

The Trouble with Standards of Proof¹

ZOË JOHNSON KING
University of Michigan

I. Preliminaries

i. Four trial outcomes

Anyone who stands trial is either innocent or guilty of the crime(s) of which they are accused. And anyone who stands trial is either convicted or acquitted. This means that each and every trial will end with one of four possible outcomes: a conviction of the guilty, a conviction of the innocent, an acquittal of the innocent, or an acquittal of the guilty.

There is a clear sense in which the first and third of these four outcomes are “correct”, while the second and fourth are “errors”; the main point of a trial system is to convict the guilty and acquit the innocent, not the other way around. Following a trend in the literature on this topic, I will use the epithet “false” to describe the erroneous verdicts and “true” to describe the correct ones, as follows:

	Guilty	Innocent
Acquitted	False acquittal	True acquittal
Convicted	True conviction	False conviction

Different trial systems may produce different *ratios* between these four trial outcomes over a given period of time. For instance, if one trial system produces 100 true acquittals and 25 false convictions over the course of a year, then it is producing these outcomes in a 4:1 ratio. If another trial system produces 100 true acquittals but only 10 false convictions, then it is producing these outcomes at a different ratio to the first system – the ratio in the second system is 10:1. But a third trial system could produce these two trial outcomes in the same *ratio* as the first system, even if the *numbers* of outcomes are different; for example, if it produces 800 true acquittals and 200 false convictions, or 200 true acquittals and 50 false convictions, or even just 4 true acquittals and 1 false conviction, then it is still producing these outcomes in a 4:1 ratio.

ii. The Ratio-to-Standard Strategy

This paper challenges a family of arguments that take these facts about ratios between trial outcomes to have tremendous import for the project of choosing a standard of proof to use in criminal trials. I will call this family of arguments the “Ratio-to-Standard Strategy”. Here, in a summative nutshell, is the strategy:

¹ I am grateful to the participants and organizers of the II Harvard Graduate Legal Philosophy Colloquium, and to Boris Babic, Sarah Buss, David Manley, and Sarah Moss, for extremely helpful discussion of earlier drafts of this paper. Special thanks to Scott Hershovitz for superlative insight and encouragement.

RATIO-TO-STANDARD-STRATEGY: We should choose a standard of proof by first settling on an ideal ratio between two or more of the four possible trial outcomes, and then choosing the standard that we can expect to produce this ratio of outcomes in the long run.

The Ratio-to-Standard Strategy is, in one sense, popular: there are a great many examples of scholars using something like this strategy in their writings on the standard of proof (e.g., Kaplan 1968; Cullison 1971; Bell 1987; Connolly 1987; DeKay 1996; Lillquist 2002; Laudan 2006, chap.3).² Perhaps the most commonly-cited ratio for these purposes is a 10:1 ratio of false acquittals to false convictions; a preference for this ratio is sometimes ascribed³ to the eminent English jurist William Blackstone, and it has become well-known since Blackstone's writings. But the Ratio-to-Standard strategy need not be based on the "Blackstone ratio". An enormous range of putative ideal ratios of false acquittals to false convictions, ranging from a modest 1:1 to a whopping 1000:1, have been floated over the course of history (see Volokh 1997 for a disarmingly broad array of intuitions on this question). The Ratio-to-Standard strategy could use any of them. Indeed, the Ratio-to-Standard Strategy is not even bound to begin with a ratio of some kind between the two trial outcomes – false acquittal and false conviction – with which Blackstone was concerned. There are lots of other ratios between pairs of trial outcomes that we might be interested in, and I will consider a wide range of them in this paper. For now it is worth bearing in mind that, as popular as the Ratio-to-Standard Strategy already is, the number and variety of possible versions of this strategy far outstrips the fairly narrow range on which the scholarly discussion has so far been concentrated.

Despite its popularity, the Ratio-to-Standard Strategy is very far from securing univocal support; it has had fierce detractors for about as long as it has had defenders. Notably, Tribe (1971) worries that any use of mathematical calculation may "operate to distort – and, in some cases, to destroy – important values which... society means to express or to pursue through the conduct of legal trials" (p.1330). He also suggests that there is something "basically immoral" about an explicit attempt to assess the comparative costs of false conviction and false acquittal (p.1373), and that such attempts can "make the legal system seem... alien and inhuman" (p.1376). Building on Tribe's remarks, Stein (2005, pp.172-83) argues that for a trial system to make explicit efforts to balance the distribution of erroneous verdicts is for it to *intentionally* convict some innocent people, a serious moral wrong. And Walen (2015, p.419) notes that our assessment of the comparative degrees of (dis)value of the four possible trial outcomes can be affected by intuitively undesirable factors, such as the deterrent effects that may result from a false conviction so long as the public believes that the defendant was guilty.

This paper raises an altogether different sort of worry for the Ratio-to-Standard Strategy. It is not a worry raised, so to speak, "from the outside" – from someone unsympathetic to the whole approach of using the

² Some of these authors prefer to imagine jurors directly intuiting the utility of a false conviction and of a false acquittal, rather than thinking of an ideal ratio between them. But these two methodologies are mathematically equivalent; one's preferred ratio of false acquittals to false convictions will be equal to what is called the "ratio of differences" between their utilities ([REF]). For example, Blackstone's 10:1 ratio is equivalent to the claim that the disutility of an additional false conviction is ten times that of an additional false acquittal.

³ This ascription uses some poetic license, though it is based on genuine comments in Blackstone's *Commentaries on the Laws of England* (1765). Some have traced the lineage of the ratio further, finding it in Genesis 18:23-32, where God agrees not to destroy the city of Sodom if at least ten righteous men can be found there. Whether this in fact embodies the Blackstone ratio depends on how many people in total were living in Sodom at the time.

mathematical tools of decision theory to set the standard of proof in a precise way. I am wholly sympathetic to this general approach. Rather, this is a worry that I raise “from the inside”. The worry is that the Ratio-to-Standard Strategy fails on its own terms. It fails because it tries to do something that we simply cannot do, whether or not we would like to do it. Moreover, the Strategy fails in a completely general way: we are not in a position to read a standard of proof off of *any* desired ratio between *any* set of trial outcomes. There is never any such thing as “the standard that we can expect to produce this ratio of outcomes in the long run”. This is because calculating such expectations would require information that we simply do not have, and that – more worryingly still – we are not in a position to acquire.

II. The Problem

i. A well-known problem

It is well-recognized in the existing literature that all versions of the Ratio-to-Standard Strategy beginning with a desired ratio of false acquittals to false convictions (like the Blackstone ratio) will face a significant hurdle. The hurdle for these versions of the Strategy is that we cannot confidently say, of any particular standard of proof, that it *will* produce a certain ratio of false acquittals to false convictions. This is because the ratio of false acquittals to false convictions is not determined by the standard of proof alone. For one thing (I will discuss another in §II.ii), this ratio depends also on the *proportion* of people on trial who are genuinely guilty as opposed to genuinely innocent. A trial system in which a high proportion of defendants are genuinely guilty can produce more false acquittals and fewer false convictions overall, hence more false acquittals *per* false conviction, than one in which fewer defendants are genuinely guilty, even if everything else – including the standard of proof – is held fixed.

We can visualize this point using a pair of diagrams:

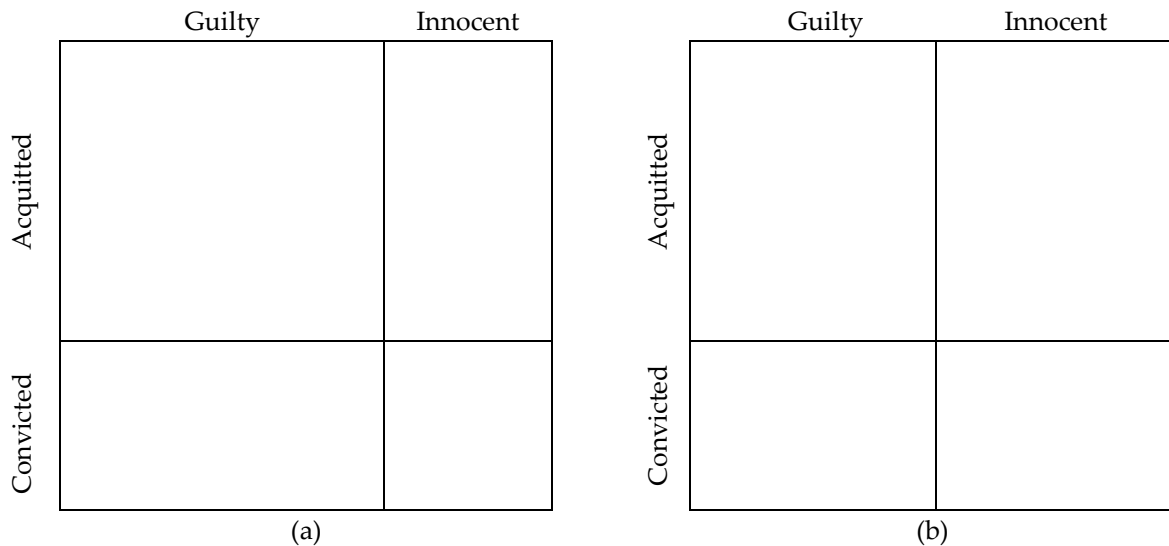


Fig. 1: Two possible sets of ratios of trial outcomes

These diagrams represent two possible trial systems, (a) and (b), whose proportions of trials with various outcomes are represented by area. In both systems, the horizontal line separating the “Acquitted” row from the “Convicted” row is roughly two-thirds of the way down, meaning that in both trial systems roughly two-thirds of trials end in acquittal and one-third in conviction. But the vertical line separating the “Guilty” column from the “Innocent” column, which represents the proportion of people on trial who are genuinely guilty as opposed to genuinely innocent, is different across these two systems. In system (a), the vertical line is also roughly two-thirds of the way along, meaning that two-thirds of defendants are genuinely guilty and the remaining one-third are innocent. But in system (b), the line is right in the middle, meaning that half of defendants are genuinely guilty and the other half are innocent.

