



# Legal proof and statistical conjunctions

Lewis D. Ross<sup>1</sup> 

© The Author(s) 2020

**Abstract** A question, long discussed by legal scholars, has recently provoked a considerable amount of philosophical attention: ‘Is it ever appropriate to base a legal verdict on statistical evidence alone?’ Many philosophers who have considered this question reject legal reliance on bare statistics, even when the odds of error are extremely low. This paper develops a puzzle for the dominant theories concerning why we should eschew bare statistics. Namely, there seem to be compelling scenarios in which there are *multiple sources of incriminating statistical evidence*. As we conjoin together different types of statistical evidence, it becomes increasingly incredible to suppose that a positive verdict would be impermissible. I suggest that none of the dominant views in the literature can easily accommodate such cases, and close by offering a diagnosis of my own.

**Keywords** Proof paradox · Legal proof · Statistical evidence · Blue Bus problem · Legal epistemology · DNA · Philosophy of law

## 1 Introduction

A question, long discussed by legal scholars, has recently provoked a considerable amount of philosophical attention: ‘Is it ever appropriate to base a legal verdict on statistical evidence alone?’ The dominant view is that merely statistical evidence

---

✉ Lewis D. Ross  
L.ross2@lse.ac.uk

<sup>1</sup> Department of Philosophy, Logic, and Scientific Method, London School of Economics, Lakatos Building, Portugal Street, Holborn, London, UK

cannot provide a proper basis for a legal verdict, even when the odds of error are extremely low.<sup>1</sup>

Here are some classic cases—sometimes referred to as ‘proof-paradoxical’ scenarios<sup>2</sup>—which illustrate the intuitive case against imposing sanctions on the basis of bare statistics:

**Blue Bus.** A bus negligently causes injury to a pedestrian, but it is not known which company the bus belongs to. On the route where the accident occurred, the Blue Bus Company runs 75% of the buses. There is no further information available to settle which company the bus belongs to. [Adapted from Tribe 1971].

**Gatecrasher.** The organizers of the local rodeo decide to sue John for gatecrashing their event. Their evidence is as follows: John attended the Sunday afternoon event—he was seen and photographed on the main ranks during the event. No tickets were issued, so John cannot be expected to prove that he bought a ticket with a ticket stub. However, while 1000 people were counted in the seats, only 300 paid for admission. [Presentation adapted from Blome-Tillmann forthcoming; original case due to Cohen 1977]

**Prisoners.** One hundred prisoners are exercising in the prison yard. Ninety-nine of them suddenly join in a planned attack on a prison guard; the hundredth prisoner plays no part. There is no evidence available to show who joined in and who did not. [Adapted from Redmayne 2008]

The ‘standard’ intuition is that it would not be appropriate in these cases to impose civil or criminal sanctions on the basis of the inculpatory statistical evidence. Such intuitions raise issues with important theoretical and practical ramifications for the law. The question of how to treat bare statistics impinges upon debates about the nature of legal proof itself,<sup>3</sup> whether it is acceptable to convict someone of a crime simply on the basis of DNA evidence,<sup>4</sup> how we should treat epidemiological evidence,<sup>5</sup> and the impermissibility of demographic profiling.<sup>6</sup> And these cases are not as artificial as one might initially suppose: the Blue Bus case has found close analogues in real case-law.<sup>7</sup>

<sup>1</sup> Many views are categorical. However, some extant views allow, or are silent about, the use of bare statistics in certain restricted circumstances. I provide a comprehensive literature survey in Sect. 4. For a succinct overview of contemporary debate, see Ross (2020a). Pardo (2019) provides a helpful introduction to the proof paradox alongside other legal paradoxes.

<sup>2</sup> For a plethora of other examples, see Gardiner (2018).

<sup>3</sup> Some suppose that aversion to bare statistics threatens ‘legal probabilism’, the idea that proof requires establishing a claim to a given level of probability. For contrasting perspectives, see Haack (2014), Hedden and Colyvan (2019) and Smith (2020). However, one can arguably reject the use of bare statistics without rejecting legal probabilism: we can justify tenets of evidence law that exclude or restrict certain classes of evidence without appealing to any particular view of the nature of legal proof.

<sup>4</sup> For example, see *R v. Lashley* [2000] EWCA 88; *Wilson v DPP* 2017 IESC 54; *State v Toomes* 191 S.W.3d 122, 129 for representative cases from different jurisdictions.

<sup>5</sup> See Ross (ms) for discussion.

<sup>6</sup> E.g., see Bolinger (2018) for an explicit connection between the Blue Bus case and issues surrounding racial profiling. Di Bello and O’Neill (2020) provide further discussion.

<sup>7</sup> E.g., see *Kaminsky v. Hertz Corporation*, 288 N.W.2d 426.

The aim of this paper is to develop a puzzle for the dominant view that we should reject bare statistics in the courtroom. The puzzle is that there seem to be compelling scenarios in which there are *multiple incriminating sources of statistical evidence*. As we conjoin together different types of statistical evidence, it becomes increasingly incredible to suppose that a positive verdict would be impermissible. This suggests that whatever *is* wrong with the evidence in familiar proof-paradoxical cases cannot simply be explained by gesturing at the statistical nature of the evidence involved. To deepen the puzzle, I show that four dominant approaches in the literature struggle to draw a principled distinction between regular proof-paradoxical scenarios and those involving statistical conjunctions. I close by outlining my own view on what explains the intuitive difference between these cases, drawing on the empirically supported ‘story model’ of legal fact-finding, and offer some reflections on where this leaves the state of the wider debate.

## 2 Civil law

Let’s begin with civil law cases, canonically represented by the Blue Bus scenario. Given that the operative standard of proof in civil law is the ‘balance of probabilities’ (sometimes called ‘preponderance of the evidence’), a reluctance to impose sanctions here is *prima facie* puzzling: surely it is more probable than not that the Blue Company caused the accident? Even though, conceptually, it seems like the evidence should straightforwardly satisfy the relevant standard of proof, it is widely held that intuition balks at this conclusion. Hence, there is apparently something amiss with the prospect of using mere statistics to settle a civil case.<sup>8</sup>

I want to suggest that our intuitions about the legal impotence of purely statistical evidence are not stable when we introduce multiple sources of statistical evidence. The end result will be a case in which the total body of evidence seems to remain purely statistical, but in which sanctioning the Blue Company is compelling.

Consider the following variation on Blue Bus, involving one additional source of statistical evidence.

Blue Bus<sup>2</sup>: A bus causes injury to a pedestrian, but it is not known which company the bus belongs to. On the route where the accident occurred, the Blue Company runs 75% of the buses and the Red Company 25% of the buses. Fresh tyre-marks are found at the scene of the accident that an investigator’s uncontested report states were caused by the offending vehicle. All parties agree these could only be made by a certain brand of bus tyre. A recent insurance application form shows that 90% of the Blue Company buses have that brand of tyre, while only 5% of Red Company buses do.

