

Summer 2020

Against the Alleged Insufficiency of Statistical Evidence

Sam Fox Krauss

Follow this and additional works at: <https://ir.law.fsu.edu/lr>

Recommended Citation

Sam F. Krauss, *Against the Alleged Insufficiency of Statistical Evidence*, 47 Fla. St. U. L. Rev. 801 (2020) .
<https://ir.law.fsu.edu/lr/vol47/iss4/8>

This Article is brought to you for free and open access by Scholarship Repository. It has been accepted for inclusion in Florida State University Law Review by an authorized editor of Scholarship Repository. For more information, please contact efarrell@law.fsu.edu.

AGAINST THE ALLEGED INSUFFICIENCY OF STATISTICAL EVIDENCE

SAM FOX KRAUSS*

I.	INTRODUCTION	801
II.	THE PROBLEM STATISTICAL EVIDENCE POSES IN TRIALS.....	806
	A. <i>The Paradox</i>	806
	B. <i>Non-Accuracy-Based Epistemic Reasons for the Insufficiency of Statistical Evidence</i>	810
	1. <i>Knowledge</i>	811
	2. <i>Belief</i>	814
III.	STATISTICAL EVIDENCE AS SUFFICIENT FOR CONVICTION OR A LIABILITY JUDGMENT	815
	A. <i>The Legal Landscape</i>	815
	1. <i>The Criminal Context</i>	817
	2. <i>The Civil Context</i>	819
	3. <i>Fairness Considerations</i>	821
	B. <i>A Psychological Explanation</i>	823
IV.	CONCLUSION	825

I. INTRODUCTION

We should take into account the relevant evidence. This is a general principle in epistemology and in evidence law and is expressly stated in the Federal Rules of Evidence.¹ Where the evidence is, or is likely to be, inaccurate, it should be discounted or excluded from consideration. And the same where the evidence is accurate but likely to mislead.² In some cases, there is controversy about what kinds of evidence are likely to be inaccurate or likely to mislead. But it is

* J.D. candidate, NYU; Ph.D., UT-Austin. For helpful conversations and suggestions, thanks to Ronald Allen, Matthias Barker, Josh Dever, Sinan Dogramaci, Larry Laudan, David Sosa, Henry Schiller, Jason Schukraft, Julia Staffel, and the staff of the *Florida State University Law Review*. Also, the author would like to thank Kenneth L. Fox, Esq. for his encouragement.

1. See, e.g., Amit Pundik, *The Epistemology of Statistical Evidence*, 15 INT. J. EVID. & PROOF 117, 137 (2011); FED. R. EVID. 402 (allowing all relevant evidence unless expressly proscribed).

2. For example, Federal Rules of Evidence Rule 403 allows the exclusion of relevant evidence “if its probative value is substantially outweighed by a danger of one or more of the following: unfair prejudice, confusing the issues, misleading the jury.” FED. R. EVID. 403. Marcello Di Bello and Collin O’Neil make a similar point, noting that admitting probative evidence can be counterproductive if factfinders mis-weight it. Marcello Di Bello & Collin O’Neil, *Profile Evidence, Fairness, and the Risks of Mistaken Convictions*, 130 ETHICS 147, 149 (2020).

uncontroversial that when evidence is inaccurate or misleading, this provides a strong, if not decisive, reason for excluding or discounting that evidence:

- Reasoning with such evidence inhibits factfinding. Call these reasons for exclusion *accuracy-based* epistemic reasons. At other times, the weight we accord evidence is not determined solely by accuracy-based.
- Reasons but rather by what might be referred to as *policy* reasons. For example, some philosophers have argued that we ought to defer, against the evidence, to friends.³

The consensus view among legal scholars and philosophers is that statistical evidence should not, on its own, be sufficient for conviction in a criminal case or a ruling in a civil case. With few notable exceptions, neither legal scholars nor philosophers give accuracy-based epistemic reasons for excluding or discounting statistical evidence.⁴ For the most part, philosophers have given *non-accuracy-based* epistemic reasons. That is, statistical evidence is deficient in some epistemic virtue, but this deficiency is not one that would inhibit factfinding. Legal scholars have mostly given policy-based reasons.⁵

I argue that we should not discount statistical evidence. Further, statistical evidence *should be* sufficient, on its own, for conviction in a criminal trial or a ruling in a civil trial.⁶ I start with a presumption

3. See, e.g., Sarah Stroud, *Epistemic Partiality in Friendship*, 116 ETHICS 498, 498–99, 523 (2006); see generally Simon Keller, *Friendship and Belief*, 33 PHIL. PAPERS 329 (2004) (arguing that we should sometimes defer, against the evidence, to friends). But see generally Jason Kawall, *Friendship and Epistemic Norms*, 165 PHIL. STUD. 349 (2012) (responding to Stroud and Keller).

4. See, e.g., Marcello Di Bello, *Trial by Statistics: Is a High Probability of Guilt Enough to Convict?*, 128 MIND 1045, 1046–47 (2019).

5. See sources cited *infra* notes 12–16.

6. I am not the first person to say that non-accuracy-based epistemic reasons do not lead neatly to a conclusion that statistical evidence should not be sufficient for legal responsibility. See, especially, David Enoch, Levi Spectre & Talia Fisher, *Statistical Evidence, Sensitivity, and the Legal Value of Knowledge*, 40 PHIL. & PUB. AFF. 197, 199 (2012); see generally David Enoch & Talia Fisher, *Sense and “Sensitivity”: Epistemic and Instrumental Approaches to Statistical Evidence*, 67 STAN. L. REV. 557 (2015) (giving policy-based reasons for thinking that statistical evidence’s insensitivity provides a perverse incentive to wrongdoers in some cases, and arguing, generally, that the court should be more concerned with fact-finding than with knowledge); Pundik, *supra* note 1, at 122 n.27 (“Proponents of the distinction [between admissible and inadmissible statistical evidence] need to provide a more refined distinction between acceptable and problematic statistical evidence, together with some explanation as to why any differential treatment of objectionable statistical evidence should not apply to types of statistical evidence they consider acceptable.”); Amit Pundik, *What is Wrong with Statistical Evidence? The Attempts to Establish an Epistemic Deficiency*, 27 CIV. JUST. Q. 461, 463 (2008) (arguing that none of the epistemic reasons “successfully establish[] an epistemic deficiency from which (only) statistical evidence suffers.”); see generally Lewis Ross, *Rehabilitating Statistical Evidence*, PHIL. & PHENOMENOLOGICAL RES., 1,

that all relevant evidence should receive its due. The main claim of this Article is that the considerations scholars give in favor of statistical evidence's insufficiency do not overcome this presumption. Even if the considerations on offer are compelling, demoting statistical evidence imposes great costs. Broad pronouncements about epistemic desiderata, I think, draw attention away from the purpose of the trial, which is, at bottom, a factfinding mission.⁷ Of course, this is not the only goal—if it were, then the Exclusionary Rule, which renders probative evidence inadmissible if it was obtained unlawfully, would not exist. But often the discussions seem to miss the cost to accuracy, which in turn is a cost to victims of crimes, their families, to future victims, and to those who suffer torts. In an article on this topic, Duncan Pritchard makes a claim about error in trials: “In short, we want a criminal justice system that excludes high levels of risk of wrongful conviction, where risk is understood modally rather than probabilistically.”⁸ It is unclear why anyone affected by the risk of wrongful conviction would prefer a *modal* conception of risk, even if it has theoretical virtues. One would think that the people involved with and affected by trials would care primarily about accuracy. The United States Supreme Court has alluded to this many times.⁹ And, in the absence of sufficient countervailing reasons to prize non-accuracy-based epistemic virtues over accuracy, it is difficult to see why those reasons ought to prevail.

The paper is organized as follows. First, I describe the problem statistical evidence presents in the law. In brief, the following sort of dilemma arises:

Factfinders in trials are charged with reaching a verdict if the evidence presented meets a specified standard of proof. It seems that purely statistical evidence can suffice for just such a level of certainty in a variety of cases where a powerful intuition is that it would nonetheless be wrong to convict the defendant, or find in favor of the plaintiff, on merely statistical evidence. So,

1 (2019) (arguing, partially because of the ubiquity of DNA evidence in trials, and partly because legal factfinders are not able to hedge in the way that individual reasons are, scholars ought to be less averse toward statistical evidence).

7. “The basic purpose of a trial is the determination of the truth.” LARRY LAUDAN, *TRUTH, ERROR, AND CRIMINAL LAW: AN ESSAY IN LEGAL EPISTEMOLOGY* 2 (2006) (quoting *Tehan v. U.S.*, 383 U.S. 406, 416 (1966)).

