

# Just Probabilities

Chad Lee-Stronach

This essay aims to resolve a longstanding paradox in legal epistemology:

## THE STATISTICAL PROOF PARADOX

PROBABILITY THRESHOLD: Any legal standard of proof is reducible to some threshold value of probability,  $t$ , such that a defendant should be found liable when and because the probability that they are liable, given the evidence, is strictly greater than  $t$ .

STATISTICAL INFERENCE: Merely statistical evidence can establish that the probability of a defendant's liability is strictly greater than  $t$ .

CONCLUSION: A defendant can be found liable on the basis of merely statistical evidence.

We will see that for standard examples of 'merely statistical' evidence, the CONCLUSION is unacceptable. To avoid it, many theorists reject PROBABILITY THRESHOLD by replacing it with alternative epistemic, moral, or procedural principles. In a slogan, they hold that legal proof involves more than just probability.<sup>1</sup>

---

1. This slogan appears to have been coined by Mike Redmayne, "Exploring the Proof Paradoxes," *Legal Theory* 14, no. 4 (2008): 282. The legal and philosophical literature on this paradox is vast. Early explorations of this paradox include: Laurence H. Tribe, "Trial by Mathematics: Precision and Ritual in the Legal Process," *Harvard Law Review* 84, no. 6 (1971): 335-399; Jonathan Cohen, *The Probable and the Provable* (Oxford: Ox-

For its part, the seemingly innocent STATISTICAL INFERENCE has largely avoided suspicion — indeed, its role in generating the paradox is rarely questioned at all. Against the majority opinion on this issue, I shall argue that PROBABILITY THRESHOLD stands falsely accused; the true culprit, as it were, is STATISTICAL INFERENCE.

My argument takes inspiration from Judith Jarvis Thomson’s (1986) causal analysis of evidence.<sup>2</sup> I develop the central idea of her analysis using formal methods of causal inference.<sup>3</sup> This approach, which I call LEGAL CAUSALISM, vindicates Thomson’s thesis that merely statistical evidence is insufficient when and because it lacks causal relevance to the defendant’s liability. It departs from her view by retaining PROBABILISTIC PROOF. According to this LEGAL CAUSALIST resolution of the STATISTICAL PROOF

---

ford University Press, 1977). Recent prominent discussions include: Martin Smith, “When Does Evidence Suffice for Conviction?,” *Mind* 127, no. 508 (2017): 1193–1218; Marcello Di Bello, “Trial by Statistics: Is a High Probability of Guilt Enough to Convict?,” *Mind* 128, no. 512 (2019): 1045–1084; Sarah Moss, “Knowledge and Persons,” chap. 10 in *Probabilistic Knowledge* (Oxford Scholarship Online, 2018); Georgi Gardiner, “The Reasonable and the Relevant: Legal Standards of Proof,” *Philosophy and Public Affairs* 47, no. 3 (2019): 288–318; Renée Jorgensen Bolinger, “The Rational Impermissibility of Accepting (Some) Racial Generalizations,” *Synthese* 197, no. 6 (2020): 2415–2431; Michael Blome-Tillmann, “Statistical Evidence, Normalcy, and the Gatecrasher Paradox,” *Mind* 129, no. 514 (2020): 563–578; Clayton Littlejohn, “Truth, Knowledge, and the Standard of Proof in Criminal Law,” *Synthese* 197 (2020): 5253–5286. For surveys of the philosophical literature on this topic, see: Redmayne, “Exploring the Proof Paradoxes”; Georgi Gardiner, “Legal Epistemology,” in *Oxford Bibliographies: Philosophy*, ed. Duncan Pritchard (Oxford University Press, 2019); Rafal Urbaniak and Marcello Di Bello, “Legal Probabilism,” in *The Stanford Encyclopedia of Philosophy*, Fall 2021, ed. Edward N. Zalta (Metaphysics Research Lab, Stanford University, 2021).

2. Judith Jarvis Thomson, “Liability and Individualized Evidence,” *Law and Contemporary Problems* 49, no. 3 (1986): 199.

3. Judea Pearl, *Causality*, 2nd edn (Cambridge University Press, 2009); Peter Spirtes, Clark Glymour, and Richard Scheines, *Causation, Prediction, and Search*, 2nd edn (MIT Press, 2001).

PARADOX, legal proof is just a matter of probability, but deriving this probability involves more than just statistics.

## 1 The Case Against Probability Threshold

When does evidence suffice to justify a finding of legal liability? Here is a straightforward but superficial answer: when the evidence meets the relevant legal standard of proof. There are many standards of proof: from the “air of possibility” or a “scintilla of proof” to “beyond a reasonable doubt” and, stronger still, “utmost certainty”. Adopting a deferential attitude to existing law and practice, we might say that these standards are met when – to put it simply – established legal practice says so.

This answer is unsatisfactory because of the question’s obvious normative dimension: what *should* it take for evidence to reach any such standard? At this point, legal doctrines proliferate in ways that can be at once obscure, piecemeal, and circular. This has led some to believe that there is no unified analysis of standards of proof.<sup>4</sup> In this context, the challenge for legal epistemology is to determine whether there exists a unifying analysis of these related, yet fragmented, standards of proof.

Of course, there *is* something that all legal standards of proof have in common: they are each defined in terms of some degree of uncertainty, given

---

4. Concerning the standard of “beyond a reasonable doubt”, in particular, see: Larry Laudan, “Is Reasonable Doubt Reasonable?,” *Legal Theory* 9, no. 04 (2003): 295–331; Larry Laudan, *Truth, Error, and Criminal Law: An Essay in Legal Epistemology*, Cambridge Studies in Philosophy and Law (Cambridge University Press, 2006), ch. 2.

the evidence available and admitted to court. Since degrees of uncertainty, given our evidence, can be measured by probabilities, the following unifying analysis is promising:<sup>5</sup>

PROBABILITY THRESHOLD: Any legal standard of proof is reducible to some threshold value of probability,  $t$ , such that a defendant should be found liable when and because the probability that they are liable, given the evidence, is strictly greater than  $t$ .

Three quick points of clarification. First, since we are concerned with the probabilities assigned by reasonable and rational finders of fact (specifically, juries and judges) in response to evidence, we are concerned with *credences*, which are probabilistic representations of their degrees of belief in a proposition, based on the available evidence and background knowledge.

Second, accepting PROBABILITY THRESHOLD does not entail that the decision procedures of juries and judges should be overtly probabilistic. It is well understood that we are cognitively bounded agents who make systematic mistakes when assessing probabilities.<sup>6</sup> Rather, PROBABILITY THRESHOLD is presented as a conceptual thesis about the nature of standards of proof in

---

5. For the theoretical origins of measuring uncertainty with probabilities, see: Frank Ramsey, "Truth and Probability," chap. 7 in *The Foundations of Mathematics and other Logical Essays*, ed. R. B. Braithwaite, 1926 (New York: Harcourt, Brace / Company, 1926), 156–198; Leonard J. Savage, *The Foundations of Statistics*, 2nd rev. (New York: Dover, 1972).

6. The systematicity of these errors, often due to 'framing effects', has been well established in the psychological literature since at least: Amos Tversky and Daniel Kahneman, "Judgement under Uncertainty: Heuristics and Biases," *Science* 185, no. 4157 (1974): 1124–1131.

legal decision-making. It provides indirect normative guidance by providing a standard by which to assess the accuracy of legal procedures and institutions.

Third, this conceptual thesis does not merely state that proof can be represented *as though* it were a threshold of probabilities. This position is widely accepted but rightly dismissed as philosophically uninteresting. Rather, the claim is also one of reducibility, which is to say that legal standards of proof depend on matters of probability. Evidence suffices to justify a finding of legal liability *when* and *because* the probability of liability, given the evidence, is above some threshold value.

PROBABILITY THRESHOLD is perhaps the most contentious element of a broader research program called LEGAL PROBABILISM, which analyses and evaluates evidence and decision-making in legal contexts by reference to probability theory.<sup>7</sup> This approach offers many theoretical benefits: it systematically defines what it means for judgements to be consistent and accurate; it provides principles for weighing evidence and updating beliefs that provably ensure ongoing consistency and accuracy. It can thus define what it means for a justice system to be truth-seeking and error-avoiding in its determinations of innocence and guilt. Since PROBABILITY THRESHOLD seems to be entailed by the basic commitments of LEGAL PROBABILISM, to reject the former is to reject the latter.<sup>8</sup>

---

7. Urbaniak and Di Bello, “Legal Probabilism”; Brian Hedden and Mark Colyvan, “Legal Probabilism: A Qualified Defence,” *Journal of Political Philosophy* 27, no. 4 (2019): 448–468.

8. Strictly speaking, PROBABILITY THRESHOLD involves both an epistemic acceptance rule and a practical decision rule. The acceptance rule can be derived from PROBABILISM

Indeed, the rejection of PROBABILITY THRESHOLD is difficult to resist. There are many cases where the probability that a defendant is liable seems to exceed the relevant threshold, yet assigning legal liability solely on that basis seems epistemically unjustified.

To start, let's consider a case where the threshold is low. Suppose the court is trying to determine whether a traffic stop was procedurally justified. This requires determining whether the officer's evidence provided them "reasonable suspicion" that the driver was involved in a felony. Suppose that this standard of proof requires only a probability threshold of 0.2 for an individual to be liable to be stopped and searched. Now consider the following evidence:

**Traffic Stop:** The police vehicle's automated surveillance camera identified a driver with a previous conviction for possession of a prohibited substance. The vehicle's in-built computer informed the officer that, based on analyses of existing police databases relating population demographics and crime statistics, it is more probable than not that the driver, if searched, will be in possession of a prohibited substance.

This is a case of profiling, whereby statistical associations between a group's features and its offending rates are used to guide judgements about whether

---

via the Lockean Thesis: see e.g. Kevin Dorst, "Lockeans Maximize Expected Accuracy," *Mind* 128, no. 509 (2019): 175–211. The decision rule can be derived using expected utility theory, John Kaplan, "Decision Theory and the Factfinding Process," *Stanford Law Review* 20, no. 6 (1968): 1065; Di Bello, "Trial by Statistics: Is a High Probability of Guilt Enough to Convict?"

a perceived member of that group has or will commit an offence. There are various explanations for why relying solely on profiling evidence can be morally wrongful.<sup>9</sup> There are also important concerns about bias in law enforcement data collection and analysis.<sup>10</sup> Assume, however unrealistically, that none of these concerns apply to this case.

Many, at this point, seem to accept that the statistical evidence establishes the probability of liability above the relevant threshold.<sup>11</sup> But this does not seem to be enough to epistemically justify the traffic stop in this particular instance. Given the facts presented, the officer's evidence seems insufficiently tailored to the individual, their actual behaviour and their particular circumstances. It is not, after all, a crime to belong to a statistical reference class.<sup>12</sup> In this sense, even if the probabilities exceeded the threshold concerning people *like* the defendant, the officer did not have a reasonable suspicion about the *defendant*. So, we seem to have a counterexample to PROBABILITY THRESHOLD.