The inculpatory evidence in this case remains purely statistical—the finding of the tyre-marks is only relevant insofar as presented alongside the reference-class (all buses) and probability (90%) linking these marks to the Blue Company. The

---

<sup>8</sup> For dissent to hostility towards statistics in the civil domain, see Krauss (2020).

impermissibility of holding the Blue Company liable is surely less obvious in this variation. After all, we must remember that choosing not to sanction leaves the injured pedestrian without compensation. Still, you may not be convinced that the evidence is yet strong enough for sanction. Nonetheless, it is apparent that we can continue devising additional sources of statistical evidence and adding them to the case. Consider a third variation:

Blue Bus<sup>3</sup>: A bus causes injury to a pedestrian, but it is not known which company the bus belongs to. On the route where the accident occurred, the Blue Company runs 75% of the buses and the Red Company 25% of the buses. Fresh tyre-marks are found at the scene of the accident that an investigator's uncontested report states were caused by the offending vehicle. All parties agree these could only be made by a certain brand of bus tyre. A recent insurance application form shows that 90% of the Blue Company buses have the implicated brand of bus tyre, while only 5% of Red Company buses do. Moreover, police find a bus hubcap on the road immediately after the crash. Only 2% of the Red Company buses were recorded as having the implicated brand of hubcap, while 96% of Blue Company buses have it.

Suppose there is no known correlation between having the implicated hubcap and tyre, and that neither party adduces evidence about how many buses belonging to each company have both. Would it be impermissible to rule against the Blue Company on the balance of probabilities, leaving the victim of the accident without compensation? I think that this is hard to accept. However, if you are not yet convinced, *we can repeat the same strategy*, continuing to Blue Bus<sup>n</sup>, conjoining ever more sources of inculpatory statistical evidence. The end result would be a body of evidence about which a defending lawyer could force the concession that it was all merely statistical, but where it becomes extremely difficult to suppose, especially since proof-paradoxical cases never involve exculpatory evidence being offered, that the case should not win the day on the balance of probabilities standard of civil proof. To maintain uniform hostility to statistical evidence, you would have to hold that there is no variation of the Blue Bus case in this vein that could license finding against the Blue Company. Such a position, leaving the victim of a negligently-caused accident uncompensated in order to preserve hostility to statistical evidence is difficult to motivate.

### 3 Criminal law

To begin the discussion of criminal law, we need to first briefly dwell on the connection between the debate on bare statistical evidence and DNA profiling.

When evaluating DNA evidence, courts rely entirely on a statistical estimate, provided by a forensic scientist, concerning the probability that the incriminating sample (blood, semen, hair, etc. found at a crime-scene) belongs to the accused person. The reason that a statistical *estimate* is provided rather than an outright assertion of a match is that it is well-understood that, given the limitations of DNA sampling techniques, there is always the chance of a *random match*. In other words,

given the level of detail provided by DNA profiling techniques, it is possible that the allele-characteristics of the incriminating sample, while very similar to those belonging to the accused, may in fact belong to some other person. Extremely improbable although it may be, apparent DNA matches can turn out to be pure coincidences.

With this in mind, it is not unusual for there to be cases that fit the following mould:

DNA: Someone is sexually assaulted in a secluded park. They cannot provide an account of the attacker's appearance. DNA evidence from the crime matches that of someone on file for some unrelated reason. The incriminating evidence is the following: a forensic scientist estimates the chance of the DNA not belonging to that person to be 1 in ten million. [Adapted from Ross 2020b]

How should we react to such cases? There is currently no consensus on this question. Both courts and legal scholars have been much more sympathetic to the legitimacy of cold hit convictions than philosophical commentators.<sup>9</sup> Within the philosophical literature, a number of theorists have explicitly extended their general antipathy towards bare statistical evidence to endorse an outright rejection of cold hit DNA convictions, and suspicion of conviction in such cases follows implicitly from many extant views.<sup>10</sup> I will return to these views below in more depth. Firstly, however, I want to demonstrate that, just as with the civil law, there are statistical conjunction cases in the criminal law that make forbearing from conviction hard to accept. Consider a conjunction of a standard 'proof-paradoxical' case introduced earlier, and an apparent DNA profile match.

Prisoners & DNA: 100 prisoners are exercising in the prison yard. Extremely grainy CCTV footage shows that 99 of them attack and kill the guard. The 100th prisoner played no role in the assault and could have done nothing to stop it. From the footage it is impossible to distinguish which prisoners were involved. The 99 murderers escape in one direction and, some time later, the 100th prisoner escapes in a different direction. One prisoner is recaptured. Upon testing, it is found that his DNA matches the most dominant DNA profile found on a discarded switch-blade at the scene of the murder. The forensic scientist estimates the chance of a random match as 1 in ten million.

I think it is hard to accept that it would be impermissible for a jury to convict in such a case. Again, though, intuitions may vary. Nonetheless, as with the civil case, it is

<sup>9</sup> See Roth (2010) for a wide-ranging discussion of the law. It is worth stressing that the case-law is not unequivocal. On the philosophical side of the debate, Ross (2020b) and Papineau (2020) offer some dissent from philosophical scepticism about bare statistics in criminal law, while Di Bello (2019) offers a decision-theoretic argument that vindicates conviction in DNA cases, but not in short-odds proof paradox cases.

<sup>10</sup> For example, see the work of Duncan Pritchard and Martin Smith. When we examine other extant views below in Sect. 4, it will be apparent why many popular theories are implicitly hostile to cold-hit DNA convictions even if their original proponents do not consider them.

entirely possible to continue the general strategy of conjoining further sources of statistical evidence to strengthen the case against the accused.

While some such further variations in the vein of Prisoners & DNA would doubtless be far-fetched, there are conceivable examples that would draw upon entirely familiar sources of evidence. Consider the following quote from two leading legal scholars about other types of forensic evidence such as fingerprints:

All forensic identification methods are probabilistic in nature. Traditional forensic examiners, such as fingerprint examiners, may treat their matches as unique, but as many have pointed out, such declarations of individualisation are fictive. The Galton points of a fingerprint, the characteristics of a person's handwriting and the striations of a ballistics match all have underlying population statistics, just like a DNA genotype [...]. The fact that the defendant's fingerprint, handwriting or gun 'matches' makes it more likely that the defendant was the source, but the inference is still probabilistic. [Cheng and Nunn 2016: 118]

It is worth stressing that not all types of forensic evidence are *currently* presented in court in the same statistical way as apparent DNA matches. Fingerprint evidence, for instance, is usually presented in a non-statistical way—an expert identifier uses their experience and intuition to testify as to whether two samples, in their view, upon examining similarities in ridge-patterns, were made by the same person.<sup>11</sup> However, this practice has come in for a fair degree of criticism. The subjectivity of fingerprint examination, according to some commentators, compares disfavouredly with the rigorous modelling techniques used to generate DNA match estimates. Moreover, some high-profile false positives provide grounds to suppose that such qualitative presentations are much less reliable than is often supposed. This debate is not my primary concern here. Rather, I simply want to note that one reasonable way that it has been suggested that we respond to such challenges is to advocate that fingerprint evidence should instead be presented in quantitative way, more in line with the way that a forensic scientist treats DNA evidence—namely by offering a statistical estimate about the probability of the incriminating sample matching the prints of the suspect.<sup>12</sup> While we do not yet have widely agreed upon statistical models concerning the likelihood of random fingerprint matches, there has been work in this direction and it is entirely conceivable that such a statistical estimate could become the norm for presenting fingerprint evidence, taking into account the estimated possibility of a random match.<sup>13</sup>

If this is a conceivable courtroom practice, as it surely is, then this raises the possibility of statistical conjunction cases involving DNA, fingerprints, and other forms of forensic evidence arrayed together to make an insurmountable case for

<sup>11</sup> See, for example, Zabell (2005) or Evett and Williams (2015) for methodological discussion.