We can see the ratios between trial outcomes in these two systems by comparing the sizes of the four boxes into which they are subdivided. The categories remain the same as those in the table in §1.i: the upper-left box represents trials in which the defendant is guilty but acquitted (i.e., false acquittal), the upper-right box represents trials in which the defendant is innocent and acquitted (i.e., true acquittal), and so on. The ratio of false acquittals to false convictions in each system is represented by the comparative size of the upper-left and lower-right boxes. It is easy to see that this ratio is different in the two diagrams; the “Guilty and Acquitted” box is bigger in system (a) than it is in (b), while the “Innocent and Convicted” box is bigger in system (b) than it is in (a). Indeed, if the proportions are as I have stipulated them, then in trial system (a) false acquittals make up 4/9 of trials and false convictions make up 1/9, whereas in trial system (b) false acquittals make up 2/6 of trials and false convictions make up 1/6. This yields a 4:1 ratio in the first system and a 2:1 ratio in the second. In other words, there are *twice as many false convictions per false acquittal* in system (b) as in system (a). And this is so despite the fact that the conviction rate is the same across the two systems. The only thing that varies is the proportion of people on trial who are genuinely guilty.

This creates an important problem for versions of the Ratio-to-Standard strategy focussed on something like Blackstone’s ratio. By changing the standard of proof, we can alter the conviction rate; lower standards will usually lead to more convictions and higher standards to more acquittals, if everything else is held fixed. But the ratio of false acquittals to false convictions is not determined *solely* by the conviction rate. As the diagrams in Fig.1 show, this ratio can be dramatically different even across systems whose conviction rate is the same. It changes depending on the proportion of defendants on trial who are genuinely guilty. Now we have an important problem. For we do not know which defendants really are guilty and which are not – if we did, then there would be no point in having trials! More importantly, even at the end of a trial it is still possible that a defendant who has been *found* guilty is *actually* innocent, or *vice versa* – this is the possibility that gives rise to worries about erroneous verdicts. Here, then, is the problem: since we do not know who is actually guilty and who is not, we do not know which of the convictions and acquittals that our trial system is producing are correct and which are erroneous. So we do not know what *ratio* of false acquittals to false convictions our system is producing. Nor, then, are we in a position to determine how any proposed change to the standard would affect this ratio. Indeed, without knowing who is actually guilty and who is not, we could not even tell whether we were in trial system (a) or trial system (b).

The implications of this point for defenders of the Ratio-to-Standard Strategy have been noted by Allen (1977), Bell (1987), Connolly (1987), DeKay (1996), Lillquist (2002), and Laudan (2006), among others. They are ably summarized, with characteristic verve, by Laudan (*ibid.*, pp.73-74):

Judges have limited say, and jurors have no say, over the proportion of truly guilty and truly innocent defendants who come to trial. That is a matter for prosecutorial discretion. These prosecutorial decisions will dramatically impact the ratio of false acquittals to false convictions that actually appears, since the number of false acquittals and false convictions depends upon the mix of truly innocent and truly guilty among the defendants... [I]t is arguable that there is *no* machinery for generating a [standard of proof] that will capture the ratio in question for every conceivable distribution of guilty and innocent defendants.

The problem is that, even if we could agree on an ideal ratio of false acquittals to false convictions, we could not use it as the basis for the Ratio-to-Standard strategy. We cannot use it because, for any given ratio of false acquittals to false convictions, there is simply no such thing as “the standard that we can expect to produce this ratio”. Whether any given standard will produce any given ratio depends on the proportion of people on trial who are genuinely guilty. This is not only something that we *don't* know, but is something that we *can't* know, given the very same epistemic limitations that gave rise to our worries about erroneous verdicts in the first place. These epistemic limitations thereby throw a spanner in the works of the Ratio-to-Standard Strategy as an attempt to respond to these worries when setting the standard of proof.

ii. Guilt and apparent guilt

In fact, the problem with versions of the Ratio-to-Standard Strategy focussed on Blackstone's ratio is more complicated than I just made it sound. It is true that the ratio of false acquittals to false convictions depends on the proportion of defendants who are actually guilty, as well as the standard of proof. (And it is true that previous presentations of the problem with Blackstone's ratio have largely focussed on this fact.) But even these two factors are not the only ones on which the ratio of false acquittals to false convictions depends. There is a third factor, which it is well worth coming to understand.

Distinguish a defendant's *actual* guilt or innocence from the degree of what we may call their “*apparent* guilt” – i.e., the degree to which it is *reasonable to believe* that they are guilty, at the end of their trial, based on the evidence that was presented. Actual guilt is an all-or-nothing matter, as each defendant either did the things of which they are accused or they did not. This is not so for apparent guilt; since a defendant's apparent guilt is the *degree* to which it is reasonable to believe that they are guilty based on the evidence presented at trial, it is clearly something that comes in degrees. Moreover, apparent guilt and actual guilt can come apart. Evidence can be misleading. If the evidence in a trial is very misleading indeed, or if there is just not very much of it, then someone who is actually guilty can appear as innocent as the rest of us. Likewise, someone who is actually innocent can be framed, can be in the wrong place at the wrong time, or for a whole host of other reasons can end up with misleading incriminating evidence against them.

We can represent degrees of apparent guilt using the standard way to represent degrees of confidence in a proposition: as numbers on a real-valued scale from 0 (certainty of falsehood) to 1 (certainty of truth).⁴ So,

⁴ This is the approach to modeling degrees of confidence that is canonical in economics, decision theory, game theory, formal epistemology, statistics, and much of philosophy of science. See e.g. Ramsey (1926), de Finetti (1937), and Savage (1956), for classic expositions of the approach.

for example, a degree of apparent guilt of 0.5 amounts to its being reasonable to believe, on the basis of the evidence presented at trial, that the defendant is equally as likely to be guilty as to be innocent. A 0.9 degree of apparent guilt amounts to its being reasonable to believe that the defendant is nine times more likely to be guilty than to be innocent (since 0.9 is nine-tenths of 1). A 0.95 degree of apparent guilt amounts to its being reasonable to believe that the defendant is *nineteen* times more likely to be guilty than to be innocent (since 0.95 is nineteen-twentieths of 1). And so on.

The real problem for Blackstone’s ratio is easiest to see if we make two more idealizing assumptions: first, that there is always a particular, precise degree of apparent guilt that is reasonable on the basis of each body of trial evidence, and, second, that jurors can always tell what this reasonable degree of belief is and use it in their decisions about whether to vote to convict. These assumptions are obviously false, but they will help us to state the problem; once the problem is clear, I will show how it can be restated without them.

Given the assumptions that each defendant has a precise degree of apparent guilt and we can always tell what it is, we could, if we wanted to, plot the *distribution* of degrees of apparent guilt among defendants in our trial system over a period of time. The distribution would be representable in a bell curve, like this:

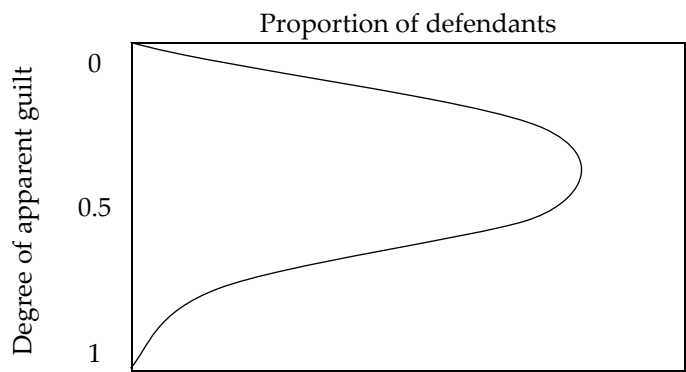


Fig. 2: A sample distribution of apparent guilt

Here’s how to understand the curve: in Fig. 2, as in Fig. 1, proportion is represented by area. To figure out the proportion of defendants that have a particular degree of apparent guilt (say, 0.75), find that point on the *y*-axis, then find the point on the curve that lies level with it. The area under that point on the curve, as a proportion of the total area under the curve, tells you the proportion of defendants with the corresponding degree of apparent guilt.