---

9. Renée Jorgensen Bolinger, "Reasonable Mistakes and Regulative Norms: Racial Bias in Defensive Harm," *Journal of Political Philosophy* 25, no. 2 (2017): 196–217; Adam Omar Hosein, "Racial Profiling and a Reasonable Sense of Inferior Political Status\*," *Journal of Political Philosophy* 26, no. 3 (2018): 1–20; Rima Basu, "What We Epistemically Owe to Each Other," *Philosophical Studies* 176, no. 4 (2019): 915–931; David Enoch and Levi Spectre, "Statistical Rsentment, or: What's Wrong with Acting, Blaming, and Believing on the Basis of Statistics Alone," *Synthese* 199, no. 3 (2021): 5687–5718.

10. Rashida Richardson, Jason M Schultz, and Kate Crawford, "Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing systems, and Justice," *New York University Law Review* 94, no. 2 (2019): 192–233.

11. For example Frederick Schauer, *Profiles, Probabilities, and Stereotypes* (Harvard University Press, 2003). Though see: Bolinger, "The Rational Impermissibility of Accepting (Some) Racial Generalizations."

12. Mark Colyvan, Helen M Regan, and Scott Ferson, "Is It a Crime to Belong to a Reference Class?," *Journal of Political Philosophy* 9, no. 2 (2001): 168–181.

This style of counterexample generalises to any probabilistic threshold of legal proof. Consider the following case that targets the proposition that the civil standard of proof — whereby liability must be supported by “the preponderance of evidence” — is  $t > 0.5$ :

**Blue Bus:** The plaintiff is seeking damages against the Blue Bus Company, claiming that during the night one of its buses damaged the plaintiff’s car. Although the plaintiff could not identify the bus, the evidence admitted to the court is that the Blue Bus Company owns 80% of the buses that circulate in that geographical area.<sup>13</sup>

Here, again, it appears that merely statistical evidence yields a probability greater than the threshold: in this case, it seems more likely than not that the Blue Bus Company is liable. This verdict is also generally deemed unacceptable and so is considered to be another counterexample to PROBABILITY THRESHOLD.<sup>14</sup>

Finally, let us see how the counterexample also applies to the claim that the criminal law’s standard of “proof beyond reasonable doubt” is reducible

---

13. This much discussed example is loosely based on a real case: *Smith vs Rapid Transit, Inc* 317 Mass. 469, 58 N.E.2d 754 (1945).

14. Tribe, “Trial by Mathematics: Precision and Ritual in the Legal Process”; Cohen, *The Probable and the Provable*; Charles R. Nesson, “Reasonable Doubt and Permissive Inferences: The Value of Complexity,” *Harvard Law Review* 92, no. 6 (1979): 1187–1125; Thomson, “Liability and Individualized Evidence.” More recent examples include David Enoch and Talia Fisher, “Sense and “Sensitivity”: Epistemic and Instrumental Approaches to Statistical Evidence,” *Stanford Law Review* 67, no. 3 (2015): 557–611; Smith, “When Does Evidence Suffice for Conviction?”; Duncan Pritchard, “Legal Risk, Legal Evidence and the Arithmetic of Criminal Justice,” *Jurisprudence* 9, no. 1 (2018): 108–119; Georgi Gardiner, “Profiling and Proof: Are Statistics Safe?,” *Philosophy* 95, no. 2 (2020): 161–183; Blome-Tillmann, “Statistical Evidence, Normalcy, and the Gatecrasher Paradox”

to a probability threshold of  $t \geq 0.99$ :

**Prison Yard:** “One hundred prisoners are in a yard under the supervision of a guard. At some point, ninety-nine of them collectively kill the guard. Only one prisoner refrains, standing alone in a corner. We know this from a video recording. The video shows that the participation ratio is 99:1, but does not allow for the identification of the ninety-nine killers. There is no other evidence. After the fact, a prisoner is picked at random and tried. If ninety-nine prisoners, collectively, killed the guard and the defendant on trial is a prisoner, the probability of his guilt is 99%.”<sup>15</sup>

According to PROBABILITY THRESHOLD, it appears that the defendant should be found liable because they are – by virtue of random selection from this group – 0.99 probable to have committed the offence, given the evidence. This verdict is widely rejected as unacceptable, leading many to conclude that PROBABILITY THRESHOLD is false.

These counterexamples differ in important ways that we will soon explore. Yet, they clearly have something in common: they rely on a particular *pattern of inference* from a statistical value to a corresponding probability value concerning the defendant’s liability. Is this inference valid? I shall argue

---

15. This formulation is given by Di Bello, “Trial by Statistics: Is a High Probability of Guilt Enough to Convict?” The original version was given by Nesson, “Reasonable Doubt and Permissive Inferences: The Value of Complexity.” This case is very similar, if not equivalent to, the famous **Gatecrasher** case. Cohen, *The Probable and the Provable*, ch. 7. The solution I propose to **Prison Yard** applies, *modulo*, to **Gatecrasher**.

that it is not. To resolve the STATISTICAL PROOF PARADOX, we should instead reject:

STATISTICAL INFERENCE: Merely statistical evidence can establish that the probability of a defendant's liability is strictly greater than  $t$ .

## 2 The Case Against Statistical Inference

Despite its importance to the case against PROBABILITY THRESHOLD, STATISTICAL INFERENCE has received surprisingly little justification or scrutiny. Some simply assert that statistical evidence alone can yield credences that are thus and so.<sup>16</sup> Most others appeal to their intuitions. Can more be said to justify it?

A common line of justification appeals to hypothetical betting behaviour. For example, Martin Smith (2017) reports that “we would sooner *bet* on the bus being a [Blue Bus] than a bus of another company”.<sup>17</sup> Call this:

*Betting Justification:* If, given the statistics, we would sooner bet on a proposition than its negation, then we are justified in assigning it a higher probability than its negation.

---

16. David Enoch, Levi Spectre, and Talia Fisher, “Statistical Evidence, Sensitivity, and the Legal Value of Knowledge,” *Philosophy and Public Affairs* 40, no. 3 (2012): p. 198. The main contention of this essay is that this stipulation cannot be justifiably made if we accept a LEGAL CAUSALIST approach to evaluating evidence.

17. Smith, “When Does Evidence Suffice for Conviction?,” 1194. See also Judith Jarvis Thomson, “Remarks on Causation and Liability,” *Philosophy and Public Affairs* 13, no. 2 (1984): p. 129; Lara Buchak, “Belief, Credence, and Norms,” *Philosophical Studies* 169, no. 2 (2014): p. 292.

This justification resembles the betting interpretation of credence, except that whereas the betting interpretation purports to elicit or describe credence via betting behaviour, this principle seeks to justify them. This difference matters: it simply does not follow that because one would bet, the bet is a good one to make. Since we are concerned with judges and juries making, as it were, sound bets on the basis of the evidence, some further justification is needed.<sup>18</sup>

What makes a bet a good one is its expected correspondence with the world, given our evidence. Could the fact *that we would bet* itself be evidence that we should bet? No. Except in special cases quite different from those we are concerned with here, we should not take our dispositions to act as evidence that we should do so.<sup>19</sup> I will suggest that since the betting justification relies on some correspondence between the statistical evidence and the world, any such justification relies on the statistics reflecting the objective chances, given our best estimation of those chances. We will return to this point shortly.

Before that, we must examine a different justification of STATISTICAL INFERENCE, given by Lara Buchak (2014):

---

18. On problems for the betting interpretation of credence, more generally, see: Lina Eriksson and Alan Hájek, “What Are Degrees of Belief?,” *Studia Logica* 86, no. 2 (2007): 183–213.

19. On whether there exist realistic cases of this phenomenon, see Arif Ahmed, *Evidence, Decision and Causality* (Cambridge University Press, 2014), ch. 4 For a helpful analysis of such cases using the causal inference methods adopted in this paper, see Christopher Hitchcock, “Conditioning, Intervening, and Decision,” *Synthese* 193, no. 4 (2016): 1157–1176.

*Narrowest Reference Class*: If it is part of your evidence that the frequency of truths in the reference class to which [the proposition]  $P$  belongs is  $x$ , and if there is no narrower or competing reference class for which you have evidence, then it is always at least rationally permissible to set  $Pr(P) = x$ .<sup>20</sup>

In cases like the above involving statistical evidence, the statistical category that the defendant falls into constitutes the narrowest reference class applicable to the proposition that the defendant is liable. Since that is the only evidence available, *Narrowest Reference Class* permits assigning the corresponding probability that the defendant being liable.

The key limitation of this justification is that although it *permits* us to assign a probability matching the statistical value, it does not *require* that this be the *only* probability that we assign. It is compatible with this principle that we assign multiple probabilities that are consistent with our limited evidence of how the events in question could have unfolded. In situations where our evidence about the case is incomplete, ambiguous or equivocal, there are multiple admissible probability functions that could assign different probabilities, some of which may be above the threshold and some below.<sup>21</sup> I shall argue that, as a class, counterexamples involving merely statistical evidence

---

20. Buchak, "Belief, Credence, and Norms," p. 288, with slight notational amendments. This principle is also known as *Frequency-Credence*, *Direct Inference*, or *Statistical Syllogism*. For discussion, see: Roger White, "Evidential Symmetry and Mushy Credence," chap. 7 in *Oxford Studies in Epistemology, Volume 3* (Oxford University Press, 2010), 161–186.

21. James M. Joyce, "How Probabilities Reflect Evidence," *Philosophical Perspectives* 19 (2005).

are of this kind: they admit multiple probability assignments concerning the defendant's liability, not simply the corresponding statistical value.

In particular, I argue that the counterexamples trade on conflating distinct interpretations of probability: relative frequencies, objective chances, and subjective credences.<sup>22</sup> This is common to the debate more generally, for example, in Schauer (2003):

[...] allowing liability on the basis of probabilistic evidence [is] no different from the routine application of standards of proof short of absolute certainty or the routine admission of nonstatistical evidence despite some likelihood of its inaccuracy.<sup>23</sup>

This conflation is problematic because our credences do not answer to statistics as such. Rather, they answer to the objective chances those statistics purport to reflect. As I shall explain, to determine whether any statistical value reflects the objective chances, we need evidence about the relevance of the statistics to the causal structure of the scenarios in question. Without such evidence, we face a choice: either we make our credences determinate but arbitrarily so, or we make them indeterminate, but for principled epistemological reasons. Opponents of **PROBABILISTIC THRESHOLD** tend to adopt the former approach. I will explain why we should instead adopt the latter: our credences should be informed by our evidence of the chances, given the

---

22. On interpretations of probability, see: Alan Hájek, "Interpretations of Probability," *Stanford Encyclopedia of Philosophy*, 2019,

23. Schauer, *Profiles, Probabilities, and Stereotypes*, p. 241.

causally relevant evidence available. I shall argue that statistical evidence lacks sufficient causal relevance, so our estimation of the chances on the basis of such evidence alone will be unavoidably indeterminate.

Why care about chances, as opposed to simply statistics or reporting our credence? Because doing so maximises the expected accuracy of verdicts.<sup>24</sup> Or, to paraphrase Thomson in a different discussion, the work of judges, juries and fact-finders more generally is not limited to what they believe to be the case: it is incumbent on them to find out what *is* the case.<sup>25</sup> In legal trials, fact-finders endeavour to find out how the events in question unfolded, so that they can determine (among other things) whether or not the defendant is liable. The best access they have to these objective facts is through estimations of the chances, as regulated by the causal structure of the events in question. This suggests, more precisely, that our degrees of credence in legal hypotheses are answerable to the well accepted epistemological principle, the:

PRINCIPAL PRINCIPLE: One's credence in a proposition should be  $x$ , given that one's estimate of the objective chance that the proposition is true is  $x$ .<sup>26</sup>

---

24. Richard Pettigrew, *Accuracy and the Laws of Credence* (Oxford University Press, 2016).

25. Judith Jarvis Thomson, *Rights, Restitution, and Risk: Essays in Moral Theory* (Cambridge, Mass.: Harvard University Press, 1986), p. 179.

