

Sensitivity, Safety, and Admissibility

1.

This paper concerns recent attempts to use the epistemological notions of sensitivity and safety to shed light on legal debates about so-called “bare” statistical evidence. These notions might be thought to explain either the outright inadmissibility of such evidence or its inadequacy for a finding of fact – two different phenomena that are often discussed in tandem, but that, I insist, we do better to keep separate. I argue that neither sensitivity nor safety can hope to explain statistical evidence’s inadmissibility, since neither offers a plausible criterion of admissibility that would exclude such evidence. The possibility remains that either notion might explain statistical evidence’s inadequacy for a finding of fact; I express some doubts about this possibility, but do not rule it out.

2.

2a. Statistical evidence

It is a starting assumption in the literature on so-called “bare” statistical evidence that courts – both civil and criminal – and regular people – that is, people who have nothing in particular to do with the law – are usually strongly averse to findings of fact based on this sort of evidence.¹² A great many examples have illustrated this phenomenon. In one of the most famous ones (see Nesson 1979), grainy CCTV footage of a prison yard shows 24 out of 25 prisoners teaming up to assault a guard while the one remaining prisoner stands idly by, but the CCTV footage is too grainy to see prisoners’ faces and thus identify the one who did not commit assault.³ For any randomly-selected prisoner who was in the yard that day, the probability that they committed assault given this evidence is 24/25, or 0.96. Nonetheless, most people have the strong intuition that there would be something deeply wrong with convicting a randomly-selected prisoner on this basis (see Wells 1992 for a study of “folk” intuitions in cases like this). And intuitions typically remain strong even if the numbers get arbitrarily high: if we imagine that the number of prisoners who committed assault is not 24 out of 25 but 99 out of 100, or 999 out of 1,000, or what have you, most people still feel that there would be something deeply wrong with convicting a randomly-selected prisoner on the grounds that the evidence makes it overwhelmingly likely that they committed assault.

¹ Not everyone accepts this starting assumption: Ross (2021) and Allen (2021), for instance, reject it as misguided. And there is scope for denying that the case history widely taken to support the assumption genuinely does support it. Ross and Allen both argue that, in the most famous cases in the literature, careful scrutiny suggests that the problem with the evidence was never supposed to be that it was statistical in nature but merely that it was insufficiently strong.

But it is this assumption on which the literature on statistical evidence is founded.

² The practice of using the terms “bare” or “naked” to refer to cases involving statistical evidence but no other evidence originates with Kaye (1980).

³ Nesson’s original case involves a far-away eyewitness rather than grainy CCTV footage. But this creates some noise, given that we might wonder whether the eyewitness is reliable as to the number of prisoners in the yard and what each of them did. Modifying the example as in the main text eliminates this noise.

These intuitions are curious, as a great many scholars have remarked, because we are often comfortable with findings of fact based on other evidence that seems to leave the probability of the disputed fact much lower. The standard contrast in the literature is between statistical evidence and eyewitness testimony. For instance, suppose that an eyewitness testifies that she saw a certain person committing assault, and suppose she is tested and found to be 95% reliable when it comes to identifying faces – that is, when she identifies a face as a certain one that she has seen before, 95% of the time it is indeed that very face. This evidence might suffice for conviction. And that is so despite the fact that it seems to be less probabilifying than any of the statistical evidence just discussed: holding fixed the prior probability of guilt, it would seem that updating on the testimony of a 95%-reliable eyewitness should leave the posterior probability lower than updating on a 96% or 99% or 99.9% statistic.⁴

One might think that our intuitions in the prison yard case reveal something about the standard of proof in criminal courts. And cases like this have indeed been used to argue that the state of being convinced “beyond reasonable doubt” resists analysis in probabilistic terms.⁵ But what appears to be the same phenomenon arises also in civil courts, where plaintiffs are only expected to prove their cases “on the balance of probabilities” or “on the preponderance of the evidence”.⁶ Here the *locus classicus* is *Smith v. Rapid Transit Inc.* In *Smith*, a poorly-driven bus caused a road accident late at night, with no eyewitnesses. The victim brought a suit against the company that owned almost all of the buses operating in the area of the accident, reckoning that this fact alone made it more likely than not that this company owned the bus that caused the accident. Mathematically speaking, that makes sense: one would think that if a certain company owns, say, 75% of the buses operating along the route of the accident, then the probability that it owns the bus that caused the accident is 0.75. But this case was ruled by directed verdict for the defendant and the verdict was upheld at appeal, with the Supreme Judicial Court of Massachusetts opining that “the ownership of the bus was a matter of conjecture” and that the fact that “the mathematical chances somewhat favor the proposition that the bus of the defendant caused the accident” was “not enough”.

All of the above concerns the *adequacy* of statistical evidence for a finding of fact. A separate issue concerns the *admissibility* of statistical evidence in court. Evidence that does raise the probability of a defendant’s guilt or liability is sometimes nonetheless ruled inadmissible, meaning that it may not be brought to trial at all. The rules of admissibility are a motley crew serving diverse purposes: some aim to ensure fairness in the fact-finding process (such as the prohibition on evidence likely to arouse strong sympathy or prejudice), while others aim to avoid creating perverse incentives elsewhere in the criminal justice system (such as the prohibition on evidence obtained illegally), and so forth. As well as being deemed insufficient for a verdict, then, statistical evidence might for some reason be deemed altogether inadmissible. And, in fact, courts’ rulings on the admissibility of statistical evidence have historically been profoundly mixed;

⁴ I do not actually accept this, for reasons given in [redacted]. But it is another starting assumption of the literature, and one that I leave unquestioned for the purposes of this paper.

⁵ For instance, Allen (1990) argues that the prison yard case shows that the criminal standard of proof requires a state of “subjective certainty” that cannot be expressed probabilistically.

⁶ Here, too, some think that our intuitions about statistical evidence tell us about the nature of legal proof. For example, Cohen (1977) takes a case that is structurally similar to the prison yard case (but involves a civil rather than a criminal violation) to show that legal verdicts cannot be calculated using standard probability theory, offering an alternative that he calls “Baconian” probability. And Nesson (1985, p.22) suggests that a case like *Smith* shows that the acceptability of a verdict “is not a simple function of mathematical probability, but rather is a complex matter of communication that depends on the nature of the issue, the process of decision, and the purposes and audiences the conclusion serves”.

some judges have deemed such evidence inadmissible, while others have been happy to allow it, with precious few discernible patterns emerging (see Koehler 2002 for an overview).⁷

This means that epistemologists interested in legal debates about statistical evidence have *two* putative explananda: the intuition that statistical evidence is an inadequate basis for a finding of fact, and the stronger intuition that it is (at least sometimes) inappropriate for courts to consider such evidence at all. Some in the literature adopt the goal of explaining statistical evidence's inadequacy, while others adopt the goal of explaining its inadmissibility, and some are noncommittal, as we will soon see. But it does matter what our explanandum is, since, obviously, different phenomena can have different explanations. Here I focus on admissibility, arguing that sensitivity and safety cannot explain statistical evidence's inadmissibility even in those cases in which it has indeed been deemed inadmissible.

2b. Sensitivity and safety

That was a primer on statistical evidence. Now, here's a primer on sensitivity and safety.

