CRIMINAL PROOF: FIXED OR FLEXIBLE?

By Lewis Ross

Should we use the same standard of proof to adjudicate guilt for murder and petty theft? Why not tailor the standard of proof to the crime? These relatively neglected questions cut to the heart of central issues in the philosophy of law. This paper scrutinises whether we ought to use the same standard for all criminal cases, in contrast with a flexible approach that uses different standards for different crimes. I reject consequentialist arguments for a radically flexible standard of proof, instead defending a modestly flexible approach on non-consequentialist grounds. The system I defend is one on which we should impose a higher standard of proof for crimes that attract more severe punishments. This proposal, although apparently revisionary, accords with a plausible theory concerning the epistemology of legal judgments and the role they play in society.

Keywords: philosophy of law, legal proof, punishment, criminal justice, beyond reasonable doubt, blame, retributivism, capital punishment.

I. INTRODUCTION

In everyday life, we seek more confidence for some decisions than others. Trivial choices—'Should I purchase a biscuit along with my coffee?'—are settled on the slightest of reasons. For decisions of moderate importance—'Should I spend my holidays here or there?'—we typically need more convincing. And decisions of great gravity—'Should I divorce my spouse?'—are things we might agonise over for years, only acting when sure the decision is the right one. Flexibility in the standards used to evaluate different decisions is a normative expectation. Someone who imposed less rigour on their decision about whether to seek divorce than whether to purchase a chocolate biscuit would be conducting their affairs in an irrational way.

Given that this flexibility is part of rational decision-making, a striking fact is that the standards used to rationalise legal decisions are rather inflexible. Although civil and criminal law each have their own standard of proof—respectively, the 'balance of probabilities' and 'beyond reasonable

doubt'—within these broad categories a diversity of cases are judged against the very same standard. For example, civil law encompasses: (i) deciding whether an overgrown hedge is encroaching the pavement and (ii) deciding whether to remove a child from the family home. Each of these decisions would be judged against an identical standard of proof.[1] Similarly, in criminal law, the same standard is used to judge guilt for (i) the theft of a chocolate bar (suppose, punished by a small fine) and (ii) a violent homicide (suppose, punished by a 20-year jail term). Of course, the law *does* have different procedures for certain crimes. For example, serious crimes may receive trial by jury rather than trial by a single judge. Nevertheless, the normative aim of the decision-maker is identical: to convict with reference to the same standard of proof.

There is something surprising about the invariance of these legal standards. The aforementioned cases clearly differ in ways that might justify imposing higher or lower standards in our everyday decision-making, not least because they involve very different consequences for those affected. Hence, the current orthodoxy requires scrutiny.

This paper investigates the fixed standard of proof in the *criminal* law. Perhaps surprisingly, there are decision-theoretic arguments for having a lower standard of proof for the most serious crimes. I reject these arguments and instead defend a flexible standard of proof on non-consequentialist grounds. The approach I defend is that all criminal trials should be judged against a relatively demanding standard, but where we moderate the standard depending on the variable costs of false positives (i.e., convicting the innocent). In short, I conclude that we should require more confidence before convicting people of serious crimes that attract harsher punishments. Thus, I endorse a modest flexibility in the standards of criminal proof, a position that avoids the extremes of consequentialist jurisprudence, while departing from official doctrine in an important way. This position will be motivated by considering the epistemology of legal judgments and how they relate to the judgements we make in our interpersonal lives.

II. STANDARDS OF PROOF

I begin with introductory remarks on standards of proof.

The *standard of proof* determines whether a body of evidence provides strong enough support to render a legal or quasi-legal judgment appropriate. The law is a thoroughly binary adjudicatory system, so all 'facts at issue' are treated as either having occurred or not, depending on whether they are established to the satisfaction of the relevant standard. To be explicit about how these standards work, legal procedure dictates: 'treat a contested fact as having occurred iff the

---

[1] See the UK Supreme Court decision in *Re. B.* [2016] UKSC 4.

entirety of the evidence establishes it beyond the threshold of the applicable standard of proof'. The *burden of proof*—with a few technical exceptions[2]—rests on the party pressing the claim: in a criminal trial, the prosecutor charging somebody with a crime. Specifically, the prosecution must establish that the accused had a culpable mental state (*mens rea*) and performed some criminal action (*actus reus*).

In my terminology, the standard of proof is *fixed* when a standard of invariant demandingness is used to adjudicate all criminal cases. On a *flexible* approach to criminal proof, the legal system avails itself of different standards for different crimes. In most legal systems—including the International Criminal Court and the common-law jurisdictions that will be my focus—the standard of proof is officially invariant. The applicable standard is 'beyond reasonable doubt' and is used to adjudicate guilt for all criminal cases: from serious crimes involving life imprisonment or the death penalty to the possession of small amounts of proscribed drugs.

Now, of course, one might wonder whether the way that 'beyond reasonable doubt' is *interpreted* by judges or juries varies between different crimes. There is little institutional endorsement of such interpretative latitude, although it is true that how to understand the beyond reasonable doubt standard is a perennial topic of jurisprudential confusion.[3] Quantificational standards[4] (e.g. .9 credence) and qualitative standards[5] (e.g. 'the standards used to make an important decision in the course of your everyday life') have waxed and waned in popularity. The current trend is for appellate courts to reject judicial attempts to further elaborate the idea of reasonable doubt. Whether judges and jurors do in fact vary their interpretation of the criminal standard is a question of psychology—and a crucially important one, to which I will return later. However, I am first and foremost interested in the normative question: should the criminal standard require stronger and weaker evidence depending on the crime?

[2] For instance, a defendant claiming a 'special defence'—e.g. insanity or compulsion—may have the burden to establish that these mitigating facts obtain.

[3] Empirical surveys on judicial interpretation of the criminal standard suggest considerable variation, with a majority viewing it as >.9 confidence (e.g. see Solan 1999). Surveys of jurors suggest wide variation (Horowitz and Kirkpatrick 1996). 'Maximalists' about criminal proof argue that beyond reasonable doubt should approximate certainty without requiring infallibility: others argue for a more permissive interpretation of reasonable doubt. For influential maximalist arguments, see, e.g. Duff, Farmer, Marshall and Tadros (2007) or Tribe (1971). For criticism of maximalism, see Walen (2015) and, more generally, see Solan (1999) or Picinali (2018).

[4] There is a serious debate about 'legal probabilism', the idea that standards of proof reduce to some confidence threshold. See Hedden and Colyvan (2019) for a defence and Allen and Pardo (2019) for the opposing view.

[5] Laudan (2006: 32–51) discusses jurisprudence from the USA. English law recently moved away from directing juries using the beyond reasonable doubt standard (e.g. see *R v Majid* [2009] EWCA Crim 2563). Now, a direction to convict if you are 'sure' of guilt is preferred.

To answer this question, we should reflect more broadly on the justification for the current criminal standard of proof.

III. THE STANDARD OF PROOF AS A UTILITY FUNCTION?

The 'beyond reasonable doubt' standard places, a demanding burden on the prosecution, far more demanding than the civil 'balance of probabilities' standard. Why such a difference between the two? Why not simply require the prosecution prove their case to be 'more likely than not'? A natural answer draws on the following idea:

Blackstone's Asymmetry: The disvalue of falsely convicting an innocent is considerably higher than mistakenly exculpating a guilty person.