Now we can complicate the picture that I presented in Fig. 1. This picture was overly simplistic because it assumed that the conviction rate will be the *same* for actually guilty defendants as it is for actually innocent defendants. There is no reason to assume this. The conviction rate in each of these groups depends on the proportion of defendants in the group whose apparent guilt is above the threshold for conviction set by the standard of proof. And there is no reason to assume that this proportion will be the same in both groups, since there is no reason to assume that the distributions of apparent guilt will be the same in both groups. The distributions could very well be different. Moreover, *the ratio of false acquittals to false convictions depends, in part, on the distributions of apparent guilt among the actually guilty and among the actually innocent*. Different distributions will lead a trial system to produce a different ratio of false acquittals to false convictions, even

if both the standard of proof and the proportion of people on trial who are genuinely guilty (the two factors that we explored in §II.i) are held fixed.

We can visualize this point, too, using a pair of (now more accurate) diagrams:

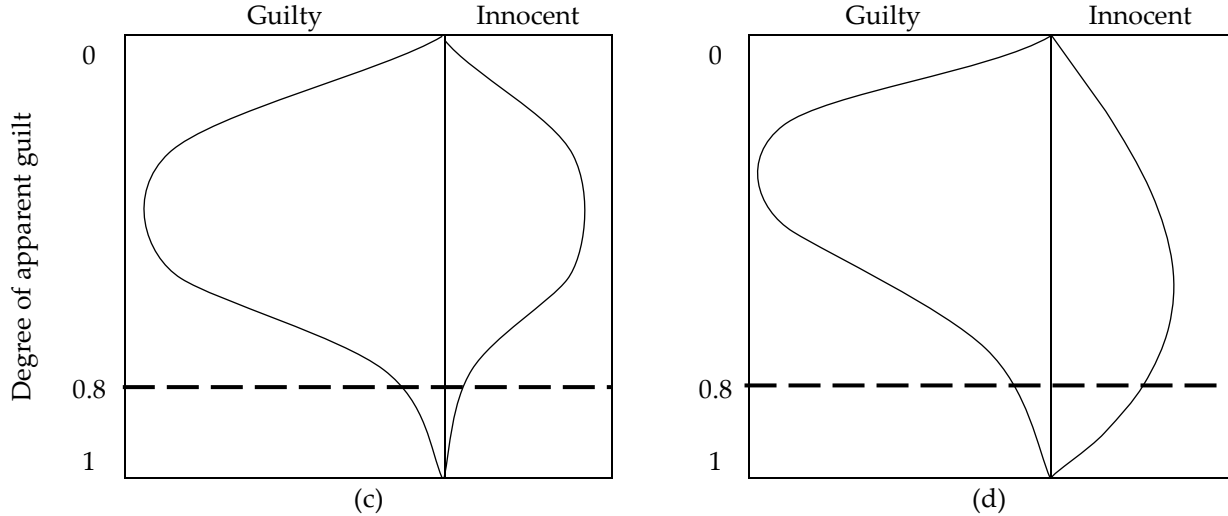


Fig. 3: Two possible trial systems

The diagrams in Fig. 3 are a combination of those in Figs. 1 and 2, so their interpretation should be handled with care. As in Fig. 1, the vertical line represents the proportion of defendants who are actually guilty as opposed to actually innocent; for the purposes of illustration, I have stuck to a 2:1 ratio (though this is just a stipulation for the sake of illustration – I make no claim as to the real proportion). As in Fig. 2, the curves represent distributions of apparent guilt, plotted on a scale from 0 to 1. What has changed is that there are now two curves, to accommodate the possibility that the distribution of apparent guilt among the actually innocent may be different from what it is among the actually guilty. Lastly, the dashed horizontal line represents the standard of proof. Again for ease of illustration, I have stipulated that the standard of proof in both of trial systems (c) and (d) is a probabilistic threshold instructing jurors to vote to convict only if the defendant’s degree of apparent guilt is at or above 0.8 – few people think that the threshold in criminal cases should be this low, but it makes the diagrams clearer and thus makes it easier to see the problem. Once the problem is clear, we can imagine how things would go if we were to raise the standard (moving the dashed line down), or if the proportion of guilty defendants were different (moving the vertical line to the left or right and changing the size of the curves accordingly), or if the distributions of apparent guilt changed (altering the shape of one of the bell curves, while holding fixed the total area under each curve).

Here’s the problem. This more accurate picture shows that the conviction rate both among the guilty and among the innocent is the result of two things: the standard of proof *and* the distribution of apparent guilt in the relevant group. The defendants who will be convicted are those whose apparent guilt meets the standard. So the proportion of defendants in either group who will be convicted can vary, even if the size of the group does not vary and nor does the standard of proof, if the distribution of apparent guilt varies. To see this, note that a far higher proportion of the defendants who are actually innocent end up getting convicted in trial system (d) than in trial system (c), even though the standard is 0.8 and one-third of all

defendants are innocent in both of these systems. This is because, in trial system (d), apparent guilt among the actually innocent is skewed toward the high end; more of the innocent defendants have a degree of apparent guilt that is sufficient to warrant conviction in this system than in trial system (c).

The ratio of false acquittals to false convictions in these diagrams is represented by the comparative area of the portion of the “Guilty” curve that is above the threshold – i.e., the proportion of trials in which a person who is actually guilty will be acquitted – and the portion of the “Innocent” curve that is below the threshold – i.e., the proportion of trials in which a person who is actually innocent will be convicted. Again, we can quickly see that this ratio differs across the two diagrams. There must be far more false convictions per false acquittal in trial system (d) than there are in trial system (c), because about the same proportion of guilty defendants are acquitted in both systems, but many more of the innocent defendants are convicted in (d) than in (c). This demonstrates that the ratio of false acquittals to false convictions can vary even if the proportion of defendants who are genuinely guilty does not vary, and nor does the standard of proof. The ratio depends not only on these two factors, but also on the two distributions of apparent guilt among the actually innocent and among the actually guilty.

This certainly does not help those who want to employ a version of the Ratio-to-Standard Strategy based on a desired ratio of false acquittals to false convictions. By complicating the problem that I set out in §I.i, we only multiply the epistemic hurdles that these theorists must leap if their strategy is to be successful. This complication adds another factor that we not only do not know, but cannot know. Exactly as before, we cannot know the proportion of defendants who are actually guilty as opposed to actually innocent, as we do not know which defendants really are innocent and which are guilty. (Again, this should not be denied by anyone who thinks that trial evidence can be misleading!) But there is something else that we cannot know. We cannot know what the two distributions of apparent guilt *among* the genuinely innocent and *among* the genuinely guilty are. Even under the dramatically idealized assumptions that I have made, according to which every body of trial evidence yields a precise degree of apparent guilt and we can always tell exactly what it is, the most that we could do is to plot the distribution of apparent guilt among *defendants as a whole*. Without knowing which defendants are genuinely guilty and which are not, we cannot know how this overall distribution subdivides into those for the two groups.

iii. It’s much worse than you might think

To be clear, the impact that the distributions of apparent guilt can have on the ratio of false acquittals to false convictions has not gone completely unnoticed. Several of the authors who have spelled out versions of the problem that I set out in §I.i have noted that apparent guilt also plays a role. Both Allen (1977) and Laudan (2006) note that apparent guilt can affect the ratios of outcomes, and DeKay (1996) has an extensive discussion of “the accuracy with which juries discriminate” (p.95) between innocent and guilty defendants.

Nonetheless, the importance of apparent guilt – and especially of our epistemic limitations in determining its distributions among the actually innocent and the actually guilty – has been drastically overlooked. The impact of this factor on the ratios of trial outcomes has far-reaching implications for the Ratio-to-Standard Strategy. The factor provides extra reason to conclude that we cannot reverse-engineer a desired standard of proof from a desired ratio of false acquittals to false convictions; it challenges the assumption that there

is such a thing as “*the* standard that we can expect to produce this ratio”, for any such ratio, by showing that the different ratios to which the same standard may lead can vary wildly. But this not just a problem for versions of the Ratio-to-Standard Strategy that are concerned with false acquittals and false convictions. Rather, it is a problem for *all possible versions* of the Strategy. It is a problem no matter which ratio between outcomes we take an interest in, and no matter what we would like this ratio to be.

This point has, I believe, been under-appreciated in the existing literature. Not everyone who discusses the problem that I summarized in §§II.i-ii sees it as a quite general problem for the Ratio-to-Standard Strategy. Many of these authors still like the strategy. Laudan, in particular, sees the problem quite clearly, but chooses to present it as a reason to use a different *ratio*, rather than a different *strategy*, when thinking about the standard of proof. In an odd passage immediately following the one that I quoted in §I.i, he says this:

But before we resign ourselves to having to pull the SoP⁵ blindly out of a hat, it is worth exploring whether there isn't a *different* ratio of interest that might serve to ground the SoP. The obvious alternative candidate is the ratio of true acquittals to false convictions... Fortunately, the SoP offers us a way to capture [this ratio] precisely. Indeed, having a SoP of (for example) 90 percent will, over the long run, generate convictions of innocent defendants no more than 10 percent of the time and, obviously, result in acquittals for at least 90 percent of the innocent persons who come to trial.