Think of someone with a belief. Consider the *basis* for this person's belief — her evidence for it, or the reason for which she holds it. To assess sensitivity, we ask: given that she holds this belief on this basis, would she still have held it even if the proposition she believes had been false? If the answer is "no", then her belief is sensitive to the truth. If the answer is "yes", then her belief is insensitive. To assess safety, by contrast, we ask: given that she holds this belief on this basis, could she easily have had a false belief? If the answer is "no", then her belief is safe from error. If the answer is "yes", then her belief is unsafe.⁸

Sensitivity and safety look similar. But they are quite different. The difference arises in light of the fact that someone who actually has a true belief usually would still believe the proposition under *some* changed circumstances in which it is false, but not others. For example, I believe that my mum just got her booster dose of the AstraZeneca vaccine. My belief is based on her testimony. One way for it to be false that my mum just got her booster dose is for her appointment to have been delayed, in which case she would not have told me that she got her booster dose and I would not now believe it. But another way for it to be false that my mum just got her booster dose is for her to be an anti-vaxxer involved in an elaborate hoax designed to get people to believe that their loved ones are vaccinated when they aren't. In this case, she would still have told me that she got her booster dose. So I would still believe this (now false) proposition, and would believe it on the same basis as I actually do, namely my mum's testimony.

⁷ One of Koehler's observations is that even courts' rulings about the admissibility of base rates intended to establish presumptive ownership have been mixed. There is *Smith*, of course, as we all know. And, in *Guenther v. Armstrong Rubber Co.*, 406 F.2d 1315, 1318 (3rd Cir. 1969), the court ruled against a plaintiff who was injured at work by a faulty tire and sued the manufacturer that makes 75% of the tires used by his company, on the grounds that, based on this evidence alone, a verdict for the plaintiff "would at best be a guess" and "could not be reasonably supported". This is very similar to the ruling in *Smith*. But what appears to be the exact opposite opinion was returned in *Kaminsky v. Hertz Corp.*, 288 N.W.2d 426 (Mich. Ct. App. 1979), in which the plaintiffs were taken to have established presumptive ownership of a truck (from which a chunk of ice had fallen and injured them) on the grounds that the truck bore a Hertz logo and Hertz owns 90% of the trucks that bear its logo. And in *Kramer v. Weedhopper of Utah, Inc.*, 490 N.E.2d 104 (Ill. App. Ct. 1986), the court ruled that the fact that a certain company supplied 90% of the bolts used by Weedhopper's aircraft kits was sufficient to permit the inference that they had supplied the bolt that sheared off the defendant's aircraft and injured him. It is very difficult to square *Smith* with *Kaminsky* and *Guenther* with *Kramer*.

⁸ For the canonical statement of sensitivity, see Nozick (1981). For the canonical statement of safety, see Sosa (1999). For a detailed comparison of the two, see Pritchard (2008).

Epistemologists deal with these issues by thinking about the *closeness of possible worlds* — that is, alternative ways the world could have been, in contrast to how it actually is. We can order possible worlds by their “closeness” to the actual world, where the “distance” of a world is a matter of how different it is from the actual world. Possible worlds in which my mum’s appointment was delayed are quite close, since things would not have had to have been very different for this to be the case; vaccine rollout is patchy in her area and NHS delays are very common. But, happily, worlds in which my mum is an anti-vaxxer involved in an elaborate hoax are remote worlds, since things would have had to have been very different for that to be the case; my mum is ideologically quite unlike the anti-vaxxers and such a hoax would be a bizarre thing to participate in even if she was one. By considering degrees of similarity to the actual world in this manner, we can determine the relative closeness of different possible worlds in which a proposition that someone believes is false.

Now we can be more precise about sensitivity and safety. To assess sensitivity, we need to see whether the agent would still believe the proposition if it were false. We do this by looking at the *closest possible world* – or set of jointly closest worlds, if there is a tie – in which the proposition in question is false. We see if the agent still believes the proposition in this world (or all of these worlds, in the case of ties). If not, her belief is sensitive to the truth. If so, her belief is insensitive. To assess safety, by contrast, we need to see whether the agent could easily have had a false belief. This idea of “easiness” is typically also understood in terms of the closeness of possible worlds: we divide the alternative possibilities into those that are close enough to the actual world to worry us – those that could “easily” have been the case – and those that are too remote for us to be bothered by what happens in them. We then see whether the agent still believes the proposition in any close worlds in which it is false. If not, her belief is safe from error. If so, her belief is unsafe. The key difference between sensitivity and safety is therefore that sensitivity directs us to look at the *closest* worlds in which the proposition that the agent believes is false, regardless of how close or remote these worlds may be, and to see whether she still believes it in these worlds, whereas safety just requires us to scan all of the *close* worlds to see whether any of them are worlds in which the proposition is false but the agent still believes it, and is uninterested in what happens at remote worlds.⁹

2c. *Joining the dots*

Some legal scholars have suggested that the notions of sensitivity and safety explain widespread suspicion of statistical evidence.

This started with a paper by David Enoch, Talia Fisher, and Levi Spectre (2012). These authors make the very weak assumption that “something in [sensitivity’s] vicinity is of genuine epistemological importance”, which they understand to mean that “with other things held equal, a sensitive belief is epistemically better than an insensitive one” (p.205). They then argue that “[by] relying on statistical evidence, we render our findings insensitive” (p.207). Their thought here is that if, for example, another company’s bus had caused the accident in *Smith*, then Rapid Transit would still have owned the majority of buses operating in the area. If that is correct, then the market share data is insensitive to the truth of the verdict that it supports. So, the authors reckon, if we are willing to find liability based on market share data alone then we would

⁹ Safety is sometimes introduced as the contrapositive of sensitivity: it is said that while sensitivity requires that the agent does not hold the belief if it is false, safety requires that it is true if she does hold it (Sosa 1999, p.142). (Note: Sosa offers multiple formulations of safety in this initial article, including one introducing the notion of “easiness”.) Testing the truth of the conditional “If the agent believes the proposition, then it is true” requires us to look at the closest worlds in which the agent still believes the proposition and make sure that none of them are worlds in which it is false.

still have found Rapid Transit liable even if the accident had been caused by another company's bus. If we base our verdict solely on insensitive statistical evidence, then, our verdict will also be insensitive.¹⁰

Michael Pardo (2018) criticizes Enoch, Fisher and Spectre's account and argues that it is not sensitivity but rather safety that explains distrust of statistical evidence. Pardo holds that "[e]pistemic safety is important for legal evidence and... plays significant roles in the law of evidence... [but sensitivity] does not (and should not) play much of a role" (p.4). (We will see why he thinks this in §3a.) Pardo also argues that safety explains our intuitions in *Smith* just as well as sensitivity. His explanation is that "[e]yewitness evidence will be safer when there are close possibilities in which similar accidents are caused by [other companies'] buses" (p.21). The thought here, I gather, is that the possible worlds in which another company's bus causes the accident but the market share data remain the same are close worlds — this could easily have happened. If that is correct, then the statistical evidence and any finding of liability on its basis is unsafe. Moreover, Pardo suggests, there is a sense in which statistical evidence is *less* safe than eyewitness testimony: worlds in which another company's bus causes the accident but an eyewitness mistakenly identifies it as a Rapid Transit bus are more remote than worlds in which another company's bus causes the accident but all else (including market share data) remains exactly the same. If that is correct, then the market share data is in a sense *less* safe than the eyewitness testimony because it leads us astray in closer worlds.

Now, recall that, when someone says that sensitivity or safety explains our distrust of statistical evidence, there are different explananda that she might have in mind. One is the intuitive inadequacy of statistical evidence for a finding of fact — the sort of intuition we have when we judge that there would be something wrong with finding guilt in the prison yard case or liability in *Smith*. Another is the stronger intuition that bare statistical evidence should not even be admissible — that it is not appropriate for courts to consider such evidence at all, regardless of their ultimate verdict.

It is not uncommon for authors in the literature to flip back and forth between talk of admissibility and talk of adequacy, as if nothing much hangs on the distinction. Meanwhile, some authors of an epistemological bent focus squarely on adequacy, exploiting the analogy between legal verdicts and individual beliefs (see e.g. Smith 2018, Littejohn 2020; and cf. Lewis 2021 for criticism of this methodology). But, curiously, those who have most extensively defended sensitivity- and safety-based approaches to the puzzles of statistical evidence have *not* focused on adequacy. Indeed, Enoch, Fisher and Spectre are explicitly non-committal as to the nature of their explanandum. Their initial statement is as follows (pp.197-198):

[T]he law typically will not be willing to base a positive finding of fact — and certainly not liability — on just... statistical evidence. Indeed, in most jurisdictions it is not even clear that such evidence would be considered admissible or relevant.