Sometimes this thought is referred to in the form of 'Blackstone's dictum', referring to a passage in which the jurist William Blackstone endorses a 10:1 ratio of false acquittals and false convictions as optimal.[6] This particular 'dictum' has no legal force. But, the underlying idea that the disvalue of mistaken convictions and acquittals is *to some degree asymmetrical* is widely accepted. Any more demanding standard than 'more likely than not' prioritises excluding false positives at the expense of generating false negatives (and vice versa for less demanding standards). Thus, we arrive at a familiar rationale for the heightened criminal standard: 'beyond reasonable doubt' respects the normative asymmetry between false convictions and mistaken exculpations by making false positives (false convictions) less likely, at the expense of false positives (mistaken exculpations).

The idea that the standard of proof should reflect the underlying evaluative asymmetry between mistaken convictions and mistaken acquittals suggests a general way to think about the criminal standard: namely, as a *function of the (dis)utility of different types of error*. In this vein, a number of theorists express sympathy with using decision-theory to think about criminal standards of proof (e.g. see Kaplan 1968; Lillquist 2002; Di Bello 2019; Ribeiro 2019). We can formalise this approach, deriving the standard of criminal proof from the disvalue of different types of error:

$$\text{SoP} = \frac{1}{\left(1 + \frac{U(\text{acquit guilty})}{U(\text{convict innocent})}\right)}.$$

This formalisation complements the traditional Blackstonian idea. For instance, if we input Blackstone's own values of a −10 utility for a false conviction

---

[6] Blackstone's (1827) *Commentaries on the Laws of England*: Book Four, ch. 27. Epps (2015) provides useful historical context.

and a –1 utility for a false acquittal, then this utility-function would yield an ∼0.9 standard of proof. In what might seem a happy result for the status quo, this ∼.9 standard corresponds with how some have thought 'beyond reasonable doubt' should be interpreted in practice.

But our task is to work out whether we are justified in having a *fixed* standard of criminal proof. To see whether the decision-theoretic approach provides such a justification, we must consider the following question: *Is it plausible to suppose that some invariant ratio reflects the (dis)value of legal errors across different types of criminality*? Let's dig a little deeper.

A key point about Blackstone's asymmetry, often overlooked, is that it must be interpreted as a *comparative* rather than *absolute* claim. In other words, any assignment of (dis)utilities to different legal mistakes must be relative to a particular crime type, rather than supposing there is some generic universal value for false acquittals and convictions. For example, it is *worse* to falsely convict somebody of murder than to falsely convict them of littering. What explains this evaluative difference? An explanation presents itself immediately: *punishment-related reasons*. It is worse to falsely convict for some crimes than others in part because the punishment meted out varies between different crimes. It is a platitude that it is worse to falsely imprison someone for 15 years (as we might for murder) than falsely impose upon them a $150 fine (as we might for littering). The greater the resultant punishment, the worse it is to mistakenly convict someone of a crime. Indeed, the radically gradable nature of punishment poses an initial challenge to the idea that we are justified in judging all criminal convictions against the same single standard of proof.

Just as the disvalue of false convictions varies in absolute terms, so does the disvalue of mistaken acquittals. For example, it is worse to *fail to punish* a murderer than fail to punish a litterbug. Once more, it is natural to look to the value of punishment—and, specifically, the value foregone by failing to punish—for the explanation. *Forward-looking* theorists of punishment emphasise the instrumental role of punishment as a means of social control: promoting good behaviours and disincentivising or disabling adverse behaviours.[7] Due to the harm inflicted by a murder, there is a stronger social imperative to deter, incapacitate, and rehabilitate potential murderers—greater than the social imperative to deter, incapacitate, and rehabilitate litterbugs. Equally, there are backward-looking *retributivist* explanations for why the disvalue of failing to punish a murderer is greater than failing to punish a litterbug. Retributivists hold that punishment is an intrinsically fitting response to wrongdoing; conversely, it is intrinsically bad to fail to punish wrongdoing. Competing varieties of retributivism defend the intrinsic value of punishment in different ways, with the common core being the connection between punishment and the

---

[7] See Wood (2010) for overview. Hoskins (2017) provides a snapshot of contemporary debate, including discussing non-consequentialist forward-looking theories centred on self-defence.

moral deserts of the wrongdoer. The crudest versions suppose it is intrinsically appropriate to inflict harm on those who inflict harm on others. Wrongdoers who impose more harm—the murderer over the litterer—deserve greater retributive penalty. More contemporary retributivisms emphasise the idea that punishment can perform a communicative function, serving as a rational entreaty to the wrongdoer which at once respects their autonomy and provides them with opportunity to performatively reconcile themselves to the moral community against which they have transgressed.[8] Under such views, certain crimes alienate individuals from the moral community to a greater extent than others, thus standing in need of more robust communicative response. Both forward-looking and retributive theories of punishment agree that it is worse to fail to punish for some crimes than others.

Since the absolute disvalue of false convictions and false acquittals varies between crimes, it quickly becomes apparent that it is exceedingly unlikely for a fixed standard of proof to optimally balance the risk of various harms. For a fixed standard to be justified on a decision-theoretic approach, there would need to be a *reliable correlation* between the disutility of false positives and false negatives *across every crime type*. For example, if it is *n* times worse to mistakenly punish for theft than release one thief unpunished, then it will be *n* times worse to mistakenly punish for murder than release one murderer unpunished, and *n* times worse to mistakenly punish for assault than release one assaulter unpunished, and so on for every type of criminality.

It would be a great coincidence if the weight of various forward and backward-looking values relevant to punishment—deterrence, incapacitating the dangerous, performing a semiotic function, providing retribution, and rehabilitation—were reliably correlated between crime-types. Indeed, theorists impressed by the forward-looking aims of punishment have argued that *some* forms of criminality have consequences rendering false acquittals particularly damaging compared with false convictions. I do not endorse these arguments in the final analysis. But, if we view the standard of proof as a utility-function, then the relevance of these forward-looking considerations to the standard of proof is hard to dispute.[9] Below, I consider some examples of these forward-looking considerations with a view to showing how they support a flexible standard of proof.

High recidivism rates. Recidivism rates vary between crime-types. Larry Laudan (2006; 2011; 2017) argues that the criminal standard of proof ought to be sensitive to harms imposed by mistakenly acquitted recidivists. The basic idea is that every time we fail to convict a violent criminal who goes on to commit further crimes, we expose our community to serious harms, harms which

---

[8] For influential examples, see Bennet (2008) or Duff (2003).

[9] See Stein (2005: ch. 6) or Walen (2015) for critical discussion of including these forward-looking values in setting the standard of proof.

we could have spared ourselves from had we used a lower standard of proof on which they would have been convicted. The higher the recidivism rate, the greater the expected cost of false negatives for that crime-type. Laudan has claimed that many egregious crimes involving interpersonal violence have strikingly high reoffending rates. As a result, he argues that we ought to *reduce* the standard of proof to maximise expected utility.[10] Importantly, Laudan is not suggesting that we punish any individual more harshly than before. Rather, the idea is to incapacitate and attempt to rehabilitate *more* recidivism-prone criminals by lowering the standard of proof.

