Legal Probabilism: A Qualified Defence

Brian Hedden[*]   Mark Colyvan[†]

Abstract

In this paper we defend legal probabilism. This is the thesis that legal standards of proof are best understood in probabilistic terms. In defending legal probabilism from a variety of objections, we highlight the most plausible forms the thesis can take and appeal to recent work in epistemology to show that the legal probabilist has more flexibility and more resources at her disposal than critics have recognised.

1   Introduction

The proper understanding of legal standards of proof is a notoriously fraught topic. The legal system currently operates with three different standards of proof. The most demanding standard, *beyond a reasonable doubt*, is operative in criminal trials. The least demanding standard, *preponderance of the evidence*, is operative in most civil trials. An intermediate standard, *clear and convincing evidence*, is operative in some civil proceedings, in particular where citizenship or child custody is at stake.

How should standards of proof such as the above be understood?Judges have generally been reluctant to provide much guidance on this issue, and when they have weighed in, their glosses have frequently been problematic (see Laudan 1996, 32–51 for a catalogue of refusals to provide clarification as well as failed attempts to do so).

In this paper, we defend *legal probabilism*, the thesis that legal standards of proof are best understood as probabilistic in form. Standards of proof should be concerned with whether the state's (in a criminal proceeding) or the plaintiff's (in a civil one) case has been established to such a degree as to justify a probability of guilt or liability above some threshold. This is, of course, only a rough statement, but the picture will become clearer over the course of the paper.

Legal probabilism has been defended by scholars associated with the so-called 'New Evidence Scholarship' which followed on the introduction of the Federal

[*]Department of Philosophy, University of Sydney, Sydney, NSW, 2006, Australia. Email: brian.hedden@sydney.edu.au

[†]Department of Philosophy, University of Sydney, Sydney, NSW, 2006, Australia. Email: mark.colyvan@sydney.edu.au

Rules of Evidence in the U.S.A. in 1975, as well as by scholars working in the tradition of economic analysis of law (Posner 1973). But it has long faced opposition, with prominent critics including Laurance Tribe (1971), L. Jonathan Cohen (1977), Leonard Jaffee (1984–5), Larry Laudan (2006), Ronald Allen (1991; 2017) and coauthors (Allen and Leiter 2001; Allen and Stein 2013), and Susan Haack (2014).

As we have characterised it, legal probabilism is a thesis only about proper standards of proof. It is not committed to the stronger claim that various other aspects of the trial process are best analysed or modelled using probability theory. In addition, legal probabilism does not claim that juries do or should explicitly reason in terms of probabilities. It does not recommend, for instance, that jurors be trained in the complexities of probability theory, drilled in Bayes' Theorem, asked to write down conditional probabilities, and the like. It only claims that at the end of the trial, the jury must reach a verdict on whether the evidence supports assigning a probability above some threshold to the defendant's being guilty or liable.

This is important, since it mitigates the worry that people do not naturally reason in terms of probabilities. Probabilistic reasoning may be somewhat alien to ordinary people, but understanding, reaching, and reporting probabilistic judgments is not. People deal with probabilities of rain, of some team's winning, of a given election result, and the like on a daily basis. It is certainly far from clear that people have any greater difficulty arriving at a judgment of whether some event is say, over 90% likely than they do arriving at a judgment of whether it is beyond a reasonable doubt. Indeed, there is evidence that people may understand explicitly probabilistic standards of proof better than the three traditional standards of preponderance, clear and convincing evidence, and beyond reasonable doubt (Kagehiro and Stanton 1985).

Finally, it is important to distinguish between descriptive and normative versions of legal probabilism. Descriptive claims concern how the world is, while normative claims concern how the world ought to be. So a descriptive version of legal probabilism would say that standards of proof, given how the legal system in fact operates, are implicitly probabilistic. A normative version of legal probabilism would say that, regardless of how standards of proof in fact operate within the confines of the existing legal system, the legal system ought to be governed by probabilistic standards of proof. Our concern is with a normative version of legal probabilism. We are less concerned with how probabilistic interpretations of standards of proof fit with existing precedent and with the actual practices of lawyers, judges, and juries, and more concerned with whether an ideal (or more ideal) legal system would use probabilistic standards of proof.

With that ground-clearing out of the way, the plan for the paper is as follows. We begin (Sections 2 and 3) by considering and rejecting some recent objections to legal probabilism raised by Ronald Allen and Susan Haack. We fault them for assuming an outdated and inexhaustive catalogue of 'interpretations of probability' (or ways to understand probabilistic claims) which lead them to ignore the possibility of integrating the explanatory and other epistemological considerations that they rightly emphasise into a probabilistic framework. One central

lesson is that probability theory is a far more flexible and adaptable framework than many scholars give it credit for. We turn then to the Conjunction Paradox (also addressed by Allen and Haack), arguing that while it presents a challenge to descriptive versions of legal probabilism, it is no objection to normative legal probabilism. Finally, we address what we consider more serious challenges to legal probabilism, the problem of how to set probability thresholds for standards of proof, and the problem of conviction (or liability) based on bare statistical evidence.

2   Allen's Objections

Ronald Allen has criticised legal probabilism in a series of papers with various coauthors, arguing instead for a 'relative plausibility' conception of standards of proof. We will not discuss this relative plausibility theory in depth here, though we will note ways in which a probabilistic framework can incorporate its insight that judgments of plausibility, explanatory power, and the like are crucial in reaching a verdict. Moreover, Allen's objections to legal probabilism are numerous, and so in this section we will only be able to address some of the most central ones (another of his main objections involves the issue of bare statistical evidence, which we discuss in Section 6).

Allen and Stein (2013) argue that legal probabilism is committed to ignoring particular facts of the case at hand as well as considerations like how well the competing cases explain the evidence presented. In their place, legal probabilism would require fact finders to pay exclusive attention to the frequencies of relevant sorts of events.