And, shortly afterwards, they say the following (p.199, emphases original):

Note that we will not here attempt a vindication of *some specific way* of discriminating between statistical and individual evidence (say, in terms of admissibility). The initial challenge is to vindicate *any distinction at all*, to show any reason at all to treat statistical evidence more suspiciously than individual evidence.

¹⁰ N.B. Enoch, Fisher and Spectre think that this epistemological story explains our *aversion* to statistical evidence, but not *what is wrong* with statistical evidence. They do not think that their epistemological story is vindicatory, as they think the law should not care about knowledge, and they offer an alternative vindicatory story (*op. cit.* pp.210-223).

This temporizing gives Enoch, Fisher and Spectre considerable breadth. They could see themselves as successful if they provide a rationale for statistical evidence's outright inadmissibility, as suggested in their parenthetical remark. But they could also see themselves as successful if they only provide a rationale for statistical evidence's inadequacy. And, indeed, they need not even attain this weaker goal. Their stated aim is simply to identify *a* reason — one reason — to be more suspicious of statistical evidence than other evidence. (It is consistent with there being such a reason that it is often, or even always, defeated or outweighed, such that statistical evidence is neither inadmissible nor inadequate all-things-considered.)

Pardo is more ambitious. He notes (p.13, fn.59) the difference between questions of adequacy and questions of admissibility. And he goes for the latter, stating that his case against the legal relevance of sensitivity is based on “three main reasons: (1) some statistical evidence is both sensitive and admissible; (2) some statistical evidence is admissible despite being insensitive; and (3) several types of [non-statistical] evidence are also insensitive and frequently admissible” (p.14). Pardo's point is that sensitivity is neither necessary nor sufficient for admissibility and thus cannot do the explanatory work cut out for it in the literature. This assumes, of course, that the work for which it is cut out is to explain the intuitive inadmissibility of bare statistical evidence.¹¹ Pardo argues further that focusing on adequacy for a finding of fact can be a “red herring” as “no single item of evidence is typically sufficient” for a finding all by itself — not even confessions (p.15, fn.70). I take it that this is why he prefers to focus on admissibility.

Enoch and Fisher (2015) have also written a companion piece, aimed more squarely at a legal audience, in which their focus is explicitly on admissibility. In this paper, describing their explananda, they say that “[i]n the first half of the twentieth century, when statistical evidence first began to appear in court, many judges responded antagonistically, deeming it inadmissible and devoid of any probative value” (p.561), but that “there are also exceptions to the general approach of inadmissibility in certain categories of cases”, such as DNA evidence (p.562). By contrast, in Enoch and Spectre's reply to Pardo, they open by observing that courts happily *find liability* based on eyewitness evidence but not statistical evidence, and stating that “the problem of statistical evidence is to explain — and, if possible, vindicate — this distinction” (2019, p.180). This concerns inadequacy rather than inadmissibility. My guess is that these authors have assumed that it does not much matter which explanandum they take on. I disagree.

In the next three subsections I will argue that, *pace* Enoch et al. and Pardo, neither sensitivity nor safety can plausibly explain statistical evidence's inadmissibility. That is because neither notion furnishes us with a plausible criterion of admissibility that would exclude such evidence. However, issues of inadmissibility and of inadequacy should not be indiscriminately plonked together, and my arguments leave open the possibility that either sensitivity or safety might explain statistical evidence's inadequacy for finding fact.¹² To be clear: I do not think that either notion does explain inadequacy, and I will raise some doubts (which I do not take to be conclusive) about that idea in the paper's final section. But I do think that, when the two issues are teased apart, it becomes easy to see just how little there is to be said for sensitivity and safety as putative explanations of statistical evidence's inadmissibility — or, for that matter, of the inadmissibility of anything else.

¹¹ Spelling this out in further detail in another footnote, Pardo says that “the key point is that neither the statistical-individual distinction nor the admissibility-inadmissibility distinction can be explained in terms of the sensitivity-insensitivity distinction. None of these three distinctions (sensitivity-insensitivity, statistical-individual, admissible-inadmissible) can be explained in terms of another” (p.16, fn.77).

¹² They also leave open that statistical evidence is indeed sometimes — or even often — inadmissible, there being some other explanation of this fact. What I argue is that sensitivity and safety cannot explain inadmissibility, not that nothing can explain inadmissibility or that sensitivity and safety cannot explain anything.

3.

3a. Counterexamples

Much of the philosophical and legal literature on sensitivity, safety, and statistical evidence up to this point has been based on discussion of alleged counterexamples. That is because, for a criterion to explain either adequacy or admissibility, it must at least be *extensionally correct*: it must divide adequate from inadequate or admissible from inadmissible evidence in intuitively plausible ways. However, most counterexamples in the literature pertain to adequacy rather than admissibility. For instance, Michael Blome-Tillmann's "opportunistic gatecrasher" and "first of the gatecrashers" (2015, pp.105-107) concern sensitive evidence that is intuitively inadequate, while Martin Smith's mistaken and devious eyewitnesses (2018, pp.1201-1203) and Lewis Ross's discussion of "cold-hit" DNA evidence (2021, pp.8-9) concern insensitive evidence that is intuitively adequate. These examples do not address issues of admissibility. Evidence cannot suffice for a verdict if it is not even admissible, of course, so intuitions of adequacy do bear indirectly on admissibility. But they do not bear directly; they are not intuitions whose content is that certain types of evidence should or should not be admitted to court. And intuitions of *inadequacy* tell us nothing about admissibility, since evidence may be admissible and yet inadequate.

Pardo's paper is an exception here, as his discussion focuses explicitly on admissibility. He, too, proceeds by counterexample: Pardo offers one example of inadmissible sensitive evidence and several examples of insensitive admissible evidence. The first is a case that he calls "Bad Lab" (p.18): a corrupt forensics lab is willing to testify to a DNA match whenever police tell the lab that their sample comes from a suspect, but on one occasion the DNA from someone who actually committed a crime is discovered at the scene and sent to the corrupt lab, which duly reports a match. Pardo reckons that if this person had not committed their crime then their DNA would not have been found at the crime scene (since there would have been no crime scene!) and so there would be no lab report. If this is correct, then the lab report is sensitive evidence. But, Pardo maintains, the testimony of a corrupt forensics lab is surely so unreliable that it should not be admitted to court.

This is offered as an example of evidence that seems intuitively inadmissible despite technically being sensitive. Even if it is successful, the example does not bolster the case for safety over sensitivity very much, as there are also examples of evidence that seem intuitively inadmissible despite technically being safe. Indeed, statistical evidence can fit this description. Statistics are usually *modally robust* — they hold across a wide range of close worlds — because they usually reflect robust underlying social phenomena (like the economic phenomena that explain market share data). But this is not true of all statistics. As the old saying goes, correlation is not causation; some statistical regularities are mere flukes, failing to hold across a wide range of nearby worlds because they do not reflect any underlying phenomena at all. And, when this is so, statistical evidence that is technically safe can still be intuitively inappropriate for courts to consider. Here is an example:

Flimsy Statistic: A study finds that women who wear lipstick are two hundred percent more likely to engage in embezzlement than those who do not. This is accurate, as far as the historical frequencies go. But it is not explained by any causal relationship between lipstick-wearing and embezzlement. It is just a fluke. The researchers found their result by "*p-hacking*": collecting and sifting through reams of data in search of a statistically significant correlation and then publishing it, whatever it is. Still, the study is introduced as evidence in the trial of a lipstick-wearing woman who is indeed guilty of embezzlement.