Especially harmful offences. Erik Lillquist begins his (2002) paper with the quote: 'The old ACLU notion that it's better to let 10 guilty men to go free than to imprison one innocent person is, in an age of suicide bombers with access to bioweapons, not just a luxury but a danger.' Some offences are more harmful than others, and some crimes are so harmful that false negatives are disproportionately costly compared with false positives. Lillquist's examples centre on terror-offences: it is exceedingly costly to mistakenly exculpate a terrorist compared with a regular criminal. Of course, we generally increase the level of punishment in response to the harmfulness of the crime. However, in all legal systems, the maximum level of punishment is *capped*: there is an upper limit on the harshness of punishment that can be meted out (even if death). Someone who kills a small number of people may well receive the maximum allowable punishment just the same as someone who blows up Times Square. When the crime in question is disproportionately harmful, we might maximise expected utility by having a lower standard of proof, as a ward against the possibility of mistakenly releasing especially dangerous criminals.[11]

Deterrence and semiotics. Some crimes have especially low conviction rates. Sexual offences are a prime example (the complaint-to-conviction rate was recently under 2 per cent in England and Wales). There are many proposals on this issue. One is that we ought to lower the standard of proof for crimes with low conviction rates. Raising conviction rates by lowering the standard of proof would, some claim, have the effect of more effectively deterring and reinforcing societal disapproval of these under-convicted crimes.[12] Again, the idea is not to punish any individual more harshly than we would otherwise (which would raise the false positives cost for the accused) but rather to convict

---

[10] Specifically, Laudan argues that we should lower the criminal standard of proof *tout court*. But the same premises recommend a flexible standard of proof on which we have a lower standard for crimes with the highest recidivism rates. Note: Laudan's interpretation of criminological statistics has been forcefully criticised in Gardiner (2017).

[11] For similar thoughts, see Ribeiro (2019).

[12] For example, see Wareham and Vos (2018). Another relevant discussion is found in Greer (2018). Epps (2015) appeals to anti-deterrent effects of the current high criminal standard, as part of an argument that it should be lowered tout court.

more people by lowering the standard of proof. According to this argument, doing so would maximise expected utility through leveraging the deterrent and semiotic functions of the criminal law to discourage sexual criminality.

This non-exhaustive list of forward-looking arguments suggests that there is little reason to think that a single fixed standard of proof is optimal from the utility-maximising perspective. Not only is there not a perfect correlation between the disutility of false positives and false negatives across different crime-types, but the values can vary significantly. For example, when it comes to crimes with exceptionally high recidivism rates, or terror-offences where recidivism is especially damaging, the cost of false negatives can be *much* worse than false negatives in other contexts. Even allowing for some coarse-graining when setting the standards of proof, a fixed standard does not find natural justification in decision-theory. Indeed, we focused only on the costs of error relevant to different *types* of crime. But the comparative utility of (in)accurate convictions and exculpations will also vary between *tokens* of various crime-types. For instance, one murder can be more heinous than another, one murderer more likely to reoffend, one murderer easier to rehabilitate, and so on. If we considered crime tokens, then the fixed standard of criminal proof would find even less support via the classic decision-theoretic approach.

Rather, these considerations instead seem to recommend a *radically* flexible approach to criminal proof: where the recommended standard varies across a wide range of evidential strengths, including below thresholds usually taken to be requirements of justice. Taken to its logical conclusions, deeply troubling upshots flow from the decision-theoretic approach. Firstly, utility-maximisation might justify a legal system in which particularly serious violent crimes—terrorism, rape, and murder—are adjudicated on a *lower* standard of proof than less serious crimes such as petty theft, because of the social benefit of deterring and incapacitating dangerous or recidivism-apt criminals. Secondly, there may be cases—for example, if certain forms of criminality had especially high recidivism rates and were especially harmful—where we would maximise utility with a very low standard of proof, such as one as low as (or even lower than!) the 'more likely than not' standard found in civil law. These conclusions strike me as unpalatable, which suggests that we have went wrong somewhere by viewing the standards of proof as a utility function.

What I want to do now is to suggest a different way to think about standards of proof, by drawing on the social role of legal verdicts.

IV. LEGAL DOXASTICISM

We need an alternative framework for thinking about standards of proof. Rather than viewing the criminal standard as a utility-function, I want to

focus instead on the nature of the legal judgment it legitimises—namely, a 'guilty' verdict.

Guilty verdicts are distinct from attributions of mere causal responsibility. Someone can be causally responsible for a homicide but not judged guilty: for example, because they lack the requisite *mens rea*, or because they have successfully advanced a fully exonerating 'special defence' such as compulsion. I assume that for core elements of the criminal law—e.g. concerning interpersonal violence, infringements of property rights, and sexual crimes—the *mens rea* requirement is a way of investigating whether the accused is morally blameworthy. In core cases of criminal law then, a guilty verdict attributes *moral responsibility* to an agent for a norm-violation.[13] Using this lens, we can rephrase our inquiry by asking: when is it appropriate to judge someone morally responsible (and praise- or blame-worthy) for a given action? A full answer to this question would encompass various conditions of moral attributability (including whether the agent enjoyed the right type of control over their action or whether they possess some excuse). But standards of proof address just the *cognitive* component of responsibility-attribution: what epistemic, doxastic state, or credal state ought to underpin the decision to hold someone moral responsibility?

One method for tackling this question is to start in the interpersonal domain. Consider the case of an individual seeking to hold another individual responsible. Clearly it is unreasonable, perhaps psychologically impossible, to judge someone morally responsible for φ-ing if you actively believe that they did not φ. There needs to be some positive evaluation of the likelihood that the agent did in fact φ. This much is a platitude, but recent work goes further and argues that it is also inapt to blame simply on the basis of a positive *credence*. One influential view is that only an outright *belief* is a legitimate foundation on which to attribute blame. Call this the *belief/blame view*. In support of this view, Lara Buchak writes:

> Whether to blame or praise someone [...] is an all-or-nothing decision based, so it seems, on what I believe (or know) about the facts concerning her and her action, such as whether she actually performed the act and whether that act was permissible. While reactive attitudes do come in degrees, the degree of blame I assign to a particular agent is based on the severity of the act, not on my credence that she in fact did it. (Buchak 2014: 299)

As Buchak points out, we do not blame *in proportion to* credence. For example, suppose your babysitter says one of your two children drew on the walls but you don't know which. You do not split the blame in half and blame each child to a .5 degree. Another standard motivation for the belief/blame view appeals to scenarios in which the only evidence indicating wrongdoing is essentially

---

[13] Here, I avoid the vexed question how to theorise about 'regulatory' offences.

statistical. These include proof-paradoxical cases in the philosophy of law[14] and cases of demographic profiling.[15] To take a famous example of the former, suppose that a number of people, in fact 75 per cent of the attendees, gate-crashed a concert.[16] Would it be legitimate to take any generic member of the audience and blame them for gate-crashing solely on the basis of this statistical evidence? Presumably not. The belief/blame view vindicates this thought, as it is commonplace to suppose that such evidence only licenses a credence in guilt rather than an outright belief.[17]

A common assumption found in older works of legal theory (e.g. Duff, Farmer, Marshall and Tadros 2007), and in more recent work, in legal episte-mology (e.g. Littlejohn 2020; Smith 2018; Gardiner 2019) is that interpersonal norms are broadly the right norms for constraining criminal judgments.[18] If this is right, we can use the foregoing reflections about interpersonal blame to infer that the criminal standard of proof should demand enough evidence to ensure an observer can rationally have an outright belief in the guilt of the accused. Connecting this to our earlier discussion, this conclusion would block any decision-theoretic argument for weakening the criminal standard below that needed to believe in guilt.

