Cambridge Elements

Philosophy of Law

The Philosophy of Legal Proof

Lewis Ross

https://doi.org/10.1017/9781009127745 Published online by Cambridge University Press

ISSN 2631-5815 (online)
ISSN 2631-5807 (print)

Contents

Introduction 1

1 Standards of Proof 1

2 Proof: Fixed or Flexible? 18

3 Should Proof Be Binary? 27

4 Legal Probabilism and Anti-Probabilism 36

5 Who Should Decide? 52

References 67

https://doi.org/10.1017/9781009127745 Published online by Cambridge University Press

~~cases that do not to fit with our general understanding. But the philosophy of civil proof is an area where there is a great deal of important philosophical work still to be done.~~

2 Proof: Fixed or Flexible?

The standards of proof appear to be surprisingly inflexible. Take criminal law. The very same standard is used to determine guilt for war crimes in the International Criminal Court and for petty theft in your local court. In each case, the standard is proof 'beyond reasonable doubt'. There is a similar inflexibility in many areas of civil law. The same standard is used to decide whether your garden hedge is overgrown and whether the state should remove a child at risk from their family home. In each case, proof is 'on the balance of probabilities'.

From one perspective, such inflexibility might seem fair and reasonable – everyone is treated equally in court, putative genocidal maniacs and petty thieves alike. But from another perspective, it is strange. Generally, when making decisions in our everyday lives, we change our approach depending on the type of decision we are making. You will want to be more certain when making a bet involving a year's salary compared to a £20 flutter on Saturday's football scores, more certain about weatherproofing when buying a house compared to renting an Airbnb.

Just like decisions in everyday life, decisions faced by criminal and civil courts vary in their gravity. It might be hard to specify all the ways in which cases can differ, but it is uncontroversial to say that some cases involve bigger risks than others. For example, a finding of guilt for petty theft might cause the accused to receive a small fine, while for murder it might lead to them receiving a long prison sentence or even the death penalty. This raises the question: should the legal system use different standards for different cases? We will focus on criminal law primarily. So, we begin by asking: should criminal justice systems use different standards for different crimes?

2.1 Different Standards for Different Crimes?

The idea that some crimes require special treatment has a long history. Consider the crime of treason, which historically came with liability to especially gruesome punishments – most infamously, being hung, drawn, and quartered. For long periods in the history of English law, you could only be convicted of treason through the testimony of *two* eyewitnesses, a rule not applied to other crimes. Curiously, this remains codified in the US Constitution – the drafting of

https://doi.org/10.1017/9781009127745 Published online by Cambridge University Press

which was influenced by English legal tradition – where it still requires the testimony of two witnesses to convict someone of treason.[26]

The eighteenth century Italian jurist Cesare Beccaria supposed that serious crimes were less common, and so we should require more evidence to be convinced that someone has committed one.[27] For example, it is presumably less common to be a serial killer than a petty thief. (But there do seem to be clear exceptions to Beccaria's claim. Sexual assaults are regarded as very serious, yet many think they are not especially uncommon.)[28]

Beccaria's larger point seems not to be that we should use a higher standard for some crimes compared to others. Rather, he suggested that the empirical distribution of criminal tendencies makes it rational have more default scepticism about the suggestion that someone has done something especially terrible. (Compare: it might be *harder* for me to be convinced that my laziest student managed to ace the exam, but this doesn't mean I should use a more demanding standard when grading their work compared to other students.) This means that everything Beccaria says is compatible with the current 'beyond reasonable doubt' standard. His point is just that doubts are more reasonable when someone has been accused of something especially nasty.

But this raises an interesting psychological question – do people tend to *interpret* 'beyond reasonable doubt' differently in different contexts? This isn't something that judges or juries are explicitly instructed to do. But it is not implausible that people will change what they count as a reasonable doubt depending on the type of crime they are considering. This raises the question, if judges or juries *do* interpret the standard differently across contexts, what factors are their interpretations sensitive to? Is it (just) the statistical rarity of different crimes? The severity of the punishment? The extent to which the victim was harmed? One's emotional reaction to the case? Something else? Some of these possibilities would seem to involve raising the standard for certain crimes, going rather further than Beccaria's suggestion to pay attention to the relative rarity of the crime being alleged.