The statistical evidence in *Flimsy Statistic* could turn out to be safe. That is because it could turn out that, in all of the close worlds in which the woman does not engage in embezzlement, the statistic does not hold. It is just a fluke, after all, so there is no reason to expect the statistic to be particularly modally robust. But, if it does turn out to be safe in this way, its safety does not make this statistic into good evidence. On the contrary, a flukey correlation found by p-hacking is precisely the sort of evidence that courts should *not* consider, since it is irrelevant to the case at hand. Indeed, in this case the very fact that makes the statistical evidence safe – the fact that it is just a fluke – is also precisely what makes it intuitively inadmissible.

If counterexamples like *Bad Lab* and *Flimsy Statistic* were the only ones available, they would not tell us much, since the most they can show is that neither sensitivity nor safety suffices for admissibility: evidence that is sensitive (like the lab's testimony) or safe (like the flimsy statistic) can still be inadmissible. Dialectically, these are weak points. They indicate that sensitivity or safety cannot be the *only* thing that matters for admissibility, but not that they do not matter at all.

However, Pardo also argues that sensitivity is not even necessary for admissibility. Indeed, he argues that a great many types of admissible evidence are often insensitive. His list is as follows:

[S]ome character evidence; prior acts admitted for a non-character purpose (such as proving motive, opportunity, intent, preparation, plan, knowledge, identity, absence of mistake, or lack of accident); most impeachment evidence; evidence of habits or routine practices; expert testimony (particularly, about general phenomena); and some admissible hearsay.

I think Pardo is correct to assume that these types of evidence are often insensitive. It will often happen that, in at least some of the closest worlds in which someone does not commit her actual crime or tort, many of these phenomena still obtain: she has performed many of the same prior acts, she has the same character, habits, and routines, the general phenomena that are the subjects of experts' testimony still obtain, and so on. The trouble for Pardo is that these very same types of evidence are also frequently unsafe. For the close worlds to which Pardo calls our attention — the worlds in which the defendant does not commit her actual crime or tort, but (some of) the forms of evidence on his list still obtain — will often be close enough to fall within the range that could “easily” have happened and with which safety is therefore concerned. The fact that someone chooses to commit a crime or tort is usually not very modally robust; people could easily have chosen to act differently than they do, especially if their actual action is somewhat impulsive. Agents' habits and routines and the general phenomena that are subjects of experts' testimony are usually more robust than this, holding across a wide range of close worlds. So it will often be easy to find close worlds in which (some of) the phenomena on Pardo's list still obtain but the agent has a different impulse and chooses not to commit her actual crime or tort. The prevalence of such worlds demonstrates that the types of evidence Pardo lists are not only frequently-insensitive but also frequently-unsafe. Pardo is therefore hoisted on his own petard: if the items on the list are indeed admissible, what this shows is that *neither* sensitivity *nor* safety is necessary for admissibility. And this, coupled with the *Bad Lab* and *Flimsy Statistic* examples, shows that sensitivity and safety are extensionally incorrect as criteria of admissibility.

3b. Factivity

In fact, we can show that neither sensitivity nor safety is necessary for admissibility without taking a stand on the particular types of evidence that Pardo lists. For there is a much more general problem in the vicinity.

The problem arises because sensitivity and safety are both *factive* criteria – that is to say, if the verdict that a piece of evidence supports is actually false, then that evidence is both insensitive and unsafe.

To see why this is so, think about the nature of sensitivity and safety again. Evidence is insensitive if it still obtains in (at least) one of the closest worlds in which the verdict that it supports is false. And evidence is unsafe if there are any close worlds in which the verdict that it supports is false, but it still obtains. But the closeness of possible worlds is a matter of their similarity to the actual world. So the actual world is always among the very closest worlds – no other world is more similar to the actual world than the actual world itself.¹³ Suppose, then, that a piece of evidence is misleading: the verdict that it supports is false. This misleading evidence obtains in one of the closest worlds in which the verdict that it supports is false – namely, the actual world. So the evidence is insensitive. Similarly, there is at least one close world in which the verdict that the evidence supports is false but the evidence still obtains – again, this is just the actual world. So the evidence is unsafe. In short: all misleading evidence is both insensitive and unsafe.

The literature already recognizes that the prevalence of misleading evidence poses a challenge for *factive* criteria of adequacy: Blome-Tillmann (2015, p.108) pointed out that such criteria entail that it is conceptually impossible for there to be verdicts that are false but, despite their falsity, are nonetheless correctly issued on the basis of strong misleading evidence. This is because *factive* criteria of adequacy entail that no misleading evidence is ever adequate, and thus that any body of evidence that supports a false verdict is therefore inadequate, regardless of how strong the evidence is. But legal systems aren't perfect, and strong misleading evidence often leads fact-finders to false verdicts that were intuitively correct for them to issue in light of their misleading evidence. This seems like something that happens all the time, says Blome-Tillmann. So our criteria of adequacy should not entail that it is a conceptual impossibility.

One might wonder whether things will fare better if we employ *factive* criteria as criteria of admissibility rather than adequacy. But that is not the case. Indeed, if anything, *factive* criteria of admissibility are even more of a disaster than *factive* criteria of adequacy. If a *factive* criterion is necessary for admissibility then misleading evidence is not just inadequate for a verdict but should not be allowed to appear in court at all. And the implications of this idea are manifestly ridiculous. This would mean that roughly half of the evidence at any trial – the evidence supporting whichever verdict is in fact false – should not have been admitted at all. What we should aim for, on this view, are trials in which the evidence is completely one-sided and the verdict that is in fact false does not get a fair hearing because no evidence in its favor is admitted. But this is nothing like what we normally think a fair trial looks like. So it is completely implausible that this is what the rules of admissibility require.

This makes it easy to find examples showing that neither sensitivity nor safety can plausibly be necessary for admissibility: all admissible misleading evidence is an example. Since plenty of misleading evidence is admissible, no *factive* criterion can be necessary for admissibility.

There is another issue with *factive* criteria of admissibility: applying these criteria would require judges to determine the verdict prior to the trial itself, at the time when they are issuing rulings as to which evidence may be admitted to court. This is because, if a *factive* criterion is necessary for admissibility, then judges must figure out what the true verdict is while they are figuring out what evidence to admit, so as to make sure that they do not admit any evidence that supports the false verdict (and is therefore inadmissible). But this gets things precisely backwards. The verdict is meant to be determined at trial on the basis of the

¹³ This principle is referred to as “strong centering” in the literature on counterfactuals. For discussion see especially Stalnaker (1968), Lewis (1973a), and Pollock (1976).

admissible evidence; it cannot be determined prior to trial in the course of deciding what evidence to admit. Any factive criterion of admissibility would make a sham of the factfinding process.