But why do they think legal probabilism is committed to such a strong and bizarre claim? Their reasoning rests on a particular, and in our view inadequate, catalogue of interpretations of probability, one which leaves out arguably the most promising way for legal probabilists to understand statements of probability. In particular, Allen and Stein claim that legal probabilists must rely on a frequentist view of probability:

> Scholars' attempts at mathematising the burden of proof follow a frequentist interpretation of probability, and for good reason. Other interpretations of the concept of "probability"—logical, propensity, and subjective beliefs—make no sense at all in the juridical context. (2013, 566)

Now, Allen and Stein do not say what is wrong with logical, propensity, and subjectivist interpretations. We concede that the first two fair poorly in the legal context. But what is the problem with adopting a subjectivist interpretation or the frequentist interpretation which they think is the legal probabilist's best bet?

Let's start with the latter. On a frequentist interpretation, we start by identifying a *reference class* for the event or proposition in question. So, for instance, the reference class relevant for the probability of this coin's landing heads on its next flip might be the class of all flips of this coin. Then, the probability that

the event has the property in question is the fraction of events in the reference class that have the property in question. So, for instance, if one half of all flips of this coin have the property of landing heads, then the probability of this coin's landing heads on the next flip is one half.

Allen and Stein object that legal probabilism, using a frequentist interpretation of probability, would require ignoring (or at least pushing to the background) particular facts of the case and paying attention primarily to general frequencies. Frequentist probability, they say, requires fact finders to abstract away from 'the actual facts of the case they are asked to resolve' and instead 'prods fact-finders to derive their decisions from the general frequencies of events' (560). This is partly true, though overstated. The actual facts of the case at hand will be of the utmost importance, even on frequentist version of legal probabilism, since these particulars determine the relevant reference class. Suppose a defendant is on trial for murder. The fact that, say, he was recently seen polishing his gun and threatening the victim with violence is relevant, since it suggests that the relevant reference class is not the class of all humans (very few of whom commit murder) but rather something like the class of all human gun-polishers who issue violent threats.

The problem of fixing the relevant reference class is a major problem in its own right (Colyvan, Regan, and Ferson 2001; Hájek 2007). In general, an event will be a member of many different reference classes, and using one reference class will often yield a different relative frequency than using another. Absent some way of choosing a privileged reference class, frequentism is thus threatened with inconsistency.

So Allen and Stein are wrong to say that the particular facts of the case largely drop out of the picture on a frequentist version of legal probabilism. But their other major objection to frequentist legal probabilism is on the mark. They rightly maintain that whether a conviction (or finding of liability) is warranted depends at least in part on explanatory considerations, that is, on how well each side's case *explains* the evidence presented in court, and perhaps on other epistemological considerations as well, such as how simple the competing cases are, how comprehensive they are, and so on. Frequentist versions of legal probabilism do not leave room for such considerations to play a role (except insofar as they are relevant to determining the relevant reference class). While we maintain that facts about relative frequencies are part of the picture and have a role to play in helping to determine which trial outcomes is warranted, we agree with Allen and Stein that they are not the whole story. Frequentist legal probabilism is untenable.

While Allen and Stein say nothing by way of criticising the next interpretation of probability, namely Subjective Bayesianism, Allen and Leiter (2001) criticise subjectivist versions of legal probabilism on the same grounds, namely that they fail to give an adequate role to explanatory and other important epistemological considerations. Subjective Bayesianism interprets at least some probability statements as expressing agents' degrees of belief (aka degrees of confidence, subjective probabilities, or credences).

But Subjective Bayesianism is also a very permissive epistemological theory.

4

It holds that an agent's subjective probabilities are rational provided that they conform to the axioms of the probability calculus and perhaps a few other formal constraints.

It is really to this permissivism that some scholars object. As Allen and Leiter note:

> Another difficulty with [subjective] Bayesian approaches to juridical evidence is that the assignments of initial probabilities, which are crucial to the application of [Bayes'] Theorem, are subjective and need respect only the conditions of consistency and summing to 1.0. That means that individuals can begin from radically different perspectives, and each, in Bayesian terms, will be operating equally rationally. (21)

So Subjective Bayesianism permits, but does not require, that agent's probability judgments be sensitive to the explanatory and other epistemological considerations that Allen and his coauthors emphasise (see also Haack 2014, 61–2). Before turning to an alternative view of probability which does incorporate such considerations, we should ask how serious this worry really is. From one perspective it is indeed serious. A standard of proof that said that jurors should vote to convict if their subjective probability for the defendant's guilt exceeds 0.95 would deem legitimate a conviction in which each juror's high confidence in guilt was intuitively irrational and wholly insensitive to explanatory considerations. This seems problematic.

But from another perspective, the issue is not terribly serious. A standard of proof that instructs jurors to vote to convict if and only if they are in fact highly confident (above 0.95, say) that the defendant is guilty will likely yield the same results as one that instructs jurors to vote to convict if and only if a proper evaluation of the evidence justifies such high confidence in guilt. After all, jurors rarely recognise a divergence between their actual attitudes and those that they rationally ought to have. So if we are concerned primarily with the practical consequences of instituting some standard of proof, putting the standard of proof in more subjective terms (about jurors' actual attitudes) will not be much worse, if at all, than putting it in more objective terms (about what attitudes are licensed by the evidence).

In any case, Allen and his coauthors ignore another interpretation of probability which is yet more favourable to the legal probabilism, namely evidential probability (Williamson 2000). The evidential probability of a proposition is the probability of the proposition, given some relevant body of evidence. It is the degree of belief that it is (most) rational to have if one has all and only that body of evidence. In the case of a trial, the relevant evidential probability would be the probability that the defendant is guilty or liable, given the admissible evidence presented at trial, along with some mundane background evidence about our physical and social world.

Evidential probabilities incorporate the sorts of epistemological considerations—explanatory quality, simplicity, comprehensiveness, and the like—that Allen and others rightly emphasise. Here is one way of filling out the picture. There is a prior probability function P whose unconditional probabilities represent, as

Williamson (2000, 211) puts it, "the intrinsic plausibility of hypotheses prior to investigation" and whose conditional probabilities reflect facts about explanatory connections, so that, for instance, the conditional probability of h given e will be high to the extent that h is a good explanation of e (see Weisberg 2009). Given a body of evidence e, the evidential probability of some hypothesis h is the result of taking that prior probability function and conditionalising it on e (i.e. $P(h|e)$).

By appealing to evidential probabilities, the legal probabilist can address the main concerns raised by Allen and his coauthors. In particular, it gives pride of place to explanatory and other epistemological considerations in determining whether a conviction is warranted, and directs fact finders' attention to the particular facts of the case at hand.