Although I find the conclusion attractive, there is a lacuna in the argument. Namely, why assume that interpersonal norms apply to the legal system?[19] After all, the analogy between individual belief and legal verdicts is strained.[20] The conditions under which a criminal verdict is appropriate is determined (and can be changed) by explicit legal rules, while the standards for rational belief are—generally taken to be—given a priori or otherwise immune from instant revision. This conceptual point reinvokes the fundamental normative question: since standards of proof are open to deliberate revision, why should we not revise them in response to the decision-theoretic arguments discussed earlier? I will hazard an answer to this question.[21]

Criminal cases are brought when a community needs to arrive at a settled judgement about whether someone has committed a crime. Courts thus per-form a *zetetic* function: they conduct *inquiries on behalf of society*. More specifically,

---

[14] See Redmayne (2008) for an overview.

[15] For discussion, see, for example, Gendler (2011) or Bolinger (2020a).

[16] Original case due to Cohen (1977).

[17] There are also psychological differences in the way we evaluate statistical evidence. For example, see Hawthorne (2003: 15–20) on 'parity reasoning' in relation to lottery-cases, and studies by Gary Wells' on juror decision-making (canonically, Wells 1992).

[18] In different ways, they all connect the standards of rational belief to the standards of legal proof.

[19] Littlejohn (2020) offers extensive sympathetic discussion of this assumption. See Thorburn (2011) for an interesting rebuttal to the individual-legal parallel.

[20] This is a point made by Ross (2021). For general scepticism about the legal epistemology project, see Enoch et. al (2021), Papineau (2021), and Fratantonio (2021).

[21] For further discussion of the relevance of individual epistemology to philosophy of law, see Ross (forthcoming).

courts provide what Philip Pettit (2010) calls 'indicative representation'. The idea of indicative representation is that the representative's opinion should be a good indicator of what you *would* think had you shared the same experience as the representative—for example, had you considered the same evidence as them. The concept of indicative representation is familiar in discussions of the criminal jury—where delegation of zetetic responsibility to members of the community is explicit—but it provides a useful perspective on the role of criminal courts more generally. Even when a criminal case is decided by a professional judge, the underlying assumption is that they will only find people guilty in situations where we would also be inclined to find the person guilty, had we evaluated the evidence available. That criminal courts act as representatives in this indicative way, I suggest, is what underpins institutional trust in the criminal justice system. We see positive criminal judgments as legitimate when they are determinations we would reach had we considered the case carefully for ourselves.

Guilty verdicts license the blame and punishment of the accused, changing the normative status of the convicted person in their community. Thus, we need to trust that the institution issuing such verdicts is only doing so when appropriate. To fulfil this trust-eliciting function, criminal verdicts must serve as a certain type of *social signal*—a social signal that inquiry has been appropriately completed. In other words, courts aim to provide a certain *finality* to inquiry. If guilty verdicts did not (generally) serve as an effective signal to stop inquiry, the wider community would be awkwardly situated in relation to the blame and punishment that follows a criminal conviction. For example, whole-hearted punishment of wrongdoers would be undermined, as there is a clear tension in punishing someone while still regarding it as an open question whether they are guilty. It is therefore a normative ideal that verdicts from a criminal court should serve as an *institutional inquiry-stopper* on the question of guilt.[22]

The inquiry-stopping function of criminal verdicts creates obligations concerning the way that trials are conducted. Some obligations relate to the quality of the inquiry: for example, requiring that parties can receive professional advice, that evidence is relevant and open for scrutiny, and that judging is impartial. But, for our purposes, the most important obligation concerns the strength of evidence required before a guilty verdict is issued. In brief, my suggestion is that the inquiry-stopping role of the court demands that the following necessary condition is satisfied:

> **Legal Doxasticism:** The evidence underpinning a guilty verdict should be strong enough so that a rational observer would *believe* that the person being tried is guilty, given the admissible evidence.

---

[22] Of course, stopping inquiry is not irreversible. Just as with inquiry in our everyday lives, we can *reopen* inquiry if we receive compelling new evidence or reason to re-evaluate old evidence.

It is important for courts to signal that *belief* in guilt is appropriate, for if they did not, guilty verdicts would not serve as an effective inquiry-stopper. There are two reasons for this. The first is a specific point about the particular nature of a guilty verdict. Criminal convictions, as we have said already, are attributions of moral responsibility for norm-violations. They license blaming the person found guilty. In light of our earlier comments about the belief-requiring nature of interpersonal blame, we then see one reason for a guilty verdict to aspire to meet the standards for rational belief: only then can it effectively signal to members of society that it is permissible to blame the person found guilty of a crime. The second line of support for **Legal Doxasticism** comes from general epistemological theory. The idea that belief serves an inquiry-stopping function has recently been defended extensively, for example, in the work of Jane Friedman.[23] It also eminently plausible from considering examples. The inquiry-stopping role of belief is clearest in cases where our credence in a proposition falls below the Lockean standard required for a full belief: for example, it is difficult to regard an inquiry into p as settled in favour of p if your confidence in p is only $\sim$0.5. However, familiar cases from the literature deepen the argument by suggesting that something similar occurs even in those cases where we lack full belief despite being highly confident. For example, if I were to suggest that 'You have lost the lottery' based only on the fact that it is—statistically—improbable for you to have a winning ticket, you would be unlikely to regard this as an inquiry-stopper on the question of whether you had won the lottery.[24] Rather, you would wait for more decisive evidence. Again, a natural diagnosis is that, until you receive decisive evidence, you lack full belief that you have lost (rather, you merely have a high credence). And something similar holds in 'proof paradoxical' cases in which the guilt of an accused is statistically—but only statistically—very likely, yet where we feel an intuitive reluctance to convict. Without belief, the question of responsibility remains open in the mind of the observer.

The relevance of theorising about the interpersonal domain to criminal justice, then, is that criminal guilty verdicts ought to serve as inquiry-stoppers within the wider community. This is achieved by issuing guilty verdicts only when underpinned by evidence that would rationalise a full belief—thus, stopping inquiry and licensing blame—in the mind of an observer. Generating belief is therefore a limiting condition on a well-functioning system of criminal justice.

V. CRIMINAL PROOF: FLEXIBLE, NOT FIXED

The core insight of the previous section is that guilty verdicts must be rationally believed. Now I want to argue for a *modestly* flexible standard of criminal proof:

---

[23] In particular, see Friedman (2019).
[24] See Ebert *et. al* (2018) for overview and empirical study.

an approach which always ensures that a guilty verdict can be rationally believed, but where the strength of evidence required varies above this belief-supporting threshold.

There are two routes to a flexible system of proof, even while accepting **Legal Doxasticism.** One route to flexibility is 'external' to doxastic theories, while another is internal to their logic. The external route begins from the simple observation that rational belief in p doesn't entail having the strongest possible evidence in p. So, there are standards of proof *stronger* than requiring that belief in guilt be rational. The next step would be to identify moral reasons *independent* from the zetetic role of the court for having stronger standards of proof for certain crimes, even while allowing that all crimes should be proven to whatever standard is required for rational belief.[25] These independent reasons might be found in decision-theory, as discussed above. Or they might be found in some deontic approach. For example, one might posit that the accused has a *right* against mistaken conviction and then suggest that this right imposes variable demands—i.e., procedural protections against being found guilty—regarding different crime-types.[26] Admittedly, combining doxastic views with extra-doxastic considerations would sacrifice theoretical unity. But this line of thought is well worth pursuing in future work. Here though, I want to elaborate on considerations *internal* to the doxastic view which favour a flexible standard.