This latter issue is related to a broader issue with so-called “externalist” criteria in epistemology — criteria according to which beliefs’ epistemic status can turn on facts to which the believer lacks epistemic access. One common criticism of externalist criteria is that they are not *action-guiding*, since they evaluate people’s beliefs on grounds that cannot be used by those people when deciding what to believe. The criticism arises here because judges cannot use sensitivity or safety criteria to guide their rulings of admissibility, given that, at the time at which those rulings must be issued (prior to the trial), judges lack access to information (about the defendant’s actual liability or guilt, or lack thereof) on which the sensitivity and safety of the evidence partially depend. But we would presumably like our criteria of admissibility to be action-guiding, since we would like them to guide judges’ decisions about whether to admit evidence.¹⁴

Mario Günther (2021) has proposed that we avoid issues with factivity by using a non-factive construal of sensitivity that he terms “epistemic sensitivity”. The crux of this idea is that judges should seek to admit, not evidence that *is* sensitive, but evidence that they *believe* to be sensitive. Since judges can be wrong, this leaves room for the possibility of verdicts that are false but nonetheless correctly issued. And, since judges do have access to their own belief states (or so Günther assumes¹⁵), this is much more action-guiding than a factive approach. So the view that *epistemic* sensitivity is necessary for admissibility avoids the two issues that I just raised for the view that sensitivity is necessary for admissibility.¹⁶

Günther’s approach is formal and detailed; he uses nested sets of possible worlds and Gärdenfors’s (1982) method of general imaging to model the ways in which agents’ belief states shift under suppositions. For present purposes, the important point is that general imaging is not *conservative*, meaning that some things of which an agent is certain can become things of which she is not certain — perhaps things that she regards as probably false — under a supposition. This enables Günther to suggest that, to assess the sensitivity of a body of evidence *e* to a proposition *p*, agents should ask themselves whether the conditionals $p > e$ and $\sim p > \sim e$ hold true in their belief states, where $p > e$ holds true in an agent’s belief state just in case *e* is true in the worlds that the agent believes most likely upon supposing that *p* (pp.7-10).¹⁷ Clearly, the evidence actually

¹⁴ Externalists have standard responses to the charge that their criteria are not action-guiding: sometimes they point out that the idea of an epistemic “action” seems out of place since we lack direct control over our beliefs (the classic statement of this view is in Alston 1988), and sometimes they point out that no criterion is perfectly action-guiding, since there is no condition such that we are always able to tell whether we are in it (for discussion see especially Srinivasan 2015, Hughes forthcoming). But these responses are not much help in the legal context. For we are not discussing beliefs but rulings of admissibility, which are actions subject to direct control. And, even if no criterion of admissibility can ever be perfectly action-guiding, it still seems a bad idea to propose criteria that it is *in principle impossible* for judges to apply at the time when they must issue their rulings.

¹⁵ We do not have infallible epistemic access to our own beliefs and credences. But we do at least have *better* access to them than judges have to the disputed facts of a case at the time when they must issue rulings of admissibility.

¹⁶ I am grateful to an anonymous referee for suggesting that I consider Günther’s position here and for helping me to understand the implications of his model. It should be borne in mind that Günther does not himself propose epistemic sensitivity as a criterion of admissibility; Günther is interested in sufficiency rather than admissibility. Still, someone could propose that we use Günther’s notion as a criterion of admissibility in order to avoid problems with factivity.

¹⁷ Here is some more detail. Günther models belief states using nested sets of possible worlds and probability functions: there is a set *S* of worlds that the agent “cannot exclude to be actual”, to which her probability function assigns some non-zero probability, and there is a larger set *S'* that includes all worlds in *S* plus some other worlds to which the agent assigns zero probability but that “may become relevant under a supposition”. “Supposing” a proposition *p* amounts

obtains. But Günther's model allows that a judge may nonetheless think that it most likely would *not* obtain on the supposition that the defendant is not guilty. And this is so, he suggests, for the testimony of a reliable eyewitness; if our judge sets aside what she knows about what the eyewitness actually says and just thinks about the fact that they are highly reliable, then she will think that worlds in which what they say is true are more likely than worlds in which what they say is false. She will then think that worlds in which the defendant is not guilty and the eyewitness does *not* testify against them ($\sim p \& \sim e$) are more likely than worlds in which the defendant is not guilty but the eyewitness mistakenly testifies against them ($\sim p \& e$). By contrast, says Günther, nothing about a bare statistic indicates that it is more likely not to obtain than to obtain on the supposition that the defendant is not guilty. So eyewitness testimony is epistemically sensitive whereas the bare statistic is not.

I am not sure that this formal framework generates the right results. The trouble is that Günther's approach requires judges to set aside what they know about whether a piece of evidence actually obtains and *then* ask themselves whether ($p \& e$)-worlds are more likely than ($p \& \sim e$)-worlds and ($\sim p \& \sim e$)-worlds more likely than ($\sim p \& e$)-worlds. But some evidence would seem improbable if we did not already know that it obtains. With evidence like this, $\sim e$ -worlds can be more likely than e -worlds even under the supposition that p . For example, consider the fingerprint of a scrupulous assassin who normally kills without leaving a trace. A judge may hear of the assassin's scrupulousness just as she may hear of an eyewitness's reliability. The assassin's scrupulousness is then akin to the eyewitness's reliability in that it makes certain sorts of possibilities generally more likely than certain others: the eyewitness's reliability makes worlds in which what they say is true more likely than worlds in which what they say is false, and likewise the assassin's scrupulousness makes worlds in which they kill without a trace ($p \& \sim e$) more likely than worlds in which they kill and leave a trace ($p \& e$). But this means that, if our judge must assess the fingerprint's epistemic sensitivity by considering whether it would most likely obtain on the supposition that the assassin is guilty, *setting aside what she knows about whether it actually obtains*, then she should say that it most likely would *not* obtain. The fingerprint is therefore epistemically insensitive on Günther's approach, and thus inadmissible if epistemic sensitivity is necessary for admissibility. But that is madness. A defendant's fingerprints found at the scene of the crime surely must be admissible evidence, however antecedently improbable their being there can be made to seem.¹⁸ For another example, consider the testimony of an eyewitness who happened to pass by the scene of the crime but is almost always elsewhere at that time of day. Again, setting aside that she knows that the eyewitness testimony actually obtains, our judge may well think that worlds in which the defendant is guilty but the eyewitness does not testify ($p \& \sim e$) are generally more likely than those in which the defendant is guilty and the eyewitness does testify ($p \& e$), since it is unlikely that the eyewitness witnesses the crime at all. This means, again, that the evidence is epistemically insensitive on Günther's approach and thus inadmissible if epistemic sensitivity is necessary for admissibility. But, again, this seems like the wrong result.

One might also think that epistemic sensitivity is simply too different from the usual notions of sensitivity and safety for it to be a viable way of developing these notions for use in debates about statistical evidence. Sensitivity and safety are usually understood as metaphysical notions. They concern relationships between

to altering one's probability function by "shifting the probability shares from the worlds in S to the worlds in S'' ", where S'' is the set of all worlds in S' in which p is true. Günther then proposes that a conditional $p > e$ holds true in an agent's belief state just in case e is true in all the most likely worlds in S'' .

¹⁸ One might think that, since the assassin is usually scrupulous, it is likely – or at least possible – that she was framed. One might then think that the fingerprint evidence does not suffice for conviction. Nonetheless, this evidence surely should be *admissible*. The mere fact that the assassin does not usually leave fingerprints should not render her actual fingerprint inadmissible on the rare occasion when (it appears that) she has finally slipped up.

propositions, described using the possible worlds framework and glossed in terms of the relative similarity of worlds or the easiness of worlds' obtaining. And these relationships hold regardless of whether judges or philosophers or anyone else believes that they do. So, if we start by thinking that evidence is admissible just in case it bears one of these metaphysical relations to the verdict that it supports, then we may well be disappointed with the view that evidence is admissible just in case the judge *believes* that it bears some such relation to the verdict. After all, judges may be poor modal reasoners. They may be quite bad at determining sensitive from insensitive and safe from unsafe evidence. Likewise, their beliefs about which worlds are most likely under various suppositions may be wildly inaccurate representations of modal space. For those who seek to explain statistical evidence's inadmissibility in terms of metaphysical relationships that it fails to bear to disputed facts, then, a move to epistemic sensitivity will feel like changing the subject. I imagine that these scholars will not want to avoid problems with factivity by giving up the game and talking about a fundamentally different sort of thing from that which they originally had in mind.

3c. Coherent formulation

The previous two subsections offered examples of evidence that is inadmissible despite being sensitive or safe and evidence that is admissible despite being insensitive and/or unsafe. Here I explore a more basic problem with using sensitivity or safety to explain statistical evidence's inadmissibility: it is difficult even to state a coherent criterion of admissibility that draws on either of these notions.