<sup>12</sup> Zabell *ibid.* provides wide-ranging discussion of the limitations of fingerprint evidence and how the practices surrounding it might be improved.

<sup>13</sup> E.g. see Zhu et al. (2007) for one such attempt. As far as I understand the model, the chance of a random match was estimated at 1 in 400,000. This is purely illustrative for my argument, nothing depends on the exact figures.

conviction, even though they are being presented in a purely statistical way.<sup>14</sup> I leave the exact details of these cases to the reader's imagination—but, surely, it is hard to accept that the conjunction of such evidence, including fingerprints and DNA, could never appropriately undergird a criminal conviction. The upshot is a simple one: if we accept that *some* cases involving conjoined sources of statistical evidence can license sanction, we cannot simply suppose that the statistical nature of the evidence involved is what explains intuitive reluctance to sanction in typical proof-paradoxical cases.<sup>15</sup>

#### 4 Four diagnoses of the proof paradox

The intuitive bad-standing of bare statistics, widely highlighted by philosophers commenting on the law, can seemingly be disrupted upon considering cases involving multiple sources of statistical evidence. In these cases, the imposition of sanctions seems to be rather compelling. Let us now turn to consider the theoretical importance of this finding.

The debate surrounding statistical evidence in the law contains a striking lacuna: there is no widely agreed statement of what makes evidence 'merely statistical'. For example, one recent paper, coming at an advanced point in the dialectic between two competing theories, states:

How do we, then, define statistical evidence? We don't. We—again, in a way that's consistent with the theoretical literature on statistical evidence—start with the examples. They clearly capture something intuitively important. We then try to understand the relevant phenomena better. If we're fortunate, we may end up with a definition, or an analysis. Or we may not [...] we may need to settle for an "I can't define it, but I know it when I see it" attitude. [Enoch and Spectre 2019: 184]

The assumption in my discussion so far has been that taking one piece of uncontroversially statistical evidence—where the incriminating element consists only in the probability of the defending party possessing some inculpatory characteristic relative to some reference-class—and conjoining it with another piece of uncontroversially statistical evidence, still leaves us with a total body of evidence which remains merely statistical.

One possibility is that a body of evidence involving multiple sources of statistics possesses some normatively important property which elevates it above the merely statistical evidence found in regular proof-paradoxical cases. It is certainly true that in some local contexts, multiple sources of statistical evidence can have a distinctive

<sup>14</sup> Di Bello (forthcoming) provides helpful discussion of fingerprint evidence in relation to the proof paradox and Martin Smith's diagnosis of why we should eschew bare statistics.

<sup>15</sup> It is worth mentioning that it would be strange if a hostility to bare statistics led to the perverse tactical incentive in the courtroom whereby forensic evidence could rightfully lead to a conviction if presented in a qualitative way but the very same evidence could not rightfully lead to a conviction if presented in a quantitative way.

justificatory effect—for instance, having multiple sources of statistical evidence may reassure us that the incriminating evidence does not exist merely due to malpractice (e.g. police misconduct) or incompetence (e.g. forensic contamination). But this cannot be the whole story. After all, in cases like the Blue Bus scenario, incompetence and malpractice are not at issue. They can even be stipulated out of the case. What we need is a more general explanation for why intuitions seem to differ between the multiple-source cases outlined above and the more familiar proof-paradoxical scenarios with which we started the paper. I will now consider four dominant theories aiming to explain what is wrong with relying upon bare statistics and will suggest that *none of them can offer a straightforward explanation* for why cases involving multiple sources of statistical evidence seem to call for different treatment.

The four types of theory I will consider are the following:<sup>16</sup>

1. *Epistemic diagnoses* argue that bare statistics fail to confer some important epistemic property—e.g. justification, knowledgeability, safety, etc.—onto legal verdicts.
2. *Moral and justice-based approaches* argue that relying on bare statistics frustrates non-epistemic normative constraints on evidence law: e.g. *backwards-looking* considerations such as respect for autonomy or due process, or *forward-looking* considerations—such as ensuring legal rules have the proper incentivising effect.
3. *Likelihood theories* argue that legal proof should be understood in terms of the comparative likelihood of competing accounts, rather than in terms of absolute probabilities.
4. *‘Phase change’ approaches* argue that bare statistical evidence becomes acceptable only when the chance of error crosses some threshold of extreme improbability.

#### 4.1 Epistemic approaches

Epistemic approaches to the proof paradox aim to identify some epistemic deficiency in bare statistical evidence. There are now a number of such theories, appealing to the absence of different epistemic properties. It is beyond a single paper to critique the details of *each* epistemic view in addition to discussing other theories, so I will here take a more general approach. I will outline and focus on a motivating analogy characteristic of epistemic views: the comparison between the proof paradox and lottery cases. A consideration of this motivating analogy will justify some scepticism about whether epistemic views can explain the difference between regular proof-paradoxical cases and those involving statistical conjunctions.

<sup>16</sup> These approaches may overlap. For instance, one might argue that the reason that certain epistemic properties are important is because rules promoting them align with the proper moral constraints on evidence law. Indeed, as discussed below, Enoch et al. (2012) relies on precisely this idea.



Epistemologists have long been concerned with statistical evidence in the form of ‘lottery propositions’, with an orthodoxy being that the highly probabilifying evidence supporting the proposition ‘I will lose a fair and large lottery’ fails to confer certain epistemic properties onto a belief in that proposition. For instance, it has been popular to suppose that, even if it is stupendously likely, one cannot *know* that one’s lottery ticket is a loser (e.g. Williamson 2000), while others suggest that one lacks certain types of *justification* to believe that one’s ticket is a loser (e.g. Smith 2016). This is a puzzling phenomenon in its own right: if I can’t know that I will lose a ten million ticket lottery, how can I claim to know many ordinary propositions about which I may be fallible? There has been a great deal of sophisticated work devoted to exploring the ramifications of lottery cases for normative epistemological theorising.

It has been widely noted by proponents of epistemic approaches to the proof paradox that these scenarios bear a striking resemblance to lottery cases. While the evidence we have in unadorned proof-paradoxical scenarios is highly probabilifying, there nonetheless seems to be something unsatisfactory about endorsing it by issuing a positive legal verdict, just as while the evidence we have that we will lose a large fair lottery is highly probabilifying, there seems to be something unsatisfactory about (say) asserting or claiming to know that you have lost. Given the similarity between proof-paradoxical cases and lottery scenarios, a number of leading epistemologists have attempted to explain the inadequacy of statistical evidence in the law with recourse to epistemic properties typically thought to be absent in lottery cases.<sup>17</sup> Some explanations appeal directly to the absence of knowledge (e.g. Moss 2016; forthcoming, Blome-Tillmann 2017, Littlejohn 2018) or certain types of justification (e.g. Smith’s 2018 normic theory of justification), while others appeal to epistemic properties that have been defended as conditions for having knowledge or justification such as: sensitivity (e.g. Enoch et al. 2012), safety (Pritchard 2015; 2018, Pardo 2018), or the elimination of relevant alternative error-possibilities (Gardiner 2020). Within this work, the analogy with lottery cases is often an explicit part of the argumentation.<sup>18</sup>

To the extent that epistemic approaches are motivated by a guiding analogy with lottery cases, they appear to face a difficult task to legitimise sanction in statistical conjunction cases.