I am not sure why Laudan thinks that having a standard of proof of 0.9 will guarantee that no more than 10% of innocent defendants are convicted – or, more generally, that having a standard of proof of 0.*x* will guarantee that no more than (100 – *x*)% of innocent defendants are convicted. But, at any rate, this is false. There is no standard that can offer such a guarantee. That is because, as I have already noted, the conviction rate among innocent defendants is not determined by the standard of proof alone. It is jointly determined by the standard of proof *and the distribution of apparent guilt among the actually innocent*. This means that one and the same standard can produce dramatically different ratios of true acquittals to false convictions in trial systems with different distributions of apparent guilt.

We can easily see this, too, by returning to Fig. 3 and comparing trial systems (c) and (d). The ratio of true acquittals to false convictions can be seen in Fig. 3 by comparing the proportion of the area under the curve in the “Innocent” column that is above the dashed line – representing trials in which an innocent person is acquitted, i.e. true acquittals – to the proportion of the area under that curve that is below the dashed line – representing trials in which an innocent person is convicted, i.e. false convictions. This ratio is visibly different in trial systems (c) and (d). More of the innocent defendants are convicted in (d) than in (c). So the *ratio* of true acquittals to false convictions, Laudan's preferred ratio, likewise differs across the two systems. And this is so despite the fact that the standard of proof in the two systems is the same. (If the reader is so inclined, she is welcome to imagine further possible shapes to the curve in the “Innocent” column, and thus further possible ratios of true acquittals to false convictions that could arise, all while keeping the standard fixed.)

⁵ In this passage Laudan uses the acronym “SoP” as an abbreviation for “standard of proof”.

It is just not true that having a standard of proof of 0.9 will “obviously” result in acquittals for at least 90% of innocent defendants. On the contrary, we have no idea what proportion of innocent defendants will be acquitted in a trial system with a 0.9 standard of proof, unless and until we know what proportion of those innocent defendants have a degree of apparent guilt that is at or above 0.9 – far from being obvious, this is something that is *impossible to determine* without knowing which defendants are actually guilty and which are innocent. Even under the idealized assumptions that every body of trial evidence yields a precise degree of apparent guilt and that we can always tell what it is, we would in a position to plot the distribution of apparent guilt among defendants as a whole but would not be able to tell how this subdivides into the two distributions among the actually guilty and among the actually innocent.

This is a real problem. The very same epistemic limitation that prevents us from estimating the impact of changes to the standard of proof on the ratio of *false* acquittals to true convictions equally undermines our capacity to estimate the impact of any such change on the ratio of *true* acquittals to false convictions. One upshot is that Laudan’s argument eats itself: the solution that he proposes to the problem that he identifies with Blackstone’s ratio is in fact subject to the exact same problem, for the exact same reasons.⁶ But things are bad even for those of us who are not Laudan. The Ratio-to-Standard Strategy’s prospects are grim. For we cannot simply switch to yet another kind of ratio between trial outcomes in response to the problems that I have identified for Blackstone’s and Laudan’s preferred ratios. We cannot do this because the same problems arise for quite literally every possible ratio between trial outcomes.

To verify that this is so, note firstly that Blackstone’s and Laudan’s ratios are not the only two ratios between trial outcomes that a proponent of the Ratio-to-Standard Strategy could be interested in. On the contrary, with four possible trial outcomes there are six possible ratios between pairs of them, any one of which could legitimately hold our interest.⁷ They are as follows:

Ratio	Interest
False convictions : False acquittals	Ratio of errors (Blackstone)
True convictions : True acquittals	“Ratio of successes”
False convictions : True acquittals	Fate of the innocent (Laudan)
True convictions : False acquittals	Fate of the guilty
False convictions : True convictions	Accuracy of convictions
False acquittals : True acquittals	Accuracy of acquittals

Just as we may be interested, à la Blackstone, in the ratio of errors produced by our trial system, so too may we be interested in what we may call the “ratio of successes”: the extent to which those of our verdicts that are *correct* tend to be acquittals of the innocent as opposed to convictions of the guilty. This ratio similarly

⁶ Laudan’s ratio does have the feature that it, unlike the Blackstone ratio, is not dependent on the proportion of people on trial who are actually guilty as opposed to innocent – since his ratio is concerned only with the proportion of innocent defendants who are acquitted, it does not matter what fraction of the *total* number of defendants this group comprises. But that feature is of no consequence in the present context. We still cannot reverse-engineer a desired standard of proof from Laudan’s preferred ratio, since, although this ratio is not dependent on the proportion of people on trial who are innocent, it is still dependent on the distribution of apparent guilt within this group.

⁷ There are also four possible ratios between trios of trial outcomes, and one ratio between all four outcomes. But I will set those aside, as, if we are unable to determine the ratios between any *pair* of trial outcomes that a standard of proof will produce, then we will not be able to determine the ratio between any trio or quad of them either.

reflects the inevitable trade-offs between the goals of our trial system; for instance, a trial system that always acquitted everyone could not be beaten when it comes to producing acquittals of the innocent but would do as badly as it is possible to do when it comes to producing convictions of the guilty. Similarly, just as we may restrict our attention, à la Laudan, to the fate of the innocent, so too may we restrict it to the fate of the guilty: this is the ratio of true convictions to false acquittals. Both the fate of the innocent and the fate of the guilty are ratios that do not depend on the overall proportion of defendants that are actually guilty, though they do each depend on one of the two distributions of apparent guilt – either among the innocent or among the guilty – so they seem roughly as promising as one another. (I am not sure why Laudan chose to focus on the innocent rather than the guilty.) Lastly, we could be interested in the ratio of true convictions to false convictions, or the ratio of true acquittals to false acquittals. We would be interested in these ratios if we just wanted to concentrate on the accuracy of our trial system’s convictions, bracketing its acquittals for a moment, or *vice versa*.

All of these are legitimate reasons to care about the ratio between a particular pair of outcomes produced by our trial system. Unfortunately, though, none of these six ratios yields a workable version of the Ratio-to-Standard Strategy. This is because all six ratios depend on the distributions of apparent guilt, either among the actually guilty, the actually innocent, or both.

I have already shown that this is true of the ratio of errors (Blackstone’s preferred ratio) and of the fate of the innocent (Laudan’s preferred ratio). Parallel reasoning shows that the same holds for the ratio of successes and for the fate of the guilty. The latter is especially easy to see: returning to the diagrams in Fig. 3, we can see that if the distribution of apparent guilt among the genuinely guilty were to be skewed toward the high end in one diagram and toward the low end in the other, then the proportion of guilty defendants who are convicted would differ, even if the standard of proof remains the same. For the ratio of successes, we need to look at the comparative area of the portion of the “Guilty” curve that is below the dashed line – the proportion of defendants who are guilty and convicted, i.e. true convictions – and the portion of the “Innocent” curve that is above the dashed line – the proportion of defendants who are innocent and acquitted, i.e. true acquittals. This ratio is already different in trial systems (c) and (d). A higher proportion of innocent defendants are acquitted in trial system (c) than in (d), while the proportion of guilty defendants who are convicted is roughly the same across the two systems. This demonstrates that the ratio of successes depends in part on the distributions of apparent guilt, since it differs across two trial systems that differ in nothing *but* their distributions of apparent guilt.

Similar remarks apply to both the accuracy of convictions and the accuracy of acquittals. In Fig. 3, these ratios are represented by the comparative size of the portions of the two curves that are below the dashed line, for convictions, or above the dashed line, for acquittals. Again, these ratios are both already different in trial system (c) than in trial system (d). System (c) acquits a higher proportion of its innocent defendants than system (d), whereas (d) convicts a higher proportion of its innocent defendants. But the two systems convict roughly the same proportion of their guilty defendants. This means that system (c) produces more true acquittals *per* false acquittal than trial system (d), and that system (d) produces more false convictions *per* true conviction. In other words, the accuracy of both convictions and acquittals is better in system (c). But, to repeat, these trial systems have the same standard of proof and the same proportion of defendants who are genuinely guilty. So these ratios, too, depend on the distributions of apparent guilt.

To sum up: the Ratio-to-Standard Strategy – the most popular way of justifying our choice of a standard of proof in the existing scholarly literature on this topic – would have us settle on a desired ratio between two or more of the four possible trial outcomes, and then attempt to reverse-engineer an argument for a certain standard from an intuition in favor of the desirability of a certain ratio. But this strategy must fail, in all its forms. We simply cannot reason our way from a desired ratio to a desired standard, since there is no ratio that is determined solely by the standard, and the necessary calculations require information that we not only do not possess, but cannot acquire.