Now, some epistemologists are skeptical of the very notion of objective evidential support. In their view, evidential support is in the eye of the beholder; it is always relative to a set of background beliefs or other assumptions. If they are correct, then obviously evidential probabilities are non-existent. A large part of the motivation for skepticism about the notion of objective evidential support stems from the seeming impossibility of characterising evidential support in precise, formal terms (Goodman 1955, Titelbaum 2010). There is much to say about this debate and one of us (Hedden 2015) has provided a defence of objective evidential support. But in the present context, it is worth emphasising two dialectical points. First, Allen's own relative plausibility view of standards of proof relies on the notion of objective evidential support, so he cannot object to the legal probabilist's appealing to the very same notion. Second, few objections to the notion of objective evidential support turn on whether rational doxastic (i.e. belief-like) states are conceived of as probabilistic in structure or as more coarse-grained (e.g., a set of on/off beliefs). For instance, it may be true that evidential probabilities cannot be uniquely specified in purely formal terms; there is no algorithm that will tell us what the evidential probability of a hypothesis is, given some body of evidence. But the same is true even if we work with more coarse-grained doxastic states. Allen and Stein admit that their own view does not admit of such a specific, algorithmic formulation: "There is no algorithm for 'plausibility;' the variables that inform judgments of plausibility are all the things that convince people that some story may be true, including coherence, consistency, coverage of the evidence, completeness, causal articulation, simplicity, and consilience" (568). If the inability to give an algorithmic characterisation is not a problem for Allen and Stein's own view, it is unclear why it should be a problem for legal probabilism.

We conclude that the objections considered so far from Allen and his coauthors fail. They mistakenly deem frequentism to be the only game in town for legal probabilists. While frequentism is indeed inadequate in the context of legal probabilism, they overstate the case against subjective Bayesian interpretations of probability, and ignore altogether the notion of evidential probability, which incorporates precisely the elements that they regard as so important, namely explanatory power, simplicity, comprehensiveness, and so on. Probability theory is a more flexible framework than they give it credit for. A serious challenge to legal probabilism must take aim at the very structure of probability, and not at

various inessential theses that some probabilists have occasionally endorsed.

3 Haack's Objections

Susan Haack also opposes legal probabilism.[1] Like Allen and his coauthors, Haack's main objection to legal probabilism is that it fails to account for the role explanatory quality and other epistemological factors should play in determining whether a conviction or finding of liability is warranted. She contrasts probabilistic considerations with these other factors: "we can't look to probability theory for an understanding of degrees and standards of proof in the law, but must look, instead, to an older and less formal branch of inquiry: epistemology" (2014, 47). Setting her unfortunate terminology aside (probability theory is clearly part of the broader discipline of epistemology), Haack rightly insists that considerations of evidential quality should be key components of our understanding of standards of proof. In the law and elsewhere, "whether and if so to what degree evidence is supportive of a claim depends on the contribution it makes to the explanatory integration of evidence-plus-conclusion—on how well the evidence and the conclusion fit together in an explanatory account" (60).

Part of her reason for thinking that legal probabilism must give no role to explanatory considerations rests on a failure to appreciate the options available to the legal probabilist. Whereas Allen and Stein think that frequentism is the only possible view of probability open to the legal probabilist, Haack considers only subjective Bayesianism. (Compare Larry Laudan's critique of legal probabilism, where he asserts that "[t]he problem is that [probabilistic] standards of proof are fundamentally subjective" (2006, 79).) But we have already seen that legal probabilists can reject Subjective Bayesianism and adopt a more objective notion of evidential probability instead.

However, Haack also argues that probability theory does not have the right structure to incorporate facts about evidential support. She give three reasons for thinking this.

First, she notes that "The mathematical probability of (p and not-p) must add up to 1; but when there is no evidence, or only very weak evidence, either way, neither p nor not-p is warranted to any degree" (62). Her talk of propositions being warranted is somewhat idiosyncratic. What does it mean for neither p nor not-p to be warranted to any degree? Presumably, she means that certain propositional attitudes one might take toward p or not-p are not warranted to any degree. And in particular, she likely means (though she does not say so explicitly) that neither belief in p nor belief in not-p is warranted (i.e. justified) to any degree. This is fair enough, but it begs the question against the probabilist. We must distinguish between whether some degree of belief (or subjective probability, or credence) is justified, and the degree to which outright belief is justified. These two things are different and can come apart. Consider a fair coin. A degree of belief of 0.5 in heads is justified, but this does not mean that one who outright

---

[1]Indeed, we have adopted the term 'legal probabilism' from her critique.

believes that heads will result is justified to any degree. The probabilist will not accept the identification of some degree of belief's being justified with some outright belief being justified to some degree. And the probabilist, unlike Haack, thinks that what is most relevant is what degrees of belief are justified, not the degree to which full, outright beliefs are justified.

Second, Haack writes that "The mathematical probability of ($p\&q$) is the product of the probability of p and the probability of q—which, unless both have a probability of 1, is always less than either; but combined evidence may warrant a claim to a higher degree than any of its components alone would do" (62).[2]

Haack is confusing evidence with hypothesis. The axioms of probability entail that the probability of a proposition must be less than or equal to that of any proposition it entails. So in particular $P(p) \leq P(p\&q)$ and $P(q) \leq P(p\&q)$. This means that when a hypothesis is the conjunction of two claims, the hypothesis as a whole cannot have a probability higher than that of either of those two claims. But this is perfectly compatible with the claim that when some *evidence* consists of the conjunction of two claims, the evidence as a whole can support a hypothesis to a greater degree than either of the two claims of which it is the conjunction. That is, the structure of probability allows the possibility that $P(h|e_1\&e_2) > P(h|e_1)$ and $P(h|e_1\&e_2) > P(h|e_2)$.

Third, Haack holds that evidential support is multi-dimensional in the sense that many different sorts of factors—explanatory integration, simplicity, comprehensiveness, and so on—together determine the extent to which some evidence supports some hypothesis. She writes that "[m]athematical probabilities form a continuum from 0 to 1; but because of the several determinants of evidential quality, there is no guarantee of a linear ordering of degrees of warrant" (62).