Firstly, epistemic approaches obviously cannot appeal to the fact that mistakes are less likely in statistical conjunction cases. After all, from the perspective of epistemological theorising, it is not usually taken to matter if the chance of winning a lottery is merely low or extremely low.<sup>19</sup> For example, whether a lottery has 100 tickets or 1,000,000 tickets is not often supposed to make a difference when it comes to the various epistemic properties found in diagnoses of the impotence of

---

<sup>17</sup> See Backes (2019) for a discussion of the other direction: using legal cases as evidence for theories in epistemology.

<sup>18</sup> In addition to the papers cited in-text, see Buchak (2014) for an early influential endorsement of this analogy.

<sup>19</sup> For empirical study of attitudes towards lottery propositions under different conditions, see Ebert et al. (2018).

statistical evidence: even the evidence provided by a truly massive lottery will not make the belief that one has a losing ticket *sensitive* (one would have the same belief even if one had a winning ticket), *safe* (there is a close counterfactual possibility in which one has won), it will not eliminate a *salient error-possibility* (i.e. that one has a winning ticket), and nor will it provide *normic justification* to the belief that one has a losing ticket (it would not be abnormal in the sense of requiring special explanation if one had a winning ticket). And, standardly, nor would one *know* that one had a losing ticket just by playing a very large lottery.

A more plausible response would emphasise the epistemic significance of there being *more than one source* of evidence in the statistical conjunction cases. There is, after all, no doubt that having more than one source of evidence is typically an epistemically good-making feature of a belief. However, as we are stipulating that the additional evidence remains purely statistical, it is in fact no easy task for epistemic views to explain why there is a qualitative difference between cases involving one inculpatory statistic and multiple inculpatory statistics. To see this, consider that it is possible to derive familiar epistemic results about the absence of knowledge—and all of the other epistemic properties mentioned—by considering scenarios involving *multiple lotteries* (i.e. I do not know that I won't win lottery 1, lottery 2, or lottery 3). As John Hawthorne explains, our intuitions about multiple lotteries may *initially* differ from those we have about more familiar cases involving a single lottery, but these are hard to defend when subjected to sustained reflection:

It is relatively easy to get ourselves in the frame of mind where we reckon ourselves to know that we will not win the New York State lottery each of the next thirty years (even if we expect to buy a ticket each year). Just ask people. They will happily claim to know that that will not happen. Now of course, with a little cognitive effort, that attitude can be disrupted. Suppose, using normal statistical calculations, the chance of winning the New York Lottery each of the next thirty years was 1 in  $n$ . We might point out to someone that if he had a ticket in one great big lottery with  $n$  tickets, he would not reckon himself able to know he would lose in that case. Intuitions would then switch. [Hawthorne 2003: 20]

Hawthorne's point can be seen even more clearly if we compare huge single lotteries with repeated iterations of smaller lotteries. It would be puzzling, if not implausible, if our best epistemological theory had the result that we cannot gain (say) knowledge that we have lost a lottery involving ten million players, but thought that we could know that we *have* lost three consecutive lotteries each only involving 100 players. Entering multiple lotteries of differing sizes *does* provide us with different sources of statistical evidence concerning lottery-relevant propositions such as 'I will never win the lottery' or 'I won't be a millionaire next Friday'. However, from the epistemic perspective, whether we have entered one single large lottery, or multiple lotteries, seems to amount to a distinction without a difference. The apparent epistemic insignificance of distributing the statistical chance of error among multiple events seems to remain true even if we imagine that these chancy events are somewhat different from each other—e.g. 'lottery 1' is a national lottery, 'lottery 2' is the church raffle, and 'lottery 3' is a workplace drawing of lots.

The combination of these two issues—(i) that the size of the lottery does not explain the absence of various epistemic properties in lottery beliefs, and (ii) the apparent insignificance of whether the statistical evidence is contained within one source or split into multiple sources—makes the statistical conjunctions cases a particularly sharp puzzle for epistemic diagnoses of the proof paradox insofar as they are motivated by the guiding comparison with lottery cases.

However, it may be the case that some epistemic theories have greater potential to explain the difference between single and multiple-source statistical evidence cases than others. For example, some approaches, such as the sensitivity theory, seem to have little flexibility in attempting to treat the cases differently. The sensitivity theory requires that we only issue verdicts on the basis of evidence that is such that we would not have issued the verdict had it been mistaken. It is difficult to see how this condition could be met in, for example, *Blue Bus*<sup>3</sup>. Consider the possibility that a *Red Bus* had in fact caused the accident: the evidence would remain unchanged, insensitive to the truth, and still overwhelmingly favour holding the *Blue Company* liable. Other theories involve a formal apparatus that allows them more room to manoeuvre. For instance, the normic justification and safety theories draw on a world-ranking framework in which there is no technical impossibility in claiming that while a verdict based on one piece of statistical evidence is unsafe or normically supported, a verdict based on two (or more) pieces of evidence *is* safe or normically supported. The difficulty for these theories is that they would not regard the analogous belief in a ‘multiple lottery’ scenario to be safe or normically supported. What we are owed is an explanation for why the legal cases outlined in this paper are epistemically different from multiple lottery cases, when the evidence involved in each seems to be fundamentally very similar.

In sum, *all* epistemic theories seem to struggle with statistical conjunction cases. But some may have more theoretical resources with which to respond than others. Given that the rival theories have hitherto been attempting to explain the same data-points—and all claim to be able to do so—considering statistical conjunction cases will be a useful spur to further adjudicate between the merits of competing epistemic theories.

## 4.2 Approaches concerning morality and justice

Diagnoses of the proof paradox concerning morality and justice can be usefully separated into backwards and forward-looking theories.

Let’s begin with the former, the *backwards-looking* views, which roughly suggest that relying on bare statistics violate certain duties that we have to defending parties. These theories, while insightful relative to certain scenarios, do not help us tell apart regular cases from those involving statistical conjunctions.

One influential version of a backwards-looking view originated in the work of David Wasserman, arguing that proof-paradoxical cases fail to respect the *autonomy* of the defending party.<sup>20</sup> Wasserman’s diagnosis was originally aimed at the

<sup>20</sup> Wasserman’s autonomy view has been usefully critiqued and developed by Pundik (2008).

Gatecrasher scenario, which was among the earliest proof-paradoxical cases discussed by legal scholars. On Wasserman's account, treating the individual attendee simply as a member of a reference-class "ignores the defendant's capacity to diverge from his associates or from his past, thereby demeaning his individuality and autonomy" (Wasserman 1992: 942–3). Appealing to autonomy is one way to vindicate the aversion to treating defending parties simply as members of a reference-class; it would be more autonomy-respecting to use evidence that is, in some way, *individualised* to the conduct of the person in question.