As presented thus far, the internal logic of doxastic views has been agnostic as to whether criminal proof ought to be fixed or flexible. This is because the standards for belief in guilt—and the associated license to blame—might be invariant (supporting a fixed standard of proof) or determined by the context (supporting a flexible standard of proof). I now develop the idea that doxastic views in fact support a flexible standard of criminal proof: this is because, plausibly, the evidential standards for appropriately blaming someone are context-sensitive.

There is a rich and entirely general view in epistemology that says: whether one should form a belief is sensitive not only to purely evidential considerations, but also to practical or moral considerations.[27] There are many examples of this view. A common view of this type is that the standards for rational belief are 'stake sensitive'—i.e., that the strength of evidence needed to make believing p rational depends on the practical costs of error (e.g. see Bolinger 2020a;

---

[25] See Pundik (2022) for a discussion of Beccaria's idea that the standard of proof that must be overcome to prove a crime should be linked to the crime's relative (in)frequency.

[26] For example, the strength of the right against mistaken conviction might vary with the severity of the punishment that would follow.

[27] The idea has also been briefly suggested by Amalia Amaya (2013: 23) that a stake-sensitive approach to criminal proof fits well with a generally attractive framework on which legal fact-finding is best accounted for in terms of coherence, as defined in terms of what belief an epistemically responsible fact-finder would form in the circumstances.

DeRose 1992; Stanley 2005).[28] A recent development of this view extends the same reasoning to moral considerations in addition to purely practical considerations (e.g. see Bolinger 2020b). The italicised idea also captures views that diverge in their theoretical detail from 'stake-sensitivity' views while having similar practical upshots: for example, the view outlined in Ross (2022) on which there can be moral and practical reasons to suspend judgement on a proposition instead of forming an otherwise rational belief.[29]

In support of this view, we can begin by noting that it is generally plausible to suppose that the standards we use when inquiring—our zetetic standards—are sensitive to the risks that follow from mistakes of different magnitudes. This, I think, is a familiar part of our interpersonal practices. For instance, suppose you are deciding whether someone can keep a secret. You will inquire to a higher standard in the case where the secret you want to tell them is momentous (e.g. 'I murdered my brother and buried him in the garden') rather than relatively trivial (e.g. 'I nap in Zoom faculty meetings while my camera is off'). This is equally true when it comes to other-regarding actions. To take a shop-worn example, you should require greater confidence about the nut-free status of meal before serving it to a guest with a fatal nut allergy than one without. One way to think of the effect of such risks is that they serve as *'threshold shifters'* which raise the bar for when one might permissibly proceed with judging that p (rather than the less plausible idea that risks count as *evidence* against p).[30] The foregoing is offered as a descriptive claim, but it also accords with the plausible normative principle: namely, that the greater the extent to which someone (including yourself) will be harmed by your mistake regarding p, the more robust your inquiry should be before settling on p.

These context-sensitive views about responsible belief-formation views enjoy wide popularity. But they are also controversial. Given the richness of the literature, I cannot mount a fully-fledged defence here. (My views are on record in any case.) What I will do is suggest that integrating these views into our understanding of **Legal Doxasticism** leaves us with an intuitive, empirically-supported, and theoretically unified explanation as to why the standard of criminal proof should be flexible.

The basic claim is that the beliefs constitutive of blaming appear to be sensitive to certain moral and practical considerations, at least in the cases of interest to criminal law. More specifically, I advocate a particular way of understanding the context-sensitivity of appropriate blame: (i) the standards for appropriately blaming *are raised by the costs of false positives*, but (ii) the standards for

---

[28] This view is often stated in terms of knowledge rather than rational belief (but many epistemologists suggest that aiming to know is a norm of rational belief).

[29] For example, there may be moral reasons to avoid demographic profiling in general, even if some negative profiles might be epistemically rational on the available evidence.

[30] The term 'threshold shifting' is borrowed from Worsnip (2021).

blaming are *not* *lowered* *by the costs of false negatives.* Stated succinctly, the proposal is that every conviction must satisfy a relatively demanding standard so that guilt can be rationally believed by the observer, but the requisite standard should vary with the risks associated with passing a positive judgement. The upshot of this position is that the criminal standard of proof ought to be modestly flexible in the following way:
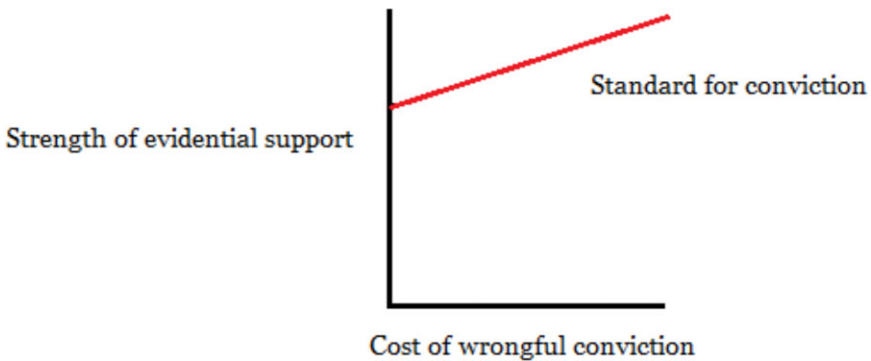
Figure 1

Different acts of condemnation expose people to greater or lesser costs, which means that the costs of false positives are higher in some situations rather than others. Criminal guilty verdicts expose those convicted to varying risks if the conviction is erroneous, primarily in the form of the punishments meted out for different crimes.[31] My suggestion is that we ought to require more confidence when the costs to the accused of erroneous blaming are greatest, given the plausible principle that we should be more cautious when the risks of error are higher. To take an example: the claim is that the reasonable observer would regard it as appropriate to demand more confidence to attribute responsibility for a crime if the punishment is death or life imprisonment compared with the case where the penalty is a $400 fine. This strikes me as a powerful intuition in favour of the proposal. It is an intuition that accords with the normative principle mentioned earlier, that we should adopt greater caution before issuing a decision that has greater costs for the person who will bear the consequences.

---

[31] A second cost of false positives are what Hoskins (2019) calls 'collateral consequences' of conviction. These are burdensome consequences that are not part of the formal sentence, but nevertheless reliably flow from conviction. For example, persons convicted find it difficult to secure employment. I am inclined to view collateral consequences as a relevant cost: they are often predictable and more burdensome than the formal sentence. A small fine might pale in comparison with the downstream effects of struggling to find employment.

There is empirical research suggesting that legal fact-finders agree. Mock jury studies have suggested a general correlation between reluctance to convict and severity of punishment, for a range of different severities. Surveys of real jurors suggest a greater reluctance to convict when there is a possibility of very severe punishment, such as the death penalty.[32] Further support is found in 'natural experiments', where we look at the effect that the introduction or abolition of severe punishments has on conviction rates.[33] For instance, Bindler and Hjalmarsson (2018) exploit a dataset involving over 200,000 historic cases to show that the removal of the death penalty and penal transportation as punishments caused a marked increase in conviction rates for the same crimes. It is worth being explicit that: (i) jury science is methodologically fraught, (ii) the relationship between severity of punishment and the evidential standard for conviction is under-studied, and (iii) alternative interpretations could be given of the evidence cited.[34] Nevertheless, existing evidence is suggestive in favour of my claim.