The idea is that evidential support depends on various epistemic values. But these values sometimes conflict. Sometimes the simplest hypothesis is not the most explanatory, for instance. Moreover, there seems to be no privileged way of weighting them so as to trade them off against each other. There seem to be no grounds for saying that simplicity (on some way of measuring it) gets exactly this much weight, and no more, for instance.

There are numerous options for dealing with this issue in a probabilistic framework. Let us begin by saying that certain ways of assigning precise weights to each of the various epistemic values are admissible, while others are inadmissible. (It may be inadmissible, for instance, to give weight only to simplicity considerations, and none at all to explanatoriness or comprehensiveness.) Each way of assigning precise weights to the various epistemic values can be represented by a prior probability function. So some prior probability functions are admissible and others are inadmissible. Let S be the set of admissible prior probability functions.

We must move, however, from claims about which probability functions are admissible to claims about which doxastic attitudes are rationally permissible. Here are some options:

First, we could say that there is a particular member of S that you ought to

---

[2]Haack misstates the relationship between the probability of a conjunction and the probabilities of the conjuncts here: $P(p\&q) = P(p) \times P(q)$ only if p and q are independent.

have as your prior probability function, but it is indeterminate which member of S it is.[3] Second, we could say that you ought to adopt some member of S as your prior probability function, but that all members of S are permissible. Third, we could give a role to so-called "imprecise probabilities" or "mushy credences" and say that your doxastic state ought to be represented not by a single probability function, but rather by a set thereof. In particular, your prior doxastic state ought to be represented by the set S as a whole, and then if you have total evidence e, your doxastic state ought to be represented by the set consisting of each member of S conditionalised on e.

These different views of the permissibility of doxastic states can then be incorporated into legal standards of proof in various ways. No matter which of the three views we adopt, we could hold that conviction is warranted only if all members of S, conditionalised on the trial evidence, assign sufficiently high probability to the hypothesis that the defendant is guilty. Or, if we adopt the second view, we could say that conviction is warranted if and only if all the members of S adopted by some member of the jury or other, when conditionalised on the trial evidence, assign sufficiently high probability to guilt. Or, if we adopt the first view, we could say that conviction is warranted if and only if the privileged member of S, conditionalised on the trial evidence, assigns sufficiently high probability to guilt (although then there will be cases where is is indeterminate whether conviction is warranted).

We do not take a stand on which of these options should be adopted by the legal probabilist. But we have shown that there are a number of attractive responses to Haack's worry about the multi-dimensional nature of evidential support.

4 The Conjunction Paradox

The so-called 'Conjunction Paradox,' first presented by L. Jonathan Cohen (1977), arises when the allegation to be proven consists of two or more claims that must each be proven in order for conviction or a finding of liability to be warranted. For instance, in a breach of contract lawsuit, in order to prove that the defendant is liable for a breach of contract and must pay damages, the plaintiff must prove both (i) that there was a valid contract in place, and (ii) that the defendant failed to fulfil the contract. In a civil suit such as this, the relevant standard of proof is preponderance of the evidence, which is most naturally understood by legal probabilists as requiring evidential probability greater than 0.5. But it is possible for each of (i) and (ii) to receive probability greater than 0.5 even though the claim as a whole (that is, the conjunction of (i) and (ii)) has a probability below 0.5. So, on a probabilistic understanding of preponderance of the evidence, it is possible for each component of the charge to be proven to the relevant standard, but for the charge as a whole to fall short.

---

[3]This indeterminacy-based view can be combined with various views about what indeterminacy amounts to. These views have been discussed extensively in the philosophical literature on vagueness. See, for example, Sorensen (2016), Williamson (1994), and Keefe (2000).

While mathematically and probabilistically-minded legal scholars have proposed numerous sophisticated ways of dissolving the paradox (Cheng 2013; Clermont 2013), our response is straightforward. The conjunction paradox may be a problem for descriptive legal probabilism, the thesis that, given how our legal system is in fact designed and how it in fact functions, standards of proof are best understood probabilistically. If existing law says that it is sufficient for a conviction or finding of liability that each component of the case meet the relevant standard of proof, and also that it is necessary for a conviction or finding of liability that the case as a whole (the conjunction of the components) meet the same standard of proof, then there really is a problem with understanding standards of proof as probability thresholds.

But our aim is not to defend descriptive legal probabilism but rather *normative* legal probabilism. Our concern is with whether an ideal (or at least more ideal) legal system would employ probabilistic standards of proof. And here, the solution is simple: We should reject the policy of holding both that proving each component of the case to the relevant standard is sufficient for conviction or liability and also that proving the conjunction of the components to the relevant standard is necessary for conviction or liability. One or the other must go. The natural thing to do is to reject the former claim. Our concern should be with whether the case as a whole, the entire allegation, meets the relevant standard of proof. So in the breach of contract suit, a finding of liability is warranted if and only if the conjunction of (i) and (ii) has probability above 0.5. It is not enough that each of (i) and (ii) has probability above 0.5 if their conjunction does not. This strikes us as a modest and sensible improvement to the legal system and one which dissolves the conjunction paradox for legal probabilism.

5   The Problem of Thresholds

Legal probabilism requires that we give probabilistic interpretations of the various standards of proof. Of course, we can just plow ahead and assign some numbers as threshold, say 0.5 probability for preponderance of the evidence, 0.7 for clear and convincing evidence, and 0.95 for beyond a reasonable doubt. But there may be good reason to resist setting sharp thresholds like this.

The problem is sharp thresholds may raise a risk of gaming the system. After all, in any case requiring clear and convincing evidence, it is all too clear what the defence must do: provide evidence that there is at least 0.3 chance that the defendant is not guilty or liable. While there is nothing wrong with this in the abstract, there is good reason to suspect that allowing such a clear target will invite various misuses of statistics such as cherry-picked data and the like. It is plausible that keeping some unclarity about the standards of proof in question makes such gaming more difficult (although certainly not impossible). Moreover, a case can be made for there being some flexibility in the interpretation of 'beyond reasonable doubt' and 'clear and convincing' on a case-by-case basis. For example, consider two criminal trials, both requiring beyond reasonable doubt as the standard of evidence, but one with a life sentence on offer and the other

attracting monetary fines. It may not be out of order for the interpretation of 'reasonable doubt' to be a little more stringent in the case with the significantly harsher penalty. But if the standard of evidence in question is stated precisely in probabilistic terms, this is not an option.