8. Duncan Pritchard, *Risk*, 46 *METAPHILOSOPHY* 436, 454 (2015).

9. “Any claim for the exclusion of evidence logically relevant in criminal prosecutions is heavily handicapped. It must be justified by an over-riding public policy expressed in the Constitution or the law of the land.” *Nardone v. U.S.*, 308 U.S. 388, 340 (1939). “[T]his admittedly drastic and socially costly course is needed to deter police from violations of constitutional and statutory protections.” *Nix v. Williams*, 467 U.S. 431, 442–43 (1984). “The exclusionary rule generates ‘substantial social costs[.]’” *Hudson v. Michigan*, 547 U.S. 586, 591 (2006).

one either has to convict with statistical evidence, in spite of an intuition that this is unsettling, or explain what (dispositive) deficiency statistical evidence has.

Second, I discuss the reasons philosophers give for statistical evidence's deficiency—mostly non-accuracy-based epistemic reasons. In general, the arguments follow this form:

For evidence to be sufficient for knowledge or belief, it must have one or another epistemic property. Statistical evidence lacks these properties, for one reason or another. The factfinder must have the relevant belief, or knowledge, to convict a defendant or award a judgment to a plaintiff. Therefore, statistical evidence cannot be sufficient for conviction or a judgment. I do not argue that the relevant philosophers are mistaken about statistical evidence lacking the properties they identify. Rather, I argue that these philosophers focus too much on the first two premises and not enough on the third—that the right kind of belief or knowledge is necessary (let alone desirable) in the legal context. Sarah Moss, for example, argues that criminal defendants have a right that the jury *knows* that they are guilty and that statistical evidence, itself, is not sufficient for that kind of knowledge.¹⁰ Even if we grant the second point, which is part of the argument of her sophisticated and well-received book, Moss offers no good reason to accept the first.

In the evidence law literature, scholars mostly give what I have called “policy” reasons against the sufficiency of statistical evidence for conviction or judgment: it undermines a defendant's right to be treated as an individual;¹¹ it can be difficult to determine *which* reference class someone ought to be considered a member of;¹² it requires an overt

10. “According to my account, defendants have the right to be convicted on the basis of nothing less than knowledge.” SARAH MOSS, *PROBABILISTIC KNOWLEDGE* 215 (2018).

11. David T. Wasserman, *The Morality of Statistical Proof and the Risk of Mistaken Liability*, 13 *CARDOZO L. REV.* 935, 943 (1991) (but noting that where the statistical evidence does not “involve an inference to the defendant's conduct from the frequency of similar conduct,” as in, e.g., fingerprint analysis, to that extent the statistical evidence may be unproblematic). For a response to Wasserman, see Federico Picinali, *Base-rates of Negative Traits: Instructions for Use in Criminal Trials: Base-rate of Negative Traits*, 33 *J. APPLIED PHIL.* 69, 73–75 (2016).

12. See Mark Colyvan, Helen M. Regan & Scott Ferson, *Is it a Crime to Belong to a Reference Class?*, 9 *J. POL. PHIL.* 168, 172 (2001).

admission of error;¹³ it provides an incentive for misconduct;¹⁴ and, for legitimacy reasons, the public has to be able to treat the verdict as a conclusion about something that *happened*, which, we have reason to think, may not be the case if the verdict is based on statistical evidence.¹⁵ I will not discuss these at length in this Article; my focus is on non-accuracy-based epistemic reasons. Some of the policy reasons require a kind of empirical analysis that is beyond the scope of this project. Moreover, they involve a sort of weighing of costs and benefits that I am not prepared to undertake.¹⁶

Third, I discuss how the philosophical literature is out of touch with the legal reality—that outside of hypothetical cases discussed (though some are based on actual cases), the courts allow statistical evidence to be sufficient in a variety of contexts. For example, in the civil context, statistical evidence can be sufficient where either the harm is collective or else it is impossible to tell who, exactly, is responsible for the harm. More controversially, so-called “risk assessments tools” are

13. “[T]here is something intrinsically immoral about condemning a man as a criminal while telling oneself, ‘I believe there is a chance of one in twenty that this defendant is innocent, but a 1/20 risk of sacrificing him erroneously is one I am willing to run in the interest of the public’s—and my own—safety.’” Lawrence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process* 84 HARV. L. REV. 1329, 1372 (1971) (quoted in Jonathan J. Koehler, *When do Courts Think Base Rate Statistics are Relevant?*, 42 JURIMETRICS 373, 377 n.18 (2002)).

14. Enoch, Spectre, and Fisher, *supra* note 7, at 220, 221–22, 223 They argue that because statistical evidence is not *sensitive*, its sufficiency for a judgment incentivizes an opportunistic actor to, for example, gate-crash. Thus, while statistical evidence’s insensitivity is not itself a reason against relying on it, there are policy-reasons arising from its insensitivity that count against doing so. Someone’s belief is *sensitive* just in case were the belief false, the person would not believe it. Some philosophers have argued that sensitivity is a necessary condition of knowledge. For an overview, see generally Jonathan Jenkins Ichikawa & Matthias Steup, *The Analysis of Knowledge*, STAN. ENCYCLOPEDIA OF PHIL., Summer 2018, <https://plato.stanford.edu/archives/sum2018/entries/knowledge-analysis/> [<https://perma.cc/E646-KBB7>]. For a brief response to this line of reasoning, see Di Bello & O’Neil, *supra* note 3, at 155–56.

15. Charles Nesson, *The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts*, 98 HARV. L. REV. 1357, 1358 (1985). As Tracy Meares describes it, Nesson argues for legitimacy even at the cost of accuracy, “where legitimacy is defined as public acceptance.” Tracey L. Meares, *Three Objections to the Use of Empiricism in Criminal Law and Procedure—and Three Answers*, 2002 UNIV. ILL. L. REV. 851, 859 (2002). For criticism of Nesson’s view, see generally H. L. HO, A PHILOSOPHY OF EVIDENCE LAW: JUSTICE IN THE SEARCH FOR TRUTH (2008); Ronald Allen, *Rationality, Mythology, and the Acceptability of Verdicts Thesis*, 66 B.U. L. REV. 541 (1986); Neil Cohen, *The Costs of Acceptability: Blue Buses, Agent Orange, and Aversion to Statistical Evidence*, 66 B.U. L. REV. 563 (1986).

16. As Koehler and Shaviro write, the determination “of whether, on balance, greater use of overtly probabilistic evidence and methods at trial is desirable . . . depends on the value attached to specific policy concerns other than verdict accuracy.” 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, 248 (1990).

used, increasingly,¹⁷ in various criminal legal contexts.¹⁸ Of course, that the practice exists in other contexts does not itself justify the practice, but to the extent that these practices have become accepted for principled reasons, it does. The various considerations I advance lead to the conclusion that when it comes to naked statistical evidence, philosophers who argue for its insufficiency have got it wrong.

II. THE PROBLEM STATISTICAL EVIDENCE POSES IN TRIALS

A. *The Paradox*

As a rough approximation, “statistical evidence” in the legal epistemology literatures refers to evidence from which the factfinder draws a statistical inference to the defendant’s guilt or liability. While it may be specious to distinguish “statistical evidence” from other kinds of evidence, I will stick to this terminology for consistency’s sake. Scholars in this literature distinguish between *individualized*, or *particularized*, evidence, on the one hand, which is said to be “about” the defendant, and *statistical evidence*, on the other, which is not. Base rates, for example, are described as “statistical evidence.” Witness statements are not. Philosophically sophisticated scholars working in this area, however, have pointed out that this taxonomy of different kinds of evidence is suspect.¹⁹

The statistical evidence is usually about a reference class. For example, assume that we have no knowledge of Simon’s shoe preferences. But he is a thirty-year-old male, and we have robust data about the shoe-buying practices of thirty-year-old males. Thus, we have statistical evidence of Simon’s shoe preferences, even though we do

17. Brandon L. Garrett, *Federal Criminal Risk Assessment*, 41 CARDOZO L. REV. 121, 124 (2019).

18. For a descriptive analysis of risk assessment in pretrial bail determinations, in particular, see SARAH L. DESMARAIS & EVAN M. LOWDER, SAFETY + JUST. CHALLENGE, PRETRIAL RISK ASSESSMENT TOOLS: A PRIMER FOR JUDGES, PROSECUTORS, AND DEFENSE ATTORNEYS 3 (2019), <http://www.safetyandjusticechallenge.org/wp-content/uploads/2019/02/Pretrial-Risk-Assessment-Primer-February-2019.pdf> [https://perma.cc/EF4L-7JKT] (cited in Garrett, *supra* note 17, at 122 n.1).

19. See, e.g., Michael S. Pardo, *The Paradoxes of Legal Proof: A Critical Guide*, 99 B.U. L. REV. 233, 262 n.128 (2019); Daniel Shavero, *Statistical-Probability Evidence and the Appearance of Justice*, 103 HARV. L. REV. 530, 530 (1989) (challenging this distinction and arguing that it is ill-defined in the literature). Ron Allen makes a more strident comment: “[T]he assumption that there are two qualitatively distinct types of evidence, statistical and non-statistical, is essentially false.” 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, 1093 (1991). Judge Posner: “All evidence is probabilistic—statistical evidence merely explicitly so.” *Riordan v. Kempiners*, 831 F.2d 690, 698 (7th Cir. 1987) (cited in Koehler, *supra* note 13, at 401 n.165.)

not know anything about Simon, individually. In contrast, if Simon had made a statement about his love of Nikes, or if someone said they heard Simon had made the statement, then we would have *particularized* evidence.

