERIALS FOR A BASIC COURSE IN CIVIL PROCEDURE* 1355-56 (10th ed. 2010). Chris Guthrie, Jeffrey J. Rachlinski, and Andrew J. Wistrich ground the distinction between statistical and individualized evidence on behavioral grounds and on the representativeness heuristic. See Guthrie et al., *supra* note 13, at 810.

52. The theoretical framework set forth in Part II draws on Enoch et al., *supra* note 30.

with the question of whether we are concerned with justifying our beliefs, our actions, or both. That is, in the context of cases like the Blue Bus Company hypothetical, we may wonder what we should *believe* about the identity of the harm-causing bus; or we may wonder what we should *do* in this regard, and how we should proceed. While the two kinds of questions may be interestingly related, the precise nature of the relation between the epistemic questions having to do primarily with the justifications of *beliefs* and the practical ones having to do primarily with the justifications of *actions* is neither obvious nor uncontroversial.⁵³ We will revisit these questions in some detail below.

Looking for an epistemic vindication of the distinction between statistical and individualized evidence amounts to insisting on some positive epistemic status that the belief that the harming bus belonged to the Blue Bus Company has in the eyewitness version of the hypothetical but that is absent in the market share version. Perhaps, for instance, the belief that the bus belonged to the Blue Bus Company is justified on the basis of eyewitness evidence, but not on statistical evidence; or perhaps it's somehow—despite the equal probabilities—*more* justified in the former than in the latter; or perhaps the belief can amount to knowledge when supported by eyewitness testimony, but not when supported by statistical evidence. There are other possible epistemic distinctions here, and different attempts at an epistemic vindication of the distinction will endorse different epistemic distinctions.

Another very different way to proceed in explaining the distinction between statistical and individualized evidence is not epistemic, but practical. When looking for practical distinctions, we may concede, for the sake of argument, that when it comes to justifying the relevant *belief*, statistical and individualized evidence are exactly alike. Nonetheless, we could still insist that we should proceed differently in practice in these cases. Perhaps, for instance, reliance on statistical evidence and reliance on individualized evidence create different incentive structures,⁵⁴ or have different implications in terms of administrative costs. If so, we should treat these pieces of evidence differently for *practical* purposes, despite their epistemic similarities.

Importantly, then, if we seek epistemic answers, the main distinction between statistical and individual evidence will be about the justifications of the *belief*; differences in the justifications of the relevant *actions* can be expected to fall outside of the epistemic distinction and in the practical realm. Perhaps, for instance, we should find against the Blue Bus Company in the eyewitness case and not in the market share case because only in the former, but not in the latter, can our *belief* that it was that company's bus amount to knowledge.

We can distinguish between epistemic and practical strategies for vindicating the distinction between statistical and individualized evidence.⁵⁵ And cru-

53. *See id.* at 200-01.

54. This will be discussed more thoroughly in our discussion in Part II.C below.

55. For somewhat similar distinctions, see Redmayne, *supra* note 3, at 285; and Schoeman, *supra* note 31, at 187.

cially, we should notice that they differ in scope. Practical justifications are highly sensitive to the relevant practical circumstances, such as the existence of incentive structures or the costs of making a decision on a certain basis, whereas epistemic distinctions seemingly apply wherever there are beliefs. Thus, practical justifications—incentive-based stories, differential administrative costs, and the like—will tend to vindicate the distinction only in legal contexts (and indeed, perhaps only in a subset of those). So if it can be shown that the problem extends beyond the legal context—that a vindication of something similar to the distinction between statistical and individualized evidence also is needed in nonlegal, and indeed in nonpractical, contexts—then the scope of practical vindications seems ill suited for the task. Consequently, what we would need is an epistemic vindication, one that applies in *all contexts* in which the problem arises.

As we will show below, the problem *does* seem to be general, and therefore an epistemic solution is needed.

B. *Some Epistemology*⁵⁶: Introducing “Sensitivity”

To demonstrate the general nature of the problem and the solution that the epistemic perspective may offer, what we need is a nonlegal, and indeed nonpractical, case where something like the statistical-individual distinction seems to be doing serious epistemic work. A version of the lottery paradox from the epistemic literature serves this purpose well.⁵⁷ Once again, we have two versions.

In the first version, you buy a ticket to a one-in-a-million lottery. You know the probabilities, of course, and perhaps on that basis you *believe* that your ticket is not a winning ticket. Prior to receiving any indication of the lottery results, do you *know* that your ticket is not a winning ticket? The almost unanimous answer here would be no. Under these circumstances, you do not possess knowledge that your ticket is not a winning ticket.

In the second version of the lottery paradox, the odds of winning the lottery to which you buy a ticket are somewhat better—one in a thousand. You purchase the ticket and wait for a day. The winning ticket is then declared, and the number made public in the newspaper does not match yours. However, a mistake in the newspaper—while unlikely—is not an impossible scenario. Suppose that with all the information made available to you taken into account (namely,

56. Some of the general epistemic discussion is further elaborated in Enoch et al., *supra* note 30.

57. *See id.* at 202-04. A more extensive presentation of the knowledge-related lottery puzzles can be found in JOHN HAWTHORNE, *KNOWLEDGE AND LOTTERIES* (2004).

This Subpart is survey-like 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, Jr.’s arguments in HENRY E. KYBURG, JR., *PROBABILITY AND THE LOGIC OF RATIONAL BELIEF* (1961).

the initial odds of the lottery and the fact that the number in the newspaper is not that of your ticket), the chances of your ticket still being a winning ticket are exactly one in a million. Now, having looked in the newspaper, do you *know* that your ticket is not a winning ticket? The almost unanimous response here would be yes. Under these circumstances, most people would hold that you do possess knowledge that your ticket is not the winning ticket.

The two cases were described in such a way that the probability of the ticket being a winning one is constant (one in a million). Yet in the first version, it seems that you do not possess knowledge (that your ticket is not a winning one), while in the second version, you do possess such knowledge. What can possibly explain this difference?

The intuitive difference between the two cases seems to be the following: In the newspaper case, it is no accident that your belief is true. It is true because in forming your belief, you're tracking the newspaper report, which in turn tracks the truth. You form your belief in the way in which a responsible believer would. And for these reasons, your belief is adequately sensitive to the truth. That is, had your ticket been the winning one, in all likelihood you would have believed that it was, and you wouldn't have continued to believe that it *wasn't* the winning ticket (because you still would have followed the newspaper report, which, in all likelihood, would have correctly reported that your ticket *was* the winning one). But in the first case, in which you base your belief (that yours is not a winning ticket) on just the statistics, none of this is true. In this case, if your belief ends up being true, it merely *happens* to be true. It's a fluke. You base your belief on the statistics alone, not on anything—such as the newspaper report—that tracks the truth in the individual case. You don't seem to be forming your belief in a responsible manner. And for these reasons, your belief is not sensitive to the truth: even had your ticket been the winning one, you would have still believed that it wasn't, because you would have based your belief on the one-in-a-million statistic, which would still be in place.

What the lottery hypothetical seems to indicate is that one of the conditions necessary for knowledge (or at least an important epistemic condition) is *Sensitivity*:

Sensitivity: *S*'s belief that *p* is sensitive =_{df} Had it not been the case that *p*, *S* would (most probably⁵⁸) not have believed that *p*.⁵⁹

The belief that your ticket is not a winning one satisfies Sensitivity in the newspaper case, but not in the mere statistics case, and this seems to explain (at least in part) their very different epistemic statuses.

58. 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 et al., *supra* note 30, at 204 n.14.

59. This definition was first proposed by Enoch et al. *See id.* at 204.

Something very similar—both in terms of the underlying intuitions and in terms of the somewhat more precise Sensitivity conditions—can be said about the Blue Bus hypothetical and the gatecrasher hypothetical.⁶⁰ When we find against the Blue Bus Company based on the reasonably reliable eyewitness, we ground our finding in the testimony, which in turn tracks the truth (imperfectly, but reasonably reliably). In other words, we show sensitivity to the truth, and accordingly, Sensitivity is satisfied: had the damaging bus not been one of the Blue Bus Company's fleet, we would have most probably not found against the Blue Bus Company (because the eyewitness would be unlikely to testify that it was a Blue Bus bus). If, however, we base a finding against the Blue Bus Company just on the basis of the market share evidence, then whether or not the finding matches the facts seems to be a matter of luck; we do not base our finding on anything that tracks the truth. And accordingly, Sensitivity is not satisfied: even had the bus not been the Blue Bus Company's, we would have still found that it was (because the market share evidence would remain the same under this scenario). The same is true for the gatecrasher case: If we rely merely on the high percentage of gatecrashers among those attending the stadium event in order to convict, we render our conviction insensitive. Even had the accused not crashed the gates, the statistics would have been highly similar, and so we would have still convicted him.