An upshot of the focus on the costs of wrongful conviction is that the appropriate stringency of standards of proof is plausibly what we might call 'code-relative'. If one criminal code has extremely severe punishments, then it may be that weightier evidence is needed to properly support a conviction than a different criminal code with less harsh punishments. This is borne out by the above-mentioned empirical evidence. It is also, I think, generally plausible: the introduction of the death penalty or whole-life tariffs is precisely the type of thing that ought to recommend greater caution before convicting those who are going to be susceptible to such severe punishment. Indeed, this is yet another reason for jurisdictions to endorse codes that avoid excessively harsh punishments.

The doxastic approach sketched provides us with a simple and elegant way to explain why the standard of criminal proof should be flexible, while still ensuring that the lower bound conforms to a fairly demanding standard. However, as has been emphasised throughout the paper, there are *two* types of risks in criminal adjudication: the risk of mistakenly condemning (and punishing) the innocent, but also the risk of failing to condemn (and punish) the guilty. These latter risks are primarily community-borne, for instance, in the form of recidivist criminality. So, why only consider the costs of false positives? Why should a responsible inquirer not take both risks into account, lowering the evidential standards in response to community-borne risks as well

---

[32] E.g. Kerr (1978) and Kaplan and Krupa (1986).

[33] E.g. Freedman (1990).

[34] An alternative reading of the evidence is that jurors are protesting against very harsh punishments by acquitting, rather than raising their evidential threshold. However, in Bindler and Hjalmarsson's case-set, there was widespread social acceptance of the death penalty as legitimate.

as raising the standards in response to accused-borne risks, thus seeking the optimum balance?

Although superficially tempting, this line of thought relies on a mistaken perspective on belief and blaming. The risk of *not* attributing a negative property to someone (were they to have that property) is not a reason to *lower* the evidential threshold for *condemning* them. An example will clarify. Take a case where someone is accused of sexual wrongdoing. Suppose, there is equivocal evidence supporting the accusation. You judge that the evidence is too weak for you to believe that they are guilty: as such, you refrain from blaming them. Suppose you are *then* reminded that the risks of failing to correctly identify a sexual wrongdoer are weighty, in a way that you had not appreciated when weighing up the evidence initially. Being reminded of these risks is a reason to *revisit* the evidence, to make sure that you are weighing it up correctly. But, I submit, appreciating these community-borne risks is not the sort of thing that justifies blaming them when you previously regarded it as unjustified do so. Responsible blaming is not instrumental in this way.

Rather, what appreciating such community-borne risks *might* do is rationalise precautionary behaviours and attitudes that fall short of blaming. For example, assigning a positive credence (short of belief) to the proposition that someone poses a danger is something that could justify taking various precautionary steps, measures which Ross (2021) refers to as *hedging*. For example, even without blame, you might take steps to protect yourself from someone against whom there are allegations supported by equivocal evidence. Many employment opportunities depend on passing background checks for which a serious allegation of sexual abuse would be disqualifying even if the person was acquitted in a criminal court. The morality of hedging is debatable.[35] However, what is notable about these protective behaviours is that they do not presuppose an outright belief in guilt and the associated attitude of blame. So, denying that the standards for belief and blame are dragged down by the cost of false negatives does not mean that these risks are irrelevant for practical decision-making. The claim is simply that coming to appreciate these community-borne risks cannot transform an unjustified act of moral condemnation into a justified one.

It is not the role of criminal courts to legitimise hedging. Indeed, this point is implicitly recognised by the way the standard of proof operates in criminal trials: it is used in such a way as to screen off reasons to hedge from the decision. To see why, consider the way criminal verdicts are generated. When the trial proceeds to consider whether the 'facts at issue' have been proved,[36] the

---

[35] For example, iterative hedging can lead to oppression against those typically the subject of adverse profiles. See Mogensen (2019) for discussion.

[36] Of course, the court may decline to proceed to this stage, even when the evidence is strong enough to prove the facts on the relevant standard. For example, if (otherwise compelling) evidence is improperly gathered, the case will be thrown out.

fact-finder issues verdicts with reference to a specific decision-rule involving the standard of proof:

> **Decision Rule: '**Find the accused "guilty" iff the standard of proof is met; return a "not guilty" verdict otherwise.'

This seems straightforward. But behind this simple rule is a theoretically important upshot: namely, *there is no standard of proof for innocence*. Rather, 'not guilty' verdicts are returned in the absence of sufficient proof of guilt—the release of the accused is an *indirect* result of the evidence being too weak to warrant conviction. Therefore, a 'not guilty' verdict is compatible with a wide range of credences in the *innocence* of the accused.[37] Criminal courts are thus not entirely analogous to individual agents. While an individual considering someone's guilt has the full menu of cognitive options open to them—to believe in guilt, to assign a positive credence to guilt, to suspend judgement, and to believe in innocence—courts have a more circumscribed task. A court only decides whether to find the accused guilty: not to judge them innocent, nor assign any credence to their guilt. Criminal courts therefore have a type of 'one-eyed' task. A basic aim of a 'guilty' verdict is to send a social signal that it is appropriate to believe in the blameworthiness of the accused. I have argued that the conditions under which a blaming belief is appropriate is modulated by the consequences of error for the accused. However, a 'not guilty' verdict cannot be viewed as sending a social signal that the evidence supports full belief in innocence, nor that interpersonal hedging is irrational.[38]

Since criminal courts do not concern themselves with the rationality of hedging, and because the risk of failing to blame someone does not justify starting to blame them, only the potential costs borne by the accused shift the threshold for criminal conviction. This means we should require more confidence to convict of more serious crimes, given the greater amount of punishment meted out for them. However, forward-looking risks of acquitting the guilty should not drag down the standard of proof. So, we vindicate a compelling intuitive pattern:

> **Accept:** Heightened confidence before convicting of *x* because the punishment for *x* is very severe (e.g. death).

---

[37] An intriguing exception is found in non-binary verdict schemes. See Picinali (2022) for a book-length treatment.

[38] See the ICC in *Prosecutor v Katanga* [ICC-01/04–01/07]: 'finding an accused person not guilty does not necessarily mean that the Chamber finds him or her innocent. Such a determination merely demonstrates that the evidence presented in support of the accused's guilt has not satisfied the Chamber 'beyond reasonable doubt'.'

Reject: Lowered confidence to convict of y because of the risks of letting a y-guilty person loose (e.g. high average recidivism rate).

Our discussion has been concerned with the theory of criminal proof. But we are left with a practical proposal that, on its face, is revisionary: instead of using the same standard of proof for all crimes, use a higher standard for crimes with more severe punishments. Interestingly, *how* revisionary this proposal is turns out to be difficult to assess. This is because there may be a mismatch between the 'official doctrine' that the same standard of proof applies across the board and the real-life behaviour of judges and juries.