It is far from clear whether the autonomy-based story provides a plausible account of the intuitive reluctance to sanction in the unadorned Blue Bus case. For instance, as is pointed out in Pundik (2008: 318), in such cases we bring a legal claim against the bus *company* rather than an individual driver. Given that the company autonomously chose to run  $n$  number of buses on a given route, it is unclear as to how holding them liable on the basis of statistics speaking to their market-share fails to treat the company as an autonomous agent. However, if such a verdict *does* fail to respect the autonomy of the bus company, then it is hard to see how this failure would not also be a feature of attributing liability in the more compelling Blue Bus<sup>3</sup> scenario. After all, by using statistical information about the frequency of certain tyres and hubcaps, we also rely on a reference-class. As such, whichever way the proponent of an autonomy-theory goes on the Blue Bus case, they lack a clear way to distinguish the standard version from the more compelling statistical conjunction variant.

A second type of backwards-looking view, due to Alexander Nunn, argues that we can explain our reluctance to sanction in proof-paradoxical cases by properly considering what is demanded by the right of *due process*.<sup>21</sup> On Nunn's account, a central prohibition that due process requires compatibility in the arguments used to justify imposing sanctions. To make this concrete, consider the prospect of convicting two separate persons of a 'lone gunman' crime. Even if the evidence against the two was, for whatever reason, compelling, there would be something perverse about convicting both of them. Nunn's diagnosis is that the perversity lies in using mutually incompatible theories—after all, both couldn't have been the lone gunman—to convict. The novelty of Nunn's approach is to go further and extend this idea to proof-paradoxical cases, suggesting that the perverse results of a (hypothetical) simultaneous conviction of the entire inculpated population explains why we ought not rely on bare statistics in certain cases. Nunn writes:

[I]f the same naked statistical evidence could be used to convict any randomly selected member of a population, and the simultaneous conviction of the entire population would constitute a due process violation (due to the mutually exclusive nature of the crime) then the conviction of *even one* of those individuals constitutes a due process violation. [Nunn 2015: 1427]

However, as Nunn himself states, the due process defence is unavailable in all proof-paradoxical cases. For example, Nunn explicitly concedes that his due process

<sup>21</sup> Nunn (2015: 1418–1421).

defence cannot explain what is problematic about sanction in the unadorned Blue Bus case. This is because there is no guarantee of error if we imagine iterated Blue Bus cases—there is no factual impossibility (only sheer improbability) in the prospect of being correct every single time if we held a Blue Bus responsible for thousands of accidents on the basis of statistical evidence alone. Hence, as the due process defence cannot explain why we go awry in the regular Blue Bus case, it will not help us discern the difference between it and those cases involving statistical conjunctions—there is equally no guarantee of error in the prospect of iterated sanction in cases such as Blue Bus<sup>3</sup>.

Before moving on I should be clear that both of these backwards-looking approaches may well be good diagnoses of specific scenarios.<sup>22</sup> However, neither Wasserman nor Nunn's view provides an explanation for what is different about cases involving statistical conjunctions from regular proof-paradoxical scenarios.

Let's move on to *forward-looking* theories. The best known forward-looking theory, developed by Enoch et al. (2012), and the one on which I focus, appeals to the role of legal incentives.<sup>23</sup> It is nearly platitudinous that a central role for both civil and criminal law is to incentivise and disincentivise different types of behaviour. From this observation, it is natural to suppose that we can consider the proper verdicts in proof-paradoxical scenarios in light of their (dis)incentivising effects.<sup>24</sup>

To illustrate the incentive approach, consider the Gatecrasher case. The incentive-based approach enjoins us to ask whether the following biconditional is true: 'only if I gatecrash, will I be sanctioned for gatecrashing'. If bare statistics are enough to carry the day, this comes out false—the attendee will be punished, given the existence of enough gatecrashers in the audience, regardless of whether they gatecrash or not. This incentivise structure is perverse insofar as rules against gatecrashing should disincentivise people from doing it. A similar story could be told about the Blue Bus case: were the companies involved aware of their respective market-share, they would be aware that their chance of being held liable in certain cases would be unrelated to their actual conduct.

What does the incentive view say about cases involving statistical conjunctions? In a separate paper, Enoch and Fisher (2015) argue that their incentive-based approach *vindicates* the prevailing legal practice of allowing convictions on the basis of bare DNA evidence but rightly rejects relying on other types of naked statistics. So, their view, at least by their own lights, promises to license sanction in

<sup>22</sup> Although a thought implicit in much of the recent philosophical literature seems to be that providing a *unified* explanation covering all proof-paradoxical cases is a desideratum insofar as the cases all seem structurally similar, for our purposes we need express no view on whether this is correct.

<sup>23</sup> For further development of this view, see Enoch and Fischer (2015) and Enoch and Spectre (2019).

<sup>24</sup> A technical note: Enoch, Spectre and Fisher's incentive view is a *hybrid* view that combines an epistemic condition concerning *sensitivity* with forward-looking consideration to do with incentives. In a nutshell, the idea is as follows: the reason that the law should care about making epistemically sensitive judgements—they explicitly warn against epistemic 'fetishism'—is because it promotes the right sort of legal incentives. Elsewhere, Enoch and Spectre (2019) clarify that the incentive component of the view is central and the epistemic component peripheral—if another epistemic condition turned out to better capture the cases, then it should be substituted for sensitivity.

the Prisoners & DNA case. However, what about the civil law cases? Take a moment to reconsider Blue Bus<sup>3</sup>. Clearly it is conceivable, given that there are only two companies competing for a given locale, that the parties may be aware—if only in very general terms—of which companies tend to use which brand of tyres and hubcaps. Suppose that they were so aware. If this were the case, then the scenario is in fact similar to the unadorned Blue Bus case insofar as the companies involved could be aware that, in a certain class of case, their chance of being held liable would be unrelated to their actual conduct. This would mean that the incentive-view cannot differentiate the unconvincing Blue Bus case from the statistical conjunction variant either.

The natural response is that Blue Bus<sup>3</sup> is simply *such* a *recherché* case that, for all practical purposes, it has no incentivising or disincentivising effect. While this is plausible as far as it goes, relying on this response begs the following question: if we can disregard scenarios such as Blue Bus<sup>3</sup> because they are marginal, should we really accept the idea that the unadorned Blue Bus case is sufficiently common so as to have a substantial incentive-changing effect? Certainly, standard rules about negligent driving incentivise companies to avoid negligence on the pain of sanction—bus companies should encourage their employees to drive carefully to avoid being held liable for accidents. However, do we think that this general incentivising effect is really disrupted by the possibility of proof-paradoxical cases? I am sceptical. We can put the worry as a choice between two options, neither of which is immediately promising for explaining what is different about statistical conjunction cases. Either the relevance of incentives is purely *theoretical*—i.e. not contingent on empirical assumptions about whether the potential for perverse incentive-schemes actually has any substantial behavioural effect—in which case Blue Bus and Blue Bus<sup>3</sup> are on par and the view can't distinguish between them. Or the relevance of incentives is *empirical* and must be plausibly linked to the actual psychology of actors in the relevant scenarios, e.g. the psychology of bus company CEOs deciding on safety standards,—in which case it is doubtful that the possibility of any proof-paradoxical case really has any substantive effect on to how agents tend to act.<sup>25</sup> This is not a knock-down argument against the incentive view, but it shows that it faces some difficult questions—and requires further elaboration—if it is to accommodate cases involving statistical conjunctions.