We can now revisit a diagnostic point from the previous Subpart. There we noted that practical, instrumental vindications of the distinction between statistical and individual evidence will tend to be law-specific: they will tend to apply only to the practical circumstances that are relevant to the specific legal arrangement. What the lottery cases teach us is that such solutions cannot be fully adequate. The problem, as these cases demonstrate, is a general one, and it applies more broadly than to just the law. And so, ideally, the solution to look for is similarly general. Put another way, because the very same problem arises in purely epistemic cases,⁶¹ where no action is at stake, the kind of solution to

60. The evidence law literature on statistical evidence (or on the proof paradoxes) has recently started to address the more general epistemic issues here, but it 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. STEIN, *supra* note 25, at 67. Mike Redmayne discusses our version of the paradox and explicitly draws an analogy between the evidence law cases and the epistemic 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.) Redmayne, *supra* note 3, at 301-02. Ho briefly mentions the similarity but fails to put it to theoretical use. HO, *supra* note 25, at 168-69. No one, so far as we are aware, discusses Sensitivity in this context 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., THOMSON, *supra* note 36, at 234-36.

61. 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 be-

look for is epistemic as well. And this is what Sensitivity amounts to—a purely epistemic, independently motivated way of distinguishing between statistical and individualized evidence.

Before getting back to the law, we want to describe another possible way from the epistemic literature of vindicating the distinction between statistical and individualized evidence. This approach starts from the intuitive distinction between mistakes that do and mistakes that do not call for explanation,⁶² and so we will refer to it as the explanatory test. While it is close in certain respects to the Sensitivity approach, it is still distinct from it and will prove useful in the application to legal doctrines later in this Article.

Good evidence, we all know, sometimes misleads; what renders it good is not the fact that it never misleads, but rather that it doesn't mislead often. Importantly, not all cases of misleading evidence are alike. In some cases in which a piece of evidence misleads us, there seems to be nothing more to say, except to note that the evidence is usually good and rarely misleading, and that this time we were unlucky. This, it seems, would be the right attitude to have if we rely on the statistical evidence in the Blue Bus Company or the gatecrasher cases and then find out that it misled us. But in other cases, the fact that evidence misled us calls for explanation. This is the case with eyewitness evidence, for instance. If we rely on an eyewitness and then find out that he or she misled us, this seems to call for explanation: Why is it, we are tempted to ask, that he was mistaken on this occasion? The question makes sense and calls for an informative answer (the lighting was not good, the other company's bus looked very similar, the witness had an interest to lie, etc.). And so we have another epistemic way of distinguishing between statistical and individualized evidence. Individualized evidence is the kind of evidence that, when it misleads, calls for explanation. Misleading statistical evidence does not call for a similar explanation.⁶³

Of course, for this thought to be fully developed, many more details need to be filled in. In particular, more needs to be said about what does and what does not call for explanation. Furthermore, it would be interesting to pursue the relationship between this explanatory test and Sensitivity. Given some plausible thoughts about the relationship between explanations and counterfactuals (perhaps, for instance, if one thing explains another, then had it not been for the former, the latter wouldn't have happened), some close relation between the

lieve that the problem is not as general as we claim in the text. See Pundik, *supra* note 25, at 4-5. 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 reluctant 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.

62. See Enoch et al., *supra* note 30, at 208-09.

63. Our argument on this point is based on Martin Smith's idea of normic support. See Martin Smith, *What Else Justification Could Be*, 44 *Noûs* 10, 17 (2010).

two ways of vindicating the distinction may not be too much to hope for.⁶⁴ But even without these further details, it is clear that this explanatory distinction—between mistakes that do and mistakes that do not call for explanation—captures something that is both intuitively important and that gets the cases we already mentioned right. When we return to discussion of legal doctrine, it will prove helpful from time to time to utilize this explanatory test as well.

C. *From Epistemology to More Practical Concerns: Should the Law Care About Sensitivity (or Knowledge)?*⁶⁵

In this Subpart we ask whether the law should care about the epistemic considerations just discussed, and, in particular, about Sensitivity. The answer that will emerge, perhaps surprisingly, is negative: whether a belief or a finding is sensitive should not be a matter of legal concern—unlike whether a belief or a finding is reliable or accurate. Recall the natural thought we mentioned above about the relationship between an epistemic vindication of the distinction between statistical and individualized evidence and its practical implications: we claimed it was reasonable to suppose that the epistemic verdict will be relevant practically as well. But it is now time to question this plausible hypothesis, at least when it comes to the law.

Assume we're right in everything thus far said. Assume, in other words, that the problem in the Blue Bus and gatecrasher cases is similar to the problem in the lottery paradox. Assume also that what is needed is an epistemic and not merely a practical vindication of the distinction that each illustrates. And assume, finally, that the relevant epistemic story involves Sensitivity, or perhaps an analysis of which mistakes call for explanation. We can still ask why any of this matters when it comes to the law. Why should it make a legal difference whether a certain belief is sensitive to the truth or qualifies as knowledge? Why does it matter, from a legal perspective, whether or not some evidence is such that if it were to mislead us, it would call for explanation? To put it bluntly: Why think that the law should care about epistemology at all? In what follows, we're going to essentially concede that it should not. But Sensitivity is going to emerge as practically important nonetheless (as will related epistemic conditions, such as the ones employed by the explanatory test). In Part II.C.1, we start by stating the puzzle more clearly and precisely. In Part II.C.2, we then proceed to present Chris Sanchirico's theory of character evidence. This theory has nothing directly to do with the epistemology and is grounded in instrumen-

64. That there is *some* close relation between explanations and counterfactuals is not plausibly deniable. But the relation is not that of identity. Not all counterfactuals indicate explanatory work in the relevant direction; think here of so-called backtracking counterfactuals, like "Had Gore become President, this would have had to be because one of the Justices in *Bush v. Gore* voted differently than he or she actually did." And it's not clear that all explanations entail similar counterfactuals; think here of the explanatory work, if any, done by necessary facts, like water's being H₂O explaining water's observed properties.

65. See Enoch et al., *supra* note 30, at 210-16.

tal reasoning about incentives. In Part II.C.3, we generalize Sanchirico's account and apply it to statistical evidence. Thus, we will present a practical, incentive-based vindication of the distinction between statistical and individual evidence. But that story, as we will also show, is not independent of the epistemic story we've been telling so far. Interestingly, our epistemic story and Sanchirico's incentive-based story are intimately related; counterfactuals that are particular instances of the general Sensitivity condition and the facts depicted by them play an important part in both stories.

1. *The remaining puzzle: why care about knowledge?*

Should it make a legal difference, then, whether a finding satisfies Sensitivity?

One way to pump intuitions about which factors should be legally significant is to construct a thought experiment in which you have to choose the legal system under which your children will live. We describe two options, *A* and *B*, and make sure that they differ only in one way—the feature whose legal significance we want to explore. Then we can ask whether you should prefer *A* or *B*, or indeed, whether you should be willing to give up other important things in order to assure that your children live under legal system *A* rather than *B*.

If we apply this test to the accuracy or reliability of judicial decisions, it seems hard to deny that these are things the law should care about. If judicial mistakes are somewhat less common in system *A* than they are in *B*, this gives some reason to prefer *A* over *B* as the system under which your children will live.⁶⁶ There may not be complete consensus as to how important it is that courts not err, which mistakes it is more important to avoid,⁶⁷ or whether error

66. See, e.g., Dale A. Nance, *The Best Evidence Principle*, 73 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 . . ."); 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 re-stated 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." (footnote omitted)). 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.

67. See Enoch et al., *supra* note 30, at 211. 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 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).

avoidance is more or less important than other social goals.⁶⁸ But it is unequivocally agreed on that courts ought to avoid too many “big” mistakes. Whatever the functions of the law, whatever good it helps achieve, its ability to do so depends on factfinding accuracy.⁶⁹ Furthermore, parties seem entitled to court procedures that will render mistakes that infringe on their rights sufficiently improbable.⁷⁰ But good statistical evidence actually *promotes* accuracy. Because statistical evidence, in the cases we’ve been focusing on, is probabilistically *good*, over the long run, excluding statistical evidence is bound to lead to less accuracy (just like other cases of ignoring genuinely probative evidence). Why should we be willing to pay this price?

A simple thought experiment suggests that we should not. Assume that the epistemic stories from above are along the right lines, so that statistical evidence is epistemically inferior compared to individualized evidence. If so, by excluding statistical evidence and relying exclusively on individual evidence, the law will render its findings more epistemically respectable; perhaps, for instance, more of them will rely on what amounts to knowledge. Still, by excluding statistical evidence, the law will render its findings overall *less* accurate, as compared to a policy of including all probative evidence. Should the law trade some accuracy or reliability in return for more epistemic respectability? We can apply our test again. Suppose that legal system *A* is the more epistemically respectable one. Perhaps, for instance, in *A* (but not in *B*) courts only base their judgments on what they know, on beliefs that are sensitive to the truth. But suppose that it is system *B* that is somewhat more accurate. Which system do you choose for your children to live under? Is it more important for you, say, to minimize the risk of your children being wrongly convicted, or to assure that they will not be convicted unless a jury’s belief fulfills the epistemic requirement of Sensitivity? Indeed, *how much* risk of a mistake are you willing to allow in order to assure the epistemic respectability of the legal system? The answer that seems plausible to us is none at all. To be willing to pay a price in accuracy in order to secure some epistemic respectability of the legal system looks to us like an objectionable kind of epistemic fetishism.