Enoch, Fisher and Spectre and Pardo's papers are conspicuously devoid of explicitly-formulated criteria of admissibility employing the notions of sensitivity and safety. So we have to guess at what they might have in mind. We can start by noting that admissibility criteria have the form "evidence is inadmissible if...". And admissibility criteria using the notions of sensitivity or safety must state that evidence is inadmissible if it still obtains in any of the closest worlds (sensitivity) or in any close worlds (safety) in which a certain condition holds. The hard part comes in stating this condition. Our task is to finish these sentences:

Sensitivity: Evidence is inadmissible if it still obtains in at least one of the closest worlds in which...

Safety: Evidence is inadmissible if it still obtains in at least one close world in which...

A natural first pass would be as follows:

Sensitivity 1: Evidence is inadmissible if it still obtains in at least one of the closest worlds in which the verdict that it supports is false.

Safety 1: Evidence is inadmissible if it still obtains in at least one close world in which the verdict that it supports is false.

The problem with these first-pass criteria is foreshadowed by Pardo's "Bad Lab" example: if a defendant is actually guilty of the crime or tort of which they are accused, then it can turn out that the close(st) worlds in which they do not commit this crime or tort are ones in which *no such crime or tort occurs at all*, rather than ones in which the actual defendant acts differently but then someone else comes along and commits a qualitatively-identical crime or tort, so that the trial and the evidence remain as they actually are rather than not happening at all. To see why, recall that the distance of alternative possible worlds is a matter of their similarity to the actual world. The exact details of this "similarity" relation are unclear; this is a complex matter on which philosophers of modality are nowhere near settled. And, although worlds in which an actual crime or tort does not occur are *ipso facto* somewhat different from the actual world, they are nonetheless worlds in which only one person's conduct differs from its actual course. By comparison,

worlds in which the actual defendant does not commit their actual crime or tort but then someone else comes along and commits a qualitatively-identical crime or tort require two people's conduct to differ from its actual course rather than just one. Is it clear, then, that worlds of the latter sort are all always *more* similar to the actual world than those of the former sort all-things-considered? Unless all actual crimes and torts are accompanied by "backup" perpetrators who stand poised and ready to do exactly what the actual defendants do if the actual defendants change their minds, it is not at all clear that this is how the details of the similarity relation will shake out. And, of course, "backup" perpetrators are in fact exceedingly rare.

This means that our first-pass sensitivity and safety criteria are not a very close fit with what we think about when we think about what is wrong with statistical evidence. In cases like *Smith*, we normally consider the possibility that another company owned the bus that caused the accident. But, if it really was Rapid Transit that caused the accident, then these worlds may be quite remote indeed. And worlds in which no accident occurs at all and the plaintiff gets home safely may be considerably closer. Thus it may turn out, perhaps quite frequently, that admissibility criteria focused on the close(st) world(s) in which the verdict that a piece of evidence supports is false are *looking at the wrong possibilities* — that is, they direct our attention toward possibilities in which no crime or tort (and so no trial) occurs at all, rather than ones in which the actual crime or tort was committed by someone other than the defendant.

We can avoid this problem by building more content into the condition that our sensitivity and safety criteria place on the possible worlds to which they direct our attention. Here is one way of going:

Sensitivity 2: Evidence is inadmissible if it still obtains in at least one of the closest worlds in which the verdict that it supports is false, *but the crime or tort still occurs*.

Safety 2: Evidence is inadmissible if it still obtains in at least one of the close worlds in which the verdict that it supports is false, *but the crime or tort still occurs*.

These second-pass criteria ensure, by stipulation, that we search modal space for worlds in which the actual crime or tort still occurs. If the actual defendant is indeed guilty, then they direct us to look for worlds in which their crime or tort is committed by someone else. However, this instruction is of dubious coherence. Its coherence depends on the identity conditions of crimes and torts across worlds — i.e., on what it is about a crime or tort taking place in an alternative possible world that makes it numerically identical to one taking place in the actual world. Again, this is a matter of the complex details of the logic of modality; specifying the modal identity conditions of crimes and torts is not going to be easier than specifying the details of the similarity relation. Moreover, at first blush, it seems plausible that crimes and torts are identified in part by their perpetrators. That is, it seems plausible that if a crime or tort in an alternative possible world is committed by a different person than one in the actual world, then these are not one and the same crime or tort. But, if that is correct, then whenever a defendant is actually guilty *there are no worlds* in which their crime or tort still occurs but is committed by someone else. And this means that our second-pass sensitivity and safety criteria are subtly incoherent when applied to all evidence supporting a true verdict.

We could instead try saying that crimes and torts are individuated by their victims.¹⁹ But this will not apply broadly enough to cover all cases. It makes sense in cases like *Smith* to say that we are interested in whatever tortious conduct led to the plaintiff's accident — conduct that could have been engaged in by a bus driver under the employ of someone other than Rapid Transit. But we cannot say anything like this in a trial for drug possession, for instance, or for trespassing. Here it cannot be taken for granted that there was some tortious conduct and some specific harm to an identifiable victim that we may look for across modal space.

¹⁹ Thanks to an anonymous reviewer for suggesting that I consider this idea.

The purpose of the trial is to determine whether any crime or tort occurred at all, rather than to determine the identity of the perpetrator of some harm that definitely occurred.²⁰ So, if in fact someone did trespass or possess drugs, then it is unclear what we are supposed to look for when we look for worlds in which the actual defendant is innocent but “the” crime or tort still occurs. Are these just worlds in which other people trespass and possess drugs? Which ones, and when? And why is what happens to these other criminals in other worlds relevant to the admissibility of our actual evidence at our actual trial?

Here is a third pass that circumnavigates these issues:

Sensitivity 3: Evidence is inadmissible if it still obtains in at least one of the closest worlds in which the verdict that it supports is false, but the basic facts of the case still obtain.

Safety 3: Evidence is inadmissible if it still obtains in at least one of the close worlds in which the verdict that it supports is false, but the basic facts of the case still obtain.

Here the phrase “the basic facts of the case” is a placeholder for a qualitative description of whatever it is in a particular case that indicates that a crime or tort has occurred. For example, the basic facts of the case may be that the plaintiff was injured in an accident caused by a poorly-driven bus, or that a certain illegal substance was found in the defendant’s car. On this approach, there will not be a single sensitivity criterion or a single safety criterion. Sensitivity and safety instead generate schema for criteria of admissibility; there will be many such criteria, each one tailored in its details to the case at hand.²¹

These third-pass criteria avoid directing us to look for worlds in which no crime or tort occurs (since they stipulate that the basic facts of the case must still obtain). And they avoid unintelligibly directing us to look for crimes and torts that are committed by different people but are nonetheless numerically identical to our actual crimes and torts (since they allow us to look for qualitatively similar crimes or torts instead) without floundering in cases of “victimless” crime (since they do not presuppose that a crime or tort occurs at all). As such, they improve considerably upon the first and second passes.

But these criteria face a curious problem of their own. For, lest we forget, these are criteria of admissibility. They state that evidence is admissible only if it does not still obtain in any of the closest worlds (sensitivity) or in any close worlds (safety) in which a certain condition holds. And we are now considering criteria that build the basic facts of the case into the stipulated condition. By stipulation, then, the basic facts of the case still obtain in the worlds we are looking at – they cannot fail to obtain in the close(st) world(s) in which the verdict that some evidence supports is false *but they still obtain*. Our third-pass criteria therefore entail, in each case, that the basic facts of the case are inadmissible. For illustration: if the sensitivity criterion for *Smith* states that evidence is admissible only if it does not still obtain in the closest possible world in which the verdict that it supports is false but it is still the case that the plaintiff was injured in an accident caused by a poorly-driven bus, then this criterion entails that the fact that the plaintiff was injured in an accident caused by a poorly-driven bus is itself inadmissible. But this is patently another disaster. The basic facts of the case cannot be inadmissible. If such facts could not be considered by the court, then it would be unclear what each trial was even about.