We are sympathetic to this line of thought. One natural response a probabilist might make is to invoke imprecise probabilities for the relevant thresholds. For example, we could set beyond reasonable doubt as, say, 'extremely high probability' (or perhaps something a bit more precise like 'about 0.95') and clear and convincing as 'quite probable' (or 'about 0.7'). This would make the most blatant gaming more difficult and would also allow for some consistency across judgements. Moreover, we could explicitly allow the relevant probability thresholds to depend on the seriousness of the penalties under consideration. This can be spelled out in decision theoretic terms and, indeed, this is plausibly the right way to look at the duty of a jury: in a criminal trial they are charged with *making a decision* about whether the defendant is to be treated as guilty or innocent. On this suggestion, then, we could require slightly higher standards of evidence when a death penalty or life in prison is on the table and slightly lower standards when the punishment is, for example, a fine.

A final worry about explicitly probabilistic thresholds (even vague ones): Unless the probability threshold is 1, setting the threshold makes explicit our willingness to tolerate some false positives and convict innocent people (Tribe 1971). Suppose that juries' probability judgments are *well-calibrated* in the sense that, for every 100 cases in which they judge the probability of an event to be n, the event obtains in $n \times 100$ of them. Thus, for a threshold for conviction of 0.95, it follows that if juries' probabilistic verdicts are well-calibrated, as many as 5% of convictions will be false.[4]

We do not find this worry troubling. First, all acknowledge that innocent people have been, are, and will be mistakenly convicted. Of course, such false convictions are troubling and should be reduced as much as feasible, but it is impossible to avoid sometimes convicting the innocent, unless we stop convicting anyone at all. Second, the relevant threshold should be thought of not as some arbitrary number pulled out of a hat, but rather as something to be debated among citizens attempting to settle the relative values or disvalues to assign to true convictions, false convictions, true acquittals, and false acquittals (DeKay 1996; Laudan 2006, 2008). The threshold is no more arbitrary than is the assignment of relative values to these different possibilities. Once we think of the threshold in that way, there should be (*contra* Tribe) no great objection to making explicit this socially, democratically determined choice.

---

[4]Note however that no conclusions about the rate of false convictions follow without the assumption of well-calibration. The rate of false convictions depends not only on the standard of proof, but also on the ratio of truly guilty to truly innocent people who are brought to trial, as well as the probability that the total evidence at trial is misleading. See DeKay (1996) and Laudan (2008) for discussion.

6 Bare Statistical Evidence

Perhaps the most serious objection to legal probabilism is that it is committed to the legitimacy of convictions and findings of liability based on *bare statistical evidence*. Statistical evidence on its own can push the probability of guilt or liability above whatever threshold is set, but it seems inappropriate to convict someone or find her liable without any 'individualised' evidence tying her to the act in question.

Consider the famous Blue Bus/Red Bus case (Tribe 1971). A pedestrian is injured by a vehicle. An eyewitness identifies the vehicle as a bus but, due to low light, cannot tell what colour it is. However, there is data showing that the Blue Bus Company operates 70 percent of the buses in the area, while the Red Bus Company operates 30 percent. In a civil case such as this, the relevant standard of proof is preponderance of the evidence, which is most naturally interpreted by the legal probabilist as a probability threshold of 0.5. It seems, then, that the bare statistical evidence puts the probability of the Blue Bus Company's liability above 0.5, licensing a finding of liability. But it seems illegitimate to hold the Blue Bus Company liable here.

In the gatecrasher case (Kaye 1978), one thousand people are found at an event for which only twenty tickets were sold. John is arrested and charged. Even though there is no individualised evidence identifying him as a gatecrasher (a videotape showing him jumping the fence, say), the statistical evidence alone seems to put the probability of his guilt around 0.98, which is surely above whatever threshold the legal probabilist might set for conviction. But it seems illegitimate to convict John on this basis alone.[5]

It is tempting to conclude from these examples that something beyond mere statistical evidence is required for conviction or liability, namely some kind of individualised evidence. There is no consensus on what exactly individualised evidence is, or why it is important. We will not attempt to survey all the proposals that have been advanced in the literature. Our concern is simply whether bare statistical evidence constitutes a problem for the legal probabilist. In what follows, we take a two-pronged strategy to defending legal probabilism. First, we argue that it is far from clear that there is something wrong *in principle* with convictions and findings of liability based on bare statistical evidence. There is admittedly an intuition that such decisions are inappropriate, but it is difficult to go beyond this intuition and identify a genuine in-principle problem with them. Second, we argue that the legal probabilist has the resources to say that there is generally something wrong *in practice* with basing a conviction or finding of liability on statistical evidence alone.

---

[5]Although the bus-company example is based on a real case (Smith v. Rapid Transport Massachusets 1945), these two examples are clearly unrealistic in various ways. There are, however, real cases where bare statistical evidence is presented as being sufficient on its own. For all their dangers, however, we will stick with these toy examples for the moment.

6.1 Is There a Problem With Bare Statistical Evidence?

Before turning to how legal probabilists might criticise certain uses of bare statistical evidence, let us suppose for the sake of argument that they are committed to the claim that bare statistical evidence can be sufficient for a finding of guilt or liability. If this were so, would it constitute an objection to legal probabilism? In this section, we consider and reject two kinds of arguments for the inappropriateness of basing conviction or a finding of liability on bare statistical evidence. The first is consequentialist in spirit and adverts to bad incentives created by allowing bare statistical evidence to be sufficient. The second is non-consequentialist and appeals to considerations of fairness and to the appropriateness of both assertions of guilt and reactive attitudes like blame.