26. This formulation aims to reflect Jenan Ismael's (2008) Generalised Principal Principle, rather than Lewis's (1980) version. Various versions of this principle can be given, depending on one's background theory of the metaphysics of chance. See: David Lewis, "A Subjectivist's Guide to Objective chance," *Ifs*, 1980, Jenann Ismael, "Raid! Dissolving the Big, Bad Bug," *Noûs* 42, no. 2 (2008): 292–307; Richard Pettigrew, "Accuracy,

If statistical evidence indicates that the chances that the defendant is liable are strictly above the threshold  $t$ , then it will follow from the PRINCIPAL PRINCIPLE that the defendant can be found liable solely on this basis. If the CONCLUSION remains epistemically unacceptable on this interpretation, then this would be enough to generate a more rigorous statement of the STATISTICAL PROOF PARADOX:

#### THE STATISTICAL PROOF PARADOX (RESTATED)

PROBABILITY THRESHOLD: Any legal standard of proof is reducible to some threshold value of probability,  $t$ , such that a defendant should be found liable when and because the probability that they are liable, given the evidence, is strictly greater than  $t$ .

PRINCIPAL PRINCIPLE: One's credence in a proposition should be  $x$ , given that one's estimate of the objective chance that the proposition is true is  $x$ .

STATISTICAL EVIDENCE: Merely statistical evidence alone can justify assigning the chances strictly above the relevant threshold.

CONCLUSION: Statistical evidence alone can be sufficient to justify finding a defendant liable.

---

Chance, and the Principal Principle," *Philosophical Review* 121, no. 2 (2012): 241–275; R.A. Briggs, "The Metaphysics of Chance," *Philosophy Compass* 5, no. 11 (November 2010): 938–952. Given the Principal Principle's wide acceptance within epistemology as a normative principle, as well as the extensive literature leading to this acceptance, I see no reason to re-litigate it in this essay.

As it turns out, however, this more precise restatement reveals a path to resolving the paradox that does not involve rejecting PROBABILITY THRESHOLD. It makes clear that the earlier-discussed STATISTICAL INFERENCE actually comprises two premises. The first, PRINCIPAL PRINCIPLE, is true. The second, STATISTICAL EVIDENCE, is false. Using methods of causal inference, and applying them to the fact-finding context of legal decision-making, I will explain why merely statistical evidence does not justify assigning chances strictly above the relevant threshold. I will call this approach:

### 3 Legal Causalism

LEGAL CAUSALISM aims to explain what it takes for evidence to sufficiently support inferences about the chances of a defendant's liability. To do so, it places additional constraints on LEGAL PROBABILISM. These constraints have independent justification. Moreover, they help to resolve the STATISTICAL PROOF PARADOX.

The first constraint applies to the *content* of the proposition whose chances are being estimated. Typical applications of LEGAL PROBABILISM adopt a very coarse-grained description of the legal decision-making process at trial, according to which courts are engaged in estimating the probability that 'the defendant is liable', given the evidence.<sup>27</sup> This approach to representing legal decision-making is problematic for two reasons.

---

27. For example, in: Hedden and Colyvan, "Legal Probabilism: A Qualified Defence."

First, it obscures the fact that liability *as such* is epiphenomenal to the fact-finding process. It is a label that supervenes on lower-level, legally-relevant causal hypotheses about the defendant's actions and state-of-mind in the contexts under investigation. Put glibly, the laws do not say "it is a crime to be guilty"; rather, they say "it is a crime to have committed such-and-such actions under such-and-such conditions". The court's role is to determine, to the best of their ability in light of the evidence, whether the individual who is the defendant actually stood in these relevant causal relations that ground liability. And yet, LEGAL PROBABILISM is standardly presented as involving inferences about whether or not 'the defendant is liable' *tout court*, treating it as a coarse-grained parameter whose value is to be estimated. It thus admits inferences at too high a level of description. In doing so, it has invited overly generous estimations of the probative value of statistical evidence.

Second, and relatedly, insofar as it permits inferences about a coarse-grained description of 'the defendant' (whoever they happen to be), LEGAL PROBABILISM permits treating the individual defendant as exchangeable with members of a statistical reference class, as well as the process that led to them being selected as the defendant, so long we abide by the axioms of the probability calculus. In general, it is not a crime to merely be the *type* of person who commits some crime.<sup>28</sup> The reigning background principle in legal contexts is that a defendant be treated as an *individual*.<sup>29</sup> The

---

28. Colyvan, Regan, and Ferson, "Is It a Crime to Belong to a Reference Class?"

29. Indeed, the idea that justice depends on non-generalized, particularistic judgement

concept of an individualized judgement has been found to be intuitive, but theoretically mysterious.<sup>30</sup> Yet the secret of individualization is hiding in plain sight: a judgement must be sufficiently *about* the individual who is the defendant.<sup>31</sup> While a person's belonging to a statistical reference class may (depending, perhaps, on how arbitrary it is) be *part* of who they are and what they do, it is clearly not all there is that person.<sup>32</sup> In recognition of this fact, the relevant target of inference in legal contexts is thus whether the *specific individual* who is the defendant actually stood in the relevant causal relations that ground liability, as defined by the substantive law.

LEGAL CAUSALISM aims to satisfy this content constraint by targeting the proposition: 'what are the chances that this particular individual actually stood in the legally relevant causal relations with respect to the alleged events?' This fine-graining of the content of inference places stringent, distinctively causal, requirements on the evidence that may be adduced against the defendant to justify a finding of liability. In order to establish a de-

---

about the individual is found in Plato's *Stranger* and Aristotle's *Rhetoric*. On the history of this particularistic thread of legal inquiry, see: Schauer, *Profiles, Probabilities, and Stereotypes*. Of course, LEGAL PROBABILISM is consistent with hyper-individualised inquiry; the point being made here is that it does not preclude overly coarse-grained inquiry of liability based on non-causal relations.

30. Michael J. Saks and Jonathan J. Koehler, "The Individualization Fallacy in Forensic Science Evidence," *Vanderbilt Law Review* 61, no. 1 (2008): 199–219.

31. Stephen Yablo, *Aboutness* (Princeton University Press, 2014); David Lewis, "Statements partly about observation," *Philosophical Papers* 17, no. 1 (1988): 1–31.

32. See: Benjamin Eidelson, "Treating People as Individuals," in *Philosophical Foundations of Discrimination Law*, ed. Deborah Hellman and Sophia Moreau (2013). For a defence of individualised judgement in legal contexts (and, I would argue, an erroneous rejection of PROBABILITY THRESHOLD), see: David T Wasserman, "The Morality of Statistical Proof and the Risk of Mistaken Liability," *Cardozo Law Review* 13 (1991): 935–947.

defendant's liability, the party that brings forth the accusation must establish certain specific facts that must form a causal structure as defined by the substantive law governing the case. When these facts form the relevant causal structure, they can ground the claim that the defendant is liable. In addition, in order to establish these facts, the party that brings forth the accusation must provide adequate supporting evidence that these facts have occurred. There must be an appropriate causal connection between these facts and the evidence, and such causal connection must itself be one of the facts to be established.<sup>33</sup> A finding of liability should then turn on whether the court's best estimation of the chances of the various parts of this proposition about the individual defendant and their actual causal role in the events, given the evidence and its causal relevance, exceeds the threshold corresponding to the relevant standard of proof.

As we shall see, this approach also helps to diagnose an intuitive shortcoming of the statistical evidence we have seen so far, suggesting a long sought after analysis of what constitutes 'merely statistical' evidence.<sup>34</sup> The diagnosis, earlier suggested by Thomson (1986), is that "what people feel the lack of, and call individualized evidence, is evidence which is in an appropriate way causally connected with the (putative) fact that the defendant caused the harm."<sup>35</sup> What is required is that there be evidence that may be ratio-

---

33. Thanks here to an anonymous Reviewer for pushing me to clarify the commitments of LEGAL CAUSALISM in this way.

34. On the extent of this gulf more generally, see: Clark Glymour et al., "Actual Causation: A Stone Soup Essay," *Synthese* 175, no. 2 (2010): 169–192.

35. Thomson, "Liability and Individualized Evidence," p. 203. This suggestion, when

nally believed to stand in an information-preserving causal relation to the events grounding liability. Such evidence, by virtue of its causal connection, is needed to provide sufficiently episode-specific information about chances that the individual actually stood in the relevant relation to the events that are the subject of litigation.<sup>36</sup> Reviewing the cases through this lens reveals that evidence is ‘merely’ statistical to the extent that it lacks the right kind of causal connection that would ground the particular defendant’s liability; it is ‘individualised’ only insofar as it is so connected.

The above constraint on the content of inquiry motivates a further, methodological constraint on LEGAL PROBABILISM. To express and make valid inferences about the chances of actual causal hypotheses, given the evidence, we cannot rely on conventional statistical models, which cannot differentiate spurious associations from causation.<sup>37</sup> Instead, we do better to employ *probabilistic causal models*, which can aid inferences about causation by imposing additional qualitative structure on the statistical models. In particular, we

---

coupled with Thomson’s commitment to knowledge as the standard of legal proof, generates well-known problems concerning factivity and luck. See: Redmayne, “Exploring the Proof Paradoxes”; Dario Mortini, “Knowledge, Individualised Evidence and Luck,” *Philosophical Studies* 179, no. 12 (2022): 1–30. LEGAL CAUSALISM, however, aims to avoid these problems by adopting PROBABILITY THRESHOLD, which permits non-factive, justified attributions of the chances of liability.

36. Here I borrow the terminology of episode-specificity from Pearl, *Causality*, Ch. 10. See also: James Woodward, *Causation with a Human Face: Normative Theory and Descriptive Psychology*, Oxford Studies in Philosophy of Science Series (Oxford University Press, Incorporated, 2021), pp. 107–114. On epidemiological evidence and specific causation: Alex Broadbent, “Epidemiological evidence in proof of specific causation,” *Legal Theory* 17, no. 4 (2011): 237–278; Sander Greenland and James M. Robins, “Epidemiology, Justice, and the Probability of Causation,” *Jurimetrics* 40, no. 3 (2000): 321–340

37. Pearl, *Causality*.

may provide a clearer articulation of the appropriate causal connection that evidence must bear to the court’s target of inquiry.

Formally, a probabilistic causal model consists of a probability distribution over variables that are related via structural equations. Informally, these variables can be understood as posing questions about events that occurred in a specific time and place.<sup>38</sup> The values of a variable can be understood as possible answers to the question posed by the variable; they tell us what happened in that event. Given probabilistic and qualitative information about the causal relations between events, we can derive estimations of the probabilities that particular sequences of events (causal hypotheses) occurred, based on actual and counterfactual probabilistic dependencies between the variables within the model.