As philosophers, we are not best equipped to answer questions of psychology.[29] But we can consider whether we should endorse or reject certain psychological tendencies that people have when approaching the standards of proof. To do this, we must ask the larger moral and political question – should we design the legal system so that it has multiple standards of proof, using different standards for different crimes? For example, some suggest that capital (death penalty) crimes should be accompanied with a higher standard of proof

[26] US Constitution, Article III, Section 3, Clause 1. [27] Beccaria 1995.
[28] This point is made by Pundik 2022, who provides a helpful discussion of Beccaria's argument.
[29] Although this often does not stop us trying.

https://doi.org/10.1017/9781009127745 Published online by Cambridge University Press

than other crimes.[30] If this is right, that we should use different standards for different crimes, what factors should justify the use of stricter or less strict standards?

A first consideration is whether there is some value in the equality of all crimes being judged against the same standard. I have heard it suggested that the right to a fair trial or treating everyone 'as equals under the law' might require using the same standard for all crimes. But giving everyone a fair trial or treating them as equals doesn't require *identical* treatment. People can share a right to a fair trial, yet this right can be realised in different ways. For example, in some jurisdictions, only comparatively serious crimes receive trial by jury, with less serious offences being decided by a single judge. It is generally true of rights that they can be realised in different ways, depending on the risks and contextual facts of the situation. A right to health, for example, might be shared among all citizens – yet, it is still appropriate to use different standards and procedures when dealing with some patients (e.g. because their disease is especially serious or contagious) than others.

A different argument for using the same standard across all cases concerns public perception. It might undermine public confidence in certain trials if it were widely known that it is easier to get convicted of (say) sexual assault compared to murder or fraud – these might come to be viewed as second-class convictions, unreliable in the eyes of the public. This is an important worry. But it is also true that public confidence is compatible with some variation in the procedures used to judge guilt. Again, some jurisdictions have both trials with juries and trials with judges. This split system – variations of which are found across the world – does not seem to fatally degrade confidence in the criminal justice system. Nor does it seem to degrade public confidence that civil law has a lower standard of proof than criminal law; people are often happy to talk as if civil cases have been 'proven' in much the same way as they do in criminal cases. So, while public confidence is a relevant concern, it isn't the end of the conversation either.

Let's revisit the traditional idea from Blackstone that helped us to analyse the beyond reasonable doubt standard (Section 1.5): 'All presumptive evidence of felony should be admitted cautiously, *for the law holds that it is better that ten guilty persons escape than that one innocent suffer.*' People often use this statement to justify a single 'fixed' BRD standard. But this famous quote is much more plausible when interpreted as *relative* to different types of crime. For example, it may be plausible to prefer letting ten guilty murderers go free than mistakenly putting one person to death after being falsely convicted of

[30] See Sand and Rose 2003.

https://doi.org/10.1017/9781009127745 Published online by Cambridge University Press

murder. But it is much less plausible to think it better to let ten guilty murderers go free than to mistakenly give somebody a £100 fine for being falsely convicted of stealing a chocolate bar. We can press the worry further. Suppose there *is* some rough number of murderers that we should prefer to go free rather than wrongly convicting one person of murder. Why should we think that this same number applies to other crimes – that the same balance appropriate for murder also applies to theft, sexual crimes, fraud, arson, and so on?

Indeed, if we return to the consequences-based arguments discussed in the previous section, the justification of a single standard becomes even less obvious. The prospective costs of different types of error – releasing the guilty and convicting the innocent – will vary across different types of case. For instance, Larry Laudan's argument for lowering the standard of proof focuses on predicted recidivism rate. This is precisely the type of thing that changes with the type of crime. For example, statistics from England and Wales suggest that 30 per cent of those convicted of knife-related crimes are reoffenders who have committed similar offences in the past.[31] By contrast, those convicted of rape tend to have comparatively low reoffending rates.[32] Although Laudan pitched his argument as lowering the standard of criminal proof across the board, it actually better supports the idea that it should be *lowered for certain crimes*. After all, his point about high recidivism rates and risk of harm only drew on considerations about violent crimes – not (say) economic crimes. Laudan's argument, if you accept his premises, really just gets you to the conclusion that we should lower the standard of proof for some crimes.

This point generalises to other arguments that try to set the standard of proof by looking at the consequences of error.[33] For example, consider the things that vary between crimes:

- Some crimes are met with higher punishments, so punishing the innocent is comparatively worse.
- Some offenders are more likely to reoffend than others.
- Some types of reoffending are more harmful to the community than others.
- Some offenders are easier to rehabilitate.
- Some crimes are more susceptible to deterrence.

[31] See GOV.UK website, 'Knife and Offensive Weapon Sentencing Statistics: July to September 2022' (Main Points, 3rd point). https://bit.ly/3TpRFI2.