Start with incentives. Enoch et al. (2012) argue that allowing bare statistical evidence to suffice for conviction creates perverse incentives. Here is their argument, based on the gatecrasher case presented above:

> Think, for instance, about John, a potential gatecrasher who is now deliberating, considering either purchasing a ticket, or perhaps gatecrashing, or perhaps going home and doing something else altogether. We are assuming, of course, that John has no influence on the behaviour of the others at and near the stadium. This means that he has almost no influence on the relevant statistical evidence: the percentage of those attending the event at the stadium who did not purchase a ticket is only to a minuscule degree influenced by the conclusion of John's deliberation. For all intents and purposes, he should think of it as already given. If so, though, our willingness to rely on statistical evidence almost entirely annihilates whatever incentive the substantive criminal law can give John not to break the law. For if the statistical evidence is strongly against him—say, because 98 percent of those attending are gatecrashers—John already knows that he will be convicted, regardless of whether or not he buys a ticket. And if the statistical evidence is not strongly against him, he knows that it will constitute strong exonerating evidence, whether or not he is guilty of gatecrashing. Either way, then, he might as well go ahead and gatecrash: whether he does or not will have very small influence on his chance of being punished. (217)

The first thing to note is that even if treating bare statistical evidence as sufficient for conviction did sometimes create perverse incentives, this fact would not by itself show that we should always require individualised, non-statistical evidence. For the creation of perverse incentives must be weighed against any potential benefits of allowing for the sufficiency of bare statistical evidence, including potential improvement in the accuracy of verdicts. Indeed, improved accuracy of verdicts would presumably yield stronger incentives not to commit crimes.

But more to the point, we deny that allowing bare statistical evidence to be sufficient creates these perverse incentives. Start with the case where the statistical evidence is strongly against some would-be offender. We contend that such persons still have an incentive not to commit the crime in question, since refraining from doing so can mean they will have exonerating evidence, or even that there's no crime to be investigated in the first place. In the gatecrasher case, for instance, Enoch et al. stipulate that John has the option of going home, in which case he'll have an alibi (and moreover will be unlikely to be arrested in connection with the gatecrashing incident in the first place). Or consider a person who is considering whether to murder her partner. She knows that if she doesn't commit the crime, she won't be convicted of murder, even if (were she to commit the murder) the bare statistical evidence would be strongly against her.

Now consider a case where the statistical evidence is not strongly against our would-be offender. Again, she still has an incentive not to commit the crime in question. This is because, even if the statistical evidence in play would not suffice for conviction, by committing the crime she would risk leaving behind individualised evidence which would lead to her conviction. So, for instance, John might be caught on CCTV hopping the fence if he decides to gatecrash. And a potential murderer might leave fingerprints and bloodstains.

Allowing for bare statistical evidence to suffice for conviction does not, therefore, yield perverse incentives: first because would-be offenders are generally in control of whether or not there's a crime committed in the first place and, second, because their actions can result in individualised evidence which has the potential to undermine or supersede the statistical evidence in play. This would lead to acquittal in a case where the statistical evidence is against them or to conviction in a case where it is in their favour.

It is true that in some cases, allowing for conviction on the basis of bare statistical evidence will not yield sufficiently strong incentives against committing a given crime. For instance, if John is dead set on going to the concert no matter what, and if he is unable to generate individualised exonerating evidence (he cannot get a receipt for the ticket and cannot get any other concert-goers to credibly vouch for his having purchased one), then he may as well gatecrash, since the statistical evidence will be against him either way.

Two points are worth making in response. First, John still has an incentive to go home rather than gatecrash, even if this incentive is outweighed by his strong desire to go to the concert. And we cannot expect any realistic legal system to be such that people always have all-things-considered sufficient reason not to commit a crime, no matter their other desires. Second, analogous problematic cases can arise even if only individualised evidence is deemed sufficient for conviction. Suppose someone is walking home after practicing his baseball swing alone in the park, when he happens upon a riot in progress. He then notices a bystander who clearly sees him right near the rioters carrying a baseball bat. If he doesn't see any way to generate exonerating evidence, he may conclude that he's likely to be convicted of rioting no matter what, given the (misleading) individualised evidence consisting of the eyewitness testimony. And so, if he has a sufficiently strong desire to join in, he may decide to do so. To be sure, such cases are

extremely rare, but then again so are gatecrasher-type cases! We conclude that while allowing bare statistical evidence would inevitably *change* the incentives facing would-be criminals, it is doubtful that it would yield weaker incentives not to commit crimes than would a system allowing only individualised evidence to suffice for conviction. Indeed, to reiterate a point made earlier, allowing statistical evidence would likely yield stronger incentives to refrain from criminal activity, to the extent that allowing statistical evidence would increase the overall accuracy of verdicts.

We now turn to less consequentialist approaches to the alleged badness of bare statistical evidence. First, it might be thought that convicting someone on the basis of bare statistical evidence is unfair and hence violates her rights. But it is not clear why exactly such a conviction would be unfair. If the person is innocent, then there is some sense in which punishing her is unfair and violates her right not to be punished for something she didn't do, but this is so regardless of whether the conviction is based on statistical as opposed to individualised evidence. Perhaps the thought is that where the conviction is based on bare statistical evidence, the conviction isn't sensitive to how she acted, and there's nothing she could have done to avoid conviction. Two things in response: For one, it's unclear why this lack of sensitivity would be unfair or constitute a violation of her rights. For another, the conviction still is sensitive to how she acted insofar as she could have done something which might have yielded exonerating individualised evidence, as we saw above (though given her innocence, this is not something she could have been expected to try to do).

Second, some opponents of bare statistical evidence have suggested that conviction is tied to assertion and to blame, and that mere high probability is not in general sufficient to warrant either. Take assertion. It is widely thought that in a fair lottery, one is not licensed to outright assert 'My ticket will lose,' no matter how large the lottery. And this is because the grounds for one's confidence that one's ticket will lose are merely statistical. Philosophers have argued instead that one is only licensed to assert propositions that one *knows* (Williamson 2000), or perhaps believes, where mere high probability is not sufficient for knowledge or outright belief. Thomson (1986, 213) and Ho (2008, 140–3) connect this thought with the inappropriateness of reliance on bare statistical evidence by saying that a conviction amounts to an assertion that the defendant is guilty.

Similarly for blame. Buchak (2014) argues that a mere high probability that someone has wronged you is insufficient to make it appropriate for you to blame her. She considers a case where you leave your iPhone in a room at a party and return to find it missing. There are only two people in the room, Jake and Barbara. If all you know is that men are 10 times more likely to steal iPhones than women (and that there is no other relevant statistical or individualised evidence), then you would not be justified in blaming Jake, even though it is very likely, given your evidence, that he stole it. By contrast, if you had some individualised evidence tying him to the theft, that would license blame, and could potentially do so even if it didn't make his guilt any more probable than in the case where you just have statistical evidence to go on. Buchak concludes that blame is warranted only if you outright believe that the person committed

the offence; mere high probability does not suffice.