### 4.3 Likelihood theory

The likelihood theory is a revisionary approach to legal proof that, according to its proponents, has the benefit of explaining our reluctance to sanction in unadorned proof-paradoxical scenarios. The central claim of the likelihood theory is we often fall prey a specific error when thinking about legal proof: namely, we focus on the absolute probabilities of something occurring rather than looking at the *comparative*

<sup>25</sup> This is not to say that a refusal to rely on bare statistics is never significant from the perspective of incentives. For example, it plausibly matters for incentivising certain behaviour whether or not we can use epidemiological statistics to establish employer liability for negligently caused diseases that would be otherwise impossible to prove. See Ross (ms) for further discussion.

likelihood of competing accounts in light of the observed evidence.<sup>26</sup> Refocusing on comparative likelihood rather than absolute probabilities is more faithful to the way that trials are conducted in practice insofar as both parties to a legal dispute typically advance their own version of events—the job of the fact-finder is to compare the relative plausibility of the competing claims. The applicability of the likelihood theory is easiest to see with respect to the civil ‘balance of probabilities’ standard of proof: a claim will be proven on the likelihood approach to proof when, given the observed evidence, it is simply comparatively more likely than the opposing claim. (A formal statement of the increased likelihood requirement applied to the civil standard would run as follows: find for the pursuer iff  $\Pr(E/H^1)/\Pr(E/H^2) > 1$  where  $E$  is the evidence adduced,  $H^1$  is the pursuer’s story, and  $H^2$  is the defender’s story).

How does the likelihood theory help us diagnose the reluctance to sanction in standard proof-paradoxical cases? Take a concrete example, the Blue Bus case. While absolute probabilities seem to favour holding the Blue Company liable, suppose that we instead compare the comparative likelihood, given the cited statistical evidence, of two competing accounts: (i) that it was a Blue Bus, (ii) that it was a Red Bus. The statistical evidence is *equally compatible* with either account—indeed, it would make no difference whatsoever to the underlying statistics whether (i) or (ii) obtained. So, according to proponents of the likelihood theory of proof, merely statistical evidence (compared with, say, an eye-witness report) does not make any substantial difference to the comparative likelihood of either account.

There is much that is insightful in the likelihood approach, but it does not seem to accommodate the thought that sanction is acceptable in cases involving statistical conjunctions. For, suppose that the Blue Company lawyer advanced a very specific rebuttal in the Blue Bus<sup>3</sup> scenario: viz. they suggest that the bus that caused the crash was a Red Company bus that: (1) was on the route at the time, (2) had the implicated hubcap, and (3) had the implicated type of tyres. Indeed, this is precisely the account—perhaps the *only* tenable rebuttal—that a defending lawyer would advance. In this case, the observed evidence is just as compatible with *that* account as it is with the alternative account on which it was a Blue Bus with each of the aforementioned properties. The likelihood theory thus predicts that there is no case against the Blue Company even in cases involving multiple sources of statistical evidence so long as a specific account is available consistent with the statistical evidence. The same feature of the likelihood theory that enables it to deal with familiar proof-paradoxical scenarios render it unable to make accommodation for similar cases involving statistical conjunctions.

#### 4.4 Phase change theory

A final theory, which I call ‘phase change’ theory after a paper due to Cheng and Nunn (2016), argues that statistical evidence involving *very high* probabilities is

<sup>26</sup> See work due to Cheng (2012) and, more recently, in a slightly different vein, by Sullivan (2019). I focus on Cheng’s approach in this section.



different in kind from the much shorter odds found in the unadorned Prisoners, Gatecrasher, and Blue Bus cases.

Cheng and Nunn develop a version of the phase change theory which is neither purely psychological nor predicated on any particular theory of legal proof.<sup>27</sup> Rather, they show that highly probabilifying evidence such a DNA match is mathematically different from regular proof-paradoxical cases: when we perform logistic regression on such evidence we find that, as the chance of an incorrect match transitions from 1 in ten million to 1 in 100 million and below, the chance of error within a given population sharply diminishes. On their view, “[this] phase change justifies treating DNA as different in kind.”<sup>28</sup>

Let’s accept, for sake of discussion, Cheng and Nunn’s mathematical assumptions: at around 1 in ten million chance of error, probabilifying evidence starts to undergo a qualitative change that distinguishes it from statistical evidence involving shorter odds. The ‘phase change’ theory deals straightforwardly with criminal law statistical conjunction cases involving certain types of forensic evidence, because it says that reliance on statistical evidence involving such long odds as we find in DNA profiling cases are different in kind from more familiar proof-paradoxical cases where the odds of error are only in the order of 1/100. However, the phase change theory does not seem to capture the fact that it seems intuitively acceptable to sanction in the civil law statistical conjunction cases discussed earlier. Suppose we take the chance of a ‘random match’ in Blue Bus<sup>3</sup> to be the chance of a given bus happening to be a Red Company bus, operating on the given route, with the implicated tyres, and the implicated hubcaps. The chance of this occurring is only in the order of 1/4000.<sup>29</sup> While the chance of error in Blue Bus<sup>3</sup> is certainly much lower than in the unadorned Blue Bus case, it is nowhere near the 1 in ten million required for the hypothesised ‘phase change’. As such, the phase change theory, although a compelling way to legitimise DNA cases and perhaps other types of forensic evidence, cannot explain the difference between unadorned proof-paradoxical cases and those involving statistical conjunctions below the phase change threshold.

## 5 Error and storytelling

Let’s briefly recap. There is an entrenched hostility to basing legal verdicts on bare statistics within the philosophical and (setting aside DNA evidence) legal literature. This paper has explored cases involving *multiple sources of statistical evidence*. These cases seem to provide a compelling basis for a civil finding of liability and, at least when some of the evidence is forensic, for a criminal conviction. I have argued that four prominent types of theory struggle to explain the relevant difference

<sup>27</sup> My thanks to Edward Cheng for clarificatory correspondence.

<sup>28</sup> 2016: 117.

<sup>29</sup>  $1/4$  (the chance of a red bus on the route) \*  $5/100$  (the chance of a red bus having X) \*  $(2/100)$  (the chance of a red bus having Y).



between cases of multiple-source statistical evidence and more familiar proof-paradoxical scenarios.

I want to close by discussing a final strategy that we might use to explain the difference between the two types of case discussed in this paper: namely, to reject the intuitions to the effect that statistical conjunction cases are relevantly different from regular proof-paradoxical scenarios. Indeed, I think that we may be able to offer a sort of error theory about the apparent difference in our intuitions. However, so I will claim, if we accept this error theory, this raises an even more fundamental worry: it gives us reason to doubt our intuitions about the original proof-paradoxical scenarios too.

A widely discussed paradigm concerning the evaluation of legal evidence is Pennington and Hastie's 'story model'.<sup>30</sup> The story model, focusing on how jurors make decisions, is an empirically supported hypothesis that supposes legal fact-finding to be driven by an exercise in narrative construction. According to the story model, jurors attempt to "impose a narrative story organization on trial information, in which causal and intentional relations between events are central."<sup>31</sup> In a nutshell, legal fact-finders look for *stories* that fit the evidence. When we can readily construct a narrative on the basis of evidence adduced, where necessary drawing supplementary causal inferences in order to do, and when that story is inculpatory, we are primed to find in favour of sanctioning the defending party. In a related and complementary stream of research, psychologists who have examined the 'Wells' effect'—the phenomenon of juror reluctance to rely on bare statistics, named after Gary Wells' and collaborators' initial study—have tested different psychological explanations for this reluctance. One explanation appeals to 'ease of simulation': legal fact-finders prefer sanctioning on the basis of non-statistical evidence to statistical evidence because the former makes it more demanding to construct mental scenarios indicating innocence, and vice versa concerning guilt.<sup>32</sup> For example, a descriptive eye-witness is more evocative than a mere statistical report, and it is harder to imagine scenarios inconsistent with an eye-witness account than with a bare statistical report.