In the legal context, a conviction or particular ruling is warranted when the government or plaintiff meets the relevant standard of proof. And it seems that statistical evidence, sometimes, is sufficient to meet this burden.²⁰ Why then, does it seem inappropriate to convict or find as the evidence suggests, when the evidence is statistical in nature? One scholar writes that doing so would be “dubious.”²¹ In the relevant literature, a few cases have come to be canonical representations of the problem. Two civil cases and one criminal case will illustrate:

Blue Bus

Suppose it is late at night . . . and an individual’s car is hit by a bus. This individual cannot identify the bus, but she can establish that it is a blue bus, and she can prove as well that 80 percent of the blue buses in the city are operated by the Blue Bus Company, that 20 percent are operated by the Red Bus Company, and that there are no buses in the vicinity except those operated by one of those two companies. . . . In these circumstances can the plaintiff recover in civil litigation against the Blue Bus Company, or, if not . . . then why not?²²

Gatecrasher

Consider a case in which it is common ground that 499 people paid for admission to a rodeo, and that 1,000 are counted on the seats, of whom A is one. Suppose no tickets were issued and there can be no testimony as to whether A paid for admission or climbed over the fence. So there is a .501 probability, on the admitted facts, that he did not pay. The conventionally accepted theory of probability would apparently imply that in such circumstances the rodeo organizers are entitled to judgment against A for the admission money, since the balance of the probability would lie in their favor. But it seems manifestly

20. See Pardo, *supra* note 19, at 253 (“What makes the examples ‘paradoxical’ is that the evidence appears on its face to surpass the applicable standard of proof, and yet the judgment of most people is that the evidence is insufficient to prove liability or guilt. This apparent inconsistency between what the applicable legal rules appear to require, on one hand, and judgments about what the correct result ought to be, on the other, creates a tension that calls out for explanation.”).

21. See Georgi Gardiner, *Legal Burdens of Proof and Statistical Evidence*, in ROUTLEDGE HANDBOOK OF APPLIED EPISTEMOLOGY 179, 180 (David Coady & James Chase eds., 2018).

22. FREDERICK SCHAUER, PROFILES, PROBABILITIES, AND STEREOTYPES 82 (2003). This case is modeled on *Smith v. Rapid Transit, Inc.*, 58 N.E.2d 754, 754–55 (Mass. 1945).

unjust that A should lose when there is an agreed probability of as high as .499 that he in fact paid for admission.²³

Prison Yard

In an enclosed yard are twenty-five identically dressed prisoners and a prison guard. The sole witness is too far away to distinguish individual features. He sees the guard, recognizable by his uniform, trip and fall, apparently knocking himself out. The prisoners huddle and argue. One breaks away from the others and goes to a shed in the corner of the yard to hide. The other twenty-four set upon the fallen guard and kill him. After the killing, the hidden prisoner emerges from the shed and mixes with the other prisoners. When the authorities later enter the yard, they find the dead guard and the twenty-five prisoners. Given these facts, twenty-four of the twenty-five are guilty of murder.²⁴

In the above cases, the relevant standard of proof seems to be met.²⁵ And yet, the intuition of most scholars, and, in some cases, the courts, is that it has not been. Describing the Blue Bus case, for example, Sean Sullivan writes, “Everyone agrees that Bayesian logic compels this result, but no one seriously thinks it is the right outcome.”²⁶ Describing the actual case on which Blue Bus is based, Andrea Roth writes that the court ruled in favor of the defendant because “a rational prediction by the jury based on the evidence ‘was not enough’ absent an ‘actual belief’ in liability ‘in the mind or minds of the tribunal.’”²⁷ In addition, this intuition is shared by the participants in several psychology studies, which find, in general, that people are reluctant to make liability determinations when the evidence is based on naked statistics.²⁸

On the other hand, the literature features an undue focus on the defendant—criminal and civil. When scholars (mostly legal scholars)

23. David Kaye, *The Paradox of the Gatecrasher and Other Stories*, 1979 ARIZ. ST. L.J. 101, 101 (1979). This case was originally presented in L.J. COHEN, *THE PROBABLE AND THE PROVABLE* 75 (1977).

24. Charles R. Nesson, *Reasonable Doubt and Permissive Inferences: The Value of Complexity*, 92 HARV. L. REV. 1187, 1192–93 (1979).

25. There is significant controversy about what standard of proofs are meant to be. I largely gloss over these issues.

26. Sean P. Sullivan, *A Likelihood Story: The Theory of Legal Fact-Finding*, 90 UNIV. COLO. L. REV., 42, 45 (2019).

27. Andrea Roth, *Safety in Numbers? Deciding When DNA Alone is Enough to Convict*, 85 N.Y.U. L. REV. 1130, 1164 (2010) (quoting *Smith v. Rapid Transit, Inc.*, 58 N.E.2d 754, 755 (Mass. 1945)).

28. Gary L. Wells, *Naked Statistical Evidence of Liability: Is Subjective Probability Enough?*, 62 J. PERSONALITY SOC. PSYCHOL. 739, 748 (1992); Edward Wright et al., *Factors Affecting the Use of Naked Statistical Evidence of Liability*, 136 J. SOC. PSYCHOL. 677, 679 (1996).

give policy-based reasons for the insufficiency of statistical evidence, they often invoke notions of fairness.²⁹ That is, fairness to the defendant. Neglected are the costs to the plaintiffs (in civil cases), society, and perhaps the victims (in criminal cases). In the criminal context, at least part of the purported justification for the criminal legal system is to keep people safe. In a related literature, scholars have debated not only what standards of proof mean, but, more provocatively, whether we are using the right standards. Using estimates of the likelihood of being the victim of a violent crime and being falsely convicted of such a crime, respectively, and the relative harms of each, Larry Laudan argues that we ought to lower the standard of proof in criminal trials.³⁰ Even those who do not agree with Laudan's analysis recognize the substantial costs to ignoring evidence.³¹

The Exclusionary Rule, for example, precludes evidence's admission when it is obtained in violation of the defendant's constitutional rights.³² This is meant to disincentivize police misconduct.³³ The Exclusionary Rule and resulting doctrines require that factfinders disregard what may be known to be accurate, relevant, incriminating, and even dispositive evidence. In general, however, there is no *right* to have unlawfully obtained evidence excluded from trial.³⁴

29. Enoch, Spectre, & Fisher, *supra* note 6 is a notable exception.

30. See generally, LARRY LAUDAN, *THE LAW'S FLAWS: RETHINKING TRIAL AND ERRORS?* (2016). For a pointed criticism of Laudan's argument, see generally Georgi Gardiner, *In Defence of Reasonable Doubt*, 34 J. APPLIED PHIL. 221 (2017).

31. Pritchard, *supra* note 9, at 454.

32. *Weeks v. U.S.*, 232 U.S. 383, 393 (1914) (incorporated to the states in *Mapp v. OH*, 367 U.S. 643 (1961)).

33. Although the Supreme Court has given different rationales for the Exclusionary Rule, over time it has all but given up on non-deterrent rationales. In *Nardone v. United States*, the Court ruled not only that unlawful phone taps were inadmissible, but also evidence derived from the unlawful tap. 308 U.S. 338, 341 (1939). To do otherwise, the Court held, would be "inconsistent with the ethical standards and destructive of personal liberty." *Id.* at 340. That was in 1939. In 1961, the court in *Mapp* called the exclusionary rule an "essential part of both the Fourth and Fourteenth Amendments." *Mapp*, 367 U.S. at 657. In *Hudson v. Michigan*, the majority writes "the exclusionary rule has never been applied except 'where its deterrence benefits outweigh 'its substantial social costs.'"" 547 U.S. 586, 594 (2006) (internal citations omitted). Note the departure from the lofty language in earlier cases. Indeed, for the majority in *Hudson*, Justice Scalia writes that the Court has revised its view on the expansiveness of *Mapp*: "[W]e have long since rejected that approach." *Id.* at 591. Justice Scalia writes that the "massive remedy" of exclusion "has never been applied except 'where its deterrence benefits outweigh its 'substantial social costs.'"" *Hudson*, 547 U.S. at 594 (internal citations omitted). In *Hudson*, we see the completion of the Court's transformation from relying on reasons of Fourth Amendment protections or judicial integrity to purely the deterrent benefit of the exclusionary rule.