68. See Twining, *supra* note 66, at 272 (“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.”).

69. For a more elaborate discussion of accuracy in legal factfinding, see Louis Kaplow, *The Value of Accuracy in Adjudication: An Economic Analysis*, 23 J. LEGAL STUD. 307 (1994).

70. Delaware v. Van Arsdall, 475 U.S. 673, 681 (1986) (discussing the right of criminal defendants to accuracy and to protection against erroneous convictions); McDonough Power Equip., Inc. v. Greenwood, 464 U.S. 548, 553 (1984) (writing that civil litigants are entitled to fair trials, though not to perfect trials). Of course, other considerations, too, may be relevant to determining the right procedures. See Robert G. Bone, *Securing the Normative Foundations of Litigation Reform*, 86 B.U. L. REV. 1155, 1162 (2006).

The point is quite general, and it applies much more widely than just to Sensitivity. Whatever the relevant epistemic factors—Sensitivity, the explanatory test, etc.—it still seems like epistemic fetishism to be willing to pay a price in accuracy in order to secure these factors. But this is precisely what excluding statistical evidence on account of its epistemic deficiency amounts to.

The point is not that the law—not even evidence law—should “care” only about accuracy. Other considerations (having to do with dignity, security, privacy, the inviolability of certain relationships, or the opportunity costs of the litigation process) may, at times, trump accuracy.⁷¹ This is true in general,⁷² and it may very well be true in our context as well. Perhaps, in other words, there are some cases involving statistical evidence in which other considerations trump accuracy. This may be the case with respect to certain profiling cases—where human dignity trumps accuracy.⁷³ Our point is merely that *epistemic* considerations alone never seem to justifiably defeat considerations of accuracy when it comes to legal policy.

In this way the story of Sensitivity as an epistemically relevant condition may be thought of not as a vindication of the distinction between statistical and individualized evidence, but as a diagnosis of the common intuitions that are suspicious of statistical evidence and perhaps even the beginnings of a debunking of these intuitions. This story helps to articulate what these intuitions track, which is something like a desire for 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 them, at least when it comes to the law. The Sensitivity-based epistemic story may render the relevant intuitions understandable but not defensible as a cornerstone of legal policy. A different story is going to have to be told if the distinction between statistical and individualized evidence is to be vindicated. But that story, we will argue, is very

71. Chris William Sanchirico, *Character Evidence and the Object of Trial*, 101 COLUM. L. REV. 1227, 1228-29 (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 privileged 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 on these competing principles lies, by most accounts, in the value—inherent or instrumental—of discovering what really happened.”). For further discussion of the trade-off between accuracy in legal factfinding and the costs of trial, see Jonathan R. Macey & Geoffrey P. Miller, *Judicial Review of Class Action Settlements*, 1 J. LEGAL ANALYSIS 167, 177-78 (2009).

72. See, for instance, Mitchell N. Berman’s interesting discussion of how the fact that many sports require indisputable or conclusive evidence to reverse the initial call after instant replay is best explained not as a concern for accuracy but by other values. Mitchell N. Berman, *Replay*, 99 CALIF. L. REV. 1683, 1725 (2011).

73. See George C. Thomas III, *Blinded by the Light: How to Deter Racial Profiling—Thinking About Remedies*, 3 RUTGERS RACE & L. REV. 39, 40 (2001) (discussing stop-and-frisk policies and other profiling techniques, and rejecting them for their degrading nature, regardless of utilitarian outcomes).

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.

2. *Sanchirico on character evidence*

In order to articulate the instrumental story we need first to consider Sanchirico's work on character evidence,⁷⁴ a type of proof that features similarities to statistical evidence⁷⁵ and is subject to equally ambivalent treatment by courts. Character evidence is typically admitted in criminal cases at the sentencing phase⁷⁶ but is inadmissible, in most contexts, at the guilt phase.⁷⁷ 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.⁷⁸ 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? Alternatively, why ban evidence of such probative value when deciding on the crucial question of guilt? Sanchirico addresses this question.

His core argument is that the rule prohibiting the use of character evidence for propensity reasons can be convincingly explained and justified only by the broader scheme underlying evidence law—namely, the creation of incentives for proper out-of-court conduct.⁷⁹ While character evidence has predictive (and, therefore, probative) value, claims Sanchirico, it has no incentive value: its presence dampens the incentives for previously convicted persons 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

74. Sanchirico, *supra* note 71.

75. Character evidence may be thought of as a type of intrapersonal statistical evidence. Just as statistical evidence feels not sufficiently directly about the relevant individual, character evidence feels not sufficiently directly about the relevant action.

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

77. 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)(1). Exceptions do exist. For example, the "mercy rule" allows the defendant to introduce evidence of good character traits inconsistent with the charged conduct. If the defendant chooses to resort to the mercy rule, the prosecution is then permitted to rebut the evidence with evidence of negative traits. *Id.* 404(a)(2)(A).

78. This underlying assumption is due to recidivist tendencies. See Alon Harel & Ariel Porat, *Aggregating Probabilities Across Cases: Criminal Responsibility for Unspecified Offenses*, 94 MINN. L. REV. 261, 280 (2009) ("The pattern-of-behavior doctrines are rooted in the premise that a person who has committed several offenses in the past is more likely to either have intended or have actually committed the offense of which that person is presently accused. . . . It is the interdependence between the past offense and the present alleged offense that provides the grounds for conviction.").

79. Sanchirico, *supra* note 71, at 1260.

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 efficient incentives, we would want the actor to know that the likelihood of his being 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.⁸⁰ The prohibition on character evidence promotes deterrence by avoiding a decrease in the marginal cost of engaging in criminal behavior.⁸¹

Sanchirico's argument underscores an important 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 process more generally.⁸² 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. 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, gatecrashing, or going home and doing something else altogether. We are assuming, of course, that John has no influence on the behavior of the other people 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. 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 ninety-eight percent of those at the stadium are gatecrashers—John already knows that he will be convicted, regardless of whether 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,

80. *See id.* at 1277.

81. 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.

82. For additional discussions of the primary-behavior approach, see Louis Kaplow, *Burden of Proof*, 121 *YALE L.J.* 738, 745-46 (2012); and Kaplow, *supra* note 69, at 348.

he might as well go ahead and gatecrash; whatever he decides will have negligible effect on the likelihood of his being punished.

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 that 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, because introduction of the statistical evidence will lower the prospects that it will be held liable for such accidents.⁸³

3. *Solution: the instrumental significance of being sensitive*

At this stage, we find ourselves in the following predicament: the scope of the resistance to relying on statistical evidence is much wider than its appearance in the law of evidence, applying even in more purely epistemic settings, where nothing resembling the instrumental considerations relevant to the law is present. An epistemic 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 due to the inadmissibility of accurate (though insensitive) statistical evidence. In the legal context, what is needed is an instrumental account, in line with Sanchirico's writing on character evidence. But of course, the instrumental story cannot assist with the lottery paradox or other nonlegal cases where talk of incentives seems out of place. Is there no way out? Furthermore, is it mere coincidence that the epistemic 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 will

83. In addition to these precautionary distortions, the admissibility of statistical evidence may also impair market competition: Each of the bus companies will be incentivized to hold less than fifty-one percent of the market share so as not to suffer from the evidentiary disadvantage that a larger market share imposes. Moreover, the company holding the larger market share will absorb higher liability costs, which may lead to a decrease in its market share and to a possible exit from the market. See Posner, *supra* note 19, at 1510.

punish me. If I don't, they won't." Suppose that John proceeds to gatecrash. Then his conditional "If I don't gatecrash the stadium, they won't punish me" picks out the same fact that can later (perhaps when John is on trial) be captured by the counterfactual "Had he not gatecrashed, we would not have punished him." This counterfactual should sound familiar: it is the relevant instance of Sensitivity. The punishment and the beliefs on which it is based are sensitive if and only if this counterfactual is true. In other words, though the epistemic 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) and for a reasonably efficient incentive structure. While the epistemic story and the instrumental story do not depend on one another, they are not totally independent of each other either, for both are contingent on Sensitivity and related counterfactuals.⁸⁴

What we end up with is the following: There is a need for an epistemic story, one that will treat lottery cases and legal cases alike. Sensitivity and its epistemic significance does so. 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 individualized evidence without resorting to knowledge fetishism. The generalization of Sanchirico's account does so.⁸⁵ But the incentive-based account derived from Sanchirico's argument also relies on the truth of relevant counterfactuals, indeed the very same counterfactuals the epistemic account relies on. Thus, Sensitivity is a part of the answer to both the epistemic and the practical questions.

Note, however, that what is relevant for policy purposes is the incentive story rather than the epistemic one. (Otherwise, we really would have a case of knowledge fetishism.) If there are cases, then, where the instrumental payoffs on which the incentive account relies 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.⁸⁶ In what follows, we will apply this theoretical struc-

84. As is often the case with explanations of coincidences, one may still ask 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.

85. 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.