Conveniently, for our purposes, this structure can be represented using directed acyclic graphs, whose nodes represent random variables and whose edges (arrows) represent the existence and direction of causal influence between the events those variables represent. These graphs depict how the value of a variable is a function of its direct causes (and probabilistically independent) error terms, yielding the familiar view that direct causes of an event *screen-off* less direct causes.<sup>39</sup> More generally, this approach encodes quantitative and qualitative information that allows us to infer, estimate and reason about the chances that particular causal hypotheses are true, given

---

38. On the semantics of such models, see: R.A. Briggs, “Interventionist Counterfactuals,” *Philosophical Studies* 160, no. 1 (2012): 139–166.

39. Hans Reichenbach, *The Direction of Time* (University of California Press, 1956).

the evidence available.<sup>40</sup>

LEGAL CAUSALISM claims that the causal hypotheses that are relevant to determining legal liability involve underlying data-generating processes that can be appropriately represented as acyclic causal structures. This type of structure provides the basis for evaluating the probative value of evidence, measured in terms of chance estimations that particular legally-relevant hypotheses are true.<sup>41</sup> As we shall see, the causal models appropriate to any particular case will rightly be a matter of interpretation and debate. Nevertheless, we shall also see that in the cases motivating the STATISTICAL PROOF PARADOX, mere statistical evidence fails to establish sufficient chances of the relevant proposition, according to a broad range of causal structures that seem most fitting to the scenarios described.

With these two methodological constraints on the table – one concerning the content of the legally-relevant propositions and the other concerning the formal methods for making inferences about them – let us now re-open the cases against PROBABILITY THRESHOLD.

### 3.1 The Profiling Case

From a LEGAL CAUSALIST perspective, an immediate challenge for profiling evidence is to establish causal relevance. In **Traffic Stop**, the evidence

---

40. Spirtes, Glymour, and Scheines, *Causation, Prediction, and Search*; Pearl, *Causality*.

41. The method I employ here – that of using directed acyclic graphs to represent legal evidential reasoning – is not untested. It has an important precedent that has been largely overlooked in the philosophical literature: David A. Schum, *The Evidential Foundations of Probabilistic Reasoning* (John Wiley / Sons, Inc., 1994).

presents a strong statistical association between features of the defendant and features of the population represented in the local and federal law enforcement database. However, as Figure 1 illustrates, such evidence does not entail any particular causal claim about the defendant.<sup>42</sup> Additional evidence about the causal structure is required to make the legally relevant kind of causal inference. Indeed, this holds true even with an infinite amount of merely associational data, which provide extensional content but lack relevant intensional content.

This simple observation is enough to block most instances where profiling evidence is presented as a counterexample to **PROBABILITY THRESHOLD**. These cases have the structure of Figure 1a, where the dashed, bidirected edge between the variables of interest represents that there is no clear causal claim to be made, given the evidence.

So, let us assume that a causal claim is being made. For the profiling evidence to be predictive in **Traffic Stop**, it must be the case that the variable **Prior** causes **Possession** (not the other way around). With this specific causal claim in hand, can we infer a probability strictly greater than 0.2, given the stipulated statistical value of the evidence?

---

42. Given the underdescription of the cases, the following causal diagrams are merely indicative of the causal structures that are compatible with the evidence. The diagrams nevertheless encode important causal information that allow us to determine what inferences are possible, given the evidence. Simplifying for present purposes, the diagrams encode this information as follows: empty nodes are unobserved variables, solid nodes are observed variables; dashed bi-directed edges denote possible causal influence, arced versions of which denote possible confounding variables, solid directed edges denote causal influence, absence of a directed edge between variables denotes lack of (direct) causal influence. See: Pearl, *Causality*, ch. 1.

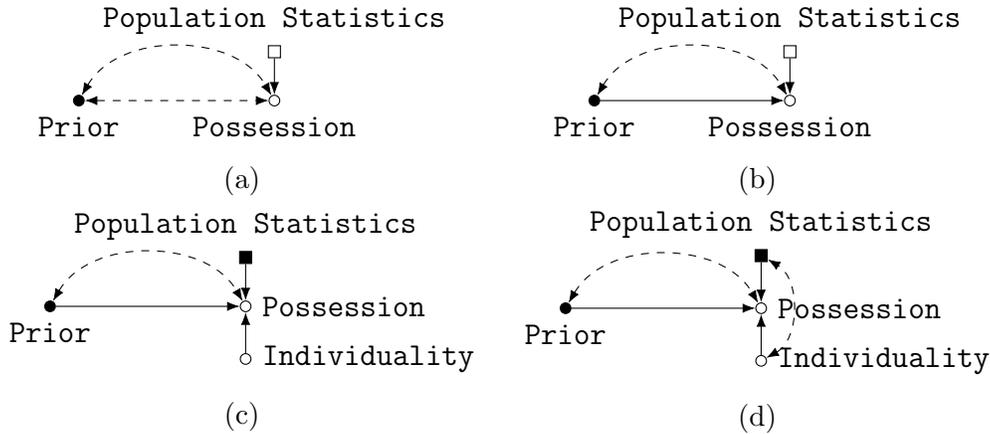


Figure 1: Some causal structures compatible with the statistical evidence provided in **Traffic Stop**. a) fails to assert a causally relevant claim concerning the defendant, admitting the possibility of a confounding variable (the dashed arc), a non-transportable statistical context (the empty box node), no determinable causal relationship between **Prior** and **Possession**. b) makes a causal claim, but provides no evidence to support transportability. c) fails to account for latent confounding or barriers to transportability, depending on how the causal properties of **Individuality** are interpreted. d) indicates a possible confounding effect: the defendant’s behaviour based on anticipation of profiling-based selection.

Again, the answer is: No. As is typical with profiling evidence, we have no evidence about whether the context from which the statistics were drawn is causally equivalent to the context to which they are being applied. Again note that this is a problem for type-level causal claims – we are not even close to the level of specificity required for actual causal hypotheses that apply to the defendant. The problem of using statistical evidence to estimate type-level causal relationships across contexts is known as the problem of *transportability*.<sup>43</sup> Without evidence concerning the causal equivalence of

43. Elias Bareinboim and Judea Pearl, “Causal inference and the data-fusion problem,”

the source and applied contexts, any probabilistic inference on the basis of such evidence will be meaningless.<sup>44</sup> This is represented in Figure 1b, where the empty square node represents potential causal non-equivalence between the contexts.

What does it take to establish that the profiling statistics are transportable? A lot, as it turns out. In principle, it is possible, if supported by additional evidence. However, in cases of profiling evidence about human behaviour, the relevant causal mechanisms must be understood at an agential level of description (allowing also that there may be subpersonal cognitive processes, such as those involved in addiction, that may also play a role in predicting and explaining behaviour). To make inferences about the actual behaviour of the defendant, the model must represent the confluence of the individual’s psychology, social context, and material circumstances. Examining how evidence might support inferences based on agent-based causal models may shed light on the much discussed, though often obscure, notion of *individuality*.

As a first approximation, individuality concerns the defendant’s beliefs, preferences, their cultural and social expectations, practices and behaviours, their available options and the situation(s) they were (allegedly) faced with.

---

*Proceedings of the National Academy of Sciences of the United States of America* 113, no. 27 (2016): 7345–7352; Judea Pearl and Elias Bareinboim, “External validity: From do-calculus to transportability across populations,” *Statistical Science* 29, no. 4 (2014): 579–595. This concept may be understood as the causal inference analogue of the concept of external validity in statistics.

44. For an alternative explanation that appeals to the Reference Class Problem, see Colyvan, Regan, and Ferson, “Is It a Crime to Belong to a Reference Class?”

Individuality is thus highly informationally demanding. This demand is mitigated in large degree by the fact that legal decision-makers are (for now at least) human, and so have background understanding of the others' agency against which to make use of the evidence provided. Profiling evidence fails to be individualised to the extent that it ignores relevant agential features and assumes the defendant is homogeneous or exchangeable with the statistical population. Importantly, this assumption is unjustified without further evidence establishing equivalence between the statistical models representing aggregate population behaviour and the individual. As such, it does not support a determinate estimation of the probability that the defendant committed the acts in question (Fig. 1c).

The possibility of feedback effects places a further obstacle to cases like **Traffic Stop**. Such effects arise when the target of prediction, in this case the profiled population, alters its behaviour in response to, or in anticipation of, the methods of prediction. For example, having a prior conviction may reasonably lead a person to scrupulously avoid the possession of such substances, in anticipation of encounters with law enforcement on the basis of their profile. At the population level, the actual or suspected use of profiling alters the incentives of agents in complex ways, making it such that the previous snapshot of statistical associations may not track the subsequent current behavioural adjustments made by that population.<sup>45</sup> Figure 1d illustrates

---

<sup>45</sup> Bernard E. Harcourt, *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age* (University of Chicago Press, 2008); Enoch, Spectre, and Fisher, "Statistical Evidence, Sensitivity, and the Legal Value of Knowledge"; Bolinger, "Reasonable Mistakes

this feedback effect, which confounds the probabilistic causal inference.

A natural objection at this point is that if we only had “better” profiling evidence – evidence that precisely models the individual as a causally determined object in the world – then it will turn out that statistical evidence alone can establish the required probabilities by demonstrating that the chances were sufficiently high. We might imagine a panoptic sci-fi world that secretly takes all possible observational data of the defendant’s life-history to build a model that perfectly maps onto and predicts their behaviour, allowing for highly accurate inferences about their past behaviour.<sup>46</sup>

In response, if such evidence does indeed establish the required chances, then according to LEGAL CAUSALISM it must be more than ‘merely’ statistical – the model generating these statistics must be, for all legal intents and purposes, causally equivalent to the actual world. Note, however, that this causal equivalence must be shown to apply to the actual case, not merely assumed. For instance, even in the sci-fi version of **Traffic Stop**, for instance, there may at some point be feedback loops that emerge where the defendant has come to suspect their panoptic predicament, and has recently been flipping a coin to make decisions in order to be unpredictable to the system (or consider, more realistically, those who reasonably suspect that they will be subject to profiling, and so alter their behaviour accordingly). Importantly, whatever these additional details turn out to be, they may be explicated and

---

and Regulative Norms: Racial Bias in Defensive Harm.”

46. Marcello Di Bello, “When Statistical Evidence Is Not Specific Enough,” *Synthese*:1–26.

tested within the framework of causal inference. Suffice it to say that in more realistic understandings of **Traffic Stop**, our suspicion of statistical evidence is justified: the success of such statistical software, where it occurs at all, is largely due to the curatorship of humans who understand and can test the causal relevance of the software model, the generative conditions of its data, and the dynamics of the context to which it is being deployed.

The conclusion of this analysis is that the profiling evidence alone fails to constrain the chances of the legally relevant causal hypotheses above the threshold. This is because it lacks sufficient causal relevance – in the ways described above – between the statistical context and the individual defendant’s case. As such, cases like **Traffic Stop** are not counterexamples to PROBABILITY THRESHOLD.

### 3.2 The Blue Bus Case

Recall that in **Blue Bus**, we aim to assess whether the statistical evidence establishes that the defendant is liable according to the preponderance of evidence, which we will interpret as “the balance of probabilities” or  $t > 0.5$ . Following LEGAL CAUSALISM, this involves determining whether the evidence is causally related to the allegation and what information that evidence gives us about the chance that the defendant is liable. The PRINCIPAL PRINCIPLE will then be employed to bridge this chance information back to probabilities. So, given the facts of the case, what is the causal relevance of the market share evidence?