34. The rights violation is thought to occur when and only when the unlawful search is conducted, not when that evidence is admitted at trial. In its decision in *Herring v. U.S.*, the Court writes, "the exclusionary rule is not an individual right[.]" They continue: "We have repeatedly rejected the argument that exclusion is a necessary consequence of a Fourth Amendment violation." 555 U.S. 135, 141 (2009). But some scholars feel differently. See, e.g.,

Notwithstanding certain disagreements, we ought to recognize the substantial costs when throwing away or discounting evidence. As Justice Cardozo famously commented on the Exclusionary Rule: “The criminal is to go free because the constable has blundered.”³⁵

Without stating a position on the purpose of the criminal legal system, or of damage awards, I can safely state an uncontroversial but rebuttable presumption: all relevant evidence should receive its due. Whatever the function of trials, it is *presumptively* the case that it is better-achieved when all relevant evidence is considered.³⁶ I do not need an argument for why there are strong reasons in favor of convicting the guilty, or of awarding damages to plaintiffs when they have been harmed. Disagreement about the function of the trial and the optimal distribution of error will affect when the presumption is rebutted, but it is important to note the cost of such a rebuttal, as the Supreme Court repeatedly does.³⁷

For the kinds of cases described above, philosophers have typically defended the common intuition that the courts ought not rely on statistical evidence. First, they give some account of the difference between statistical evidence and individualized evidence. Then they explain what (decisive) normative implications this difference has. Without questioning the distinctions that others have drawn between statistical and individualized evidence, I argue that the normative implications they draw are either misguided, or else, not decisive. That is, the difference between statistical evidence and individualized evidence could be relevant but not decisive in determining whether a factfinder can rest a particular finding entirely on statistical evidence. It is good to keep in mind that shortcomings in statistical evidence, whatever they are, are not necessarily decisive against its sufficiency.

B. Non-Accuracy-Based Epistemic Reasons for the Insufficiency of Statistical Evidence

Philosophers have offered a wealth of views to defend the insufficiency of statistical evidence. Mostly, the rationalizations have

Richard Re, *The Due Process Exclusionary Rule*, 127 HARV. L. REV. 1885, 1907 (2014) (locating the right to exclusion in the 5th Amendment’s Due Process clause).

35. *People v. Defore*, 150 N.E. 585, 587 (N.Y. 1926).

36. See generally Larry Laudan, *The Social Contract and the Rules of Trial: Re-Thinking Procedural Rules* (Feb. 25, 2008) (unpublished draft), <http://ssrn.com/abstract=1075403> [<https://perma.cc/UG4R-XPWZ>].

37. *Nardone v. United States*, 308 U.S. 388, 340 (1939) (“Any claim for the exclusion of evidence logically relevant in criminal prosecutions is heavily handicapped. It must be justified by an over-riding public policy expressed in the Constitution or the law of the land.”); *Nix v. Williams*, 467 U.S. 431, 442–43 (1984) (“[T]his admittedly drastic and socially costly course is needed to deter police from violations of constitutional and statutory protections.”); *Hudson v. Michigan*, 547 U.S. 586, 591 (2006) (“The exclusionary rule generates ‘substantial social costs[.]’”).

relied on what I have called non-accuracy-based epistemic considerations. Broadly, philosophers have argued either that statistical evidence lacks a property necessary for knowledge or that it lacks a property necessary for belief.

Conceptual analysis of propositional knowledge is often dated to Plato's *Theaetetus*, in which Socrates asks Theaetetus what knowledge is. One of Theaetetus' answers—that knowledge is justified, true belief³⁸—remained (mostly) popular until Edmond Gettier's influential 1963 article, *Is Justified True Belief Knowledge?*, answered the titular question in the negative.³⁹ So-called Gettier cases are instances in which the subject has a justified, true belief and yet the intuition is that they lack knowledge. Prefiguring Gettier, Bertrand Russell offers his famous stopped clock case, in which a person walks by a stopped clock that happens to display the correct time.⁴⁰ We are meant to think that, though the person who walks by is justified in believing the time based on what the clock reads and *truly* believes that it is the time the clock displays, he nevertheless fails to *know* the time. Gettier, in his article, offers similar cases, and in the half-century that has followed, scholars have both offered new cases and tried to determine the missing necessary condition for knowledge, or else to make more robust the justificatory condition. So-called lottery cases, in which one assigns very high probability to their having lost the lottery, but fails to believe or know it, have occasioned a similar kind of response.⁴¹

The important question we should ask is whether the kinds of answers to Gettier and lottery type cases, even if correct in that context, are relevant to the legal case. I believe that they are largely not.

1. Knowledge

Judith Jarvis Thomson argues that, for evidence to be sufficient for conviction, there has to be a causal connection between the evidence presented and the crime. Like knowledge, a verdict should not rest on luck. As Thomson puts it, someone's "reason for believing that p is true

38. One respected translation has it as "true judgment with an account." PLATO, *Theaetetus*, in PLATO: COMPLETE WORKS 157, 202 (John M. Cooper ed., M.J. Levetrans., 1997).

39. Edmund Gettier, *Is Justified True Belief Knowledge?*, 23 ANALYSIS 121, 121 (1963). But see, Julien Dutant, *The Legend of the Justified True Belief Analysis*, 29 PHIL. PERSP. 95 (2015) (giving a revisionist history of the Justified True Belief account of knowledge).

40. BERTRAND RUSSELL, HUMAN KNOWLEDGE: ITS SCOPE AND LIMITS 154 (1948).

41. See, e.g., Sinan Dogramaci, *A Problem for Rationalist Responses to Skepticism*, 168 PHILOS. STUD. 355, 359 (2014) (arguing, in part, that the statistical inference in lottery cases does not suffice for knowledge).

must ensure, or *guarantee*, that *p* is true.”⁴² For example, in Blue Bus, the bus’s blueness causally explains the testimony.⁴³ In contrast to statistical evidence, she writes, “individualized evidence for a defendant’s guilt is evidence which is in an appropriate way causally connected with the (putative) fact that the defendant is guilty, and hence (putatively) guarantees the defendant’s guilt.”⁴⁴ For example, if the witness testified that they had seen a red cab cause the accident, then if the red cab *actually* did cause the accident, *that* would causally explain why the witness seemed to see the red cab cause the accident.⁴⁵

In a response to the Thomson-type argument, Ferdinand Schoeman points out that the law makes no such distinction between evidence that is and is not causally relevant. Focusing on Thomson’s distinction between “internal” and “external evidence,” he writes, “She does not explain why we should require evidence to be internal before we regard it as reliably probative.”⁴⁶ Thomson herself recognizes that her causality requirement is intuitive but not easy to explain.⁴⁷ She essentially compares it to an anti-luck condition for knowledge. She argues that it is unjust for the jury to convict when it is just by luck if the jury gets things right.

But what Thomson does not do, as she says it is difficult to do, is give a principled reason for thinking that this causal connection, even if it is necessary for knowledge, is necessary in legal settings.

In her book, *PROBABILISTIC KNOWLEDGE*, Sarah Moss applies her novel account of knowledge—that the object of knowledge is probabilistic content, rather than propositional—to the statistical evidence paradox.⁴⁸ Moss notes, however, that even if the knowledge requirement seems apt for juries in criminal cases, where the standard of proof is very high, it seems inapt in civil cases, where the standard

42. Judith Jarvis Thomson, *Liability and Individualized Evidence*, 49 LAW & CONTEMP. PROBS. 199, 208 (1986).

43. See Gardiner, *supra* note 22, at pt. 3, which offers a treatment of Thomson’s argument. Importantly, Gardiner worries Thomson’s account of causality is underexplained, and argues that statistical evidence *can* play the relevant causal role Thomson desires. Second, Gardiner worries about Thomson’s *guarantee* condition, especially as it would pertain to DNA evidence.

44. Thomson, *supra* note 43, at 214.

45. *Id.* at 203.

46. Ferdinand Schoeman, *Statistical vs. Direct Evidence*, 21 NOÛS 179, 190 (1987).

47. “If we had individualized evidence. . . then we would feel considerably less reluctant to impose liability on Red Cab. Why is that? That seems to me a very hard question to answer.” Thomson, *supra* note 43, at 205. She adds, implicating others: “Friends of the idea that individualized evidence is required for conviction have not really made it clear why this should be thought true.” *Id.* at 206.

48. See generally MOSS, *supra* note 11. For a summary and review of Moss’s main contributions, see Kenny Easwaran, *Sarah Moss: Probabilistic Knowledge*, NOTRE DAME PHIL. REV. (2018) (book review), <https://ndpr.nd.edu/news/probabilistic-knowledge/> [<https://perma.cc/P37M-8UAH>].

is far lower. As she puts it, “Suppose the criminal standard of proof requires a justified full belief that the defendant is guilty. If this is correct, then what attitude does the civil standard [the applicable standard in *Gatecrasher*] require?”⁴⁹ It cannot be full belief; it has to be something like *more likely than not*. But if this were the case merely statistical evidence would suffice.