²⁰ [Footnote redacted for anonymity]

²¹ Alternatively, we could generate the same results by sticking with the first-pass sensitivity and safety criteria but stipulating that the similarity relation should be specified in such a way as to minimally alter the basic facts of the case. Thanks to an anonymous reviewer for this observation.

These third-pass criteria also face a second problem, which is less conceptual and more practical: they require us to decide which facts of the case are “basic” (in order to determine what to hold fixed when we look across worlds). There is no clear method for dividing the facts of a case into basic and non-basic ones. But the way we choose to execute our divisions matters. For example, suppose that Rapid Transit has the largest market share along the route of the accident primarily because it operates all of the late-night buses. In that case, it might turn out that in the close(st) worlds in which another company’s bus causes an accident *late at night*, the market share data are different (since another company owns some of the late-night buses), whereas in the close(st) worlds in which a bus owned by another company causes an accident *during the day*, the market share data remain the same. This would mean that whether the market share data are admissible depends on whether we hold fixed the fact that the accident occurred late at night. So, is this one of the “basic” facts of the case, or not? It is hard to think of a principled way to answer the question.

To sum up: it falls to theorists who posit sensitivity or safety criteria of admissibility to explicitly formulate workable versions of these criteria that do not generate incoherent, irrelevant, or otherwise obviously bad results across a broad range of cases. To do so, they must be clear about what judges are supposed to hold fixed when they look across modal space for worlds in which the verdict that a piece of evidence supports is false. I have argued that holding nothing fixed, holding fixed the actual crime or tort, and holding fixed a qualitative description of the basic facts of the case are all unworkable. These options are not exhaustive, but they are the most natural candidates. Until I see a coherent criterion of admissibility formulated in terms of sensitivity or safety, then, I will have my doubts as to whether this can be done.

4.

We have seen that neither sensitivity nor safety can plausibly explain statistical evidence’s inadmissibility, since neither criterion can provide a plausible condition of admissibility for statistical evidence to then fail to meet. But perhaps it was a mistake to focus on admissibility in the first place. Perhaps philosophers and legal theorists should stick with the weaker explanandum: why statistical evidence is inadequate for a finding of fact.

I doubt that sensitivity or safety can explain inadequacy. That is partly because of the counterexamples mentioned briefly in §3a. But it is also because, intuitively, sensitivity and safety are *just not the right sort of thing* to explain what makes evidence adequate or inadequate. The adequacy of a body of evidence for a finding of fact is a matter of whether it proves the disputed fact to be true, to the relevant standard of proof. And evidence’s probative value depends on explanatory relationships – causal or constitutive relationships – between the evidence and the disputed fact. But sensitivity and safety are not concerned with explanatory relationships. They are metaphysical notions, as we have seen, but this does not mean that the relationships that they explore are explanatory. Instead, they are simply concerned with patterns of facts across worlds.

For illustration, consider the following inferences:

- P. Two plus two is four.
- C. Therefore, Paris is the capital of France.

- P. The U.S. Women’s National Team won the world cup in 2019.
- C. Therefore, Nina Simone ate eggs for breakfast on her twenty-fifth birthday.

An inference is sensitive iff, in all of the closest worlds in which its conclusion fails to hold, at least one of its premises also fails to hold. And an inference is safe iff, in all close worlds in which its premises still hold,

its conclusion also still holds. The second of these inferences, then, could turn out to be sensitive: it could turn out that in the closest worlds in which another team wins the 2019 women's world cup, Nina Simone eats something else for breakfast on her twenty-fifth (or she skips breakfast that day, or she does not exist, or...). And the first of these inferences is clearly safe: in all close worlds it remains the case that two plus two is four and that Paris is the capital of France. But the safety and sensitivity of these inferences does not indicate that the premises have high probative value in an investigation as to the truth of the conclusions, nor that the premises are good evidence for the truth of the conclusions. On the contrary, these inferences' safety and sensitivity does not redound to the credit of the premises at all. The premises and conclusions have nothing to do with one another. The inferences are safe and sensitive just because a pattern of facts happens to exist across possible worlds.

It is true that philosophers working within the possible worlds framework often speak of propositions as "ruling out" or "distinguishing between" possibilities, thereby "narrowing down" modal space and allowing us to determine which possible world is actual. But, as the above examples illustrate, this sort of "ruling out", "narrowing down", and "distinguishing" is a sort that can obtain even when the propositions in question have nothing to do with one another. It depends on relations between propositions that are made true simply by the shape of modal space; sensitivity and safety concern whether propositions are true or false in the same worlds, rather than whether there is any good sense in which one is true *because* the others are true or they are both true *because* of the same further facts.²²²³

Now, when courts express doubts about statistical evidence's admissibility, they often say that the problem is that such evidence is not specific enough to the case at hand: that "it is difficult... to understand how statistical information would assist a trier of fact in reaching a determination as to guilt in an individual case" (*Stephens v. State*, 774 P.2d 60 (Wyo. 1989) at 64); that "the relevancy of this evidence is not discernible" (*State v. Maule*, 667 P.2d 96 (Wash. Ct. App. 1983) at 99); that it is "distractive" (*Hall v. State*, 692 S.W.2d 759 (Ark. Ct. App. 1985), at 773). It is this inchoate sense of *specific-ness* that the literature on statistical evidence aims to make precise – we use the stipulative term "individualized" for evidence that is intuitively specific enough to be both relevant and adequate, and the goal of the literature is to spell out exactly what it is for evidence to be individualized. But it would be highly surprising for it to turn out that the notions of sensitivity and safety pick out the sort of specific-ness in question. For these notions point us toward relations that can hold between propositions that have nothing whatsoever to do with one another. The modal relationship between <two plus two is four> and <Paris is the capital of France> is surely not what we are looking for in a relationship between a body of evidence and a verdict. We need something much *more* specific than this.

²² For lengthier discussion of causation as it relates to the statistical evidence debate, see Thomson (1986) and Wright (1985, 1988). Blome-Tillmann (2015) offers some pushback and a shrewd counterexample.

²³ One might think that appealing to causal relations poses no threat to criteria of admissibility that look for patterns of facts across worlds, because one might think that causal relations just *are* patterns of facts across worlds. And one might appeal to Lewis's (1973b) account of causation in support of this view. However, Lewis's view does not support the conclusion that just any old pattern of facts across worlds – such as those in the two examples above – constitutes a causal relation. On the contrary, Lewis was well-aware of the fact that these patterns do not distinguish E1's causing E2 from E2's causing E1 and from E1 and E2 being effects of a common cause. **It is in light of these problems that Lewis revised his account by adding a no-backtracking condition and introducing the notion of chains of causal dependence;** i.e., precisely the sort of chain that does *not* obtain between the premises and conclusions of the above inferences. (Lewis 2000 further revised the account in recognition of deeper problems for counterfactual theories of causation, chiefly involving cases of preemption.) Thanks to an anonymous referee for suggesting that I mention Lewis's account here.

Of course, sensitivity and/or safety might still be necessary for adequacy. But, if they are, it is presumably because they are necessary for the more specific relationship between a body of evidence and a finding of fact that we are really looking for when we say that evidence should be “individualized” (whatever it is). And, in that case, it isn’t really sensitivity or safety that *explains* why certain bodies of evidence are inadequate for certain findings of fact. Rather, the absence of this more specific relationship does the explaining. The insensitivity or unsafety of the evidence is just a corollary.

To be clear: I do not know what this more specific relationship is. The literature on the topic is sprawling and fraught.²⁴ So I do not take myself to have decisively refuted the claim that sensitivity or safety explains inadequacy, since I am not offering a positive argument for an alternative view. But I do take my remarks in this section to have provided some reasons to be skeptical about these notions’ explanatory prospects.

5.