For the inference to succeed, we need to establish the causal relevance of the evidence. Is the causal system of buses in the street structurally equivalent to balls in a randomised urn or chambers in a revolver used in a game of Russian Roulette? If that can be established then we may accept that statistical evidence provides chance information that may justify finding the defendant liable. However, even a cursory inspection of the facts of the case reveals that such an inference is beyond the statistical evidence provided. Indeed, this appears to be true across a range of background methodological assumptions. These assumptions permit different causal models of the facts in **Blue Bus**, as depicted in Figure 2. Importantly, however, according to none of these models does the probability of liability strictly exceed the threshold.

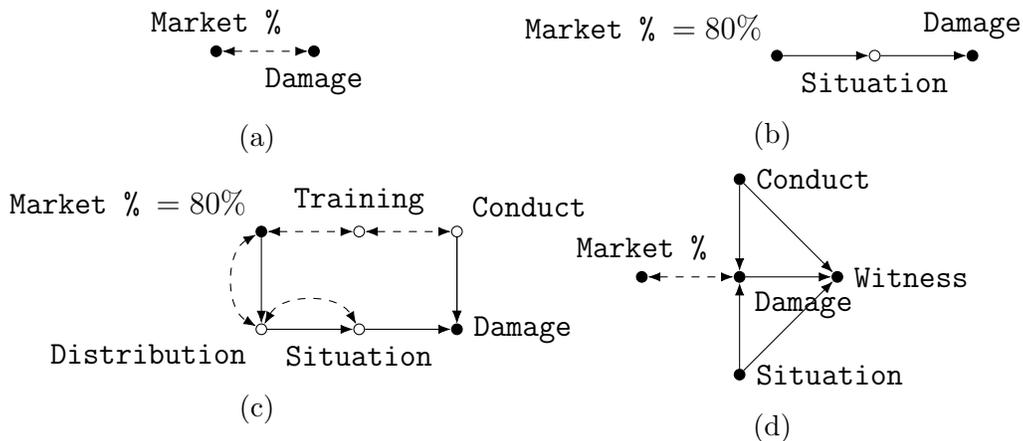


Figure 2: Causal scenarios compatible with the statistical evidence of **Blue Bus**. a) assumes no causal relations beyond evidence given. b) assumes some causal relation exists, albeit indirectly. c) makes expansive assumptions of relevant causal relations. d) demonstrates the causal relevance of eye-witness evidence.

The first causal model is highly austere in its causal assumptions. In Figure 2a, we see that there is no established causal relationship between the market share evidence and the incident. This is the interpretation suggested by Thomson (1984), who held that the statistical evidence is “external” to the causal history of the incident, which is to say, *causally irrelevant*.<sup>47</sup> For reasons given in the previous case, on this austere interpretation of the evidence, **Blue Bus** fails to be a counterexample to PROBABILITY THRESHOLD.

An alternative understanding of the case adopts more profligate causal assumptions. Figure 2c depicts a variety of factors surrounding the incident that are causally relevant but for which we have no evidence. For example, driver training, driver conduct, circumstances (like whether the bus was late, causing the driver to speed). In this understanding of the case, these unobserved causal variables are *confounders*. Without further evidence, it is not possible to estimate how the market share variable affected the chance of the incident, which is necessary in this case for determining the probability of liability. On this interpretation of **Blue Bus**, the statistical evidence is not enough because it does not control for unobserved competing causes.

Suppose, instead, that we very carefully assume that the statistical evidence is causally relevant but also that other confounding factors are not causally relevant. Could this justify a probability determinately above the threshold? In Figure 2b, we see that the confounders have been removed, but the statistical evidence is still remote. This is as it should be: even if

---

47. Thomson, “Remarks on Causation and Liability.”

the statistical evidence is assumed to be causally relevant, it is only indirectly so. To increase one's market share does not, in itself, cause traffic collisions. There are mediating factors that stand between **Market Share** and **Damage**. We will represent these mediating causes with the placeholder variable **Situation**, which screens-off **Market Share** from the incident. In probabilistic terms, this screening-off means that the probabilities of the defendant's liability, given the statistical evidence, are equal to the probabilities given the mediating causes. Since the mediating causes are unobserved, a causal inference method known as *mediation analysis* is required to estimate the probability of the defendant's liability.<sup>48</sup>

The first step involves assessing the causal effect of **Market Share** on the mediating variable, **Situation**. The second step involves assessing the direct causal effect of **Market Share** on **Damage**, which we already know to be zero (market share by itself, recall, does not directly cause traffic accidents). The third step brings these quantities together, estimating the causal effect within some interval: that is, the probabilities will be indeterminate, depending on the possible values the mediating variable could have taken. Note that if the evidence concerning chance hypotheses is incomplete, then by the **PRINCIPAL PRINCIPLE**, our credences should mirror this.

Given the facts of the case, however, we cannot assess the first step. As such, the evidence is not enough: there may, under additional causal assump-

---

48. Pearl and Bareinboim, "External validity: From do-calculus to transportability across populations."

tions regarding this case, be enough information to estimate the bounds of probability such that the lower bound exceeds the threshold. It is difficult to see how statistical evidence that does not causally track the events will fill these informational gaps. Examples of such evidence will have clear causal connection to the case: say, a dent in the plaintiff's car flecked with blue paint or, perhaps, learning that one of its buses due at that time also had maintenance issues. Without such evidence, the mere statistical evidence presented does not yield probabilities above the threshold.

These analyses might not be exhaustive, but they indicate the difficulty of extracting determinate chances from merely statistical evidence. As further support for the LEGAL CAUSALIST analysis given here, let us now briefly see how it vindicates a judgement made in a variation of **Blue Bus** involving not statistical evidence but an imperfectly reliable eye-witness:

**Blue Bus Witness:** Suppose now there is no causal claim made by the statistical evidence. Instead, the evidence is that an impartial eye-witness, who took contemporaneous notes, attests that it was a bus belonging to the defendant. In such cases, experimental evidence suggests that eye-witnesses are 0.8 probable to make a correct identification.<sup>49</sup>

There are three features of the evidence in this variation that are importantly different to the above case (see Figure 2d). One is that the reliability of the

---

49. It is worth mentioning that conventionally-held confidence in eye-witness testimony is difficult to justify. Gary L. Wells and Elizabeth A. Olson, "Eyewitness Testimony," *Annual Review of Psychology* 54 (2003): 277–295. Nevertheless, there is an important causal relevance to this case that is important for the present analysis.

eyewitness testimony is supported by experimental, as opposed to merely observational, data. Experimental approaches generally involve identifying, and intervening on or controlling for relevant causal variables, allowing in principle for an estimation of the possible individual variation relating to the particular eye witness in this case. The chance that the eye-witness is correct is therefore determinately above the threshold.

The second feature relates to the location of the evidence within the causal chain of events in question. As Thomson (1986) notes, the eyewitness testimony is backwards-looking: the content of the testimony is causally explained by the event.<sup>50</sup> This establishes an aspect of causal relevance, but there is a further feature that is important, which the market share evidence lacks: directness.

The incident is, by assumption, a direct cause of the content of the eyewitness's testimony. Since there is no intermediate cause (the eye-witness is not learning about the event by hearsay, for instance), the chance that the eyewitness testimony would identify the Blue Bus Company, given the incident, can be determinately estimated.<sup>51</sup> Following Thomson, we might say that these causal features of **Blue Bus Witness** evidentially *guarantee* that the probability of liability of the Blue Bus Company is above the relevant threshold, even though it may nevertheless be mistaken.<sup>52</sup> In this way,

---

50. Thomson, "Liability and Individualized Evidence," p. 204.

51. Of course, under more realistic assumptions about the case, there may be confounding factors and degradations of memory and so forth that would make a determinate estimate more difficult.

52. Thomson, "Liability and Individualized Evidence."

LEGAL CAUSALISM resolves the puzzle of why eye-witness evidence, unlike statistical evidence, can suffice.

### 3.3 The Prison Yard Case

On the face of it, **Prison Yard** presents the most difficult challenge to the LEGAL CAUSALIST analysis, for a couple of reasons.

First, the evidence is statistical, but not ‘merely’ so. To the contrary: the video evidence is causally explained by the events and also provides all of relevant causal information about the attack – except for the identity of the specific ninety-nine individuals who actually conducted the attack. It is an impartial eye-witness with an indelible though unfortunately unspecific memory.

Second, the random selection of the defendant appears to render the case structurally equivalent to a lottery with ninety-nine winning tickets and one losing ticket. As such, it seems that we should take the chance that the defendant committed the attack, given the video footage evidence, to be 0.99. So, we should find the defendant liable according to PROBABILITY THRESHOLD. Case closed, right?

Objection! A more careful examination of the causal structure reveals that the evidence fails to provide chance information at the relevant level of description: that is, concerning the individual defendant and the legally relevant, actual causal hypotheses concerning the individual defendant that ground liability. Operating at this level of granularity, the court relies on

evidence to discriminate between possible causal hypotheses that would determine the defendant’s liability. It asks: given the evidence, what are the chances that *this individual* actually committed the alleged acts? However, the statistical evidence given by **Footage** is not informative at this level of granularity. The evidence only probabilifies part of what is relevant to the individual defendant: that they, being a prisoner, were in the room with other prisoners where attack occurred.<sup>53</sup> It does not allow us to distinguish their actions in terms of their actual conduct: it is equally possible, given the evidence, that had the defendant acted otherwise the footage would show 98 or 100 attackers.<sup>54</sup> To differentiate these hypotheses, we need further information about the prisoners, and the defendant in particular.

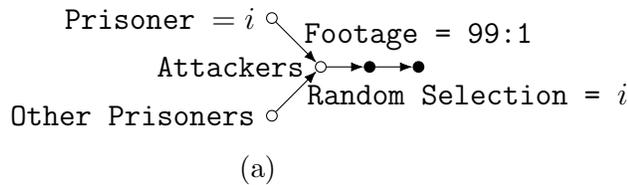


Figure 3: The random selection based solely on the footage screens-off the causal influence of the defendant and other prisoners.

Interestingly, randomised selection closes the door to this more fine-grained inquiry. The output of the randomised selection is, by definition, causally uninfluenced by the events of interest.<sup>55</sup> It intervenes on the causal chain of events, rendering the selection of the defendant dependent on their

<sup>53</sup>. On partially relevant confirmation, see: Yablo, *Aboutness*, ch. 6.

<sup>54</sup>. Di Bello, “Trial by Statistics: Is a High Probability of Guilt Enough to Convict?,” p. 6.

<sup>55</sup>. Spirtes, Glymour, and Scheines, *Causation, Prediction, and Search*, p. 224.

group membership, rather than their actions. Granted, the randomisation ensures the chance that ‘the prisoner selected for trial by the randomization procedure did it’ is 0.99. However, this is not informative of the chance that ‘the defendant  $i$  did it’, even though they are extensionally equivalent. Given that randomization is insensitive to actual conduct of defendant  $i$ , it fails to provide information about the proposition that ‘the defendant  $i$  is liable because they committed such-and-such actions’.