The advantage of her *probabilistic* account of knowledge, she says, is that it preserves the requirement that the jury *know* something, which many others have argued for in the criminal context. Legal proof, she says, requires knowledge, no matter the standard of proof.⁵⁰ With her new, probabilistic account of knowledge, Moss argues that proof beyond a reasonable doubt requires that the factfinder “know a certain probabilistic content, namely that it is beyond a reasonable doubt that the defendant is guilty.”⁵¹ And, whereas the civil standard cannot require that the factfinder *know* the defendant is liable—because the standard of proof is merely preponderance of the evidence—as Moss puts it, “[P]roof of liability by a preponderance of the evidence requires that the factfinder know *that the defendant is probably liable*.”⁵²

As with Thomson, though, Moss’s discussion lacks an argument for the view that legal proof requires knowledge. *Why* should we think that defendants have a *right* that the jury *know* that they are probably liable? Such a right has not been recognized. One reason to be skeptical of a right that the jury know the defendant is probably liable, or any even remotely similar statement, is that it is absent from a sampling of model civil jury instructions on the preponderance of the evidence standard from several U.S. circuit courts.⁵³

49. MOSS, *supra* note 11, at 207.

50. *Id.* at 211.

51. *Id.* at 211.

52. *Id.* at 210.

53. KEVIN O’MALLEY, JAY GRENIG & WILLIAM LEE, *FEDERAL JURY PRACTICE AND INSTRUCTIONS* 101:41 (6th ed. 2019):

3rd Circuit: “[Plaintiff] has the burden of proving [his/her/its] case by what is called the preponderance of the evidence. That means [plaintiff] has to prove to you, in light of the all the evidence, that what [he/she/it] claims is more likely so than not so. To say it differently: if you were to put the evidence favorable to [plaintiff] and the evidence favorable to [defendant] on opposite sides of the scales, [plaintiff] would have to make the scales tip somewhat on [his/her/its] side. If [plaintiff] fails to meet this burden, the verdict must be for [defendant]. If you find after considering all the evidence that a claim or fact is more likely so than not so, then the claim or fact has been proved by a preponderance of the evidence.”

5th: “[P]laintiff must prove every essential part of his claim by a preponderance of the evidence. A preponderance of the evidence . . . means evidence that persuades you that the plaintiff’s claim is more likely true than not true. . . . If the proof fails to establish any essential part of the plaintiff’s claim by a preponderance of the evidence, you should find for the defendant as to that claim.”

8th: “You will have to decide whether certain facts have been proved [by the greater weight of the evidence]. A fact has been proved [by the greater weight of the evidence], if you find

2. *Belief*

Another strategy in this field focuses on degrees of belief, or *credence*, where a credence is the subjective probability one assigns to the truth of a proposition. There is much debate over the relationship between credence and belief—in particular, whether the latter is reduceable to the former. One might think that belief just is credence past a certain threshold—whether stable or context-dependent.

Lara Buchak weighs in on this debate. She first argues that belief does not reduce to credence, and that this is explained, at least in part, by the unique role belief plays in assigning blame.⁵⁴ She gives two cases that are meant to elicit the intuition that those cases, both of which license the same degree of belief, do not both license *belief*, because in only one of the cases is blame appropriate. In the cases Buchak gives, the only important difference is that the evidence in one case is statistical and individualized in the other. But, she says, “[W]hat is interesting about statistical evidence is that it is often by itself not enough to produce a belief that *p*, even when [the credence] is very high.”⁵⁵ Belief and credence are sensitive to different kinds of evidence, she writes. And *blame* is sensitive to belief, but not credence.⁵⁶ She articulates the subjective version of the “BLAME NORM”: “Blame someone if and only if you believe (or know) that she transgressed.”⁵⁷ Because juries are called on not only to determine the facts, but also to “take a stand about whether [the defendant] is guilty,” and because (by the BLAME NORM) assigning blame requires belief, *and* because statistical evidence cannot give rise to belief, statistical evidence alone cannot license a jury’s verdict.⁵⁸

Andrea Roth makes a similar claim. She writes that the reasonable doubt standard “requires factfinders to reach an ‘actual belief’ in, rather than an acknowledgment of a high probability of, the defendant’s guilt.”⁵⁹ Her explanation is largely historical, and draws from the

that it is more likely true than not true. You decide that by considering all of the evidence and deciding what evidence is more believable.”

9th: “When a party has the burden of proving any claim...by a preponderance of the evidence, it means you must be persuaded by the evidence that the claim ... is more probably true than not true.”

54. See generally Lara Buchak, *Belief, Credence, and Norms*, 169 PHIL. STUD. 285 (2014).

55. *Id.* at 292.

56. See *id.* at 296–97.

57. *Id.* at 299.

58. *Id.* at 301.

59. Roth, *supra* note 28, at 1159.

common law notion of “moral certainty,” which, at least historically, must be based on testimony and perception, and which the “reasonable doubt” language replaced in the 19th Century.⁶⁰

There are a few ways to dispute Buchak’s claims. First, we might bite the bullet and contest her description of the intuitions she has in the cases—perhaps we *would* count ourselves as having a full belief in the statistical cases she gives. Second, we might contest her description of blame as requiring belief and not merely high credence. That is, we might think to ourselves *well, even if I would not have had the belief, my high credence does license blame*. But perhaps the most profitable response we could make to Buchak is to dispute that the trial’s role in the assignment of blame, at least in the way she describes. One might say that if the trial’s role is to assign blame, it is something different from what she requires for interpersonal blame in the cases she presents. And, even if she is right in the criminal context (I do not believe that she is), it is far from clear that the purpose of a civil trial is to assign blame in any sense reminiscent of the reactive attitudes we have toward each other.⁶¹ For example, one prominent strain of tort theory—the economic approach—does not countenance blame, in any moral sense, at all.⁶²

III. STATISTICAL EVIDENCE AS SUFFICIENT FOR CONVICTION OR A LIABILITY JUDGMENT

A. *The Legal Landscape*

Philosophers who argue for the insufficiency of statistical evidence in the legal context often overstate the extent to which the courts agree with them.⁶³ (Though, in their defense, it is often difficult to track the

60. *Id.* at 1160.

61. See Buchak, *supra* note 55, at 304.

62. Judge Calabresi describes tort law in the following way: “[It is] axiomatic that the principle function of accident law is to reduce the sum of the costs of accidents and the costs of avoiding accidents. . . .” Richard A Posner, *Guido Calabresi’s “The Costs of Accidents”: A Reassessment*, 64 MD. L. REV. 12, 15–16 (2005) (quoting GUIDO CALABRESI, *THE COST OF ACCIDENTS: A LEGAL AND ECONOMIC ANALYSIS* 26–28 (1970)).

63. “It is important to note that the statistical evidence is not *inadmissible*; rather, it is *insufficient* on its own.” Buchak, *supra* 55, at 291 (discussing the evidence in the Blue Bus case, in particular).

“In [a version of the Gatecrasher case,] . . . courts will find for the defendant.” Michael Blome-Tillmann, *Sensitivity, Causality, and Statistical Evidence in Courts of Law*, 4 THOUGHT J. PHILOS. 102, 103–04 (2015) (adding that “[t]he intuitive distinction between individual and bare statistical evidence can be found in a large number of court judgments and is drawn frequently, with more or less rigour, in the legal and philosophical literature”).

“We would never convict someone of a crime based on statistical evidence alone.” Elizabeth Jackson, *Belief, Credence, and Evidence*, SYNTHESIS, Oct. 4, 2018 at 5.

courts' less than systematic approach to statistical evidence.⁶⁴) But philosophers do not engage, to any significant degree, with the cases in which statistical inference is sufficient for liability in the civil context. While it may be true that courts would regard as insufficient the statistical evidence available in the canonical cases presented in the philosophical literature, there are many, many kinds of cases in which the courts permit statistical evidence to be sufficient. For example, statistical evidence is sufficient in employment discrimination cases and in market share liability cases.⁶⁵ Slightly more controversially, statistical evidence is used, decisively, in bail and sentencing determinations, although these are not subject to the same standard of proof considerations as determinations of guilt or liability.⁶⁶

Below, I highlight several instances in which the courts have held that statistical evidence is sufficient for liability judgments, and argue that, were we to countenance the arguments made by many philosophers, remedies would be largely unavailable to different kinds of injured parties.

"Law courts would not adjudicate in favor of the claimants. . . ." Gardiner, *supra* note 21 (giving versions of five canonical cases, including Gatecrasher, Blue Bus, and Prison Yard). Smith and Dant, respectively, make somewhat weaker claims, which seem a bit misleading nonetheless (though I don't mean to suggest intentionally so): "Indeed, it seems generally true that courts are reluctant to base affirmative verdicts – verdicts of guilt or liability – on evidence that is purely statistical in nature." Martin Smith, *When Does Evidence Suffice for Conviction?*, 127 MIND 1193, 1195 n.3, 1213 n.19 (2018) (adding "courts' general reluctance to rely on purely statically evidence" has not been entirely consistent and noting that "the legal treatment of statistical evidence has not been entirely consistent");

Mary Dant, *Gambling on the Truth: The Use of Purely Statistical Evidence as a Basis for Civil Liability*, 22 COLUM. J. L. SOC. PROBS. 31, 33 (1988) ("Courts and commentators often defend the traditional view that statistical evidence is alone insufficient to support a verdict by appealing to the injustice of imposing liability based on statistical data.").