86. The extent of the overlap between the epistemic 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) that weigh 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

ture to the legal doctrine—to demonstrate its capacity for solving some doctrinal puzzles—and offer prescriptions for legal reform.

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.⁸⁷ Statistical inferences based on genetic profiling and 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.⁸⁸ DNA is often used for purposes of identification⁸⁹ in criminal trials⁹⁰ and family law disputes alike.⁹¹ 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.⁹² Courts also allow statistical assessments of a similar nature to be admitted in voting,⁹³ jury selection,⁹⁴ and constitutional and human rights cases.⁹⁵ In tort law cases,

the epistemic 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.

87. For further discussion of the origins and use of statistical evidence in trials, see DAVID W. BARNES & JOHN M. CONLEY, *STATISTICAL EVIDENCE IN LITIGATION: METHODOLOGY, PROCEDURE, AND PRACTICE* 3-14 (1986); and Michael I. Meyerson & William Meyerson, *Significant Statistics: The Unwitting Policy Making of Mathematically Ignorant Judges*, 37 PEPP. L. REV. 771 (2010).

88. Stephen E. Fienberg & Miron L. Straf, *Statistical Assessments as Evidence*, 145 J. ROYAL STAT. SOC'Y: SERIES A 410, 410-11 (1982).

89. 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. Koehler, *supra* note 6, at 388.

90. Richard A. Nakashima, *DNA Evidence in Criminal Trials: A Defense Attorney's Primer*, 74 NEB. L. REV. 444, 444-45 (1995).

91. *See, e.g.*, *Everett v. Everett*, 201 Cal. Rptr. 351, 356 (Ct. App. 1984) (discussing DNA testing in a paternity suit).

92. *See, e.g.*, *Hazelwood Sch. Dist. v. United States*, 433 U.S. 299, 303, 307-09, 312-13 (1977) (noting that statistical evidence can be used to prove employment discrimination on racial grounds); Scott Baker, Comment, *Defining "Otherwise Qualified Applicants": Applying an Antitrust Relevant-Market Analysis to Disparate Impact Cases*, 67 U. CHI. L. REV. 725, 728-31, 730 nn.29 & 32 (2000) (collecting Supreme Court disparate impact cases).

93. *See, e.g.*, *United States v. Blaine Cnty., Mont.*, 363 F.3d 897, 909-14 (9th Cir. 2004).

94. *See, e.g.*, *Batson v. Kentucky*, 476 U.S. 79, 96-98 (1986); *see also* Michael O. Finkelstein, *The Application of Statistical Decision Theory to the Jury Discrimination Cases*, 80 HARV. L. REV. 338, 347-48 (1966) ("The opinions of the Court in underrepresentation cases leave little doubt that disparity between the proportion of Negroes on venires and in the population generally is evidence of discriminatory selection if it is large, continuing, and unexplained.").

statistical evaluation is used to prove causation.⁹⁶ It is also used for the calculation of economic damages—for example, when courts use base rates that identify incomes lost by similarly situated individuals.⁹⁷ Such evidence is likewise used in the context of competition law.⁹⁸

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 contexts similar to the Blue Bus and gatecrasher hypotheticals: namely, cases in which the statistical evidence provides a base rate for the defendants' liability⁹⁹ (as opposed to a factual characteristic that is isolated from the ultimate legal question of guilt or liability) and establishes a conclusion that is itself nonstatistical. More specifically, we will focus on situations in which the defendant's liability-triggering, or guilty, conduct is inferred based on reference to membership in a particular population or reference class.¹⁰⁰

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 theo-

95. See, e.g., *Adarand Constructors, Inc. v. Peña*, 515 U.S. 200, 212 (1995); *City of Richmond v. J.A. Croson Co.*, 488 U.S. 469, 479-80 (1989).

96. See, e.g., *In re Hanford Nuclear Reservation Litig.*, 292 F.3d 1124, 1131-32 (9th Cir. 2002). Statistical evidence of this nature played a central role in the tobacco litigation. See, e.g., *Blue Cross & Blue Shield of N.J., Inc. v. Philip Morris, Inc.*, 113 F. Supp. 2d 345, 372-73 (E.D.N.Y. 2000) (permitting statistical evidence of this nature in tobacco litigation). But this type of evidence was rejected in other mass tort cases. See, e.g., *In re Fibreboard Corp.*, 893 F.2d 706, 709-12 (5th Cir. 1990); see also Amit Pundik, *The Epistemology of Statistical Evidence 1 & n.1* (Mar. 30, 2011) (unpublished manuscript), available at <http://ssrn.com/abstract=1134655> (describing the case law, including *Blue Cross* and *Fibreboard*, as reflecting a "lack [of] a systematic approach to statistical evidence"). In product liability cases (when the market share liability doctrine is implemented), statistical evidence is used to determine market share. See *In re "Agent Orange" Prod. Liab. Litig.*, 597 F. Supp. 740, 823-26 (E.D.N.Y. 1984), *aff'd*, 818 F.2d 145 (2d Cir. 1987); *Sindell v. Abbott Labs.*, 607 P.2d 924, 937 (Cal. 1980).

97. Koehler, *supra* note 6, at 398-99; see also *Wilson v. B.F. Goodrich Co.*, 627 P.2d 1280, 1285 (Or. Ct. App. 1981), *aff'd*, 642 P.2d 644 (Or. 1982); Raymond S. Strangways & Michael T. Zugelder, *General Versus the Specific: Forecasting Wage Growth in Injury and Death Cases*, J. LEGAL ECON., Fall 1998, at 1, 3.

98. See, e.g., *United States v. Columbia Pictures Corp.*, 25 F.R.D. 497 (S.D.N.Y. 1960).

99. 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. See *Sindell*, 607 P.2d at 936-38.

100. 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. Such cases are often termed in the literature as involving "naked statistical evidence." Koehler & Shaviro, *supra* note 1, at 264. We, in contrast, will discuss the use of statistical evidence from both the admissibility and sufficiency perspectives.

retical framework can explain the different legal regulation of these types of statistical evidence under prevailing law.

A. *DNA Evidence*

DNA evidence is an interesting illustration of the adaptability of our theory to the legal arena. Despite the statistical nature of DNA evidence, which identifies the frequency with which genetic profiles occur in reference populations, courts seem rather happy to rely on it.¹⁰¹ 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 distinction between statistical and individualized 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,¹⁰² leading to numerous convictions and hundreds of postconviction exonerations.¹⁰³ DNA testing was depicted by one court as “the single greatest advance in the ‘search for truth’ . . . since the advent of cross-examination”¹⁰⁴ and has been analogized to “the finger of God.”¹⁰⁵ Like other courts around the world, American courts—at both the federal and state levels—sweepingly admit DNA evidence¹⁰⁶ in paternity suits¹⁰⁷ and in criminal trials. In 1988, Florida’s Fifth District Court of Appeal was the first appellate court in the United States to uphold the admission of DNA evidence in a criminal proceeding.¹⁰⁸ This was followed by the *United States v. Jakobetz* case,¹⁰⁹ which marked the first approval of the admission of DNA evidence by a federal

101. Andrea Roth, *Safety in Numbers? Deciding When DNA Alone Is Enough to Convict*, 85 N.Y.U. L. REV. 1130, 1132-33 (2010) (discussing the phenomenon of “pure cold hit” cases).

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

103. For further information about DNA exonerations, see *DNA Exonerations Nationwide*, INNOCENCE PROJECT, http://www.innocenceproject.org/Content/DNA_Exonerations_Nationwide.php (last visited Feb. 23, 2015).

104. *People v. Wesley*, 533 N.Y.S.2d 643, 644 (Cnty. Ct. 1988), *aff'd*, 589 N.Y.S.2d 197 (App. Div. 1992), *aff'd*, 633 N.E.2d 451 (N.Y. 1994).

105. Bolden, *supra* note 22, at 410 (quoting *DNA Links Convict to 21-Year-Old Slaying; Evidence Likened to “the Finger of God,”* RECORD (N.J.), Mar. 14, 2005, at A5) (internal quotation marks omitted).

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

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

108. *Andrews v. State*, 533 So. 2d 841, 850 (Fla. Dist. Ct. App. 1988); see *id.* at 843 (noting the absence of previous appellate decisions).

109. 955 F.2d 786, 789, 795 (2d Cir. 1992).

appellate court. By 1990, thirty-eight states had admitted DNA evidence,¹¹⁰ and by the mid-1990s, most states were allowing DNA test results to be admitted as evidence in criminal trials.¹¹¹ Whether the *Frye* standard, *Frye-Kelly* standard, or *Daubert* standard is applied, DNA evidence is currently almost universally accepted in both federal and state courts.¹¹²

Another dimension of the role of DNA evidence in court is the evidentiary weight ascribed to it. 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 that will be specified below,¹¹³ courts and juries are more prone to convict on the basis of DNA evidence when it is corroborated by other types of evidence, but most courts do not rule out the possibility of convicting on DNA alone.¹¹⁴ Indeed, when the probabilities of guilt given the DNA match have been sufficiently high, courts have convicted solely on the basis of DNA evidence.¹¹⁵ Moreover, at least one court has declared that DNA evidence alone is a sufficient basis for conviction even in the face of conflicting eyewitness evidence.¹¹⁶

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 evidence. Rather, it will assume a

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

111. *See id.*; *see also* George Bundy Smith & Janet A. Gordon, *The Admission of DNA Evidence in State and Federal Courts*, 65 FORDHAM L. REV. 2465, 2478-81 (1997).