This is the trouble with standards of proof.

iv. A subjective version

Some readers may be silently fuming at the two idealizing assumptions I made earlier, namely that each body of trial evidence always yields a precise degree of apparent guilt and that jurors can always tell what it is. That is reasonable – the assumptions are extremely unrealistic. Fortunately, we need not make them.

I take it that few people really believe that there are no rational constraints whatsoever governing jurors' reasoning, and that jurors should be free to weigh the evidence presented at trial in absolutely any manner they like. For example, few of us would be content with a juror who regarded clear video evidence of the defendant committing the crime as something that makes it *less* likely that she is guilty, and few would be content with a juror who sees the existence of a cast-iron alibi as something that *increases* the probability of guilt. There are evidently some standards here. Nonetheless, it is equally plausible that these standards are not so precise as to identify, for each and every possible body of trial evidence, a precise real-valued degree of confidence that is the uniquely rational degree of confidence in guilt to have on the basis of this body of evidence.⁸ Some of the business of moving from a body of evidence to a degree of apparent guilt may be up to individual jurors to carry out as they see fit. Moreover, even if – as I agree is implausible – there *are* epistemic standards that are so precise as to pin down a precise real-valued degree of apparent guilt for each body of trial evidence, jurors are manifestly not aware of these standards and their edicts. Jurors often disagree about the defendant's degree of apparent guilt. And this is obviously not because some of them are maliciously ignoring the dictates of the precise standards of which they are perfectly well aware. Rather, it is because, if there are such precise standards, we are unaware of them.

So let's jettison the two idealizing assumptions. No substantial part of the argument in this paper has to change if we do so; mostly, the argument just needs to be rephrased. To jettison the two assumptions is to admit that our trial system's outcomes depend, not on the degrees of confidence in guilt that are *reasonable* to have on the basis of the evidence, but on the degrees of confidence that jurors *will in fact have* – rightly or wrongly. But on this picture, the ratios between trial outcomes over time still depend on three things: the standard of proof, the proportion of people on trial who are actually guilty, and the distribution of degrees of confidence in the defendant's guilt that jurors *in fact have* at the end of each trial. This latter distribution

⁸ Almost everybody working in contemporary epistemology denies that each body of evidence pins down a unique, real-valued degree of rational credence for each proposition, even in normal, everyday cases in which the stakes are much lower because the cost of error is much less steep. Notable exceptions are White (2005, 2013), Horowitz (2014), Meacham (2014) and Greco and Hedden (forthcoming). For a summary see Titelbaum and Kopec (2016).

is what replaces our earlier notion of “apparent guilt”. But it is not worlds apart from that notion; we are still talking about how much it seems like the defendant is guilty. If we like, we can even continue to use the phrase “apparent guilt”, keeping in mind that it now has a subjective rather than an objective construal.

The argument of §§II.ii-iii can now be restated in subjective terms. It remains true that we will always be able to tell, at the end of a trial, what the defendant’s degree of apparent guilt is. In fact, this is now much easier to establish – we just ask jurors how confident of the defendant’s guilt they themselves are. So the bell curves in Figs. 2 and 3 can be reinterpreted as plotting degrees of apparent guilt, *subjectively construed*. But it is also still true that, though we could plot the distribution of apparent guilt among defendants as a whole (as in Fig. 2), we would have no way to determine how this single curve subdivides into the two distributions among the actually innocent and among the actually guilty, as we still would not know who is actually guilty and who is not. Going subjective about the norms on jurors’ reasoning will not help in this regard, since it does nothing to bridge the gap between actual and apparent guilt. Lastly, and crucially, it remains true that every possible ratio between trial outcomes is determined in part by at least one of the distributions of apparent guilt among the actually innocent and among the actually guilty. These ratios depend not only on who is guilty and what the standard of proof is, but on what proportion of defendants in the two groups are thought by jurors to be sufficiently likely to be guilty to meet that standard. So it remains true that, since we do not know the two distributions of apparent guilt, we cannot predict any ratio between trial outcomes to which any standard will lead. The trouble with standards of proof remains intact.

In fact, there is one respect in which the trouble with standards of proof becomes even more acute once we begin to construe apparent guilt subjectively. There may be additional reason to expect that apparent guilt and actual guilt can come apart – sometimes far apart – given the reams of information that we have learnt from psychology and sociology to the effect that jurors may be subject to errors and bias in their reasoning. The more of these errors and biases there are, the more reason we have to expect jurors to think that the actually innocent appear guilty, or that the actually guilty appear innocent, at least in certain types of cases. And there is plenty of evidence for such things. For example, a study by Levinson, Cai and Young (2010), using a version of the Implicit Attitudes Test, which measures the strength of implicit associations between racial categories and certain words, found that participants unconsciously associated black faces with the word “Guilty” and white faces with the word “Innocent”. (A follow-up study in 2013 also found that participants associated black faces with words indicating worthlessness and expendability and white faces with words indicating value.) Rachlinski *et al* (2009) found another strong white preference on the Implicit Attitudes Test among white judges. And Levinson and Young (2010) gave participants bodies of fake trial evidence to evaluate, which were identical across conditions except for the race of the perpetrator depicted in a photograph (which had been altered using photoshop – the photographs were otherwise identical), finding that participants who saw a dark-skinned perpetrator were more likely to evaluate the evidence as overall indicating guilt than those who saw a light-skinned perpetrator. So, if we are construing apparent guilt in subjective terms, we have reason to expect that apparent guilt will be higher for black defendants than for white ones, whether or not they are actually guilty. And this is just one example; the varieties of implicit bias are many and varied, and research into their effects is ongoing.

Here’s the rub: moving from an objective to a subjective way of thinking about apparent guilt does nothing to alleviate the problems that I set out in §§II.i-iii. If anything, it makes them worse.

v. Generalizing further

One might go so far as to say that this problem affects the entire project of construing standards of proof as probabilistic thresholds. That would be a step in the right direction. But we should go further still. For the very same problem still arises even if we construe the standard of proof in non-probabilistic terms, and if we describe the state of mind that jurors must be in before voting to convict as a state that does not come in degrees. The problem is not a problem with probabilistic models, but with the Ratio-to-Standard Strategy. And it affects *every* possible version of this strategy, whether probabilistic or non-probabilistic.

Here's why. Suppose that somebody wanted to construe the standard of proof in non-probabilistic terms – say, as a state of “moral certainty” (see *Commonwealth v. Webster*, 59 Mass. 295, 1850) or “abiding conviction” (114 S. Ct., at 1244) or “wholehearted conviction” (*Lau ms*), or the absence of “a doubt based on reason and common sense” (see *Commonwealth v. Ferreira*, 373 Mass. 116, 1977) or of a “plausible story” consistent with the defendant’s innocence (see Allen 1993) – but that they still liked the Ratio-to-Standard strategy as a way of arguing that we should insist that jurors be in the non-probabilistic state that they describe before voting to convict. There is nothing about the Ratio-to-Standard strategy that confines it to those who see standards of proof as probabilistic thresholds, so this is one way that the Strategy could go. Someone who wanted to pursue such a version of the Ratio-to-Standard strategy would then have to argue that requiring jurors to be in their favored non-probabilistic state before voting to convict will produce a desired ratio between two or more of the four possible trial outcomes. This theorist could try to claim, for example, that requiring jurors to be convinced “to a moral certainty” will produce a 10:1 ratio of false acquittals to false convictions.

The problem with all such claims is apparent: there is no way to argue convincingly for them. At least the probabilistic models had mathematics on their side (though their proofs required assumptions to which we are not in fact entitled). By contrast, any claim to the effect that using a *qualitatively* described mental state as the standard of proof will produce a particular, precise ratio of trial outcomes just seems absurd. This move from a qualitatively-described input to a precise quantitative output evidently goes far beyond what we are in a position to establish.

Ultimately, though, the trouble with standards of proof is the same trouble, whether the standard is thought of in probabilistic or non-probabilistic terms. We can even run the same argument (from §§II.ii-iii), using the same diagrams. The reason why we can do this is that, even if a mental state is described qualitatively, jurors can be uncertain about *what that state is and whether they are in it*. This much is evident from the fact that jurors often request clarification as to the nature of the standard even when a qualitative description of it has already been offered. We are imperfect introspectors, and we are not completely transparent to ourselves. So the question of whether it is appropriate to describe ourselves as being convinced “to a moral certainty”, or having an “abiding conviction”, or etc., is not always easy to answer. And, like all uncertainty, jurors’ uncertainty as to whether they have attained the qualitatively-described mental state that is deemed necessary for conviction will come in degrees. But they will vote to convict only if they are sufficiently confident – confident to a sufficient degree – that they are in the relevant mental state. So we can formulate a completely general version of the trouble with standards of proof by construing the bell-curves in §II.ii’s diagrams as representing jurors’ degrees of confidence *that the evidence presented at trial has been sufficient to put them in the relevant mental state*, whatever this state may be. The reason why we cannot argue that any

qualitatively-described mental state will yield a precise ratio between two trial outcomes is that we simply have no idea, for any such mental state, what these curves will look like in our trial system.