The problem is that the selection procedure implies, without justification, that each prisoner is statistically exchangeable: that each defendant is as likely as any other to have actually committed the attack. However, from the point of view of causal inference, “judgments of exchangeability are related . . . to judgments about uniformity of causal structure, and . . . an explicit account of the interaction of causal beliefs and probabilities is necessary to understand when exchangeability should and should not be assumed.”<sup>56</sup> Absent further additional evidence demonstrating the exchangeability of the individuals, the selection procedure invalidates inferences to the chances of the individual defendant. Thus, the evidence in **Prison Yard** does not establish that the chances of this particular defendant being liable are above the relevant threshold.<sup>57</sup>

---

56. Christopher Meek and Clark Glymour, “Conditioning and Intervening,” *The British Journal for the Philosophy of Science* 45, no. 4 (1994): n. 19; Pearl, *Causality*, p. 179.

57. Note that randomization is not the only way in which the selection process may block the relevant statistical inference: for example, if the prosecutors picked a prisoner simply because they disliked them, this would render the selection insufficiently sensitive to the actual conduct of the prisoner. Thanks to Kevin Dorst for this observation.

Suppose, however, that one remains unpersuaded regarding the requirement of targeting chance propositions at the level of actual causal hypotheses concerning the individual defendant. That is, suppose that one holds that coarse-grained proposition, ‘the defendant is liable’, is the relevant target of inference. Can we now say that the chance of the defendant being liable is equal to the statistical value? The answer is: Yes, but at the cost of arbitrariness or injustice.

Let’s start with arbitrariness. We begin by again examining the causal structure of the scenario. As illustrated in the simplified diagram, Figure 4, let the variable **Prisoner  $i$**  refer to the whether or not the defendant took part in the attack; **Other Prisoners** denotes the other prisoners in the room and whether they attacked, and **Attackers** refers to the specific combination of prisoners who conducted the attack; **Footage** records the combined number of prisoners who took part in the attack, but does not provide visual information that differentiates them.

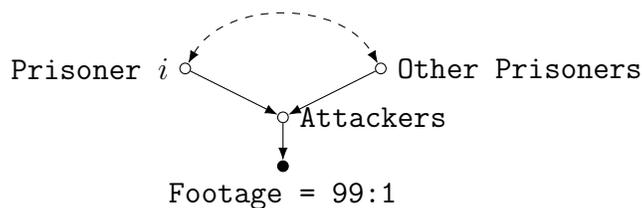


Figure 4: In **Prisoner**, the statistical evidence (recorded in the video footage) is a collider variable. Conditioning on this evidence creates a negative association between the causal hypothesis that the defendant is liable as opposed to some other prisoner(s).

We note that **Footage** is a descendant of a collider variable: it is a vari-

able that is directly influenced by multiple other causal variables. The reason to model it as a collider variable is to isolate hypotheses about the defendant's conduct from those whose actions may be alternative explanations of the events. To model it otherwise, say, at the level of combinations of prisoners, would be to engage in a mass trial of multiple prisoners at once. The aggregate variable `Prisoners` would not allow for sufficiently fine-grained inferences about the defendant on trial.<sup>58</sup>

And now here's how the arbitrariness arises. Because `Footage` is a collider variable, when we conditioning our beliefs on it to make an inference about `Prisoner i`, whether we should find them liable will depend on the *order* in which they were selected, even though we know – indeed, by stipulation – that the evidence is symmetric with respect to each defendant and the order of selection is evidentially irrelevant. Granted, if the defendant is the *first* selected, then our best estimate of the chances that they are liable is indeed 0.99, and so will exceed the threshold. However, bizarrely, this helps to acquit the other prisoners. The extent to which it does so depends on the epistemic significance of finding a previous defendant liable. Or, in legal terms, it depends on the deliberative significance of the principle of *res judicata*: that matters decided by the court cannot be reconsidered barring

---

58. One way to see this is via counterfactual reasoning. If, for example, we were to learn that the defendant's cell doors were malfunctioning – either stuck open or shut closed with some probability – we would need to update our estimation for these hypothetical possibilities by breaking the causal linkage between the defendant and the events, while maintaining the causal relations between the other defendants and the events. Pearl, *Causality*, ch. 7.

exception circumstances, such as new evidence or a procedural defect in the prior decision.<sup>59</sup> There are three possibilities.

The first treats finding **Prisoner**  $i$  liable as evidence about which we can be certain when deliberating about **Prisoner**  $j$ , if they are selected. More precisely, this is equivalent to Bayesian conditionalization. The second possibility, which seems more plausible, maintains that the court allow uncertainty about whether the defendant  $i$  is objectively liable when assessing defendant  $j$ . This is equivalent to Jeffrey Conditionalization.<sup>60</sup> Both approaches straightforwardly imply that the chance that we select a prisoner who is liable is dependent on the number of previous defendants we have found liable. In general, the probability of selecting a liable prisoner by lottery without replacement follows a hypergeometric distribution that is influenced by the number of trials and prior (expected) successes. For brevity, a numerical example should suffice to illustrate. If we treat liability as certain evidence, then probability that the second defendant is guilty is  $98/99 = 0.9898 < t$ . If we treat liability as uncertain evidence, the probability assigned to the second defendant =  $(98/99)(0.99) + (99/100)(0.01) = 0.9899 < t$ , which is also below the threshold. Either way, the case against the defendant, given the statistical evidence, turns on the order and outcome of the randomised selection, which is evidentially irrelevant to whether or not they are actually

---

59. Kevin M. Clermont, “Res Judicata as Requisite for Justice,” *Rutgers University Law Review* 68, no. 3 (2016): 1067–1142.

60. Richard C. Jeffrey, *The Logic of Decision*, Second (Chicago and London: The University of Chicago Press, 1983), ch. 11

liable. Moreover, to almost certainly secure their acquittal, the defendant may rightly plead: ‘Randomise again!’ This, I submit, is unacceptably arbitrary.

Suppose we ignore the idea that finalized legal decisions should provide some measure of deliberative (if not logical) closure. In fact, suppose that the court opts for the third option of assigning no weight to its previous decisions. It finds liable all to whom the evidence applies, even when it is certain, given the evidence, that not all of them could have performed the actions question. In the present case, it could find guilty, on the basis of the evidence, *all* of the prisoners, since it applies to them equally at the coarse-grained level of description. This option, however, is widely considered to be “palpably”<sup>61</sup> and “manifestly”<sup>62</sup> unjust, since the evidence clearly shows that it is certain that one of the prisoners is innocent.<sup>63</sup>

If we accept LEGAL CAUSALISM, then this problem dissolves. This is because of the need to ensure that the selection procedure does not screen-off the defendant’s individual contribution to the events. Suppose, for example, that DNA samples were collected from clothing of the prisoners, with a finding that ninety-nine of the prisoners had traces of the guard’s blood. Suppose also that there is evidence indicating that the only way such traces could have been found on their clothing is through the attack.<sup>64</sup> Given its specificity,

---

61. Smith, “When Does Evidence Suffice for Conviction?,” 3

62. Cohen, *The Probable and the Provable*, 75

63. But doesn’t the use of a decision-theoretic threshold imply sacrificing innocent defendants in this way? I reconvene discussion on this point in **Part 4**.

64. On this point, and the issue of specificity more generally, see Di Bello, “When Sta-

the DNA evidence can be disaggregated for each individual that it picks out. As illustrated in Figure 5, it can provide information about the probability that the defendant is liable without triggering arbitrary estimations of the probability that the other prisoners were also involved. This does not arise because we do not condition on the collider variable, but rather only on the basis of evidence that causally connects the selection of the defendant to the events in question. The footage evidence may be used to determine *how many* prisoners to select, but not to determine *who* to select in particular.

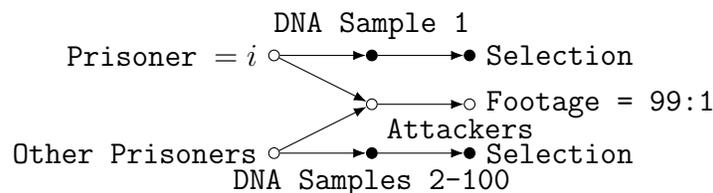


Figure 5: In this version of **Prison Yard**, by relying on DNA evidence about specific individuals, instead of the statistical evidence, collider bias is avoided.

To recap: from an epistemic point of view, the statistical evidence in the **Prison Yard** case, coupled with the selection procedure, lacks specificity at the level of description demanded by legal inquiries concerning liability. Adopting a higher level of description creates arbitrary statistical estimations of the selected defendant’s liability, since the defendant may equally have been selected at a different order in the process and assigned a different probability of liability on that basis, either above or below the threshold. As such, either way we cannot say that the probability of liability is strictly tistical Evidence Is Not Specific Enough.”

above threshold. Thus, is **Prison Yard** case is not a counterexample to PROBABILITY THRESHOLD.

## 4 Cross-Examining Legal Causalism

LEGAL CAUSALISM gives an account of merely statistical evidence and explains why it is not enough to establish the chances above the relevant probability threshold. An objection to this position is that it seems to violate the principle that we should use all available evidence when making decisions.<sup>65</sup> If our only available evidence is statistical, why should we not use it to guide our judgements?

Though initially plausible, this objection is mistaken. The approach presented here is compatible with the view that we should use all available evidence.<sup>66</sup> It simply argues for a further maxim: do not go beyond your evidence. Merely statistical evidence is consistent with multiple causal models and causal hypotheses. To nevertheless assign a determinate probability is to go beyond our evidence: it imposes more precision on a causal scenario than the evidence permits.

Moreover, we should resist the stipulation made in these cases that there is no other evidence. From an epistemological point of view, this is false.

---

65. Kasper Lippert-Rasmussen, ““We Are All Different”: Statistical Discrimination and the Right to be Treated as an Individual,” *Journal of Ethics* 15, nos. 1-2 (2011): 47–59.

66. Note that this principle, known as the Principle of Total Evidence, is not true for all cases. See: I. J. Good, “A Little Learning Can Be Dangerous,” *The British Journal for the Philosophy of Science* 25, no. 4 (1974): 340–342; Arif Ahmed and Bernhard Salow, “Don’t Look Now,” *British Journal for the Philosophy of Science* 70, no. 2 (2019): 327–350.

When investigating empirical hypotheses, we should always accept that there may be further evidence that could overturn them – no statement in this domain is immune to revision.<sup>67</sup> The stipulation thus slides the debate into the territory of policy-making, where we must account for the costs and benefits of attempting to gather information in the service of more accurate decision-making. If the court cannot feasibly admit more evidence at the time or in the foreseeable future and must proceed immediately, then the legal decision then becomes one of overriding social policy. In doing so, the underlying justification of its decision concerning the defendant is practical, not epistemic.

Granted, the choice of threshold may be justified by policy considerations about the relative importance of seeking truth and avoiding error in determinations of legal liability. We should be careful, however, not to base our assessments of the probative value of evidence on vague appeals to policy considerations. One might think, for example, that if we assume that the probability threshold is set correctly in **Prison Yard**, then we should find each prisoner liable, regardless of whether the evidence adequately tracks their individual causal role in the events. However, this would be a mistake.