112. *See* Frank B. Ulmer, Note, *Using DNA Profiles to Obtain "John Doe" Arrest Warrants and Indictments*, 58 WASH. & LEE L. REV. 1585, 1598 (2001) (asserting 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 courts. 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, 895-99 (Mich. Ct. App. 2000).

113. *See infra* Part III.D.

114. Roth, *supra* note 101, at 1155.

115. *Id.* at 1143 & n.54; *see also* Brooke G. Malcom, Comment, *Convictions Predicated on DNA Evidence Alone: How Reliable Evidence Became Infallible*, 38 CUMB. L. REV. 313, 315 & n.17 (2008) (listing court decisions ruling that DNA is sufficient for conviction). This is also the case in England. *See R v. Adams*, [1996] 2 Crim. App. 467 at 470 (Eng.) (per curiam) ("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."). For a discussion of the *Adams* case, *see* Roth, *supra* note 101, at 1154-55.

116. *See People v. Rush*, 630 N.Y.S.2d 631, 633-34 (Sup. Ct. 1995), *aff'd*, 672 N.Y.S.2d 362 (App. Div. 1998); *cf.* *State v. Toomes*, 191 S.W.3d 122, 124, 129, 131 (Tenn. Crim. App. 2005) (upholding a conviction based solely on DNA where a rape victim "was unable to identify the defendant—or anyone else—as her assailant"); *Roberson v. State*, 16 S.W.3d 156, 159, 167-68, 172 (Tex. Ct. App. 2000) (upholding a conviction based on DNA where a rape victim "was unable to identify her assailant").

low probability of a false positive match. Indeed, we are going to restrict our attention to just those cases in which the probability that the accused is guilty, given that there is a DNA match, is extremely high, though not quite 1.¹¹⁷ 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 from the scene of the crime and run against some database, and a match was found; the suspect was not first pinpointed and then tested for a DNA match.) And third, we will be restricting our attention to the use of DNA evidence as evidence for the prosecution (in a criminal case).¹¹⁸ With these stipulations in place, then, can anything be said in favor of using DNA evidence, especially given the background of suspicion toward 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 to high-probability evidence like DNA, the value of accuracy is much greater in cases in which the probability is so high. This very high probability underlying the evidence is the core of the difference between the ruling in *Kaminsky v. Hertz Corp.* (in which the appellate court found a triable issue of fact where market share evidence showed that the Hertz Corporation owned ninety percent of all yellow trucks bearing a Hertz logo)¹¹⁹ and the ruling in *Smith v. Rapid Transit, Inc.* (in which the court rejected the market share evidence, which pointed to a smaller market share held by Rapid Transit).¹²⁰ In *Kaminsky*, as in DNA cases, considerations related to the value of accuracy outweighed the standard reasons for not relying on statistical evidence.¹²¹

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 Compa-

117. In symbols: $P(G|M) = 1 - \epsilon$, for a positive but arbitrarily small ϵ .

118. The legitimacy of DNA evidence as exonerating evidence, at least in the criminal context, is clear enough not to be interesting; it is often nonstatistical in nature, and, in any case, the relevant high probability of accuracy certainly suffices for reasonable doubt.

119. 288 N.W.2d 426, 427 (Mich. Ct. App. 1980).

120. 58 N.E.2d 754, 754-55 (Mass. 1945).

121. Jurors, such as those the appeals court permitted to hear the claim in *Kaminsky*, tend to convict on the basis of high-probability statistical evidence. Kevin Jon Heller, *The Cognitive Psychology of Circumstantial Evidence*, 105 MICH. L. REV. 241, 301 (2006) (“[A]lthough jurors are extremely sensitive to deviations away from certainty, research indicates they are generally willing to convict on the basis of probabilistic evidence that . . . establishes a 0.995 likelihood of the defendant’s guilt.”).

ny.¹²² This line of reasoning relies on the assumption 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 or 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* matches. 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 seventy percent. In such a case, too, the incentive story collapses, for the effect of the evidence *ex post* on the imposition of punishment is relatively low, and therefore its ability to disincentivize engagement in the relevant criminal conduct *ex ante* is similarly low. Yet the intuitive reluctance to rely on statistical evidence is still as strongly present.¹²³

122. *See supra* Part II.C.2.

123. 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 a brief suggestion in our discussion above: that at least 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 in which we indict all those attending the stadium event.) This problem of a guaranteed false decision doesn't seem to be relevant to DNA evidence, which can be systematically relied on without yielding 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 about distinguishing between a guarantee and a ridiculously high probability that we mentioned above in Part I.H, 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 event escaped before the police arrived. Indeed, suppose (again, as we did in Part I.H) 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. Thus, DNA remains special.

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 using such nonepistemic stories as guides to legal policy. But the Sensitivity requirement can add an additional dimension to the explanation of the DNA conundrum, for DNA is a unique type of statistical evidence from the Sensitivity perspective, due to possible-world considerations—to which we now turn.

Consider the following illustration. 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 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 in which the antecedent is true.¹²⁴ There is considerable intuitive pressure to think that a possible world in which *A* is innocent and yet the DNA sample matches his DNA is *further away* from the actual world than is 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 in 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 in which 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.¹²⁵

In sum, the theoretical account developed here offers an explanation for the exceptional treatment of DNA, in contrast to the usual wariness accorded to statistical evidence. Unlike other types of statistical evidence, DNA evidence seems to be sensitive; had the defendant not committed the offense, we would, in all likelihood, not have found DNA that could match hers on the scene, and so we would, in all likelihood, not have convicted her. Moreover, the incentive

124. For further discussion of the leading philosophical account regarding possible-world talk, see Enoch et al., *supra* note 30, at 204 n.14.

125. 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. DAVID LEWIS, *Counterfactual Dependence and Time's Arrow*, in 2 PHILOSOPHICAL PAPERS 32, 47 (1986). This fits nicely with the points that we make above: 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.

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.

B. *Propensity-for-Crime Evidence at the Guilt Phase of Trial*

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. 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 yet another offense. 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.¹²⁶ Yet despite the probative value of propensity-for-crime evidence, under prevailing doctrine it is generally disallowed at the guilt phase of trial.¹²⁷

This can be justified by the adverse effects, in terms of deterrence, that would result from submission of propensity-for-crime evidence at the guilt phase of trial. At first glance, propensity-for-crime evidence could be mistakenly conceptualized as the mirror image of evidence that reconstructs past events.¹²⁸ But in the context of providing incentives for engaging in certain

126. For further discussion of this “predicting violence” hypothetical, see generally Redmayne, *supra* note 3, at 283, 292-96. 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 is trace evidence, whereas the fact that *A* is a carpenter is predictive evidence. With predictive evidence, the direction of causation goes from the evidenced phenomenon to the conduct in question. With trace evidence, the causal relationship runs in the reverse direction. See Sanchirico, *supra* note 71, at 1234-35.

127. See, e.g., Hall v. State, 692 S.W.2d 769, 773 (Ark. Ct. App. 1985); Stephens v. State, 774 P.2d 60, 64 (Wyo. 1989). According to Koehler, this type of base-rate evidence is rejected most frequently in child abuse cases. Koehler, *supra* note 6, at 384.

128. Cf. Denise Meyerson, *Risks, Rights, Statistics and Compulsory Measures*, 31 SYDNEY L. REV. 507, 521 (2009) (“Although it makes sense to say that the expected incidence of re-offending in the group to which the individuals belong is 0.75, strictly speaking it does not make sense to say of any particular individual that he or she has a 75 per cent risk of re-offending or that he or she is likely to re-offend.”); Note, *Prediction Markets and Law: A Skeptical Account*, 122 HARV. L. REV. 1217, 1229 (2009) (“Most of adjudication’s fact determinations are retrospective and so may at first glance seem different from predictions about an uncertain future.”).

forms of primary behavior (as well as from the perspective of Sensitivity), the past-future distinction carries great significance. In line with Sanchirico's 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 to engage 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.¹²⁹ Inadmissibility of such evidence, therefore, promotes deterrence.¹³⁰

As discussed above, Sanchirico applies this basic intuition to the context of character evidence, arguing that admission of evidence of bad character or prior acts 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 subcategory 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 first offense. 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.

129. Sanchirico, *supra* note 71, at 1278 n.128 (“If the law grounds a current conviction on the bare fact of past convictions, rather than the underlying facts, it risks creating a sort of ‘propensity bubble.’ Suppose an individual is convicted once, perhaps correctly. Suppose that the next time, however, he actually did not commit the crime, but his past record leads to a new conviction. Now he has two past convictions and is an easy mark for prosecutors, which leads to a third conviction, which makes him an even easier mark for a fourth, etc. This feedback loop is avoided if the court refers to the evidence that produced each conviction, not just the fact of conviction. The discussion in the text assumes that the law does not make this even greater mistake.”).