The trouble with standards of proof runs deep. It does not depend on either of the idealizing assumptions that I made earlier in this paper; the trouble remains even if we do not make them. And even if we do not assume that the standard of proof should be construed probabilistically, the trouble remains. Rather, this problem is a problem for the Ratio-to-Standard Strategy, and versions of the very same problem will arise for any conceivable version of the strategy, no matter how it is pursued. No way of conceptualizing jurors' reasoning can bridge the gap between actual and apparent guilt. But unless this gap is bridged, the Ratio-to-Standard Strategy fails on its own terms.

vi. Why be so skeptical?

The defender of the Ratio-to-Standard Strategy is likely to object to my argument in this paper by claiming that I have been unduly pessimistic and skeptical. Sure, she will say, we do not *know* what the distribution of apparent guilt is, either among the actually innocent or among the actually guilty. But we can and should optimistically assume that they are within some *range* – one that reflects our evidence's being moderately probative and jurors' being reasonably good at reasoning. We should have a little faith in the justice system. After all, we can expect that our practices of gathering and presenting evidence in criminal cases are at least moderately truth-conducive, and this means that apparent guilt will tend to be higher for the actually guilty than it is for the actually innocent. Isn't it more likely, in general, that someone who is actually guilty gets found out than that someone who is actually innocent gets framed or misjudged? And isn't it more likely, in general, that somebody who has very little evidence against them did not commit the crime(s) of which they are accused than that we have simply failed to come up with the facts that would incriminate them? So shouldn't we take an optimistic approach, proceeding on the assumption actual and apparent guilt cannot come very far apart?

I have three lines of response to this objection. The first is a defensive argument, the second is an offensive argument, and the third is a general critical comment.

Here is the defensive argument. Once we have acknowledged the possibility of erroneous trial outcomes, we should not then insist that the number of erroneous outcomes must nevertheless be low, in the absence of any further information to support this claim. Why think that there must not be many mistaken verdicts? After all, the prevalence of mistaken verdicts is not a function of our personal level of optimism about the reliability of our trial system. It is a function of the extent to which defendants' degrees of apparent guilt stray from their actual guilt – either because the evidence (including absence of evidence) presented at trial is misleading, or because jurors do not reason well on the basis of this evidence. This is not something to speculate about from the armchair. If we are going to conclude that apparent guilt cannot stray too far from actual guilt, then we need some concrete, positive reason to think so. Optimism does not provide evidence for what has been optimistically assumed, so the defender of the Ratio-to-Standard Strategy must say more to support her case. But she must tread carefully here. She must tread carefully because it is the possibility of erroneous verdicts that encouraged us to think about the four trial outcomes and their comparative degrees of (dis)value in the first place, and that thereby lent the Ratio-to-Standard strategy its initial appeal.

So the defender of the Strategy walks a fine line between motivating her approach and undermining it, as the considerations doing the motivating are the same ones that do the undermining.

Here is the offensive argument. Suppose that the defender of the Ratio-to-Standard Strategy agrees to roll up her sleeves and engage in a substantive debate about the prevalence of misleading evidence and jurors' tendencies to reason poorly. Detractors of the Strategy, like me, have plenty of material on which to draw in this debate. When it comes to jurors' tendencies to reason poorly, we can cite the empirical work on bias and error mentioned in §II.iv. And when it comes to the prevalence of misleading evidence, we can look at information obtained outside of trial contexts that suggests that evidence in trials may be way off-kilter, at least for certain types of trials – ones involving a particular type of defendant, or type of victim, or type of crime, or type of evidence. Here are some examples:

- *Police brutality.* Some of the recent high-profile cases of police brutality give us reason to think that apparent guilt among genuinely guilty police officers might be skewed toward the low end. For example, video footage of police officer Michael Slager shooting citizen Walter Scott (after a traffic stop) depicts Slager shooting Scott six times in the back while Scott fled, and then dropping a small object by Scott's body. Slager later claimed that Scott stole his taser. If the small object in the video is the taser, then this is video footage of a police officer planting false exculpatory evidence at the scene of his crime. Information like this highlights the fact that police officers who commit crimes often have both motive and opportunity to alter the amount or quality of evidence available against them. But they will not always be caught on video, so we cannot be sure what the extent of this confabulation is. (For more information and a copy of the video, see Schmidt and Apuzzo 2009.)
- *Arson.* We can also learn information outside of trial contexts to suggest that apparent guilt among genuinely innocent defendants has historically been skewed toward the high end. For example, in 2004 Cameron Todd Willingham was executed by the state of Texas – after twelve years on death row protesting his innocence – shortly after new expert opinion had concluded that *none* of the putative evidence of arson in Willingham's original trial had any sound scientific basis whatsoever. This opinion has since been repeatedly corroborated. Current scientific opinion is that the factors that for decades were treated as evidence of arson by fire investigators, including puddle-shaped stains on floors and V-shaped marks on walls traditionally thought to be explicable only by the hypothesis that someone had added a liquid accelerant, are in fact equally likely to occur if the fire started by accident. That which has been taken as evidence of guilt for years in fact has no probative value whatsoever. This provides us with reason to think that in arson cases, apparent guilt among the genuinely innocent has been artificially high. (For Willingham's story, see Grann 2009).
- *Forensic evidence.* There could well be more types of evidence that continue to be relied on in court, but that are in fact distorting the distributions of apparent guilt. The National Academy of Sciences recently published a scathing report calling into question almost all forms of traditional forensic evidence as unsupported by sound scientific research. The report challenged fingerprint, hair, ballistics, handwriting, and especially bite mark analysis, stating that “[w]ith the exception of nuclear DNA analysis... no forensic method has been rigorously shown to have the capacity to consistently, and with a high degree of certainty, demonstrate a connection between evidence and a specific individual or source” (National Research Council 2009, p.7). Further research into all of these types of evidence is, of course, needed. But in the meantime, we should be wary in all trials involving such evidence, as these may be further types of trial in which the distribution of apparent guilt among actually innocent defendants is artificially high.

- *Plea bargaining.* The vast majority of criminal cases in the USA are resolved using plea bargaining. But there is an increasingly well-supported worry that this may result in increased guilty verdicts for the actually innocent, as even the innocent may prefer to accept a bargain than face the prospect of a criminal trial and the possibility of a heftier sentence. For example, Dervan and Edkins (2013) attempted to simulate “the innocent defendant’s dilemma” by accusing actually innocent students of cheating on a math test, and found that 56.4% of them chose to falsely incriminate themselves in an official report rather than having to face an academic review board and possibly being forced to enroll in time-consuming remedial Ethics classes. The fact that over half of the actually innocent accepted the “bargain” is already worrying. But we may worry even more given that plea bargains often ask defendants to testify against others in exchange for a lighter sentence. Further research is needed to determine the extent to which we may reasonably worry about innocent parties being bullied or cajoled into falsely incriminating *others*, in addition to themselves. But these preliminary findings suggest another type of trial in which apparent guilt among the actually innocent may be unduly high: trials in which one of the witnesses has agreed to testify as part of a plea bargain.

If the defender of the Ratio-to-Standard Strategy is to maintain her note of cheerful optimism in the face of information like this, she must think either that the relevant phenomena arise in only a very small number of cases, or that the myriad ways in which evidence can be systematically misleading somehow cancel each other out. The former reply gets less plausible with each new item we learn suggesting that the evidence in certain types of trials is likely to be misleading in a certain way. And the latter – the suggestion that these pernicious influences on the probativeness of trial evidence ultimately cancel each other out – is, on its face, wildly implausible. So if she is to rationalize her cheerful optimism in the face of this positive evidence that it is unwarranted, the defender of the Ratio-to-Standard Strategy has some work to do.

Here is a general critical comment. The defender of the Ratio-to-Standard Strategy has a difficult task ahead of her, and she will need hard information and mathematical calculation, not mere optimism, to complete it. Remember what she is trying to do. The Ratio-to-Standard Strategy attempts to argue for our adoption of a *particular* standard of proof, as opposed to a rough range of possible standards that are somewhat close to one another. And it approaches this task by settling on a desired ratio between trial outcomes, and then attempting to reverse-engineer an argument for the desirability of the standard from an intuition in favor of the desirability of the ratio. It does this by claiming that its preferred standard is the standard that we can expect to produce our desired ratio. There is room for a little margin of error here, but not a lot of room. If we are trying to justify our adoption of a *particular* standard, then one of the things that we have to rule out is the adoption of a standard that is very close to the one we favor, but ever so slightly higher, or a little bit lower. And the more loosey-goosey we are in our estimations of the distributions of apparent guilt, the less we will be able to generate precise expectations about the ratios between trial outcomes to which any given standard of proof will lead. So if we have at our disposal only a general hope that the distributions of apparent guilt are within some vague range, this will not be enough for the Ratio-to-Standard Strategy to yield specific recommendations.