One might use this line of thought to argue against the use of bare statistical evidence. After all, if conviction constitutes a way for the state to blame the defendant for the crime, or if it is intended to signal to other citizens that such blame is appropriate, then it should be based on grounds that could at least potentially suffice to warrant such blame.

We are prepared to concede (at least for the sake of argument) that mere high probability does not suffice to warrant assertion or blame. Insofar as existing legal practice does treat conviction as an assertion of the defendant's guilt, or a way of blaming the defendant for the crime, then this would constitute an objection to a descriptive version of legal probabilism. But, as we have emphasised, we are defending a normative version of legal probabilism, that is, the claim that an ideal (or, at least, better) legal system would use probabilistic standards of proof. And we think it entirely appropriate for such a legal system to treat conviction not as an outright assertion that the defendant is guilty, but rather as an assertion that the defendant is *very probably* guilty. We also think it appropriate to move away from a retributivist picture on which conviction involves blaming the defendant and giving her her just deserts.

Absent a strong argument in favour of tying conviction to assertion and blame, we conclude that there is no good in-principle reason against permitting conviction on the basis of bare statistical evidence. (It also bears noting that even if criminal conviction is tied to blame and an assertion of guilt, and hence requires individualised evidence, this would not yield an argument that findings of liability in civil cases require individualised evidence. After all, civil liability is not tied to blameworthiness, and the fact that a finding of civil liability typically requires only preponderance of the evidence means that such a finding cannot be understood as an outright assertion that the defendant committed the tort in question.)

Having said that, there are good in-practice reasons against permitting such convictions. In the next section, we show that the legal probabilist has the resources to criticise reliance on bare statistical evidence in real-world cases.

6.2 Must Legal Probabilists Always Embrace Bare Statistical Evidence?

The legal probabilist is committed to holding that there is no *in principle* problem with convictions or findings of liability based on bare statistical evidence. But such finds may often be problematic *in practice*. And so the legal probabilist has the resources to account for the intuition that something beyond bare statistical evidence is typically necessary.

In particular, when given statistical evidence supporting guilt or liability, it is important to ask after the statistical model used to derive the probabilities in question. In toy examples, things can be set up in such a way that we can move smoothly from the statistical evidence in question to the probability of guilt or liability. But in more realistic cases, things are not so straightforward. For

instance, we may have frequency data for multiple reference classes, in which case we face the question of how to amalgamate all that frequency data into a single probability judgment. (Why can't we just pick the smallest reference class as the relevant one? There may be no single smallest reference class for which we have frequency data, and even if there is a smallest, it may not be the most relevant one.) We may also have doubts about the reliability of that frequency data.[6]

Consider a realistic statistical evidence case in which there is a DNA match between a crime-scene sample and a person in a DNA data base (e.g. People v. Pucket, California 2009, as described in Roth 2010). It is important that the person matched with the DNA is not a suspect for any other reason, otherwise it would not be a case of pure DNA evidence.[7] In such a case, the probability of a false positive is reputedly of the order of 10^{-11}. Therefore, it seems that the DNA match must meet any reasonable probabilistic standard of proof, including a probabilistic understanding of beyond reasonable doubt. Hence the probabilist is seemingly committed to considering a DNA match as sufficient for a conviction. Others feel that something more is required.

But the probabilist should question our confidence in the model that delivered the probability of a false positive. After all, the figure of 10^{-11} assumes many things. For instance, it assumes that the sample was indeed from the crime scene in question, that the crime scene was not contaminated, that DNA samples were not mixed up at the lab, that the lab work was carried out properly, and so forth. In short, false positives are very rare so long as nothing has gone wrong. But now we need to know the probabilities of errors of the sorts just listed. It is noteworthy that any such errors do not simply raise the probability of a false positive a little; errors of the sort under consideration undermine the very statistical model used to deliver the figure of 10^{-11}. That is, in a case of crime-scene contamination, say, we do not start with 10^{-11} as the prior probability and then update. Rather, the contaminated crime scene means that the false positive figure of 10^{-11} was delivered by a flawed statistical model and all bets are off.

It is thus perfectly reasonable for a probabilist to hold out for further evidence in such cases. The probabilist need not be always satisfied with bare statistical evidence. It is interesting that what is asked for here is not simply more (first-order) probabilities, nor is it something non-probabilistic. What is required is a kind of meta-probability: evidence that the statistical model used to calculate the probability of a false positive is reliable. Such evidence could be provided by statistics about crime-scene contaminations, mixed-up samples in the lab, and the like. Less obvious forms of validation of the statistical model can come

---

[6]See Steele and Colyvan (to appear) for more on the problems with such toy cases; and see Colyvan et al. (2001) and Colyvan and Regan (2007) for discussion of a real case of statistical evidence in the court room and how uncertainty abut the statistical model, specially reference-class considerations, can undermine the bare statistical evidence.

[7]Although DNA evidence may seem specific to the person, such matches are purely statistical and thus are cases of bare statistical evidence in the required sense. Indeed, there is substantial debate over the use of purely DNA evidence in the courtroom for precisely the reasons under discussion here. See Balding and Steele (2015) for an excellent discussion of DNA evidence, its use in the court room, examples of such cases, and some of the problems associated with such use.

from showing that there is independent reason for being suspicious of the person matched with the DNA. Circumstantial evidence such as finding that the matched person had motive, means, and opportunity to commit the crime would help shore up our faith in the bare statistics. So would some specific evidence such an eyewitness placing the suspect at the crime scene at the time of the crime. But such evidence shores up the bare statistics not by lowering the probability of a false positive to something below 10^{-11}. For 10^{-11} is already sufficiently low. This extra circumstantial or specific evidence can be thought of as increasing our confidence that the statistical model itself is reliable. It thus gives us reason to believe that the probability of a false positive is indeed something like 10^{-11}. The extra evidence is thus a kind of meta-evidence but no less probabilistic for this.

The probabilist is thus not compelled to act on the basis of bare statistical evidence, although, under certain ideal conditions, they may. Whether bare statistical evidence suffices will depend on the case. And it is noteworthy that when the probabilist does hold out for more evidence, the evidence they could reasonably hold out for (e.g., motive, means and opportunity) is precisely the kind of further evidence that many critics of probabilism see as absent from the probabilist's account.