[32] For example, Freedom of Information data suggests that 4 per cent of rape convictions between 2013–2017 were against those with previous convictions. See Full Fact 2019.

[33] For consequences-based arguments for having different standards for different crimes, see Lillquist 2002 and Ribeiro 2019.

https://doi.org/10.1017/9781009127745 Published online by Cambridge University Press

- Some crimes have unusually low conviction rates.
- Some crimes are higher-profile and convictions send out a stronger public message.

If we take a consequences-based approach, given the many ways crimes can vary it seems unlikely that requiring the same fixed level of confidence before convicting would optimise the risks in every single case. The consequences-based approach instead supports a much more flexible approach to proof, where the standard used for a given crime type is sensitive to the listed factors. If we accept this way of thinking, then we should use different standards for different crimes, depending on the expected consequences associated with each.

Is this a plausible approach? Earlier, we tested the merits of consequences-based thinking by taking the view to its logical conclusions. We can do the same for the consequences-based perspective on using different standards of proof for different crimes.

Suppose that some very serious crime type like murder or rape happened to be high-profile, have very high reoffending rates, be amenable to rehabilitation and deterrence through long punishment, and be particularly upsetting to the community. In this case, the benefits of convicting the guilty are raised compared to the costs of punishing the innocent. On the line of thinking we are considering, this would mean that the standard of proof for these serious crimes should become *lower* relative to other crimes. (Convicting more people would allow us to pluck the low-hanging fruit of the benefits I just stipulated.) However, it seems perverse to have a legal system where what we traditionally regard as the most serious crimes that come with the harshest punishments are assessed against *weaker* standards than more trivial crimes such as petty theft.[34] Something seems to have gone wrong with this way of thinking, if the consequences-based approach could license using a lower standard for murder or rape than for petty theft. We need a different perspective.

2.2 Another Perspective on Flexible Proof

In Section 1, I outlined a view on which the criminal standard of proof should support rational belief in the guilt of the accused. Can we use this belief-based theory to make progress on the debate about using multiple standards of proof? To answer this question, we need to consider the relationship between rationally believing something and strength of evidence.

[34] I have made this argument in Ross 2023b, where I develop my view on multiple standards of proof in more depth, arguing that we should use more demanding standards for crimes with the severest punishments.

https://doi.org/10.1017/9781009127745 Published online by Cambridge University Press

There is some threshold of evidence required for it to be rational to believe something (given the evidence: don't believe in Flat Earthism; do believe in anthropogenic climate change). But does this threshold change depending on the situation? Epistemologists – those who study belief and the evaluative concepts used to assess beliefs – are engaged in long-running debate about this question. A key fault line in their debates is whether the rationality of believing something depends on the practical consequences of being right or wrong.[35]

One view is that practical consequences make *no* difference to what it takes for a belief to be rational; only 'intellectual' considerations determine whether you should believe something. Some call this view *intellectualism* about belief. The opposite view is that the rationality of believing something *is* affected by the practical consequences associated with the belief (and what happens if you have the wrong belief). Take a mycological example. Will a sensible person require *stronger evidence* to rationally believe that a red-and-white dotted mushroom is safe when they plan to eat it, compared to just believing it is safe while walking past it on a countryside ramble? Cases like this motivate some to argue that the practical consequences (or 'stakes') influence the strength of evidence needed to rationally believe something – the greater the harm if your belief is wrong (e.g. eating the mushroom and dying), the more evidence it takes to rationally believe (e.g. the mushroom is safe). Some call such views *stake-sensitive* theories of belief.

If intellectualist views about belief are right, then the belief-based theory of criminal proof isn't particularly supportive of using different standards for different crimes – since none of the practical consequences matter when forming a belief, the different benefits and risks of legal error don't impinge on the standards required to believe the person is guilty. But *if* stake-sensitive views are correct, then a belief-based approach to criminal proof ought to be sensitive to some of the practical differences between convictions and acquittals for different types of crime. In my view, stake-sensitive views are attractive because they consider the close relationship between believing something and being in a position to act on it.[36] A canonical role that belief plays is to guide our behaviour, so it's natural to think that the standards for believing rationally should be influenced by the results of the action that the belief will license.

When it comes to convicting people of crimes, the practical 'stakes' can be divided into two categories, corresponding to different mistakes we can make. The first is the risk to the accused of convicting them when they are actually

[35] See Kim 2017 for a summary. My contribution to this debate is outlined in Ross 2022.
[36] There are also difficult challenges for stake-sensitive views. For example, see Worsnip 2021.

https://doi.org/10.1017/9781009127745 Published online by Cambridge University Press

innocent. The second is the risk to the community if we release someone who is actually guilty.