To sum up: it is true that, if our practices of gathering and presenting evidence in criminal cases are at least moderately truth-conducive, then apparent guilt will tend to be higher for actually guilty defendants than it is for actually innocent defendants. But this is nowhere near enough grounds for optimism to rehabilitate the Ratio-to-Standard Strategy. To use this strategy, we need to make precise estimates of the distributions

of apparent guilt. So, the defender of the Ratio-to-Standard strategy may find skeptical worries about the prevalence of misleading evidence distasteful, but she does need to address them – and to respond to the reams of information suggesting that apparent guilt *is* likely to be frequently askew – if she is to claim that her estimates embody reasonable expectations rather than her own blind faith.⁹

III. Solutions

i. **Keep doing what we're doing**

One way to respond to the argument of this paper, short of abandoning the Ratio-to-Standard Strategy, is to knuckle down and do the empirical research that would be required for us to rationally have faith in our estimations of the distributions of apparent guilt. Although we are not in a position to establish what these distributions are – and this is inevitable, since we do not know who is actually innocent and who is guilty – we may nonetheless make more or less educated guesses. And there is in principle nothing stopping us from responding to the problem that I have identified by working to make our guesses as well-educated as possible. We might try to retain both the current standard and its current justification, admit their now-clear flaws, and then simply muddle on as best we can, rationalizing our optimism by doing the hard work of integrating empirical information about the prevalence of misleading evidence and about errors or biases in jurors' reasoning into our calculations about which standard of proof to adopt. With better acknowledgement of how misleading the evidence presented at trial can be, and of how easily misled jurors can be, we may be able to use the Ratio-to-Standard Strategy more accurately. In effect, this solution to the problem that I have posed is: keep doing what we're doing, but try to do it better.

It is edifying to think about what this approach would require of our legal system. Were we to adopt it, and to take both the Ratio-to-Standard Strategy and the problems I have raised for this strategy seriously, we would certainly have our work cut out for us. This approach asks us to develop an empirically-informed legal system, sensitive to recent developments in psychological and sociological research and to scientific information about the probative value of the kinds of evidence typically relied on in court. Indeed, if we are to take proactive steps to improve our estimations of the degrees of apparent guilt, we must respond to each new piece of information suggesting that apparent guilt may be affected by some irrelevant factor by commissioning new research into this factor, attempting to discover precisely how much and what kind of an influence it has. And we must integrate the findings of all such research in an ongoing, regularly revised estimate of the distributions of apparent guilt among the actually innocent and actually guilty defendants in our trial system. We'll need to recalculate regularly, and to revise our standards accordingly.

One interestingly dramatic revision to our current practices that fits well with this approach would be to start to use different standards of proof in different types of criminal case. When we learn new information suggesting that the distributions of apparent guilt are unusually skewed for a certain type of crime, or a certain type of defendant or victim, or when a certain type of evidence has been presented, we could react by using a different standard of proof in cases involving this type of crime, defendant, victim, or evidence. If the distributions of apparent guilt systematically vary across types of criminal case, then using the same

⁹ I am grateful to Professor Mark Tushnet for pressing the objection that led me to write this section of the paper.

standard for all of them will result in different ratios across these different case types. So we may vary the standards of proof that we use in an attempt to make outcomes more uniform in the long run; this would amount to an empirically-informed attempt, on the part of our trial system, to counteract the effects of misleading evidence and to bring the ratios of trial outcomes closer to where we would like them to be. For example, if we have reason to believe that the apparent guilt of genuinely guilty police officers tends to be unusually low, then we could lower the standard of proof for crimes allegedly committed by police officers, attempting to reduce false acquittals and to build the rate of true convictions back up to where we want it. Similarly, if we have reason to believe that the apparent guilt of genuinely innocent black defendants is skewed toward the high end, then we could raise the standard of proof for cases in which the defendant is black, in an attempt to keep false convictions under control.

I am quite sympathetic to this approach. It would certainly be a marked improvement on current practices. Nonetheless, there are two serious problems with it. One is practical, the other more philosophical.

The practical problem is a reference-class problem (see Reichenbach 1949, Hájek 2007, and cf. Schauer 2003 for legal implications). On this approach, when we are trying to determine an appropriate standard of proof for an individual trial, we will be trying to determine what the defendant's degree of apparent guilt is likely to be if they are actually guilty, and what it is likely to be if they are actually innocent. It seems completely reasonable that we could improve these estimates by taking account of some specific information about the defendant and her alleged crime, placing her in a narrower reference-class than the class of all defendants. But once we start narrowing down, it is notoriously difficult to know when or how to stop. For example, if we think that trials whose defendants are police officers tend to result in low apparent guilt even for the actually guilty, whereas trials whose defendants are black tend to result in high apparent guilt even for the actually innocent, what should we expect in a trial whose defendant is a black police officer? Each particular trial has a great many individuating features, and our best empirical research might associate each of these features with different distributions of apparent guilt. This is a widespread problem in probability theory that has not yet been solved. So, the further we progress down the empirically-informed route, the more difficult it will become to apply an ever-proliferating set of data to the real-life, individual cases with which we are faced. This is a major practical problem with the empirically-informed approach.

The philosophical problem is a reprise of the same skeptical worries that motivate this paper. It is this: no matter how complex our calculations of the "right" standard of proof for each trial may become, they will not ultimately solve the problem that I posed in §II. This is because no amount of prior research gives us access to the actual facts about which defendants really are guilty and which are innocent. Though we *might* be making increasingly accurate guesses as to the distributions of apparent guilt among these two groups as a result of our empirical research, we would have no way to verify that this is indeed the case. So our calculations might instead be making our estimations of the distributions of apparent guilt *less* accurate than before. This could happen if the information that our research produces is itself misleading, or if it is poorly used; for example, if we over-generalize from too few examples, or the results of our psychological studies do not reflect what really happens given the dissimilarity of the lab environment from real trials, or we simply misidentify the variables affecting our results. If any of these problems arises, we could end up altering the standard of proof in a direction that actually makes the ratios between trial outcomes *worse*, while our best estimates continue to suggest that we are getting better. But, since we can never get beyond our estimates to the facts, we would never come to appreciate the depth of these mistakes.

ii. No standards of proof?

If we do not want to just keep doing what we're doing, what other options are there?

We could continue with the current standard of proof, or randomly pick a new one, while admitting that it has no rationale at all. Or we could eliminate standards of proof altogether, allowing jurors to cast their votes however they like.

But I doubt that either of these would be popular options. They would do a terrible job of maintaining the apparent legitimacy of our trial system's verdicts, which is important if we are to encourage respect for and compliance with the law (see the related remarks in Nesson 1985, pp.1366-1368). Of course, it is even better for our verdicts to actually *be* legitimate, and known to be so, in order to justify the coercive imposition of state apparatus into the lives of autonomous citizens. For these reasons, to respond to the problems that I have raised in this paper by just giving up seems like a very bad idea.

Moreover, if we do have a desired ratio between trial outcomes, we should not expect that eliminating standards of proof altogether and allowing jurors to cast votes however they like will yield this ratio. There is no reason to expect that to happen. Indeed, even in a completely homogenous society in which there was unanimous agreement that a certain ratio of trial outcomes was ideal, the trial system would not necessarily produce this ratio. As always, whether it would do so or not depends on the distributions of apparent guilt. For example, if a particular jurisdiction had such shoddy police and prosecutorial practices that apparent guilt were low for literally every defendant, then jurors in this jurisdiction would have to keep acquitting everybody, no matter what ratio they wanted to bring about. Examples like this show that there is no direct relationship between the ratios of trial outcomes that our jurors regard as ideal and the ratios to which their decisions will lead; this relationship is, unsurprisingly, mediated by the distributions of apparent guilt.

iii. Find a new rationale

A third option is to find an alternative strategy to use when thinking about the justification for our standard of proof. The problem that I have set forth challenges every conceivable version of the Ratio-to-Standard Strategy. But it does not challenge any other strategies. So perhaps another strategy can succeed where the Ratio-to-Standard Strategy fails.

I have nothing against this proposal, though nor am I able to offer any promising suggestions as to how it might go. It is difficult to think of an alternative rationale that does not turn out to implicitly depend on an assessment of the comparative (dis)value of the four possible trial outcomes, although the criteria for this assessment may vary. I suspect that a rights-based approach to choosing a standard of proof will be the most promising alternative, if we can find some rights such that the rationale behind them does not depend, even in part, on an implicit assessment of the desirability of the four trial outcomes, *and* the extent to which a standard respects the relevant rights does not depend, even in part, on the distributions of apparent guilt.

I leave it to the reader to see whether she can come up with something.

IV. Conclusion

The outcome of each trial depends on three things:

- (1) What the standard of proof is.
- (2) Whether the defendant's apparent guilt is high enough to meet this standard.
- (3) Whether they are actually guilty or innocent.