It is a familiar point that there is a difference between: (i) the decision made by a judge or jury in a given case, and (ii) the decision made when designing the institution in which the fact-finder operates.[39] Here, I addressed the issue from the second perspective—what choice should be made by the designer when setting the standard of criminal proof? There are different ways in which a designer can implement their answer. One is to create precise rules for those who decide on guilt. Another is to offer less determinate guidance, relying on the judge's or jury's natural facility to vary their approach depending on the case at hand. In many legal systems, it appears that the latter route is predominant. Jurisprudential and statutory orthodoxy declines to offer close guidance on how to interpret the prevailing 'beyond reasonable doubt' standard. Rather, the correct application is left to the judge or jury to decide for themselves. For example, English criminal law now instructs juries to simply ask themselves 'Are you sure?' that guilt has been proven.[40] To my mind, this sounds little like a precise rule and more like reliance on the discretion of the jury.[41] It is therefore possible that fact-finders vary how they interpret 'reasonable doubt' depending on the case at hand, even though not formally instructed as such.[42] If this the case, then we have given a theoretical justification for something that may already be an important yet underappreciated and opaque aspect of legal adjudication.

To state things more schematically, there are two ideas worth distinguishing. This paper has argued for one version of a general normative claim:

(A) Stronger evidence should be required before convicting of certain crimes.

Specifically, I argued, on non-consequentialist grounds, for a version of this claim on which the standard should vary with the potential costs to accused. This is distinct from:

---

[39] See, prominently, Rawls (1955).

[40] The official position in the *Crown Court Compendium* (5–3) is that 'being sure' is equivalent to 'beyond reasonable doubt'.

[41] McKeown (2022) discusses empirical work on this instruction.

[42] See Loeb and Reyes (2022) for the radical view that standards of proof do not aim to specify *any* evidential threshold.

(B) 'Beyond reasonable doubt' is and should be interpreted differently in different situations.[43]

This is a narrower jurisdiction-specific claim about how fact-finders deal with a particular way of formulating the standard of proof. **(B)** is one way to *operationalise* the general normative claim—**(A)**—argued for in this paper. As an operationalisation it has advantages: not least that it allows us to avoid delicate questions about how to formulate distinct standards of proof and direct fact-finders accordingly. It is not the only operationalisation. (An alternative approach would be to explicitly formulate different standards of proof and design mechanisms for choosing which explicitly distinct standard ought to govern the criminal trial at hand.) We have already discussed empirical evidence and a theoretical framework that would explain why **(B)** may effectively operationalise **(A)**. But we must await further empirical results to better understand whether this is right. The current state of jury science is fragile and suffers from real questions concerning ecological validity, partly stemming from the use of mock juries.[44] We need to better understand the dynamics of criminal adjudication—preferably studying real juries—to get a better grip on the best way to implement the general normative perspective on criminal proof argued for here. Much depends on whether real fact-finders do interpret 'beyond reasonable doubt' in line with the general normative perspective outlined.

A final issue arises. Since the standard ought to vary with the putative cost to the accused, should juries be given information about the typical punishments for different crimes? As far as I am aware, there is a paucity of empirical evidence on how accurate jurors are in their assessment of likely punishments. There is nevertheless a modest yet persistent line of work in legal theory that precisely argues that juries should be made aware of the likely punishment range for the crime they are considering (e.g. see Cassack and Huemann 2007). I agree. And there is in fact a simple argument for this idea, beyond the general motivation already offered. Most criminal cases are already decided by fact-finders who know the likely range of punishments: the paradigm fact-finder is a professional judge well versed in punishment practices, and who will have a view as to what punishment they might eventually impose. This means there is already in many jurisdictions an under-discussed asymmetry between adjudication by professional judge and adjudication by jury. Informing juries about punishments removes a substantial asymmetry in criminal procedure.

A more specific and derivative issue is whether juries ought to be given information about the likely punishment that would be given to the *particular* defendant. For example, repeat offenders will usually be liable to harsher

---

[43] For further discussion of whether 'reasonable doubt' might be interpreted in a variable way, see Lillquist (2002) or Picinali (2013).

[44] Ross (2023) outlines these shortcomings.

punishments than first-time offenders. Should the former, then, be afforded the protection of a higher standard of proof? This is a delicate question, to which I lack a conclusive answer. However, in many cases, the point is moot: it will not be in the defendant's interests to have past convictions revealed. Thus, the best protection to the repeat offender will often be secured by keeping their past convictions hidden from the jury, as happens in many legal systems.

VI. CONCLUSION

Should we use the same standard of proof to judge guilt for all crimes? Or should we vary the standard depending on the crime?

Rejecting popular decision-theoretic analyses, I instead offered a non-consequentialist perspective on criminal proof. I argued that we should require stronger evidence to convict of crimes that attract harsher punishments. So, for example, we should require stronger evidence to convict somebody of murder than petty theft. This moderate flexibility in our legal standards respects the fact that the risks to the accused are greater when more severe punishments are on the table. This follows from a general perspective on which we ought to only convict on evidence strong enough to underpin a belief in the blameworthiness of the accused, where the appropriateness of blaming beliefs isare sensitive to contextual facts about the costs of false positives. Imposing a doxastic requirement on conviction, I suggested, is what allows the criminal court to effectively fulfil its social role of settling questions of guilt on behalf of the community.

This paper ranged only over criminal proof, leaving proof in civil law untouched. Future work would do well to address analogous civil law questions: civil cases vary in many—but not all—of the same ways as criminal cases.[45]

REFERENCES

Allen, R. J., Pardo and M. S. (2019) 'Relative Plausibility and Its Critics', *International Journal of Evidence and Proof*, 23: 5–59.

Amaya, A. (2013) 'Coherence, Evidence, and Legal Proof', *Legal Theory*, 19: 1–43.

Bindler, A. and Hjalmarsson, R. (2018) 'How Punishment Severity Affects Jury Verdicts: Evidence from Two Natural Experiments', *American Economic Journal: Economic Policy*, 10: 36–78.

Blackstone, W. (1827) *Commentaries on the Laws of England*, vol. 2. Oxford: CreateSpace Independent Publishing Platform.

Bolinger, R. (2020a) 'The Rational Impermissibility of Accepting Racial Generalizations', *Synthese*, 197: 2415–31.

——. (2020b) 'Varieties of Moral Encroachment', *Philosophical Perspectives*, 34: 5–26.

Buchak, L. (2014) 'Belief, Credence, and Norms', *Philosophical Studies*, 169: 1–27.

Cassak, L. and Heumann, M. (2007) 'Old Wine in New Bottles: A Reconsideration of Informing Jurors about Punishment in Determine-and Mandatory-Sentencing Cases,' *Rutgers Journal of Law & Public Policy*, 4/2: 411–96.

Cohen, J. L. (1977) *The Probable and the Provable*. Oxford: OUP.

DeRose, K. (1992) 'Contextualism and Knowledge Attributions', *Philosophy and Phenomenological Research*, 52/4: 913–29.

Di Bello, M. (2019) 'Trial by Statistics: Is a High Probability of Guilt Enough To Convict?', *Mind*, 128: 1045–84.

Duff, A. (2003) 'Penance, Punishment and the Limits of Community', *Punishment & Society*, 5: 295–312.

Duff, A., Farmer, L., Marshall, S. and Tadros, V. (2007) *The Trial on Trial: Towards a Normative Theory of the Criminal Trial*. London: Hart Publishing.

Ebert, P., Smith, M. and Durbach, I. (2018) 'Lottery Judgments: A Philosophical and Experimental Study', *Philosophical Psychology*, 31/1: 110–38.