Given the second type of risk, it might seem that thinking about the stake-sensitive nature of belief just recreates Laudan's argument: that it is rational to believe something (that someone is guilty of a violent crime) on the basis of weaker evidence, if the consequences of error (mistakenly acquitting a criminal) are particularly bad (they are likely to reoffend and harm society). If this is the right way to understand how the stake-sensitive theory applies to criminal proof, then we haven't made much progress; we have just recreated the implausible conclusion that we could find someone guilty of (say) murder based on weaker evidence than convicting them of petty theft.

Fortunately, this is not the best interpretation of what the stake-sensitive view means for criminal proof. I agree that some risks should make us require stronger evidence before judging someone guilty. In particular, if someone is due to be punished particularly harshly, I think this should make us want stronger evidence. The harsher the punishment, the more confident we should be before we convict someone, given the increased risk of harming them if they are innocent.[37] But my view is that the risks of false negatives (mistakenly releasing the guilty) are *not* a reason to lower our standards. So, the fact that someone might be at an increased risk of reoffending does not make it rational to blame them on the basis of weaker evidence than you would normally require. This is a bit abstract, so it might help to have an example. The view I defend generates the following pattern:

Accept: Greater confidence before convicting someone of x because the punishment for x is death.

Reject: Lower confidence when convicting someone of y if y-guilty people have a high average recidivism rate.

One way to test whether this idea is plausible is to consider the following thought experiment. Compare a really harsh legal system with a more lenient system. Suppose country A cuts off a hand for theft while country B only imposes a fine. Should judges in country A require stronger proof before convicting thieves than judges in country B? My judgement is that the answer is yes.[38] But why?

[37] For more detail, see the argument developed in Ross 2023b.

[38] Some 'natural experiments' might support this view. When transportation and the death penalty were removed as sentences in England, the conviction rate increased (see Bindler and Hjalmarsson 2018). However, this evidence is ambiguous. Another explanation, to which we will return in Section 5, is that these were cases of jury nullification of unjust punishment rather than the jury not being convinced that the evidence was strong enough.

https://doi.org/10.1017/9781009127745 Published online by Cambridge University Press

The underlying rationale for this judgement is that the risk of false negatives is not directly relevant to the belief-eliciting role of criminal courts. When a court convicts someone, they are sending out a signal that it is acceptable to blame and punish the person in question. This means that the court must be interested in the possibility that they are inducing a community to blame and punish someone undeservedly. The more severe the blame and punishment, the more cautious the court should be – hence a higher standard of proof should be used. However, when a court finds someone *not* guilty, the court isn't in fact saying anything about whether people should believe the person will not commit crimes in the future.[39] Rather, the court is *only* saying that the evidence isn't strong enough to blame and punish the person for the particular thing they are being accused of. So, the risk of false negatives is in this sense irrelevant to the social signalling role performed by a criminal court. In Section 3 I will problematise this view by exploring the idea that courts should proactively 'clear the name' of the accused and support the belief that they are definitively innocent – rather than taking the more restricted role of just saying the evidence isn't strong enough to blame the accused in the current case. This is a question for the next section. Nevertheless, given the *current* social role of the court, we can deny that the risk of releasing the possibly guilty is relevant to the beliefs that courts are tasked with supporting among the community.

2.3 Radically Flexible Standards?

Our discussion has focused on what is at stake in trials about different *types* of crime. However, there is an even deeper challenge we have not yet discussed.

Not only do different *types* of crime differ in the prospective costs of different errors, but so do *instances* of the same crime type. For example, two people accused of murder might have different risk profiles for the future, they might face different degrees of punishment (e.g. because one is a repeat offender), one might be more amenable to rehabilitation, or be more likely be harmed by a lengthy punishment (suppose they have young children). Not only do the characteristics of suspects change, but so do the characteristics of the crimes they are accused of. One murder can be much more violent or high-profile than another; one might come with a particularly harsh sentence or be particularly egregious in the community in which it occurs. Since every criminal accusation has its own unique profile, this poses an even deeper challenge to the idea that the law should use a one-size-fits-all approach to setting the standard of proof.