When we try to explain something, it matters what we are trying to explain, since considerations that are good explanations of some facts are not thereby good explanations of all facts. The literature on statistical evidence is not always clear about what it is trying to explain. But it matters which explanandum we take on. In particular, the two modal conditions on knowledge that have attracted much recent attention are better-positioned to explain why we deem statistical evidence inadequate for a verdict than they are to explain why we sometimes feel as though statistical evidence should not be admitted to court at all. And that is because these conditions cannot provide plausible general criteria of admissibility for statistical evidence to violate. Sensitivity and safety criteria might explain parts of the law, for all I have said here, but rules of admissibility are not among them.

²⁴ For some classic treatments of the topic that I have not already mentioned, see Tribe (1971)’s scathing attack on the use of mathematical reasoning in the legal process, Brilmayer and Kornhauser (1978) and Colyvan, Regan and Ferson (2002)’s individuality-based arguments, Posner (1972) and Brook (1982)’s cumulative-impact-based arguments, Nesson (1985)’s public-trust-based argument, Wasserman (1991) and Pundik (2008)’s autonomy-based arguments, and Stein (2005)’s “case specificity” argument, and compare Koehler and Shaviro (1991)’s defense of the use of naked statistical evidence by courts.

REFERENCES

- Allen, Ronald J. (1990). "On the Significance of Batting Averages and Strikeout Totals: A Clarification of the Naked Statistical Evidence Debate, the Meaning of Evidence, and the Requirement of Proof beyond a Reasonable Doubt". *Tulsa Law Review* 65:1093-1110.
- Alston (1988). "The Deontological Conception of Epistemic Justification." *Philosophical Perspectives* 2:257–99.
- Blome-Tillmann, Michael (2015). "Sensitivity, Causality, and Statistical Evidence in Courts of Law". *Thought: A Journal of Philosophy* 4(2):102-112.
- Brilmayer, Lea & Lewis Kornhauser (1978). "Review: Quantitative Methods and Legal Decisions." *University of Chicago Law Review* 46:116-149.
- Brook, James (1982). "Inevitable Errors: The Preponderance of the Evidence Standard in Civil Litigation". *Tulsa Law Review* 18:79-109.
- Cohen, L. Jonathan (1977). *The Probable and the Provable*. Clarendon Press, Oxford.
- Colyvan, Mark, Helen M. Regan and Scott Ferson (2002). "Is it a Crime to Belong to a Reference Class?" *The Journal of Political Philosophy* 9(2):168-181.
- Enoch, David, Levi Spectre and Talia Fisher (2012). "Statistical Evidence, Sensitivity, and the Legal Value of Knowledge". *Philosophy and Public Affairs* 40(3):197-224.
- Enoch, David, and Talia Fisher (2015). "Sense and 'Sensitivity': Epistemic Approaches and Instrumental Approaches to Statistical Evidence." *Stanford Law Review* 67:557-611.
- Enoch, David and Levi Spectre (2019). "Sensitivity, Safety, and the Law: A Reply to Pardo". *Legal Theory* 25(3):178-199.
- Gärdenfors, P. (1982). "Imaging and Conditionalization". *The Journal of Philosophy* 79(12): 747-760.
- Guenther v. Armstrong Rubber Co.*, 406 F.2d 1315, 1318 (3rd Cir. 1969).
- Günther, Mario (2021). "Epistemic Sensitivity and Evidence". DOI: 10.1080/0020174X.2021.1936158
- Hall v. State*, 692 S.W.2d 759 (Ark. Ct. App. 1985).
- Hughes, Nicholas H. (forthcoming). "Epistemology Without Guidance".
- Kaminsky v. Hertz Corp.*, 288 N.W.2d 426 (Mich. Ct. App. 1979).
- Kaye, David (1980). "Naked Statistical Evidence". *Yale Law Journal* 89:601.
- Kaye, David (1982). "The Limits of the Preponderance of the Evidence Standard: Justifiably Naked Statistical Evidence and Multiple Causation." *American Bar Foundation Research Journal* 7(2):487-516.

- Koehler, Jonathan J. and Daniel N. Shaviro (1990). "Veridical Verdicts: Increasing Verdict Accuracy Through the Use of Overtly Probabilistic Evidence and Methods." *Cornell Law Review* 75:247-279.
- Koehler, Jonathan J. (2002). "When Do Courts Think Base Rate Statistics Are Relevant?", *Jurimetrics* 42:380-85.
- Lewis, David (1973a). *Counterfactuals*. Cambridge, MA: Harvard University Press.
- Lewis, David (1973b). "Causation". *Journal of Philosophy* 70:556-67.
- Lewis, David (2000). "Causation as Influence". *Journal of Philosophy* 97:182-97.
- Littlejohn, Clayton (2020). "Truth, Knowledge, and the Standard of Proof in Criminal Law". *Synthese* 192(12): 5253-5286.
- Nesson, Charles (1979). "Reasonable Doubt and Permissive Inferences: The Value of Complexity." *Harvard Law Review* 92(6):1187-1225.
- Nesson, Charles (1985). "The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts." *Harvard Law Review* 98(7):1357-1392.
- Nozick, Robert (1981). *Philosophical Explanations*. Harvard University Press.
- Pardo, Michael (2018). "Safety vs. Sensitivity: Possible Worlds and the Law of Evidence." *Legal Theory* 24(1):50-75.
- Pollock, John (1976). *Subjunctive Reasoning*. Dordrecht: D. Reidel Publishing Co.
- Posner, Richard (1972). *Economic Analysis of the Law*. Boston: Little, Brown.
- Pritchard, Duncan (2008). "Safety, Sensitivity, and Anti-Luck Epistemology". In *The Oxford Handbook Of Skepticism*, John Greco (Ed.). Oxford University Press.
- Pundik, Amit (2008). "Statistical Evidence and Individual Litigants: A Reconsideration of Wasserman's Argument from Autonomy." *International Journal of Evidence and Proof* 12:303-331.
- Ross, Lewis (2021). "Rehabilitating Statistical Evidence", *Philosophy and Phenomenological Research* 102(1): 2-23.
- Smith v. Rapid Transit Inc.*, 317 Mass. 58 N.E.2d 754 (Mass. 1945).
- Smith, Martin (2018). "When Does Evidence Suffice for Conviction?" *Mind* 127 (508): 1193-1218.
- Sosa, Ernest (1999), "How to Defeat Opposition to Moore". *Philosophical Perspectives* 33: 141-153.
- Srinivasan, Amia. (2015). "Normativity Without Cartesian Privilege". *Philosophical Issues* 25 (1): 273-299.

Stalnaker, Robert (1968). "A Theory of Conditionals". In *Studies in Logical Theory*, Nicholas Rescher (ed.), Oxford: Basil Blackwell, 98–112.

State v. Maule, 667 P.2d 96 (Wash. Ct. App. 1983).

Stephens v. State, 774 P.2d 60 (Wyo. 1989).

Tribe, Lawrence (1971). "Trial By Mathematics: Precision and Ritual in the Legal Process." *Harvard Law Review* 84(6):1329-1393.

Thomson, J. J. (1986). "Liability and Individualized Evidence." *Law and Contemporary Problems* 49(3): 199-219.

Wasserman, David T. (1991). "The Morality of Statistical Proof and the Risk of Mistaken Liability." *Cardozo Law Review* 13 935:942-43.

Wells, Gary L. (1992). "Naked Statistical Evidence of Liability: Is Subjective Probability Enough?" *Journal of Personality and Social Psychology* 62(5):739-752.

Wright, Richard W. (1985). "Causation in Tort Law." *California Law Review* 73:1735-1828.

Wright, Richard W. (1988). "Causation, Responsibility, Risk, Probability, Naked Statistics, and Proof: Pruning the Bramble Bush by Clarifying the Concepts." *Iowa Law Review* 73(5):1001-1078.