130. 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 to avoid situations that might lead to their arrest in light of the greater probability of conviction once indicted. *See, e.g.,* Park & Saks, *supra* note 25, at 1016. Our answer to this critique is that while this effect may take place, as long as there are instances when it is counterbalanced by the chilling effect suggested by Sanchirico, both should be taken into account. And of course, one must also bear in mind that overdeterrence is also a problematic result.

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 time frame is crucial in this context, for when focusing on the legal payoff prior to the first crime, a rule that permits information about 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 involving propensity-for-crime evidence based on gender, age, ethnicity, or other 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 do they reduce social welfare. Exclusion of propensity-for-crime evidence relating to such nonvoluntary, 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 to which our use of his model does not.

In sum, the inadmissibility of propensity-for-crime evidence relating to features of a nonvoluntary 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 evidence of bad character at the guilt phase of trial. Although Sanchirico's analysis was intended to explain the inadmissibility of character evidence, it actually offers better reasoning for more paradigmatic types of propensity-for-crime statistical evidence, such as predictions based on age, gender, or ethnicity.

The incentive-based analysis can be complemented by the epistemic perspective of our theory. Propensity-for-crime evidence, it has been claimed, poses a fundamental challenge to Thomson's analysis because it often satisfies the causal-connection requirement.¹³¹ 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.¹³² Due to this

131. See Redmayne, *supra* note 3, at 296-97.

132. Sanchirico, *supra* note 71, at 1234-35.

reverse direction of causality (for example, 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 epistemic story, that is relevant for policy purposes, the two accounts align here.

Predictive evidence is routinely admitted into court during the sentencing phase of trial, but since this Article deals with the use of statistical evidence for purposes of imposing liability, we will only briefly address this point. Predictive evidence used at sentencing includes the defendant's age, rehabilitative potential, and other proxies for future dangerousness.¹³³ Other important evidence relating to the offender's character that is admissible during sentencing under the U.S. Federal Sentencing Guidelines includes her criminal record.¹³⁴ Alongside offense severity, prior convictions are the weightiest factor in sentence gravity.¹³⁵ State guidelines likewise incorporate prior-conviction provisions.¹³⁶

Unlike the guilt phase of trial, admitting predictive evidence at the sentencing phase is expected to increase 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 deadweight 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" on them. Tailoring criminal punishments to the "deterrence proneness" of individuals enhances the deterrence level per given cost of punishment. Predictive evidence regarding a defendant's age, 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—which is to say, for those individuals who are less easily deterred. At the same time, it allows for less costly deterrence for low-propensity offenders.

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 reoffending, 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 on them may be effectively eroded: individuals with criminal records have lower opportunity costs, there is a decrease in

133. See 28 U.S.C. § 994(d) (2013).

134. *Id.* § 994(d)(10).

135. Lee, *supra* note 76, at 571 ("In the United States, the most important determinant of punishment for a crime, other than the seriousness of the crime itself, is the offender's criminal history." (italics omitted)).

136. See Richard S. Frase, *Untitled Essay, in Is Guided Discretion Sufficient? Overview of State Sentencing Guidelines*, 44 ST. LOUIS U. L.J. 425, 425, 434 (2000) ("[N]ewer guidelines systems are likely to incorporate recidivist and 'dangerous offender' provisions, especially for sex crimes.").

the marginal cost of imprisonment years, and the additional reputational 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 is not 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 on race, sex, national origin, creed, or socioeconomic status are not incorporated into the Federal Sentencing Guidelines. But these are exceptions; as a general matter, admitting statistical evidence of a predictive nature at the sentencing phase of trial amplifies incentives and furthers deterrence, while admitting this type of evidence at the guilt phase of trial tends to impair deterrence.

C. *Incriminating Versus Exonerating Statistical Evidence*

With the exception of DNA evidence, the use of statistical evidence for conviction purposes is extremely uncommon and very controversial. Note that not every evidentiary use of statistics for incriminating purposes is analogous to the Blue Bus hypothetical. Often, statistical pieces of information are used as 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 explanations one explanation is more ontologically or mathematically simple, is more in line with what we already have reason to believe, is more enlightening, or explains a wider range of phenomena, we have reason to believe that it is the better explanation.¹³⁷ And sometimes, probabilities influence how good an explanation is. Thus, if between 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 (nonstatistical!) pieces of evidence well, whereas the hypothesis that the gardener did it leaves many details hostage to unlikely 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 statistical probability that the coincidence “explanation” is correct. But this is not what is going on in the Blue Bus or gatecrasher cases. In those cases, there is no sense in which the

137. There is much discussion in epistemic contexts of inference to the best explanation, and some of it is quite critical. For an overview and many references, see Igor Douven, *Abduction*, STAN. ENCYCLOPEDIA PHIL. (Mar. 9, 2011), <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.

relevant hypothesis (that it was a blue bus; that the accused crashed the gates) better explains, compared to alternative hypotheses, a relevant phenomenon. Rather, it is the pure use of the statistical evidence itself that is at stake. The distinction between these two ways of using statistical information—as part of a seemingly legitimate inference to the best explanation, or as purely statistical reasoning—is important, as the discussion that follows shows.

A notorious criminal law case in which statistics were used for conviction purposes was the 1899 trial of Alfred Dreyfus, a Jewish captain in the French Army.¹³⁸ Among the evidence used to convict Dreyfus were letters he had written, which, according to the prosecution, were cipher messages.¹³⁹ The prosecution attempted to substantiate its claim with statistical evidence, by showing a disproportionate frequency of certain letters of the alphabet relative to the standard pattern of French prose.¹⁴⁰ Dreyfus was convicted at trial.¹⁴¹ It should be noted, however, that the *Dreyfus* case is not analogous to the Blue Bus or gatecrasher cases either. Rather, it is a case of inference to the best explanation premised on statistical data: within the context of the case, the court sought the hypothesis that best explained the data, one hypothesis being that Dreyfus was a spy, the other hypothesis being that the unlikely letter frequency was mere coincidence. The (supposedly) low probability of the coincidence, which served to show that the first explanation was better, was subsequently challenged for the inaccuracy of the data (as opposed to its actual use).¹⁴²

Another well-known and very problematic use of statistics in a criminal trial occurred in *People v. Collins*.¹⁴³ An eyewitness account of the robbery of an elderly woman described a blonde woman and a bearded African American man fleeing the scene in a yellow car.¹⁴⁴ A man and his wife, who generally fit the eyewitness description, were arrested and brought to trial. In their trial, the prosecution offered statistical evidence that the overall likelihood of a couple meeting these criteria was 1 in 12,000,000.¹⁴⁵ The prosecution argued that given this low probability, the defendants must be guilty.¹⁴⁶ This, in other words,

138. D.H. Kaye, *The Admissibility of "Probability Evidence" in Criminal Trials* (pt. 1), 26 JURIMETRICS J. 343, 344 (1986). For a discussion of the Dreyfus trial in the statistical evidence context, see *id.* at 344-45; and D.H. Kaye, *Revisiting Dreyfus: A More Complete Account of a Trial by Mathematics*, 91 MINN. L. REV. 825, 829-35 (2007) [hereinafter Kaye, *Revisiting Dreyfus*].

139. Kaye, *Revisiting Dreyfus*, *supra* note 138, at 829-30.

140. *Id.*

141. *See id.* at 832-33.

142. *Id.* at 831-33.

143. 438 P.2d 33, 36-40 (Cal. 1968) (en banc); see Michael O. Finkelstein & William B. Fairley, *A Bayesian Approach to Identification Evidence*, 83 HARV. L. REV. 489, 490-96 (1970); Tribe, *supra* note 25, at 1334-38.

144. *Collins*, 438 P.2d at 34.

145. *Id.* at 36-37.

146. In his closing argument, the prosecutor advised the jury that the chance that another couple was the guilty couple given the 1:12,000,000 probability was "something like one

is another case of inference to the best explanation analogous to the Dreyfus trial. The jury returned a guilty verdict, but the conviction was reversed on appeal by the Supreme Court of California, partly due to its inadequate statistical basis and partly due to the very resort to statistical evidence.¹⁴⁷

The 1992 *United States v. Shonubi* case is an example of the use of incriminating statistical evidence of the Blue Bus type in a criminal trial, albeit for sentencing purposes. Shonubi, a Nigerian citizen residing in the United States, was arrested at John F. Kennedy International Airport for smuggling heroin by way of “balloon swallowing.”¹⁴⁸ 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 possessing and importing heroin. 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 427.4 grams by eight, resulting in an aggregate amount in excess of the 3000-gram threshold for sentence enhancement (corresponding to a base offense level of thirty-four).¹⁴⁹

Shonubi appealed, and the Second Circuit Court of Appeals vacated the sentence and remanded the case to the trial court for resentencing.¹⁵⁰ At the resentencing, the state presented customs service data relating to the amounts of heroin found in the possession of 117 other Nigerian drug smugglers who had transported heroin via the same balloon-swallowing method in the time between Shonubi’s first and last smuggling trips. According to this data, there was a ninety-nine percent 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.¹⁵¹ Shonubi appealed yet again, and the sentence was again 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.¹⁵² Consequently, back in the district court, Shonubi was sentenced for only the 427.4 grams of heroin found on him upon

in a billion.” George Fisher, *Green Felt Jungle: The Story of People v. Collins*, in EVIDENCE STORIES 7, 14 (Richard Lempert ed., 2006) (internal quotation marks omitted).