64. See the difference in the Supreme Court's holdings in *Tyson Foods v. Bouphakeo* and *Wal-Mart Stores, Inc. v. Dukes* as shown in Robert G. Bone, *Tyson Foods and the Future of Statistical Adjudication*, 95 N.C. L. REV. 607 (2017) (discussing *Tyson Foods, Inc. v. Bouphakeo*, 136 S. Ct. 1036 (2016); *Wal-Mart Stores, Inc. v. Dukes*, 564 U.S. 388 (2011)). See also Amit Pundik, who notes that "case law seems to lack a systematic approach to statistical evidence." Pundik, *supra* note 1, at 117. *But see* Koehler, *supra* note 13 (giving a systematic assessment of when courts seem to approve or disapprove of the sufficiency of statistical evidence).

65. See *infra* notes 87–88.

66. Distinct from the use of statistical evidence in trials, the use of so-called "risk-assessment" in bail and sentencing is thought by some to be subject to racial bias, which would constitute both an accuracy-based epistemic reason and a policy reason against its use. See generally Sandra Mayson, *Bias In, Bias Out*, 128 YALE L.J., 2218 (2019); see also Christopher Slobogin, *Principles of Risk Assessment: Sentencing and Policing*, 15 OHIO ST. J. CRIM. L. 583, 589–93 (2018) (arguing that, properly guided by the relevant principles of fit, validity, and fairness, risk-assessment algorithms are to be preferred over individualized professional judgment).

1. *The Criminal Context*

As forensic science has become more sophisticated, DNA evidence's use in criminal trials has increased, as have debates about its proper use.⁶⁷ In a *pure cold hit*, the DNA match between the defendant and the relevant material is the only evidence. Some criminal defendants have attempted to argue that cold hit DNA evidence cannot be sufficient for conviction because it puts them in a class of suspects rather than uniquely identifying them as the culprit. But, as Andrea Roth puts it, appellate courts have “uniformly rejected” these arguments.⁶⁸ This is broadly in line with Jonathan Koehler's descriptive analysis of when courts tend to allow the sufficiency of statistical evidence: DNA cases represent instances where judges tend to allow the sufficiency of statistical evidence because they “rebut[] the argument that the match happened by chance.”⁶⁹

In most cases, the courts have relied on the extremely high probability of the DNA evidence's accuracy.⁷⁰ As Roth sees it, the court's willingness to countenance this is because the probabilities of correct DNA matches can be incredibly high—so high as to license “actual belief”, or moral certainty. Putting it slightly differently, Roth writes, “[W]hen source probabilities are high enough, they are effectively transformed into statements of certainty rather than of probability.”⁷¹

If, as Roth argues, the difference between DNA evidence and other statistical evidence is the extremely high likelihood of a correct match, this is inconsistent with the explanations for the intuitions expressed about Gatecrasher-style cases in most of the philosophical literature.⁷² As one pair of scholars put it, “It will be interesting to see whether such legal theories will be challenged by the sheer statistical power of the probabilities generated by forensic DNA matches, which, some might say, make the DNA database ‘a system not of evidence but of proof.’”⁷³ But the point of the Prison Yard style cases, as I took it, is to show that no matter how many prisoners there are, no matter how

67. Roth, *supra* note 28, at 1140–41; *see generally* David Wasserman, *Forensic DNA Typing*, in A COMPANION TO GENETHICS 349 (Justine Burley & John Harris eds., 2004); ANDREI SEMIKHODSKII, DEALING WITH DNA EVIDENCE: A LEGAL GUIDE 136 (2007).

68. Roth, *supra* note 28, at 1150. *See also* Di Bello, *supra* note 5, at 1074 n.46 (citing *Missouri v. Abdelmalik*, 273 S.W.3d 61 (Mo. Ct. App. 2008)).

69. Koehler, *supra* note 14, at 388–89.

70. Roth, *supra* note 28, at 1150.

71. *ID.* at 1158–59.

72. Marcello Di Bello is one notable exception. He argues, for accuracy-based epistemic reasons, that we ought not rely on statistics in the Prison Yard-type cases but we may be able to in DNA evidence cases. Di Bello, *supra* note 4, at 1048.

73. Simon A. Cole & Michael Lynch, *The Social and Legal Construction of Suspects*, 2 ANN. REV. L. SOC. SCI. 39, 51 (2006).

confident the jury should be in the defendant's guilt, when that high confidence is based on statistical evidence alone this does not suffice for conviction.

Relatedly, Martin Smith's argument for the insufficiency of statistical evidence in the legal domain relies on his "normic support" condition, where "a body of evidence *E* *normically supports* a proposition *P* just in case the circumstance in which *E* is true and *P* is false would be less normal, in the sense of requiring more explanation, than the circumstance in which *E* and *P* are both true."⁷⁴ Thus, even where the probability of some event is low it may not call out for much explanation, like, for example, winning the lottery: "The fact that there are 100 tickets in the lottery and only one winner does not normically support the proposition that ticket #72 has lost."⁷⁵ Thus, we cannot say we *know* that our ticket has lost a large, fair lottery, even though we could know, say, by testimony of a mostly-reliable witness, that the defendant committed the crime. In the Blue Bus case, for example, if the generally reliable witness had got things wrong—if she had hallucinated, or the Yellow Bus company had for some reason painted its busses blue and put "Blue Bus Company" signs on them, then we would expect some sort of explanation. As Smith puts it, "It can't 'just so happen' that the testimony was wrong. But it *could* just so happen that the bus involved was not a Blue-bus in spite of the fact that 90% of the buses operating in the area on the day in question were Blue-Bus busses. While this might in a sense be surprising, given the proportions involved, it clearly wouldn't demand any kind of further explanation."⁷⁶

But as he notes, "A normic standard of proof would block pure cold hit DNA convictions," which tend to be allowed by the courts.⁷⁷ How should we resolve this discrepancy, does he think? "The clash with the normic standard could be portrayed as a reason for being critical of such convictions, but could also be seen as a reason for resisting the standard and seeking an alternative solution to the legal puzzle of statistical evidence."⁷⁸ (Though he seems not to see this as a sufficient reason to do so.) Moreover, Smith's account would have trouble countenancing other kinds of cases in which the courts routinely rely on statistical evidence. Would it *call out for explanation* if what seemed to be a pattern of discriminatory behavior in an employment scenario, for example, was not intentional? On the normic support view, it seems

74. Smith, *supra* note 64, at 1208.

75. *Id.*

76. *Id.*

77. *Id.* at 1214.

78. *Id.*

like it would not. And, while Sarah Moss addresses the sufficiency of DNA evidence for knowledge, she does not quite say whether, on her view of probabilistic knowledge, when (if ever) it could be.⁷⁹

One promising response is offered by Enoch and Fisher, who effectively highlight the often-confused way in which scholars discuss “statistical evidence.” They describe DNA evidence as “statistical evidence,” but argue that one draws not a statistical inference from such data, but rather an inference to the best explanation. Rather than inferring from the defendant’s membership in some class that they are likely guilty, an inference from DNA evidence posits that the best explanation for the DNA match is that the defendant committed the crime.⁸⁰ In this sense, while the evidence does appear statistical in nature, the inference involving it is not properly analogous to the paradoxical cases commonly given. This line of argument is one promising way for those who dislike statistical inference for legal liability to preserve the sufficiency of DNA evidence.⁸¹

2. *The Civil Context*

In the civil context, statistical evidence is widely admitted into evidence and is often sufficient for the plaintiff to win.⁸² This is not to say that the Gatecrasher and Blue Bus cases do not have persuasive force. But to use these hypotheticals, or real-world approximations of them, as evidence that we ought not, or that the courts do not, consider similar evidence sufficient is improper. Courts have held that statistical evidence is sufficient for a particular ruling when so-called individualized evidence is impossible to gather.⁸³ Here again, this is broadly in line with Koehler’s analysis of courts’ inclination toward or against the sufficiency of statistical evidence. He finds that courts permit the evidence to be sufficient, specifically where there exists no particularized information.⁸⁴ In general, courts have allowed for the sufficiency of statistical evidence when, otherwise, it would be difficult, if not impossible, for genuinely injured parties to collect damages.

79. MOSS, *supra* note 10, at 218.

80. See Enoch & Fisher, *supra* note 7, at 596, 597–601, 602.

81. Thanks to Sinan Dogramaci for bringing this to my attention. As far as I could tell, Enoch and Fisher are the only authors, in either the legal or philosophical literature, to make this distinction.