[39] For example, the court isn't saying that it would be inappropriate to take precautionary steps in future dealings with the person (see Ross 2021a for more discussion).

https://doi.org/10.1017/9781009127745 Published online by Cambridge University Press

Initially, it is tempting to dismiss this challenge by saying that – although philosophically interesting – the practical obstacles to *explicitly* formulating different standards for each trial render this issue merely academic. Convenience, economy, and the limitations of human effort must place a cap on how tailored criminal justice can be. However, the rise of artificial intelligence and algorithms challenge even this thought. A huge literature has arisen on the pros and cons of using algorithmic tools in criminal justice. They promise to eliminate the idiosyncratic biases of human reasoners,[40] but at the same time have been claimed to risk entrenching demographic biases in sentencing and parole decisions.[41] One thing such tools might allow is for decision-making to be customised in ways that is currently impossible. A question for the future is whether this is a road we wish to travel.

To come full circle, let's return to our earlier comments on *interpreting* the standards of proof. It might be that juries and judges interpret 'beyond reasonable doubt' differently, depending on the context. Given this possibility, perhaps accounting for variation between cases might not require *explicitly* formulating different standards of proof. Since the law uses rather general phrases like 'beyond reasonable doubt' when instructing judges or juries, there is some latitude for the judge or jury to approach the case at they see fit; they are deciding the case against a vague standard, not an extremely precise test. It might be that the current approach is *already* a very flexible one – where cases are judged under the broad banner of 'beyond reasonable doubt' but where the meaning of this is interpreted differently depending on the case and the judgement of the judge or jury about what contextual features matter.

This illustrates two very different ways to think about setting the standards of proof. One is a 'top-down' approach, where the legal system evolves or is designed so that individual judges and jurors are explicitly instructed in what the right standard to use is – where the aim is for cases to be judged as uniformly as possible. Another approach rejects the top-down method of deciding in advance what standard should be used in a given case, but rather leaves the decision about the appropriate standard *in the hands of those who are adjudicating the case*.[42] In this latter sense, there is no prior fully determinate decision about the standard of proof, but rather the standards of proof emerge from a patchwork of individual decisions made by fact-finders on the ground hearing cases every day.[43] From this patchwork of individual decisions made by those who

[40] For example, see Sunstein 2021.
[41] One example is the extremely high-profile critical report found in Angwin et al. 2023.
[42] For further discussion, see Loeb and Molina 2022.
[43] 'Fact-finder' is the legal term for the person who decides whether the standard of proof is met – the judge or the jury depending on the type of case.

https://doi.org/10.1017/9781009127745 Published online by Cambridge University Press

decide particular trials, the standards of proof emerge. This 'bottom-up' way of thinking about proof entrusts a great deal of responsibility to those who decide cases, relying on their judgement about what the appropriate way to think about the case is. To return to the phrase with which we began this section, it would mean that each judge or jury has to make up its own mind about what is required for an accusation to be proved to a 'moral' certainty.

~~3 Should Proof Be Binary?~~

~~Legal proof is usually conceived as a binary. Consider the words of Lord Hoffmann, an influential English judge:~~

> ~~'If a legal rule requires a fact to be proved . . . a judge or jury must decide whether or not it happened. There is no room for a finding that it might have happened. The law operates a binary system in which the only values are 0 and 1. The fact either happened or it did not.'~~[44]

~~The first paragraphs of this Element were highly sympathetic to Hoffman's description. I suggested that courts cannot shrug their institutional shoulders and that 'maybe' was no use in legal adjudication. In criminal cases, which we continue to focus on, trials end with somebody being convicted or acquitted. Although we've discussed how criminal proof might be flexible by using stronger or weaker standards in different situations before convicting somebody, we haven't yet questioned the *binary nature* of legal decisions. However, there are deep questions about whether this binary system is the best way to approach legal proof.~~

~~3.1 Binary versus Non-Binary Systems~~

~~Binary legal systems predominate across the world. These are systems in which criminal trials end in either a single 'Guilty' verdict or a single verdict of acquittal (usually phrased as 'Not Guilty').~~

~~But not all jurisdictions use a binary system of proof. In Scotland, for example, a 'third verdict' called *not proven* has existed for several hundred years.~~[45] ~~*Not proven* exists alongside *guilty* and *not guilty* as a separate option for the court. Not proven is the verdict returned when the judge or jury is not satisfied with returning a not guilty verdict, but also cannot find the accused guilty beyond a reasonable doubt. In this sense, not proven is what we can call an *intermediate verdict*. The not proven verdict officially leads to the same outcome as not guilty – the accused is released without punishment. However,~~

[44] ~~*Re B* [2009] 1 AC, para. 2.~~

[45] ~~There are plans to abolish this verdict, and at the time of writing it seems likely that it will be abolished.~~

https://doi.org/10.1017/9781009127745 Published online by Cambridge University Press