Enoch, D., Fisher, T., Spectre and L. (2021) 'Does Legal Epistemology Rest on a Mistake? On Fetishism, Two-tier System Design, and Conscientious Fact-Finding', *Philosophical Issues*, 31: 85–103.

Epps, D. (2015) 'The Consequences of Error in Criminal Justice', *Harvard Law Review*, 128: 1065–151.

Fratantonio, G. (2021) 'Evidence, Risk, and Proof Paradoxes: Pessimism about the Epistemic Project', *The International Journal of Evidence & Proof*, 25: 307–25.

Freedman, J. L. (1990) 'The Effect of Capital Punishment on Jurors' willingness To Convict', *Journal of Applied Social Psychology*, 20: 465–77.

Friedman, J. (2019) 'Inquiry and Belief', *Noûs*, 53/2: 296–315.

Gardiner, G. (2017) 'In Defence of Reasonable Doubt', *Journal of Applied Philosophy*, 34: 221–41.

——. (2019) 'The Reasonable and the Relevant: Legal Standards of Proof', *Philosophy & Public Affairs*, 47/3: 288–318.

Gendler, T. (2011) 'On the Epistemic Costs of Implicit Bias', *Philosophical Studies*, 156: 33–63.

Greer, G. (2018) *On Rape*. London: Bloomsbury.

Hawthorne, J. (2003) *Knowledge and Lotteries*. Oxford: OUP.

Hedden, B., Colyvan and M. (2019) 'Legal Probabilism: A Qualified Defence', *Journal of Political Philosophy*, 27: 448–68.

Horowitz, I. A., Kirkpatrick and L. C. (1996) 'A Concept in Search of a Definition: The Effects of Reasonable Doubt Instructions on Certainty of Guilt Standards and Jury Verdicts', *Law and Human Behavior*, 20: 655–70

Hoskins, Z. (2017) 'Punishment', *Analysis*, 77, 3: 619–32.

——. (2019) *Beyond Punishment? A Normative Account of the Collateral Legal Consequences of Conviction*. New York, NY: OUP.

Kaplan, J. (1968) 'Decision Theory and the Factfinding Process', *Stanford Law Review*, 20: 1065–92.

Kaplan, M. F., Krupa and S. (1986) 'Severe Penalties under the Control of Others Can Reduce Guilt Verdicts', *Law & Psychology Review*, 10: 1–18.

Kerr, N. L. (1978) 'Severity of Prescribed Penalty and Mock Jurors' Verdicts', *Journal of Personality and Social Psychology*, 36: 1431–42.

Laudan, L. (2006) *Truth, Error, and Criminal Law: An Essay in Legal Epistemology*. Cambridge: CUP.

——. (2011) 'The Rules of Trial, Political Morality, and the Costs of Error: Or, Is Proof beyond a Reasonable Doubt Doing More Harm than Good?', in L. Green, and B. Leiter (eds.) *Oxford Studies in the Philosophy of Law I*. Oxford: OUP.

——. (2017) *The Law's Flaws: Rethinking Trials and Errors?* London: College Publications.

Lillquist, E. (2002) 'Recasting Reasonable Doubt: Decision Theory and the Virtues of Variability', *UC Davis Law Review*, 36: 85–197.

Littlejohn, C. (2020) 'Truth, Knowledge, and the Standard of Proof in Criminal Law', *Synthese*, 197: 5253–86.

Loeb, D. and Molina, S. R. (2022) 'Standards of Proof as Competence Norms', *Jurisprudence*, 13: 349–69. https://doi.org/10.1080/20403313.2022.2049077.

McKeown, P. (2022) 'Dreaded Questions, Doubtful Answers & the Trouble with "Sure"', *New Law Journal*, 172: 13.

Mogensen, A. (2019) 'Racial Profiling and Cumulative Injustice', *Philosophy and Phenomenological Research*, 98/2: 452–77.

Papineau, D. (2021) 'The Disvalue of Knowledge', *Synthese*, 198: 5311–32.

Pettit, P. (2010) 'Representation, Responsive and Indicative', *Constellations*, 17: 426–34.

Picinali, F. (2013) 'Two Meanings of "Reasonableness": Dispelling the "Floating" Reasonable Doubt', *The Modern Law Review*, 76: 845–75.

——. (2018) 'Can the Reasonable Doubt Standard be Justified? A Reconstructed Dialogue', *Canadian Journal of Law & Jurisprudence*, 31: 365–402.

——. (2022) *Justice In-Between*. Oxford: OUP.

Pundik, A. (2022) 'Should Murder Be More Difficult To Prove than Theft? Beccaria and Differential Standards of Proof', in A. Bois-pedain and S. Eldar (eds.) *Re-reading Beccaria*. Bloomsbury.

Rawls, J. (1955) 'Two Concepts of Rules', *Philosophical Review*, 64: 3–32.

Redmayne, M. (2008) 'Exploring the Proof Paradoxes', *Legal Theory*, 14/4: 281–309.

Ribeiro, G. (2019) 'The Case for Varying Standards of Proof', *San Diego Law Review*, 56: 161–219.

Ross, L. (2021) 'Rehabilitating Statistical Evidence', *Philosophy and Phenomenological Research*, 102: 3–23.

——. (2022) 'Profiling, Neutrality and Social Equality', *Australasian Journal of Philosophy*, 100:808–24, doi: 10.1080/00048402.2021.1926522.

——. (2023) 'The Curious Case of the Jury-Shaped Hole: A Plea for Real Jury Research', *The International Journal of Evidence & Proof*: https://doi.org/10/grk2rg.

——. (forthcoming) 'The Foundations of Criminal Law Epistemology', *Ergo*.

Smith, M. (2018) 'When Does Evidence Suffice for Conviction?', *Mind*, 127/508: 1193–218.

Solan, L. (1999) 'Refocusing the Burden of Proof in Criminal Cases: Some Doubt about Reasonable Doubt', *Texas Law Review*, 78: 105–47.

Stanley, J. (2005) *Knowledge and Practical Interests*. Oxford: OUP.

Stein, A. (2005) *Foundations of Evidence Law*. Oxford: OUP.

Thorburn, M. (2011) 'Criminal Law as Public Law', in A. Duff and S. P. Green (eds.), *Philosophical Foundations of Criminal Law*. Oxford: OUP.

Tribe, L. H. (1971) 'Trial by Mathematics: Precision and Ritual in the Legal Process', *Harvard Law Review*, 1329–93.

Walen, A. (2015) 'Proof beyond a Reasonable Doubt: A Balanced Retributive Account', *Louisiana Law Review*, 76: 355–446.

Wareham, C. and Vos, J. (2018) 'Why Rape Cases Should Not be Subject to beyond Reasonable Doubt', *Aeon*: https://aeon.co/ideas/why-rape-cases-should-not-be-subject-to-reasonable-doubt, accessed January 2023.

Wells, G. L. (1992) 'Naked Statistical Evidence of Liability: Is Subjective Probability Enough?', *Journal of Personality and Social Psychology*, 62: 739–52.

Wood, D. (2010) 'Punishment: Consequentialism', *Philosophy Compass*, 5: 455–69.

Worsnip, A. (2021) 'Can Pragmatists Be Moderate?', *Philosophy and Phenomenological Research*, 102: 531–58.

London School of Economics, UK