82. Stephen E. Fienberg, *Gatecrashers, Blue Buses, and the Bayesian Representation of Legal Evidence*, 66 B.U. L. REV. 693, 699 (1986) (pointing out that although statistical evidence is rarely the only relevant evidence, where it is used it “far outweighs” the other evidence presented).

83. See, e.g., *In re Agent Orange Prod. Liab. Litig.*, 597 F. Supp. 740 (E.D.N.Y. 1984).

84. Koehler, *supra* note 14, at 374, 388 (adding that market share liability cases “provides another context in which general base rates are sometimes regarded as relevant, admissible, and even dispositive”). *Id.* at 399.

In toxic torts cases, for example, plaintiffs sue whoever is responsible for exposing them to toxins that they allege are the cause of their medical malady. Because it is often impossible to show that, say, a company's toxic dump *caused* their disease, plaintiffs rely on epidemiological evidence to show causation between the company's dump and their illness by showing an increase in disease relative to the normal incidence of the same disease in the population.

As the court wrote in a representative and much discussed case, in toxic tort cases: "[Because] the chance that there would be particularistic evidence is in most cases quite small, the consequence of retaining the requirement [for particularistic evidence] might be to allow defendants who, it is virtually certain, have injured thousands of people and caused billions of dollars in damages, to escape liability."⁸⁵

Relatedly, in market share liability cases, for example—cases involving an injury from taking a generic drug produced by one of a small number of manufacturers—the courts have ruled in favor of plaintiffs even where they cannot show particularized evidence that the pill that made them ill came from any one manufacturer. Rather, the manufacturers of the drug are held liable in proportion to their share of the relevant market.⁸⁶

Courts have ruled in a similar manner in employment discrimination cases, where, for example, it is alleged that an employer hires too few minority employees given the number of qualified minority employees in the relevant population.⁸⁷ And in antitrust

85. *In re Agent Orange*, *supra* note 84, at 836. For discussion, see Dant, *supra* note 64, at 61–69.

86. See *Sindel v. Abbot Labs.*, 607 P.2d 924, 937 (Cal. 1980); *Hymowitz v. Eli Lilly & Co.*, 539 N.E.2d 1069, 1076 (N.Y. 1989). For discussion, see generally David Kaye, *The Limits of the Preponderance of the Evidence Standard: Justifiably Naked Statistical Evidence and Multiple Causation*, 7 AM. BAR FOUND. RES. J. 487 (1982) (arguing that, in multi-defendant cases, an interpretation of the preponderance of the evidence standard consistent with the reasoning of e.g., *Sindel*, is appropriate); Koehler, *supra* note 14, at 399–400 (observing that *Sindel* is indicative of the courts' willingness to rely on statistical evidence when it would be implausible for plaintiffs to offer individualized evidence).

Sara Moss discusses this case very briefly. Moss's view, she writes, can account for the differing intuitions about the ability of statistical evidence to suffice for knowledge. But, it is not clear what Moss has to say about these kinds of cases. She writes that her "account of statistical evidence has an unusual strength: it can explain why verdicts against defendants [who are persons,] in *Prison Yard* and *Gatecrasher* seem especially intolerable in comparison with other verdicts [like *Sindel*] that *might or might not be* licensed by statistical evidence." MOSS, *supra* note 11, at 219 (emphasis added).

87. See *Bone*, *supra* note 65, at 612 ("disparate impact in a Title VII [of the Civil Rights Act of 1964, prohibiting employment discrimination on various grounds, including race, sex, and religion] is essentially a statistical concept calling for statistical proof."); Koehler, *supra* note 13, at 386; *Intl. Brotherhood of Teamsters v. U.S.*, 431 U.S. 324 (1977); *Ricci v. DeStefano*, 557 U.S. 557 (2009); *Hazelwood School District v. U.S.*, 433 U.S. 229 (1977) (cited in Koehler, *supra* note 14, at 386 n.68.).

law.⁸⁸ And securities class action cases.⁸⁹ And in class action lawsuits: In *Tyson Foods v. Bouphakeo*,⁹⁰ for example, plaintiffs were workers at a meat processing plant who sued when they were not paid for time spent putting on and taking off protective work gear. The court allowed statistical sampling (using *average* times it took workers to dress) to determine whether workers were underpaid.⁹¹

3. Fairness Considerations

The sufficiency of statistical evidence for liability gives some reason to worry about fairness. Not only does it appear unfair to punish a person based on evidence that (admittedly) leaves a large chance of error and has nothing to do with that person, in particular, but it also leads to an odd conclusion: it allows for the double-counting of evidence. Take the Gatecrasher case, for example: If the evidence that 501 attendees went into the rodeo without a ticket is sufficient to convict any one of the 1000 guests, then it is also sufficient to convict *all* of the attendees, 499 of whom *did* buy a ticket. And this seems unfair, if not downright ridiculous. How can one resolve this apparent difficulty with the use of statistical evidence? There are at least three options.

First, one might bite the bullet here and argue that while this seems odd, there is no contradiction or rights violation. Because of the standard of proof, decisions in civil cases admit of a high error rate. Why think there is any principled difference between one case with a high chance of error and a large set of cases with the same, as set by the standard of proof? The latter is just what we have now. Some objections to this kind of thinking are similar to the legitimacy worries

88. See Bone, *supra* note 64, at 612 (“statistical modeling is used in antitrust suits to determine damages when it is impossible to know directly what the counterfactual market free from the antitrust violation would have looked like.”) (adding, in the antitrust case, “statistical evidence is the obvious—and often the only—way to prove the issue and generate a reasonably correct substantive result for each individual case”).

89. See, e.g., *Halliburton Co. v. Erica P. John Fund, Inc.*, 134 S.Ct. 2398, 2414–16 (2014) (cited in Michael S. Pardo, *The Paradoxes of Legal Proof: A Critical Guide*, 99 B.U. L. REV. 233, 264 n.143 (2019)); Michelle Burtis, Johnah Gelbach & Bruce Kobayashi, *Error Costs, Legal Standards of Proof, and Statistical Significance*, 25 SUPREME COURT ECON. REV. 1, 4–5 (2017).

90. *Tyson Foods, Inc. v. Bouphakeo*, 136 S.Ct. 1036, 1042 (2016).

91. *Id.* at 1048; for discussion, see generally Bone, *supra* note 65; Pardo, *supra* note 90. Although, there is reason to think that this is a case of statistical *sampling*, and ought to be distinguished from the use of statistical evidence, see e.g., base rates, Bone, *supra* note 65, at 612. “[The Supreme Court] treats the case as one involving statistical evidence and employee-specific inferences when it actually involves substituting statistical averages for employee-specific fact finding. This makes it more like a case of statistical adjudication than a case of statistical evidence.” *Id.* at 610. Noting, however, “The distinction between statistical adjudication and statistical evidence is not always perfectly clear or precise.” *Id.* at 613.

that come up in the legal literature—what I have called *policy* reasons against the use of statistical evidence, and which I have discussed, briefly, above.

Second, one could point out that this already happens in other domains in the law.⁹² In *Bradshaw v. Stumpf*, for example, the state used the same evidence to try two men for murder, even though the prosecutor knew at most one was guilty.⁹³ On remand to consider due process violations, the Sixth Circuit held that no rights violation occurred: the prosecutions merely derived two conclusions from the same body of evidence.⁹⁴

Anne Poulin argues that this use of evidence is common, and that such use *does* constitute a due process violation.⁹⁵ But she notes that courts have considered and rejected a number of challenges to prosecutors using this kind of evidence.⁹⁶ Her style of argument does raise some interesting questions about when such a violation occurs, though. According to Poulin, it occurs at the time the prosecutor asserts the second position.⁹⁷ It is at least preliminarily suspect to think that a due process violation occurs, as it were, outside the scope of one defendant's trial. That is, how could it be that whether a due process violation occurs at Nancy's trial depends on what happens, say, six months later, in Ron's trial? But put this oddness aside. Some courts *have* ruled that it is a due process violation for a prosecutor to use one bit of evidence to secure mutually exclusive verdicts.⁹⁸ Alex Nunn argues that the using statistical evidence is a due process violation for a similar but distinct reason: the mere fact that some bit of evidence *could be* used to secure the conviction of any number of people constitutes a due process violation if that evidence is used, even against one person.⁹⁹

As a last response to the problem of double-counting evidence, tort law could move wholesale to a sort of modified proportional liability scheme. In the market share liability cases discussed

92. This is the general strategy in Pundik. *See generally*, Pundik, *supra* note 7.

93. 545 U.S. 175, 189 (2005). For a description of the case in the popular press, see Ken Armstrong, *Two Murder Convictions for One Fatal Shot*, THE NEW YORKER, Nov. 6, 2017.