Even accepting a decision-theoretic justification for setting the threshold, there is no good reason to constrain the social utility-information to exclude the effects of *how* trials are conducted. As others have argued, legal evidential

---

67. W. V. Quine, “Two Dogmas of Empiricism,” *Philosophical Review* 60, no. 1 (1951): 20–43.

standards and procedures themselves have effects on the fairness, accuracy, and efficiency of law enforcement institutions more generally. In particular, the use of group-level evidential procedures that are insensitive to individuals' actions may impose additional unfairness, induce ratcheting effects that distort the accuracy of the legal system, and may incentivise more crime, thus being socially inefficient.<sup>68</sup> These considerations are not detailed in the description of **Prison Yard**, nor can they be assumed without more careful argumentation explaining how the threshold is invariant to the social utility implications of a policy of finding everyone in a group liable when there is known to be an innocent party. The LEGAL CAUSALIST analysis aims to side-step this complication by taking the question of individual liability as the accepted grounds for determining the probative value of evidence.

On the epistemic side, a different concern may be that LEGAL CAUSALISM admits too much uncertainty, since there could always be some causal hypothesis according to which there is a sufficiently low probability that the defendant is liable. At least two responses are available here.

The first response is, in a sense, merely formal.<sup>69</sup> Within the framework

---

68. On fairness-considerations: Barbara D. Underwood, "Law and the Crystal Ball: Predicting Behavior with Statistical Inference and Individualized Judgment," *The Yale Law Journal* 88, no. 7 (1979): 1408; Bolinger, "Reasonable Mistakes and Regulatory Norms: Racial Bias in Defensive Harm"; Di Bello, "Trial by Statistics: Is a High Probability of Guilt Enough to Convict?"; Marcello Di Bello and Collin O'Neil, "Profile evidence, fairness, and the risks of mistaken convictions," *Ethics* 130, no. 2 (2020): 147–178. On accuracy-considerations: Harcourt, *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age*. On policy incentives: Harcourt, *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age*; Enoch, Spectre, and Fisher, "Statistical Evidence, Sensitivity, and the Legal Value of Knowledge."

69. Thanks to Kieran Setiya for this objection.

provided, we can specify the acceptance rule to require that some proportion of admissible probabilities be above the threshold, depending on the threshold.<sup>70</sup> In the case of civil cases, we may require that the majority of admissible probabilities be above 0.5. In criminal cases, we might hold that what it means for there to be a *reasonable doubt* is that there is *some* admissible probability function that falls below the threshold. So, there is flexibility within the LEGAL CAUSALIST framework to avoid this objection.

However, the more substantive response addresses the underlying epistemological issue inherent in the objection: it is a general feature of any broadly probabilistic approach to epistemology: probability, after all, is measure that is defined relative to a space of possibilities. In rough and ready terms, we may distinguish two uses of evidence that are often conflated in the debate. In its *wide-scope* usage, evidence tells us which causal hypotheses are relevant to the investigation. It helps us form, as it were, the background assumptions of the investigation. Given a set of background assumptions and causal hypotheses, evidence may serve a *narrow-scope* role of changing the relative probability of the hypotheses. Statistical evidence may be used in either way: it may serve a wide-scope use through causal discovery techniques;<sup>71</sup> or, it may serve to update our credences in the chances of a pre-defined set

---

70. The approach might draw on the idea of caution formalised in Isaac Levi, *Gambling with Truth: An Essay on Induction and the Aims of Science* (New York: Knopf, 1967); Isaac Levi, *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance* (MIT Press, 1980).

71. Spirtes, Glymour, and Scheines, *Causation, Prediction, and Search*; Nevin Climenhaga, Lane DesAutels, and Grant Ramsey, "Causal Inference from Noise," *Noûs* 55, no. 1 (2021): 152–170

of hypotheses. The important fact is that statistical evidence of the kind presented as counterexamples to PROBABILITY THRESHOLD does neither.

So how do we determine which probability values are permitted by the evidence? The same way that we infer probabilities of causation more generally: through observation, intervention, simulation, and various epistemic principles that have been proposed elsewhere in the debate by those who have (mistakenly, I have argued) rejected PROBABILITY THRESHOLD.

These concepts include assessments of counterfactuals, normalcy, relative plausibility, and relevant alternatives.<sup>72</sup> These accounts are correct to identify the importance of these concepts in examining the epistemological features of the evidence. However, they are too general because they do not constrain these concepts with a causal analysis; in fact, they appear to deliberately avoid this commitment. Perhaps this is to afford their theories more generality. Or perhaps it is because the causal interpretation is widely, but unspokenly, already assumed to be correct. In any case, without causal concepts playing a more central role, these accounts have difficulty providing

---

72. On the idea that statistical evidence lacks counterfactual sensitivity or safety, see (respectively) Enoch, Spectre, and Fisher, “Statistical Evidence, Sensitivity, and the Legal Value of Knowledge”; Enoch and Fisher, “Sense and “Sensitivity”: Epistemic and Instrumental Approaches to Statistical Evidence” and Pritchard, “Legal Risk, Legal Evidence and the Arithmetic of Criminal Justice.” On normalcy and the importance that evidence explain a hypothesis of liability, see Smith, “When Does Evidence Suffice for Conviction?”; Martin Smith, “More on Normic Support and the Criminal Standard of Proof,” *Mind* 00 (2021): 361–365. On relative plausibility and inference to the best explanation, see: Ronald J. Allen and Michael S. Pardo, “Relative Plausibility and Its Critics,” *International Journal of Evidence and Proof* 23, nos. 1-2 (2019): 5–59. On principles for discerning relevant alternatives to the hypothesis that the defendant is liable, see Gardiner, “The Reasonable and the Relevant: Legal Standards of Proof.”

a principled way of constraining their analysis.

For example, the normic account appeals to quantities of explanation of particular events, but it leaves open what type of explanation would suffice. Arguably, in legal contexts, the relevant type of explanation will be a causal one. Whether or not we interpret normalcy in terms of explanation, in order to isolate actual causation, it seems that some delimiting concept of normalcy will be required.<sup>73</sup> Importantly for present purposes, this normic notion would not constitute a complete or rival account to LEGAL CAUSALISM; it would be an auxiliary set of principles in conjunction with broader methods of causal inference.

The importance of causality also applies to counterfactual analyses of evidence. These analyses tend to rely on similarity comparisons across possible worlds, holding that the value of evidence is at least partly determined by its ability to track the liability or non-liability of the defendant across the most similar worlds to the actual one. One might ask: similarity with respect to which properties? And, if we are estimating the probability that the defendant's actions were a necessary condition for the resulting harm (the *sine qua non* test), how do we come to a principled estimation of this quantity? LEGAL CAUSALISM offers a principled set of methods for addressing these otherwise seemingly intractable questions.

---

73. On approaches that rely on normalcy assumptions for identifying actual causation, see: H. L. A. Hart and Tony Honore, *Causation in the Law* (Oxford: Oxford University Press, 1985); Ned Hall, "Structural Equations and Causation," *Philosophical Studies* 132, no. 1 (2007): 109–136; Joseph Y. Halpern, *Actual Causality* (UK: MIT Press, 2016).

The concept of relative plausibility has been explicitly promoted as an alternative to probability theory in these contexts. What is clear about this concept is that it emphasises the kind of explanatory information that LEGAL CAUSALISM demands. What is unclear about this concept is how it differs from Bayesian reasoning, which LEGAL CAUSALISM (being a constrained version of LEGAL PROBABILISM) ascribes to.<sup>74</sup> Insofar as it does differ, it is doubtful whether this is a desirable feature of a normative theory of legal epistemology.

Last, but not least, the relevant alternatives approach, too, gets at something important: it focuses on how evidence can “eliminate” error possibilities, an idea that is kept largely figurative, but has its roots in early causal inference.<sup>75</sup> Absent a causal interpretation, we may ask: how, exactly, does evidence eliminate error possibilities? To best answer this question, we seem to require the conceptual resources developed in philosophical and formal theories of causal inference.

Clearly, however, these competing (or compatible?) views deserve a fairer hearing than I can give them here. For now, I hope to have shown that there exists a distinct and defensible LEGAL CAUSALIST approach to resolving the STATISTICAL PROOF PARADOX that, unlike its rivals, maintains PROBABILITY THRESHOLD.

---

74. Here I endorse: David S. Schwartz and Elliott Sober, “What is Relative Plausibility?,” *International Journal of Evidence and Proof* 23, nos. 1-2 (2019): 198–204.

75. John Stuart Mill, *A System of Logic*, 4th (London: John W. Parker / Son, 1856).

## 5 Conclusion

Legal standards of proof are reducible to probability thresholds. However, this reduction does not entail that statistical evidence is sufficient for determining liability. In fact, examining the causal relevance of statistical evidence reveals that, in general, such evidence has long been given too much credence. The STATISTICAL PROOF PARADOX can be resolved along the lines suggested by Thomson (1986), with the benefit of recent developments in formal methods of causal inference. These methods, more generally, appear to provide a transparent and systematic approach to determining when evidence – statistical or otherwise – is enough.<sup>†</sup>

## References

- Ahmed, Arif. *Evidence, Decision and Causality*. Cambridge University Press, 2014.
- Ahmed, Arif, and Bernhard Salow. “Don’t Look Now.” *British Journal for the Philosophy of Science* 70, no. 2 (2019): 327–350.
- Allen, Ronald J., and Michael S. Pardo. “Relative Plausibility and Its Critics.” *International Journal of Evidence and Proof* 23, nos. 1-2 (2019): 5–59.
- Bareinboim, Elias, and Judea Pearl. “Causal inference and the data-fusion problem.” *Proceedings of the National Academy of Sciences of the United States of America* 113, no. 27 (2016): 7345–7352.

---

<sup>†</sup>. Forthcoming in *Noûs*. For especially helpful conversations and comments I wish to thank Chris Bottomley, Ray Briggs, Lara Buchak, Marcello DiBello, Kevin Dorst, Phil Dowe, Sina Fazelpour, Branden Fitelson, Bruce Glymour, Alan Hajék, Renée Jorgensen, David Kinney, Kieran Setiya, Jeremy Strasser, Mike Titelbaum, Rafal Urbaniak, Pat Williams, audiences at Stanford University’s McCoy Family Center for Ethics in Society, Marcello DiBello’s Artificial Intelligence and Ethics graduate seminar at Arizona State University, Northeastern University’s Artificial Intelligence and Data Ethics Working Group, Branden Fitelson’s 2021 Northeastern Epistemology Workshop, and a PhilSoc Seminar at the Australian National University. Not least, I wish to thank various anonymous reviewers for their insightful and constructive feedback.