This means that the ratios between trial outcomes produced by any given trial system over a period of time will also depend on three things:

- (1) The standard of proof.
- (2) The distributions of apparent guilt among the actually guilty and among the actually innocent.
- (3) The ratio of actually guilty to actually innocent people on trial.

The first and third of these are well-understood, but the importance of the second has been overlooked. Recognizing the role of the two distributions of apparent guilt amounts to recognizing that we do not know to what extent, if at all, our altering the standard of proof would affect the conviction rate either among the actually guilty or among the actually innocent. This scuppers the whole project of choosing a standard of proof with a view to bringing about a desired ratio between false acquittals, true convictions, true acquittals and false convictions produced by our trial system in the long run. There is no ratio between any of these trial outcomes that is determined by the standard of proof alone. Some depend also on the proportion of defendants who are actually guilty, and *all* depend also on at least one of the distributions of apparent guilt. This problem for the Ratio-to-Standard Strategy runs deep, remaining whether we think of apparent guilt in subjective or objective terms, and remaining no matter how we choose to describe the state of mind that jurors must be in before they vote to convict.

It is difficult to know how to respond to this problem. We cannot just do away with standards of proof, nor can we admit that they have no clear rationale. An alternative rationale is yet to be developed. And we can try to use empirical evidence to estimate the distributions of apparent guilt, but, since we still will not know which defendants are actually innocent and which are actually guilty, we will be unable to tell whether our estimates are accurate. Ultimately, the empirical approach does little more than to introduce an additional layer of appearances – we might call it “*apparent* apparent guilt”! – on top of the real distributions, the particular degree of apparent guilt of each defendant, and the question of whether they are actually guilty. Our basic epistemic limitations remain. We cannot get beyond appearances to tell whether our strategy is working.

To conclude: We are in trouble.

REFERENCES

- Allen, Ron (1977). "The Restoration of *In Re Winship*." 76 Mich. L. Rev. 30.
- Allen, Ron (1993). "Constitutional Adjudication, the Demands of Knowledge, and Epistemological Modesty." 88 Nw. U. L. Rev. 436.
- Bell, Richard (1987). "Decision Theory and Due Process: A Critique of the Supreme Court's Lawmaking for Burdens of Proof". *Journal of Criminal Law and Criminology* 78, 557-585.
- Blackstone, William (1765). *Commentaries on the Laws of England*. Oxford: Clarendon.
- Commonwealth v. Webster*, 59 Mass. 295, 1850
- Commonwealth v. Ferreira*, 373 Mass. 116, 1977
- Connolly, Terry (1987). "Decision Theory, Reasonable Doubt, and the Utility of Erroneous Acquittals". *Law and Human Behavior*, 11(2), 101-112.
- Correll, J., B. Park, C. Judd, & B. Wittenbrink (2002). "The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals". *Journal of Personality and Social Psychology*, 83, 1314-1329.
- Correll, J., B. Park, C. Judd, B. Wittenbrink, M. Sadler, & T. Keesee, T. (2007). "Across the thin blue line: Police officers and racial bias in the decision to shoot." *Journal of Personality and Social Psychology*, 92, 6, 1006-1023. doi: 10.1037/0022-3514.92.6.1006
- Cullison, A. (1971). "The Model of Rules and the Logic of Decision". In Stuart Nagel, ed., *Modeling the Criminal Justice System*. Beverley Hills, California: Sage.
- Dreier, James (2011). "In Defense of Consequentializing". *Oxford Studies in Normative Ethics*, 1, 97-119.
- DeKay, Michael (1996). "The Difference Between Blackstone-like Error Ratios and Probabilistic Standards of Proof", *Law & Social Inquiry*, 21, 95-132.
- Dervan, Lucien and Vanessa Edkins (2013). "The Innocent Defendant's Dilemma: An Innovative Study of Plea Bargaining's Innocence Problem". 103 J. Crim. L. & Criminology 1.
- Dunbar v. U.S.*, 156 U.S. 185 (1984).
- de Finetti, Bruno (1937). "La prevision: ses lois logiques, ses sources subjectives". *Annals de l'Institut Henri Poincare*, 7, 1-68. Translated into English in H. E. Kyburg and H. E. Smokler, eds. (1984), *Studies in Subjective Probability*. New York: Wiley, pp.95-158.
- Franklin, Benjamin (1904). *Works*, edited by John Bigelow. Albany, N.Y: The Knickerbocker Press.

- Grann, David (2009). "Trial by Fire", published online by *The New Yorker*. Url: <http://www.newyorker.com/magazine/2009/09/07/trial-by-fire>
- Greco, Daniel and Brian Hedden (forthcoming). "Uniqueness and Metaepistemology". Forthcoming in *Journal of Philosophy*.
- Grofman, Bernard. (1981). "Mathematical Models of Juror and Jury Decision-Making: The State of the Art". In Bruce D. Sales, ed., *The Trial Process*. New York: Plenum.
- Hájek, Alan (2007). "The Reference-Class Problem is Your Problem Too". *Synthese*, 156(3), 563-585.
- Horowitz, Sophie (2014). "Immoderately Rational". *Philosophical Studies* 167, 41-56.
- Joyce, James (1998). "A Nonpragmatic Vindication of Probabilism". *Philosophy of Science*, 65(4), 575-603.
- Laudan, Larry (2006). *Truth, Error and Criminal Law*. New York, NY: Cambridge University Press.
- Kaplan, John. (1968). "Decision Theory and the Factfinding Process". 20 *Stan. L. Rev.* 1065.
- Kopec, Matthew and Michael G. Titelbaum (2016). "The Uniqueness Thesis". *Philosophy Compass* 11(4), 189-200.
- Lau, Arthur (*ms.*) "Wholehearted Conviction". Draft of April 2016.
- Laudan, Larry (2006). *Truth, Error, and Criminal Law*. New York, NY: Cambridge University Press.
- Levinson, Justin D., Huajian Cai, and Danielle Young (2010), "Guilty by Implicit Racial Bias: The Guilty/Not Guilty Implicit Association Test". 8 *Ohio St. J. Crim. Law* 187.
- Levinson, Justin D. and Danielle Young (2010), "Different Shades of Bias: Skin Tone, Implicit Racial Bias, and Judgments of Ambiguous Evidence". 112 *W. Va. L. Rev.* 307.
- Levinson, Justin D., Robert J. Smith, and Danielle M. Young (2013), "Devaluing Death: An Empirical Study of Implicit Racial Bias on Jury-Eligible Citizens in Six Death Penalty States". 89 *NYU L. Rev.* 513.
- Lillquist, Erik (2002). "Recasting Reasonable Doubt: Decision Theory and the Virtues of Variability". 36 *U. C. Davis L. Rev.* 85.
- Meacham, Christopher (2014). "Impermissible Bayesianism." *Erkenntnis* 79, 1185-217.
- National Research Council (2009), "Strengthening Forensic Science in the United States: A Path Forward". PDF document available from the National Academies Press, at <http://www.nap.edu/catalog/12589.html>

Nesson, Charles (1985). "The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts". 98 Harv. L. Rev. 1357.

Rachlinski, Jeffrey J., Sheri Johnson, Andrew J. Wistrich, and Chris Guthrie (2009). "Does Unconscious Racial Bias Affect Trial Judges?" *Cornell Law Faculty Publications*. Paper 786.

Ramsey, Frank P. (1926). "Truth and Probability". In Ramsey, 1931, *The Foundations of Mathematics and other Logical Essays*, edited by R.B. Braithwaite, pp.156-198. London: Kegan, Paul, Trench, Trubner & Co.

Reichenbach, Hans (1949). *The Theory of Probability*. University of California Press.

Savage, Leonard (1954). *The Foundations of Statistics*. New York, NY: John Wiley & Sons.

Schauer, Fred (2003). *Profiles, Probabilities, and Stereotypes*. Cambridge, Mass.: Harvard University Press.

Schmidt, Michael S. and Matt Apuzzo (2015). "South Carolina Officer is Charged with Murder of Walter Scott", published online by *The New York Times*. Url: <http://www.nytimes.com/2015/04/08/us/south-carolina-officer-is-charged-with-murder-in-black-mans-death.html? r=0>

Stein, Alex (2005). *Foundations of Evidence Law*. Oxford: Oxford University Press.

Tribe, Lawrence (1971). "Trial by Mathematics: Precision and Ritual in the Legal Process". 84 Harv. L. Rev. 1329.

U.S. v. Hall, 854 F.2d 1036 (7th Cir. 1988)

Volokh, Alexander (1997). "*n* Guilty Men". 146 U. Penn. L. Rev. 193.

Walen, Alec (2015). "Proof Beyond Reasonable Doubt: A Balanced Retributive Account". 76 La. L. Rev. 355.

White, Roger (2005). "Epistemic Permissiveness". *Philosophical Perspectives* 19, 445–59.

White, Roger (2013). "Evidence Cannot Be Permissive". In *Contemporary Debates in Epistemology*, 2nd ed. Eds. M. Steup, J. Turri, and E. Sosa, pp.312–23. Hoboken: John Wiley & Sons.