7 Probabilism Survives

We have defended legal probabilism—the view that legal standards of proof are best understood in probabilist form—against various objections. In so doing, we have also helped highlight some of the benefits of thinking of legal evidence in these terms. We have not, however, provided a positive argument for legal probabilism. Our main concern here has been to defend the view against various objections.

A common theme in our replies to the objections has been that the legal probabilist has more resources at her disposal than critics appreciate. For example, the legal probabilist need not interpret probabilities as relative frequencies or subjective degrees of confidence, but can instead interpret probabilities as degrees to which the evidence supports a given claim, which will be sensitive to the explanatory considerations emphasised by opponents of legal probabilism. And the legal probabilist can make room for imprecise probabilities and refuse to endorse sharp, precise probability thresholds for the various standards of proof. Additionally, the legal probabilist need not ignore uncertainty about the statistical models used to deliver the statistical evidence. To be fair to the critics though, many of our responses have relied on fairly recent developments in formal epistemology. The criticisms in question may well be problematic for older, less sophisticated understandings of probabilism. Our discussion thus helps to highlight the importance of a continued dialogue between legal epistemology and formal epistemology.[8]

References

Allen, R. 1991. 'The Nature of Juridical Proof', *Cardozo Law Review*, 13(2–3): 373–342.

Allen, R. 2017. 'The Nature of Juridical Proof: Probability as a Tool in Plausible Reasoning', *International Journal of Evidence and Proof*, 21(1–2): 133–142.

Allen, R. and Leiter, B. 2001. 'Naturalized Epistemology and the Law of Evidence', *Virginia Law Review*, 87: 1491–1550.

Allen, R. and Stein, A. 2013. 'Evidence, Probability, and the Burden of Proof', *Arizona Law Review*, 55: 557–602.

Balding, D.J. and Steele, C.D. 2015. *Weight-of-Evidence for Forensic DNA Profiles*, 2nd edition, Chester UK: Wiley.

Buchak, L. 2014. 'Belief, Credence, and Norms', *Philosophical Studies*, 169: 285–311.

Cheng, E. 2013. 'Reconceptualizing the Burden of Proof', *Yale Law Journal*, 122(5): 1254–1279.

Clermont, K. 2013. 'Death of Paradox: The Killer Logic beneath the Standards of Proof', *Notre Dame Law Review*, 88(3): 1061–1138.

Cohen, L. Jonathan. 1977. *The Probable and the Provable*, Oxford: Oxford University Press.

Colyvan, M. and Regan, H.M. 2007. 'Legal Decisions and the Reference-Class Problem', *International Journal of Evidence and Proof*, 11(4): 274–285.

Colyvan, M., Regan, H.M. and Ferson, S. 2001. 'Is it a Crime to Belong to a Reference Class?', *The Journal of Political Philosophy*, 9(2): 168–181. Reprinted in H.E. Kyburg and M. Thalos (eds.), *Probability is the Very Guide of Life*, Open Court, Chicago, 2003, pp. 331–347.

DeKay, M.L. 1996. 'The Difference Between Blackstone-Like Error Ratios and Probabilistic Standards of Proof', *Law and Social Inquiry*, 21(1): 95–132.

Enoch, D., Spectre, L. and Fisher, T. 2012. 'Statistical Evidence, Sensitivity, and the Legal Value of Knowledge', *Philosophy and Public Affairs* 40 (3): 197–224.

Goodman, N. 1955. *Fact, Fiction, and Forecast*, Cambridge, Ma: Harvard university Press.

Haack , S. 2014. *Evidence Matters: Science, Proof, and Truth in the Law*, New York: Cambridge University Press.

Hájek, A. 2007. 'The Reference Class Problem is Your Problem Too', *Synthese*, 156(3): 563–585.

Hedden, B. 2015. 'A Defense of Objectivism about Evidential Support', *Canadian Journal Of Philosophy*, 45(5–6): 716–743.

Ho, H.L. 2008. *A Philosophy of Evidence Law*, New York: Oxford University Press.

Jaffee, L.R. 1984–5. 'Of Probativity and Probability: Statistics, Scientific Evidence, and the Calculus of Chances at Trial', *University of Pittsburgh Law Review* 46: 925–1083.

Kagehiro, D. and Stanton, W. 1985. 'Legal vs. Quantified Definitions of Standards of Proof', *Law and Human Behavior* (9): 159–178

Kaye, D. 1978. 'The Paradox of the Gatecrasher and Other Stories,' *Arizona State Law Journal*, 101–143.

Keefe, R. 2000. *Theories of Vagueness*, Cambridge: Cambridge University Press.

Laudan, L. 2006. *Truth, Error, and Criminal Law: An Essay in Legal Epistemology*, Cambridge: Cambridge University Press.

Laudan, L. 2008. 'The Elementary Epistemic Arithmetic of Criminal Justice', *Episteme*, 5(3): 282–294.

Posner, R.A. 1973. *Economic Analysis of Law*, Boston: Little, Brown and Company.

Roth, A. 2010. 'Safety in Numbers? Deciding When DNA Alone is Enough to Convict', *New York University Law Review*, 85: 1130–1185.

Sorensen, R. 2016. 'Vagueness', in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), URL = https://plato.stanford.edu/archives/win2016/entries/vagueness/.

Steele, K. and Colyvan, M. To Appear. 'Crime, Punishment and Inadequate Evidence'.

Titelbaum, M. 2010. 'Not Enough There There: Evidence, Reasons, and Language Independence', *Philosophical Perspectives*, 24(1): 477–528.

Thomson, J. 1984. 'Remarks on Causation and Liability', *Philosophy and Public Affairs*, 13 (2): 101–133.

Tribe, L.H. 1971. 'Trial by Mathematics: Precision and Ritual in the Legal Process', *Harvard Law Review*, 84: 1329–1393.

Weisberg, J. 2009. 'Locating IBE in the Bayesian Framework', *Synthese*, 167(1): 125–143.

Williamson, T. 1994. *Vagueness*, London: Routledge.

Williamson, T. 2000. *Knowledge and Its Limits*, Oxford: Oxford University Press.