This empirical research naturally suggests a reason for why cases involving statistical conjunctions seem more compelling to us, namely because: (i) they generate an easy-to-simulate inculpatory narrative in the mind of the assessor, and (ii) they make it more demanding to simulate an exculpatory narrative. Let me explain.

Regular proof-paradoxical cases are generally lacking in narrative structure: e.g. they simply tell us that it was likely that the bus was blue, or that the gatecrasher lacked a ticket, and that is all there is to it. Conversely, the extra complexity of statistical conjunction cases makes them more evocative, and require us to take more steps to imagine that the defending party was not at fault. Take Blue Bus<sup>3</sup>. In

<sup>30</sup> See Pennington and Hastie (1991, 1992, 1993) for a representative sample of papers outlining both the theoretical and empirical support for the story model.

<sup>31</sup> *Ibid.* (1992: 189).

<sup>32</sup> See Wells (1992) for the original study and Neidermeier et al. (1999) for further empirical work in support of the ease of simulation hypothesis.

the version discussed, there is statistical evidence concerning not only the frequency of buses on a route, but also concerning a hubcap found at the scene, and tyre marks found on the road. This is ripe for the imposition of a narrative structure; it is easy to imagine a blue bus causing a crash, *that bus* making marks on the road, and a hubcap *belonging to that bus* coming off and rolling into the bushes. It is easy to imagine a nexus of physical causality that explains all of the adduced evidence just by imposing a temporal narrative structure onto the evidence. Moreover, rejecting that narrative takes more cognitive effort than the original Blue Bus case. We must assume that not only has one improbable coincidence occurred, but three, or more, at different points in the story. This, I think, makes for a central difference between cases involving statistical conjunctions and those involving only one source of evidence. We treat the former as a compelling story, the rejection of which would require the effortful adoption of a sceptical stance at multiple points of the narrative. This makes error seem like a much more distant possibility than in cases involving single-source statistical evidence, and primes us to find against the defending party.

This amounts to an entirely *psychological* diagnosis of why we are more inclined to find sanction acceptable in statistical conjunction cases. The harder question is this: what is the *normative* relevance of this psychological account? I suggest that our psychological diagnosis can facilitate an argument for a sort of error theory, i.e. the claim that our intuitions about statistical conjunction cases should not be trusted. Suppose we begin, naturally enough, from a veritistic perspective—what relevance does narrative and ease of simulation have for getting to the truth? Certainly, the ease of imposing narrative structure onto evidence and simulating an exculpatory scenario can be a helpful heuristic for truth. The mind has a general facility for discerning the plausible from the implausible. However, in other cases, the role of evocative narratives in generating judgements of plausibility and implausibility can amount to a cognitive bias. Take the intuitive shiftiness noted by Hawthorne earlier, that people are initially more inclined to say that they know they won't win multiple smaller lotteries but not that they won't win a single giant lottery where the probability of winning is identical to winning multiple times in a row. The prospect of winning the lottery multiple times in a row is an evocative and incredible-sounding story—it appears implausible to the imagination, and involves accepting that multiple incredible things will happen in a row. So, we are apt to discount the possibility and reject the attribution of knowledge. But, from the perspective of veritistic judgement, we should evidently not treat that possibility any differently from the judgement about losing the single larger lottery. To do otherwise is to be led astray by the imagination and fall prey to a type of bias. Framed this way, we might worry that the same bias is manifesting itself when we find statistical conjunction cases more persuasive than those involving a single source of evidence.

While these thoughts are somewhat preliminary, we can see the outlines of an error theory that would recommend the following conclusion: to the extent that intuitions about statistical conjunction cases differ from unadorned proof paradoxical scenarios, they should not be trusted—rather, if we are concerned with accuracy, we would be better off discounting the fact that the statistical evidence is dispersed over multiple sources. If we accepted such a theory, then one might suppose that the theories discussed earlier in the paper actually face no issues with

statistical conjunction cases. Indeed, we might suppose that it is actually a mark in their favour if they treat single and multiple-source bare statistics on par. According to this error theory, once we understand the aetiology of the intuitions involved, we should not be perturbed about refusing to sanction even in cases involving the conjunction of many different types of statistical evidence.

*However*, this error theory is in fact double-edged. The same psychological diagnosis that undermines intuitions about statistical conjunction cases can also undermine intuitions about the *original* proof-paradoxical scenarios. A standard lesson drawn from the proof paradox is that there is a disfavourable contrast between bare statistical evidence (bad) and familiar non-statistical evidence such as eye-witness testimony (good). It is easy to see why relying on the error theory outlined above is problematic for those who wish to uphold our intuitions about the original proof paradox. Unlike bare single-source statistics, eye-witness accounts are narratively rich and it is cognitively demanding to generate consistent exculpatory accounts. The very same idea, that we should not allow the ease of simulation heuristic to lead us astray when evaluating evidence from a veritistic perspective, can also be used to call into question the disfavourable contrast between statistical and non-statistical evidence. So, to the extent that an error theory can challenge intuitions about cases involving statistical conjunctions, the intuitions about the original proof paradox are open to challenge on the same basis.

To close, it is worth pointing out that one might attempt to find some consideration that could vindicate treating narratively rich evidence differently *even if* our intuitions about it are driven by a potentially bias-inducing heuristic. One easy but perhaps not altogether satisfying reason for the law to prefer narratively rich evidence follows from the importance of the perception of fairness in the law. One role for any legal system is to inspire public confidence in its workings. This explains, for example, why cases can be overturned for being decided in an *apparently* biased way even if it is accepted that there was in fact no actual bias tainting the decision.<sup>33</sup> If we can simply note that a normal observer *does* find sanction reasonable in statistical conjunction cases but not in unadorned proof-paradoxical scenarios, then this creates its own reason to treat the cases differently. Appearances matter in the law, even if we cannot provide an underlying theory to vindicate the difference between the unadorned proof paradoxical cases and those involving statistical conjunctions. This is perhaps a somewhat conservative and quietist response to the puzzle outlined in this paper. If we want a better solution, we must throw down the gauntlet to those hostile to bare statistics—what *is* different about cases involving multiple sources of statistical evidence?<sup>34</sup>

<sup>33</sup> One of the most striking examples in modern times is *R v Bow Street Metropolitan Stipendiary Magistrate, ex parte Pinochet Ugarte (No 2)* [2000] 1 AC 147. The extradition of Augusto Pinochet from the United Kingdom to Spain was set aside because the spouse of one of the judges had links to Amnesty International, an intervener in the case.