94. Andrew Pollis, *Trying the Trial*, 84 GEO. WASH. L. REV. 55, 87 (2016).

95. Anne Bowen Poulin, *Prosecutorial Inconsistency, Estoppel, and Due Process: Making the Prosecution Get Its Story Straight*, 89 CALIF. L. REV. 1423, 1425 (2001).

96. *Id.* at 1425 n.4.

97. *Id.* at 1425.

98. G. Alexander Nunn, *The Incompatibility of Due Process and Naked Statistical Evidence*, 68 VAND. L. REV. 1407, 1418–20, 1418 n.61–63 (2015).

99. “[I]f the same naked statistical evidence could be used to convict any randomly selected member of a population, and the simultaneous conviction of the entire population would constitute a due process violation (due to the mutually exclusive nature of the crime), then the conviction of *even one* of those individuals constitutes a due process violation.” *Id.* at 1427.

above, companies were held liable to the extent that their product was represented in the market. At bottom though, the judgment against any company was determined by the likelihood that they caused the harm, and, therefore, the total damage award was capped by the extent of the damage. Part of the worry with relying on statistical evidence in the Gatecrasher case, as discussed above, is that if statistical evidence is sufficient for conviction, the rodeo owner could recover more than they are owed. If, for example, the tickets each cost \$1, and 499 guests paid and 501 guests crashed, the owner is owed \$501. But if statistical evidence alone is sufficient for a judgment, they could collect \$1 from each of the one thousand guests in 1000 individual trials. This is more than they are owed.

In a modified proportional liability scheme, instead, the owner might be allowed to collect only \$.51 from each guest. Thus, the damage award is capped. In a similar manner, in an individual case with statistical evidence, the plaintiff could instead recover damages in proportion to the weight of the evidence. This would require a more precise analysis of the evidence, and perhaps a more nuanced jury deliberation, but it may be a workaround to some of the problems statistical evidence poses. This is not too far from what Poulin hints at, as a solution. She suggests, as an option to deal with the uncertainty: “[The prosecutor] may embrace the uncertainty, acknowledge that she cannot prove which of the two pulled the trigger, and adjust her charge and sentencing goals downward.”¹⁰⁰

B. A Psychological Explanation

Both philosophers and legal scholars have offered sophisticated, if greatly varied, arguments for the insufficiency of statistical evidence as proof of guilt or liability. In some cases, the courts have agreed, though in far from a systematic way. Insights from psychology may provide an explanation.

Several psychology studies have found that people are reluctant to make judgments about legal responsibility (civil or criminal) when the evidence is based on naked statistics.¹⁰¹ This tendency is known as the “Wells Effect” and is named for the author of the first study on

100. Poulin, *supra* note 96, at 1424–25.

101. See generally Wells, *supra* note 29; Wright et al., *supra* note 29; see e.g., Ori Friedman & John Turri, *Is Probabilistic Evidence a Source of Knowledge?*, 39 COGNITIVE SCI. 1062 (2015); Keith E. Niedermeier, Norbert L. Kerr & Lawrence A. Messé, *Jurors’ Use of Naked Statistical Evidence: Exploring Bases and Implications of the Wells Effect.*, 76 J. PERSONALITY SOC. PSYCHOL. 533 (1999); Deanna L. Sykes & Joel T. Johnson, *Probabilistic Evidence Versus the Representation of an Event: The Curious Case of Mrs. Prob’s Dog*, 21 BASIC APPLIED SOC. PSYCHOL. 199 (1999); Hal R. Arkes, Brittany Shoots-Reinhard & Ryan S. Mayes, *Disjunction Between Probability and Verdict in Juror Decision Making: Disjunction Between Probability and Verdict*, 25 J. BEHAV. DECISION MAKING 276 (2012).

this topic. A representative statement from a follow-up study captures the phenomenon: “[P]eople judge that knowledge is less likely to result from probabilistic evidence than from perception and, moreover, . . . people deny that knowledge is gained from probabilistic evidence.”¹⁰²

In describing the phenomenon, the experimenters have dismissed as explanations of reticence: an inability to understand the standard of proof; causal relevance; that the evidence is not sufficient to raise subjective probabilities to the necessary level; and fairness. In one study, Niedermeier et al. identify what they call the “ease-of-simulation” effect: that jurors are more willing to acquit defendants when they can more easily imagine the situation in which the defendant is not guilty—often, but not always, when the evidence is circumstantial.¹⁰³

Andrea Roth and Mike Redmayne agree that this can explain often-disparate treatment of DNA evidence.¹⁰⁴ This explanation appears similar to Smith’s account of normic support.¹⁰⁵ And, it is supported by Koehler’s finding that courts downplay the probative weight of statistical evidence when there is individualized evidence, in addition.¹⁰⁶ When the context does not merely involve the use of base rates or other statistical evidence, but also involves issues of morality, things get even muddier. In a now-famous study, Tetlock et al. show that when reasoning about sensitive moral issues, people are (even) less willing to rely on base rates.¹⁰⁷ This finding is echoed in another

102. Friedman and Turri, *supra* note 102, at 1066–67.

103. “When probabilistic evidence of a defendant’s guilt contains information that can be used to build a possible (even if unlikely) scenario in which another party is responsible, jurors will be more reluctant to use that evidence to convict.” Niedermeier, Kerr, and Messé, *supra* note 102, at 541–42. Adding, however, “when probabilistic evidence of a defendant’s guilt contains little or no information that can be used to construct such an exonerating scenario, jurors will be more willing to rely on that evidence to convict.” *Id.* at 542. See also Kevin Heller, *The Cognitive Psychology of Circumstantial Evidence*, 105 MICH. L. REV. 241, 290–98 (2006), for an expanded discussion.

104. “While juries tend to discount DNA match statistics when they can actually envision examples of other potential suspects in the population who might match, they will treat the match as ‘compelling proof’ of guilt when they can no longer envision such examples.” Roth, *supra* note 28, at 1168.

This ‘ease of simulation’ explanation for the data also accounts for results in experimental research on DNA evidence. Here it has been found that mathematically equivalent ways of expressing the probative force of a DNA match have different effects on subjects: subjects think guilt more likely when told that ‘the probability that the suspect would match the blood drops if he were not the source is 0.1 percent’ than when told ‘1 in 1,000 people in Houston who are not the source would also match the blood drops.’ This seems to be because the latter formulation makes the possibility of a match with an innocent person easier to imagine.

Mike Redmayne, *Exploring the Proof Paradoxes*, 14 LEGAL THEORY 281, 304 (2008) (footnote omitted).

105. Smith, *supra* note 64, at 1206–11.

106. Koehler, *supra* note 14, at 401.

107. Unparsimonious though it may strike those who aspire to create universal theories of social cognition, the current findings suggest that people place a complex host of

study about statistical evidence at trials, noting the decidedly moral nature of the trial—especially criminal trials.¹⁰⁸ Indeed, there is some reason to be cautious about relying on intuitions when it comes to statistical evidence: base-rate neglect, for example, has been well-documented.¹⁰⁹

The psychological studies, however, do not tell us what to do with their findings. What we learn is that the behavior of juries is broadly (but not entirely) in line with philosophical theory: for non-accuracy-based epistemic reasons, juries, and to some extent, judges, are less likely to attribute responsibility, or knowledge, when the evidence is statistical. But is there good reason for doing so? As far I could determine, there is no mention of a “Wells *Fallacy*” in the literature, as there is, for example, in descriptions of base rate neglect. The psychological literature merely describes a tendency. Perhaps it offers an explanation for why one philosophical theory has largely tended one way on the question of statistical evidence. But it does not justify the theory.

IV. CONCLUSION

I have argued that explanations for the insufficiency of statistical evidence, even if they provide solid non-accuracy-based epistemic reasons, do not convincingly show why legal factfinders should care about these reasons. Indeed, in many civil contexts, where otherwise remedies to injured parties would be impossible to obtain, statistical evidence’s sufficiency for liability is necessary. While there may be documented reticence to ascribe blame or liability using statistical inference, it is far from clear that our intuitions here ought to guide legal factfinding.

superficially ad hoc content constraints on how they execute trade-offs, use base rates, and apply causal schemata to narratives. People who function like intuitive scientists or economists in one setting can be quickly transformed into intuitive moralist-theologians when provoked by assaults on sacred values.

Philip E. Tetlock et al., *The Psychology of the Unthinkable: Taboo Trade-Offs, Forbidden Base Rates, and Heretical Counterfactuals.*, 78 J. PERSONALITY SOC. PSYCHOL. 853, 865–67 (2000).

108. “[W]hat the laws of probability indicate is likely to have occurred—is generally viewed as an unacceptable basis for holding the defendant liable for the actual event.” Sykes and Johnson, *supra* note 102, at 211 (discussing Charles Nesson’s argument that even juries who ascribe great weight to the evidence nevertheless resist that the defendant committed the crime).

109. See generally Amos Tversky & Daniel Kahneman, *Judgement under Uncertainty: Heuristics and Biases*, 185 SCI. 1124 (1974).