147. *Collins*, 438 P.2d at 33, 39-40.

148. 802 F. Supp. 859 (E.D.N.Y. 1992).

149. *Id.* at 860-61.

150. *United States v. Shonubi*, 998 F.2d 84, 91 (2d Cir. 1993).

151. *United States v. Shonubi*, 895 F. Supp. 460, 464, 468, 530 (E.D.N.Y. 1995).

152. *United States v. Shonubi*, 103 F.3d 1085, 1092-93 (2d Cir. 1997).

arrest.¹⁵³ Here, again, the rejection of the statistical evidence was focused on the application of the data in the specific circumstances of the case as well as the very resort to statistical evidence.

Another well-known use of statistics at criminal trial—albeit of the *Dreyfus* type—is the 1999 English case of Sally Clark.¹⁵⁴ 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 (SIDS). After the death of her second baby, 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 73,000,000. This calculation was later found to be statistically flawed. Regardless, Clark was convicted, and the Court of Appeal upheld her conviction.¹⁵⁵ 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).¹⁵⁶ However, in this case too, it was the controversial statistical quality of the data that was found to be lacking; there was no objection 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 major exception being DNA evidence. The majority of the difficulty and controversy in the cases in which statistical evidence was used in trial relates to the faulty statistical basis of the evidence and to questions about its 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.¹⁵⁷

Accuracy considerations can help explain the suspicion toward statistical evidence for conviction purposes. As argued earlier, the rules of evidence are designed with the purpose of promoting factfinding accuracy.¹⁵⁸ Statistical evidence seems, at first glance, to do just that: minimize the overall risk of error.

153. *Id.*; see also *United States v. Shonubi*, 962 F. Supp. 370, 371 (E.D.N.Y. 1997). For an elaborate discussion and analysis of the Shonubi trials, see Peter Tillers, *Introduction: Three Contributions to Three Important Problems in Evidence Scholarship*, 18 *CARDOZO L. REV.* 1875, 1879-89 (1997).

154. A. Pundik, *Using Statistical Evidence in Courts: A Case Study, or What Went Wrong in the Case of Sally Clark?* 1 (Feb. 10, 2007) (unpublished manuscript), available at http://ben-israel.rutgers.edu/711/Sally_Clark.pdf. For further discussion of the *Clark* case, see Adam Wilson, *Expert Testimony in the Dock*, 69 *J. CRIM. L.* 330 (2005).

155. *R v. Clark*, [2000] EWCA (Crim) 54, [256]-[258] (Eng.).

156. *R v. Clark*, [2003] EWCA (Crim) 1020, [2], [180]-[181] (Eng.).

157. See Redmayne, *supra* note 3.

158. See *supra* Part II.C.1; see also Kaplow, *supra* note 69, at 307-08 (“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.”).

But the notion of error avoidance is only one component of accuracy in the criminal trial setting. Error allocation is also of great importance, for 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 higher than those associated with false acquittal.¹⁵⁹ Minimizing the aggregate social costs of error in criminal proceedings thus entails lowering the incidence of false convictions, even by way of increasing the prevalence of false acquittals.¹⁶⁰ 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.¹⁶¹ 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 at the expense of errors in favor of the prosecution.¹⁶²

This gives rise to a possible distinction between incriminating statistical evidence and exonerating statistical evidence in terms of admissibility in 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 (even if it compromises the certainty of the innocence of the acquitted).¹⁶³

159. See Alexander Volokh, n *Guilty Men*, 146 U. PA. L. REV. 173, 174-77 (1997). This utilitarian calculus is reflected in William Blackstone's famous maxim that "it is better that ten guilty persons escape, than that one innocent suffer." 4 WILLIAM BLACKSTONE, COMMENTARIES *352.

160. Alex Stein, *The Refoundation of Evidence Law*, 9 CAN. J.L. & JURIS. 279, 294-95, 324-25 (1996) (noting that the risks of error must be allocated between the defense and the 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, 400-01, 408-17 (1973); and Frederick Schauer & Richard Zeckhauser, *On the Degree of Confidence for Adverse Decisions*, 25 J. LEGAL STUD. 27, 34 (1996).

161. 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." (footnote omitted)).

162. See Posner, *supra* note 160, at 410-15; Schauer & Zeckhauser, *supra* note 160, at 33-35.

163. 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-defendant statistical evidence in the civil trial arena. In the civil context as well, the rules of evidence 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 Shaviro, *supra* note 26, at 532 n.12; see also STEIN, *supra* note 25, at 219; Ariel Porat & Alex Stein, *Liability for Uncertainty: Making Evidential Damage Actionable*, 18 CARDOZO L. REV. 1891, 1956 (1997) ("[E]very dollar undeservedly lost, either by the plaintiff or by

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 statistical 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 "mercy rule," whereby the defendant may submit evidence of good character 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 general prohibition on the prosecution against introducing evidence of the defendant's bad character. As claimed by Sanchirico, incentive considerations dictate abolishing the mercy rule.¹⁶⁴

We disagree with Sanchirico, as far as the mercy rule is concerned, for reasons similar to the ones explored in Part III.B above—namely, that Sanchirico's analysis focuses exclusively on the legal payoff in the period of time following the act or acts suggesting positive character traits (or, in the case of bad character, suggesting negative traits). Indeed, the legal payoff faced by the potential defendant during this time period would be suboptimal if information as to her positive character traits were admitted into court at the guilt phase of subsequent trials. Such evidence may dampen her incentives not to engage in criminal activity during this specific time frame. But as we argued above, there is no reason to focus exclusively on the incentive structure in the period of time following the initiation of the good act or acts. Individuals should also be incentivized to begin with and to acquire a positive reputation. Allowing the submission of evidence of good character, thereby lowering the prospects of subsequent convictions, may incentivize individuals to engage in positive acts and to possess a good character to begin with. This positive incentive effect may outweigh (at least in certain circumstances) the negative impact—in terms of incentives—that the mercy rule would create at the later time frame, on which Sanchirico focuses. In other words, the prism through which the proper incentive structure is viewed in Sanchirico's model and in his critique of the mercy rule is not wide enough. At the same time, as we elaborated in Part III.B above, similar claims can be formulated with respect to the submission of evidence of bad character. Thus, a distinction between exonerating and incriminating evidence cannot be accommodated from an incentive-based perspective.

the defendant, ought to be regarded as equally regrettable.”). This is what justifies and even necessitates the allocation of the risk of error between plaintiffs and defendants in a roughly equal manner. See Stein, *supra* note 160, at 333-35. 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.

164. Sanchirico, *supra* note 71, at 1305.

At first sight, it may seem that epistemic Sensitivity considerations also cannot deliver 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 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.¹⁶⁵ But the crucial point here is 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 (because conviction requires belief, and Sensitivity is an epistemic-respectability property of beliefs), applying Sensitivity to exonerating evidence does not make sense (because 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 distinguish rather strongly between incriminating and exonerating statistical evidence.

D. *Admissibility Versus Sufficiency of Statistical Evidence*

In his empirical article *When Do Courts Think Base Rate Statistics Are Relevant?*, Koehler extrapolated from the case law the conditions under which appellate courts are more likely to view statistical evidence as relevant and admissible. One of the conditions he identified refers to cases in which individualized evidence is impossible to obtain. Under Koehler's account, appellate courts show a tendency toward rejecting statistical evidence and viewing it as irrelevant and inadmissible in cases in which alternative individualized evidence could have been obtained.¹⁶⁶ And even where individualized evidence is not readily accessible, there is a general reluctance to base judgments on "naked statistical evidence," as opposed to statistical evidence corroborated by individualized evidence.¹⁶⁷

165. And of course, sometimes acquittal is called for even in the face 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.

166. Koehler, *supra* note 6, at 401.

167. See Tribe, *supra* note 25, at 1350 (arguing that the fact that statistical evidence "taken alone can rarely, if ever, establish [guilt] . . . does not imply that such evidence—when properly combined with other, more conventional, evidence in the same case—cannot supply a useful link in the process of proof"); Note, *supra* note 128, at 1229 (discussing the

The incentive story may offer an explanation for this apparent preference for individualized evidence: 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 of 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 doing so 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 of relying on individualized evidence—even individualized 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 (for example, of the relative contribution of the statistical evidence to the litigation payoff). When statistical evidence is uncorroborated by individualized evidence, the incentive-corrupting effect is at 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 tends to diminish.¹⁶⁸ The result is that the greater the likelihood that the decision will be based solely on 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 individualized evidence and thereby reduce the erosive effect of statistical evidence on incentives.