<sup>34</sup> Anonymous reviewers and Dario Mortini provided instructive written feedback on this paper—my thanks to them.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Backes, M. (2019). Epistemology and the law: Why there is no epistemic mileage in legal cases. *Philosophical Studies*. <https://doi.org/10.1007/s11098-019-01337-6>.
- Blome-Tillmann, M. (2017). 'More likely than not'—Knowledge first and the role of statistical evidence in courts of law. In G. Carter & B. W. Jarvis (Eds.), *Knowledge first—Approaches in epistemology and mind* (pp. 278–292). Oxford: Oxford University Press.
- Blome-Tillmann, M. (forthcoming). Statistical evidence, normalcy, and the gatecrasher paradox. *Mind* 129(514), 563–578.
- Bolinger, R. J. (2018). The rational impermissibility of accepting racial generalizations. *Synthese*. <https://doi.org/10.1007/s11229-018-1809-5>.
- Buchak, L. (2014). Belief, credence, and norms. *Philosophical Studies*, 169(2), 1–27.
- Cheng, E. K. (2012). Reconceptualizing the burden of proof. *Yale LJ*, 122, 1254.
- Cheng, E. K., & Nunn, G. A. (2016). DNA, Blue Bus, and phase changes. *The International Journal of Evidence & Proof*, 20(2), 112–120.
- Cohen, L. J. (1977). *The Probable and the Provable*. Oxford: Clarendon Press.
- Di Bello, M. (2019). Trial by statistics: Is a high probability of guilt enough to convict? *Mind*, 128, 1045–1084.
- Di Bello, M. (forthcoming). Proof Paradoxes and Normic Support: Socializing or Relativizing? *Mind* fzz021.
- Di Bello, M., & O'Neil, C. (2020). Profile evidence, fairness, and the risks of mistaken convictions. *Ethics*, 130(2), 147–178.
- Ebert, P. A., Smith, M., & Durbach, I. (2018). Lottery judgments: A philosophical and experimental study. *Philosophical Psychology*, 31(1), 110–138.
- Enoch, D., Spectre, L., & Fisher, T. (2012). Statistical evidence, sensitivity, and the legal value of knowledge. *Philosophy & Public Affairs*, 40(3), 197–224.
- Enoch, D., Spectre, L., & Fisher, T. (2015). Sense and 'sensitivity': Epistemic and instrumental approaches to statistical evidence. *Stanford Law Review*, 67, 557.
- Enoch, D., Spectre, L., & Fisher, T. (2019). Sensitivity, safety, and the law: Reply to Pardo. *Legal Theory*, 25, 178–199.
- Evetts, I., & Williams, R. (2015). A review of the sixteen point fingerprint standard in England and Wales. *Journal of Forensic Identification*, 65(4), 557–580.
- Gardiner, G. (2018). Legal burdens of proof and statistical evidence. In D. Coady & J. Chase (Eds.), *The Routledge handbook of applied epistemology*. London, England: Routledge.
- Gardiner, G. (2020). The reasonable and the relevant: Legal standards of proof. *Philosophy & Public Affairs*, 47, 288–318.
- Haack, S. (2014). *Evidence matters: Science, proof, and truth in the law*. New York: Cambridge University Press.
- Hawthorne, J. (2003). *Knowledge and lotteries*. Oxford: Oxford University Press.
- Hedden, B., & Colyvan, M. (2019). Legal probabilism: A qualified defence. *Journal of Political Philosophy*, 27, 448–468.
- Krauss, S. (2020). Against the alleged insufficiency of statistical evidence. *Florida State University Law Review*, 47. <https://ssrn.com/abstract=3440138>.

- Littlejohn, C. (2018). Truth, knowledge, and the standard of proof in criminal law. *Synthese*. <https://doi.org/10.1007/s11229-017-1608-4>.
- Moss, S. (2016). *Probabilistic knowledge*. Oxford: Oxford University Press.
- Moss, S. (forthcoming). Knowledge and legal proof. In *Oxford studies in epistemology* (Vol. 7). Oxford: Oxford University Press.
- Niedermeier, K., Kerr, N., & Messe, L. (1999). Jurors' use of naked statistical evidence: Exploring bases and implications of the Wells effect. *Journal of Personality and Social Psychology*, 76(4), 533–542.
- Nunn, A. (2015). The incompatibility of due process and naked statistical evidence. *Vanderbilt Law Review*, 68(5), 1407.
- Papineau, D. (2020). The disvalue of knowledge. *Synthese*. <https://doi.org/10.1007/s11229-019-02405-4>.
- Pardo, M. S. (2018). Safety vs. sensitivity: Possible worlds and the law of evidence. *Legal Theory*, 24, 50–75.
- Pardo, M. S. (2019). The paradoxes of legal proof: A critical guide. *Boston University Law Review*, Vol. 99, 2019; U of Alabama Legal Studies Research Paper No. 3293023.
- Pennington, N., & Hastie, R. (1991). A cognitive theory of juror decision making: The story model. *Cardozo Law Review*, 13, 519.
- Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology*, 62(2), 189–206.
- Pennington, N., & Hastie, R. (1993). The story model for juror decision making. In R. Hastie (Ed.), *Inside the juror: The psychology of juror decision making* (pp. 192–221)., Cambridge series on judgment and decision making Cambridge: Cambridge University Press.
- Pritchard, D. (2015). Risk. *Metaphilosophy*, 46(3), 436–461.
- Pritchard, D. (2018). Legal risk, legal evidence and the arithmetic of criminal justice. *Jurisprudence*, 46(3), 436–461.
- Pundik, A. (2008). Statistical evidence and individual litigants: A reconsideration of Wasserman's argument from autonomy. *The International Journal of Evidence & Proof*, 12(4), 303–324.
- Redmayne, M. (2008). Exploring the proof paradoxes. *Legal Theory*, 14(4), 281–309.
- Ross, L. D. (2020a). Recent work on the proof paradox. *Philosophy Compass*. <https://doi.org/10.1111/phc3.12667>.
- Ross, L. D. (2020b). Rehabilitating statistical evidence. *Philos Phenomenol Res*. <https://doi.org/10.1111/phpr.12622>.
- Roth, A. (2010). Safety in numbers—Deciding when DNA alone is enough to convict. *NYU Law Review*, 85, 1130.
- Smith, M. (2016). *Between probability and certainty: What justifies belief*. Oxford: Oxford University Press.
- Smith, M. (2018). When does evidence suffice for conviction? *Mind*. <https://doi.org/10.1093/mind/fzx026>.
- Smith, M. (2020). Against legal probabilism. In J. Robson & Z. Hoskins (Eds.), *Truth and trial*. Abingdon: Routledge.
- Sullivan, S. P. (2019). A likelihood story: The theory of legal fact-finding. *University of Colorado Law Review*, 90, 1.
- Tribe, L. (1971). Trial by mathematics: Precision and ritual in the legal process. *Harvard Law Review*, 84(6), 1329–1393.
- Wasserman, D. T. (1992). The morality of statistical proof and the risk of mistaken liability'. *Cardozo Law Review*, 13, 935.
- Wells, G. L. (1992). Naked statistical evidence of liability: Is subjective probability enough? *Journal of Personality and Social Psychology*, 62, 739–752.
- Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford University Press.
- Zabell, S. L. (2005). Fingerprint evidence. *Journal of Law and Policy*, 13(1), 143–179.
- Zhu, Y., Dass, S. C., & Jain, A. K. (2007). Statistical models for assessing the individuality of fingerprints. *IEEE Transactions on Information Forensics and Security*, 2(3), 391–401.