- Basu, Rima. "What We Epistemically Owe to Each Other." *Philosophical Studies* 176, no. 4 (2019): 915–931.
- Blome-Tillmann, Michael. "Statistical Evidence, Normalcy, and the Gate-crasher Paradox." *Mind* 129, no. 514 (2020): 563–578.
- Bolinger, Renée Jorgensen. "Reasonable Mistakes and Regulative Norms: Racial Bias in Defensive Harm." *Journal of Political Philosophy* 25, no. 2 (2017): 196–217.
- . "The Rational Impermissibility of Accepting (Some) Racial Generalizations." *Synthese* 197, no. 6 (2020): 2415–2431.
- Briggs, R.A. "Interventionist Counterfactuals." *Philosophical Studies* 160, no. 1 (2012): 139–166.
- . "The Metaphysics of Chance." *Philosophy Compass* 5, no. 11 (November 2010): 938–952.
- Broadbent, Alex. "Epidemiological evidence in proof of specific causation." *Legal Theory* 17, no. 4 (2011): 237–278.
- Buchak, Lara. "Belief, Credence, and Norms." *Philosophical Studies* 169, no. 2 (2014): 285–311.
- Clermont, Kevin M. "Res Judicata as Requisite for Justice." *Rutgers University Law Review* 68, no. 3 (2016): 1067–1142.
- Climenhaga, Nevin, Lane DesAutels, and Grant Ramsey. "Causal Inference from Noise." *Noûs* 55, no. 1 (2021): 152–170.
- Cohen, Jonathan. *The Probable and the Provable*. Oxford: Oxford University Press, 1977.
- Colyvan, Mark, Helen M Regan, and Scott Ferson. "Is It a Crime to Belong to a Reference Class?" *Journal of Political Philosophy* 9, no. 2 (2001): 168–181.
- Di Bello, Marcello. "Trial by Statistics: Is a High Probability of Guilt Enough to Convict?" *Mind* 128, no. 512 (2019): 1045–1084.
- . "When Statistical Evidence Is Not Specific Enough." *Synthese*:1–26.
- Di Bello, Marcello, and Collin O’Neil. "Profile evidence, fairness, and the risks of mistaken convictions." *Ethics* 130, no. 2 (2020): 147–178.
- Dorst, Kevin. "Lockeans Maximize Expected Accuracy." *Mind* 128, no. 509 (2019): 175–211.
- Eidelson, Benjamin. "Treating People as Individuals." In *Philosophical Foundations of Discrimination Law*, edited by Deborah Hellman and Sophia Moreau. 2013.
- Enoch, David, and Talia Fisher. "Sense and "Sensitivity": Epistemic and Instrumental Approaches to Statistical Evidence." *Stanford Law Review* 67, no. 3 (2015): 557–611.
- Enoch, David, and Levi Spectre. "Statistical Rsentment, or: What’s Wrong with Acting, Blaming, and Believing on the Basis of Statistics Alone." *Synthese* 199, no. 3 (2021): 5687–5718.
- Enoch, David, Levi Spectre, and Talia Fisher. "Statistical Evidence, Sensitivity, and the Legal Value of Knowledge." *Philosophy and Public Affairs* 40, no. 3 (2012): 197–224.

- Eriksson, Lina, and Alan Hájek. “What Are Degrees of Belief?” *Studia Logica* 86, no. 2 (2007): 183–213.
- Gardiner, Georgi. “Legal Epistemology.” In *Oxford Bibliographies: Philosophy*, edited by Duncan Pritchard. Oxford University Press, 2019.
- . “Profiling and Proof: Are Statistics Safe?” *Philosophy* 95, no. 2 (2020): 161–183.
- . “The Reasonable and the Relevant: Legal Standards of Proof.” *Philosophy and Public Affairs* 47, no. 3 (2019): 288–318.
- Glymour, Clark, David Danks, Bruce Glymour, Frederick Eberhardt, Joseph Ramsey, Richard Scheines, Peter Spirtes, Choh Man Teng, and Jiji Zhang. “Actual Causation: A Stone Soup Essay.” *Synthese* 175, no. 2 (2010): 169–192.
- Good, I. J. “A Little Learning Can Be Dangerous.” *The British Journal for the Philosophy of Science* 25, no. 4 (1974): 340–342.
- Greenland, Sander, and James M. Robins. “Epidemiology, Justice, and the Probability of Causation.” *Jurimetrics* 40, no. 3 (2000): 321–340.
- Hájek, Alan. “Interpretations of Probability.” *Stanford Encyclopedia of Philosophy*, 2019.
- Hall, Ned. “Structural Equations and Causation.” *Philosophical Studies* 132, no. 1 (2007): 109–136.
- Halpern, Joseph Y. *Actual Causality*. UK: MIT Press, 2016.
- Harcourt, Bernard E. *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age*. University of Chicago Press, 2008.
- Hart, H. L. A., and Tony Honoré. *Causation in the Law*. Oxford: Oxford University Press, 1985.
- Hedden, Brian, and Mark Colyvan. “Legal Probabilism: A Qualified Defence.” *Journal of Political Philosophy* 27, no. 4 (2019): 448–468.
- Hitchcock, Christopher. “Conditioning, Intervening, and Decision.” *Synthese* 193, no. 4 (2016): 1157–1176.
- Hosein, Adam Omar. “Racial Profiling and a Reasonable Sense of Inferior Political Status\*.” *Journal of Political Philosophy* 26, no. 3 (2018): 1–20.
- Ismael, Jenann. “Raid! Dissolving the Big, Bad Bug.” *Nous* 42, no. 2 (2008): 292–307.
- Jeffrey, Richard C. *The Logic of Decision*. Second. Chicago and London: The University of Chicago Press, 1983.
- Joyce, James M. “How Probabilities Reflect Evidence.” *Philosophical Perspectives* 19 (2005).
- Kaplan, John. “Decision Theory and the Factfinding Process.” *Stanford Law Review* 20, no. 6 (1968): 1065.
- Laudan, Larry. “Is Reasonable Doubt Reasonable?” *Legal Theory* 9, no. 04 (2003): 295–331.
- . *Truth, Error, and Criminal Law: An Essay in Legal Epistemology*. Cambridge Studies in Philosophy and Law. Cambridge University Press, 2006.
- Levi, Isaac. *Gambling with Truth: An Essay on Induction and the Aims of Science*. New York: Knopf, 1967.

- Levi, Isaac. *The Enterprise of Knowledge: An Essay on Knowledge, Credal Probability, and Chance*. MIT Press, 1980.
- Lewis, David. "A Subjectivist's Guide to Objective chance." *Ifs*, 1980.
- . "Statements partly about observation." *Philosophical Papers* 17, no. 1 (1988): 1–31.
- Lippert-Rasmussen, Kasper. "'We Are All Different': Statistical Discrimination and the Right to be Treated as an Individual." *Journal of Ethics* 15, nos. 1-2 (2011): 47–59.
- Littlejohn, Clayton. "Truth, Knowledge, and the Standard of Proof in Criminal Law." *Synthese* 197 (2020): 5253–5286.
- Meek, Christopher, and Clark Glymour. "Conditioning and Intervening." *The British Journal for the Philosophy of Science* 45, no. 4 (1994): 1001–1021.
- Mill, John Stuart. *A System of Logic*. 4th. London: John W. Parker / Son, 1856.
- Mortini, Dario. "Knowledge, Individualised Evidence and Luck." *Philosophical Studies* 179, no. 12 (2022): 1–30.
- Moss, Sarah. "Knowledge and Persons." Chap. 10 in *Probabilistic Knowledge*. Oxford Scholarship Online, 2018.
- Nesson, Charles R. "Reasonable Doubt and Permissive Inferences: The Value of Complexity." *Harvard Law Review* 92, no. 6 (1979): 1187–1125.
- Pearl, Judea. *Causality*. 2nd edn. Cambridge University Press, 2009.
- Pearl, Judea, and Elias Bareinboim. "External validity: From do-calculus to transportability across populations." *Statistical Science* 29, no. 4 (2014): 579–595.
- Pettigrew, Richard. *Accuracy and the Laws of Credence*. Oxford University Press, 2016.
- . "Accuracy, Chance, and the Principal Principle." *Philosophical Review* 121, no. 2 (2012): 241–275.
- Pritchard, Duncan. "Legal Risk, Legal Evidence and the Arithmetic of Criminal Justice." *Jurisprudence* 9, no. 1 (2018): 108–119.
- Quine, W. V. "Two Dogmas of Empiricism." *Philosophical Review* 60, no. 1 (1951): 20–43.
- Ramsey, Frank. "Truth and Probability." Chap. 7 in *The Foundations of Mathematics and other Logical Essays*, edited by R. B. Braithwaite, 156–198. 1926. New York: Harcourt, Brace / Company, 1926.
- Redmayne, Mike. "Exploring the Proof Paradoxes." *Legal Theory* 14, no. 4 (2008): 281–309.
- Reichenbach, Hans. *The Direction of Time*. University of California Press, 1956.
- Richardson, Rashida, Jason M Schultz, and Kate Crawford. "Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing systems, and Justice." *New York University Law Review* 94, no. 2 (2019): 192–233.
- Saks, Michael J., and Jonathan J. Koehler. "The Individualization Fallacy in Forensic Science Evidence." *Vanderbilt Law Review* 61, no. 1 (2008): 199–219.

- Savage, Leonard J. *The Foundations of Statistics*. 2nd rev. New York: Dover, 1972.
- Schauer, Frederick. *Profiles, Probabilities, and Stereotypes*. Harvard University Press, 2003.
- Schum, David A. *The Evidential Foundations of Probabilistic Reasoning*. John Wiley / Sons, Inc., 1994.
- Schwartz, David S., and Elliott Sober. “What is Relative Plausibility?” *International Journal of Evidence and Proof* 23, nos. 1-2 (2019): 198–204.
- Smith, Martin. “More on Normic Support and the Criminal Standard of Proof.” *Mind* 00 (2021): 361–365.
- . “When Does Evidence Suffice for Conviction?” *Mind* 127, no. 508 (2017): 1193–1218.
- Spirtes, Peter, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*. 2nd edn. MIT Press, 2001.
- Thomson, Judith Jarvis. “Liability and Individualized Evidence.” *Law and Contemporary Problems* 49, no. 3 (1986): 199.
- . “Remarks on Causation and Liability.” *Philosophy and Public Affairs* 13, no. 2 (1984): 101–133.
- . *Rights, Restitution, and Risk: Essays in Moral Theory*. Cambridge, Mass.: Harvard University Press, 1986.
- Tribe, Laurence H. “Trial by Mathematics: Precision and Ritual in the Legal Process.” *Harvard Law Review* 84, no. 6 (1971): 335–399.
- Tversky, Amos, and Daniel Kahneman. “Judgement under Uncertainty: Heuristics and Biases.” *Science* 185, no. 4157 (1974): 1124–1131.
- Underwood, Barbara D. “Law and the Crystal Ball: Predicting Behavior with Statistical Inference and Individualized Judgment.” *The Yale Law Journal* 88, no. 7 (1979): 1408.
- Urbaniak, Rafal, and Marcello Di Bello. “Legal Probabilism.” In *The Stanford Encyclopedia of Philosophy*, Fall 2021, edited by Edward N. Zalta. Metaphysics Research Lab, Stanford University, 2021.
- Wasserman, David T. “The Morality of Statistical Proof and the Risk of Mistaken Liability.” *Cardozo Law Review* 13 (1991): 935–947.
- Wells, Gary L., and Elizabeth A. Olson. “Eyewitness Testimony.” *Annual Review of Psychology* 54 (2003): 277–295.
- White, Roger. “Evidential Symmetry and Mushy Credence.” Chap. 7 in *Oxford Studies in Epistemology, Volume 3*, 161–186. Oxford University Press, 2010.
- Woodward, James. *Causation with a Human Face: Normative Theory and Descriptive Psychology*. Oxford Studies in Philosophy of Science Series. Oxford University Press, Incorporated, 2021.
- Yablo, Stephen. *Aboutness*. Princeton University Press, 2014.