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

prohibition against the use of “naked statistical evidence” (quoting FREDERICK SCHAUER, *PROFILES, PROBABILITIES, AND STEREOTYPES* 81 (2003)) (internal quotation marks omitted); see also SCHAUER, *supra*, at 80-81; Nance, *supra* note 66, at 290 n.300.

168. Posner, *supra* note 19, at 1482. For a related claim, see Harel & Porat, *supra* note 78, at 291 (“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.”).

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.¹⁶⁹ 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 judgment-proof and so not worth suing), or that he has simply not bothered to conduct an investigation.¹⁷⁰

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 individualized evidence, maintains Posner, weakens the probative value of the market share evidence.¹⁷¹ And of course, here there is room to differentiate among cases according to the availability and cost of such alternative individualized evidence.¹⁷² The less costly and more readily available the alternative individualized evidence, the stronger the signal sent by its exclusion from trial. Alternatively, the easier it is to obtain individualized evidence, the stronger the assumption that the relevant party failed to search for that evidence or that her search yielded individualized 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 individualized evidence could have been obtained.¹⁷³

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 individualized evidence may signal to the court that the relevant party failed to uncover such evidence

169. See discussion *supra* Part I.A.

170. Posner, *supra* note 19, at 1509.

171. *Id.* at 1508-10.

172. Some writers have claimed that individual evidence of liability is almost always available. See, e.g., Ronald J. Allen, *A Reconceptualization of Civil Trials*, 66 B.U. L. REV. 401, 429 (1986) ("In fact, those cases where courts have not allowed verdicts based upon statistical evidence make much more sense if viewed as involving a sanction for discovery violations or an inference drawn from the failure to produce *available evidence*." (emphasis added)); Kaye, *supra* note 10, at 104 n.19; Richard Lempert, *The New Evidence Scholarship: Analyzing the Process of Proof*, 66 B.U. L. REV. 439, 457 (1986) ("The reason is that in the real world we will never be sure if the conditions of the hypothetical are met; much more often than not they won't be met, and *more information will be available to the plaintiff*." (emphasis added)). This is an empirical matter lying outside the scope of this Article.

173. See Koehler, *supra* note 6, at 401.

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 individualized evidence, the incentives to the parties in the evidence-seeking process 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 individualized evidence, rather than the reason or justification for that distinction.¹⁷⁴ 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.

It is interesting to comment on the possible interaction between statistical evidence of the DNA type and individualized evidence where both are available. Think of cases in which the evidence for the prosecution includes both cold-hit DNA evidence and some paradigmatically individualized evidence—say, the testimony of an eyewitness.¹⁷⁵ There are several possible cases: One possibility is that the two pieces of evidence (the DNA evidence and the eyewitness testimony) were obtained independently—say, by two independent, noncommunicating police departments. In such a case, if the eyewitness testimony is sufficiently strong to justify conviction regardless of the DNA evidence, the case is easy and immaterial from a statistical evidence perspective. If, at the other extreme, the statistical evidence is very strong and the individualized evidence rather weak, then our previous discussion of DNA evidence can be applied,¹⁷⁶ including the claim that DNA is a unique type of statistical evidence from the Sensitivity perspective due to possible-world considerations. A more challenging case is one in which no piece of evidence suffices on its own for conviction, and only the combined weight of the evidence (assuming we are happy to rely 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 because the individualized 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 sta-

174. On the other hand, this dismisses the case for *ab nihilo* implementation, but under the current regime, wherein courts accord unequal evidentiary weight to statistical and individual evidence, this consideration may count against equalization.

175. Of course, if the prosecution has *only* cold-hit DNA evidence, then the discussion above applies.

176. See discussion *supra* Part III.A.

tistical evidence would still have been available, but the individualized evidence (the eyewitness 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,¹⁷⁷ statistical evidence should be admitted and relied on in mixed cases of the kind just described. And indeed, in such a case, relying on the cold-hit DNA evidence to enhance the evidential weight of the eyewitness testimony does not seem as intuitively objectionable as relying on statistical evidence alone. The above-mentioned disinclination of courts and the legal doctrine to base judgments on naked statistical evidence,¹⁷⁸ as opposed to statistical evidence corroborated by individualized evidence, echoes this intuitive difference, which Sensitivity nicely explains.¹⁷⁹

But all the cases discussed in the previous paragraph are cases where the statistical evidence (cold-hit DNA evidence) and the individualized evidence (eyewitness testimony) were independently obtained. What if obtaining one of the relevant pieces of evidence depended on the other? Two directions are possible.

Suppose that first we get the eyewitness testimony, and then we check the person the eyewitness 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 eyewitness 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 eyewitness 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 individualized evidence is more interesting. Suppose, then, that we first obtain the cold-hit DNA evidence. Then we place the suspect in a confirming lineup, and the eyewitness identifies him. If we convict in such a case, is our conviction sensitive? The details of the answer (but not its bottom line) depend 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 likelihood we would not have had available the DNA evidence against him, nor would we (consequently) have the eyewitness testimony against him, and so we would not have convicted. If this

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

178. See sources cited *supra* note 167.

179. The exception in this regard is DNA evidence; courts tend to convict on the basis of cold-hit DNA evidence that is not corroborated by other evidence. See, e.g., Roth, *supra* note 101, at 1154-55.

is the relevant counterfactual, then the conviction is sensitive. What would have happened had the defendant not committed the crime but we still had the DNA evidence against him? Well, in such a case, we would have still held the lineup, but presumably the eyewitness wouldn't have identified him (for it is a reliable eyewitness, 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 it's problematic at all.¹⁸⁰ 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 confirming lineup. If the eyewitness 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 fact, we wouldn't have convicted, for in such a case, we would have held the lineup, and our reliable eyewitness 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 individualized evidence, to the distinction between the admissibility and sufficiency of statistical evidence, and also to the distinction between the appropriate role of statistical evidence at trial and as an investigative tool.

IV. PRESCRIPTIVE ANALYSIS

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 how implementing our model would open the door to categorical distinctions currently not made by the law. We already showed that

180. COLIN AITKEN ET AL., *FUNDAMENTALS OF PROBABILITY AND STATISTICAL EVIDENCE IN CRIMINAL PROCEEDINGS* 13-26 (Royal Statistical Soc'y, *Communicating & Interpreting Statistical Evidence in the Admin. of Criminal Justice, Practitioner Guide No. 1*, 2010) (discussing the central and growing role of statistical evidence and probabilistic reasoning in criminal investigations).

the formal and informal legal distinctions presently governing the use of statistical evidence in court refer to the type of statistical evidence (DNA versus propensity-for-crime evidence), its probability level (DNA evidence as an example of a uniquely high probability level), the type of legal proceeding (the criminal trial being a uniquely problematic arena), and the availability of alternative individualized evidence (the admissibility versus sufficiency issue). However, the proposed theoretical model also supports distinguishing among the different uses of statistical evidence based on additional criteria that are currently unrecognized by the legal doctrine. One, which we discuss next, is the type of misconduct with which the defendant is charged.

Under prevailing legal doctrine, the use of base-rate evidence is not conditioned on 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 murder of epistemologists' spouses.¹⁸¹ 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 threatens the primary-behavior incentives the law provides and should, therefore, be excluded from trial.

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, and personal-context cases, in which 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.

181. We would like to thank Mitch Berman for this point.

This result is at least somewhat counterintuitive. Sensitivity can 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 of murdering her.¹⁸² 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 option is the relevant counterfactual and 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.

CONCLUSION

We set out to dispel the confusion underlying the “bitter debate”¹⁸³ regarding the use of statistical evidence at trial. 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 epistemic 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 that should guide the law are dictated (at least partly) by its incentive-creating function. The incentive-based story, however, also supports the initial suspicion toward statistical evidence at trial, as 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 on Sensitivity-like counterfactuals.

The proposed theoretical framework—in epistemic 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

182. 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 is 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 world in which John does not murder Sara, no one else does, nor do we convict anyone of murder.

183. Koehler & Shaviro, *supra* note 1, at 247; *see also id.* at 247 n.1.

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 epistemic and incentive-based accounts align in the DNA context, both pointing toward admissibility of such statistical evidence (although, under our theory, it is the incentive-based perspective—rather than the epistemic 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 on defendant characteristics of a nonvoluntary 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 epistemic 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 extends 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 toward using statistical evidence for the purpose of conviction. The social costs of wrongful convictions are considered to be 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 use 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 the absence of a 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 individualized evidence. From an incentive-based perspective, which is the perspective pertinent for policy purposes, 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 individualized evidence, the incentive-corrupting effect is at its peak. The epistemic perspective requires finer distinctions about the precise nature of the interaction between the statisti-

cal and the individualized evidence, distinctions that shed interesting light both directly on the differential attitude toward statistical evidence when accompanied by individualized 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 on 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 on party characteristics, as the incentive structure may differ across corporations and individual litigants. Similar distinctions may apply with respect to statistical evidence relating to past versus future events. The solution we suggest to the statistical evidence puzzle—in terms of a partly epistemic explanation of the relevant intuitions and an incentive-based account of the relevant policy considerations—can facilitate further discussions and guide investigation of these and similar issues, which we leave for future research.