Philosophical Foundations of the Law of Evidence

Chapter 21: Burdens of Proof

By Mark Spottswood

Keywords: burden, proof, persuasion, Bayesian, probability, inference to the best explanation, likelihood ratio, evidential weight, preponderance, reasonable doubt

--

1. Introduction

Burdens of proof remain a lively subject of debate within the academic literature on evidence and juridical inference. Burdens consist of rules that guide judges and juries in determining what level of factual proof suffices to support an award of sanctions. In essence, they are functions that map some set of inputs, such as varying levels of confidence with respect to a defendant's guilt, onto some set of consequences, such as guilty or not-guilty verdicts.

One can describe many legal judgments that hinge on some underlying facts as involving a burden, whether or not it is commonly so labelled. For instance, *burdens of production* require a party to produce sufficient evidence to permit a reasonable jury to rule in their favor. Passing this threshold permits the party to present its case to a jury, while failing to pass it leads to the court finding against the party as a matter of law.[1] But this chapter, like the vast bulk of evidence scholarship on burdens, will focus exclusively on the *burden of persuasion* at trial. This burden determines when a fact-finder should sustain or deny a plaintiff or prosecutor's claims against a defendant. In American litigation, these burdens of persuasion take a few familiar forms: the *preponderance of the evidence* burden that is used for most civil cases, the intermediate *clear and convincing evidence* burden that is used for a smaller group of civil cases, and the high standard of *proof beyond a reasonable doubt* burden that governs in criminal cases.

Scholars have debated many kinds of questions with respect to burdens. Some have explored the underlying psychological mechanisms that jurors use when they analyze evidence and arrive at verdict decisions. Others have focused on describing the assumptions that are expressed or implicit in the existing law regulating burdens. Many have analyzed which sorts of burden rules are best from a normative standpoint, in terms of their abilities to generate fair decisions or to maximize social welfare. And some have tried to construct theories that bridge the different levels, providing accounts that balance descriptive accuracy with normative desirability.[2]

The remainder of this chapter will set forth some of the claims about burdens that have been made on each of these levels. For ease of analysis, we will divide the key questions into two parts.[3] First, we will explore the *inputs* that a burden of proof rule should draw upon, including a

---

[1] Nance 2016, 2-3.
[2] *Cf.* Spottswood 2019, 75-76.
[3] *See* Tuzet 2020.

survey of potential candidates that scholars have identified and a consideration of factors that might help us assess their suitability for differing purposes. Then, we will turn our attention to the ways that proof burdens might transform those inputs into recommended levels of sanctions in civil and criminal cases, giving special attention to the normative suitability of differing burden structures.

2. Burden Inputs

The first step of applying a burden is to identify which particular case-related facts most directly determine its outcomes. This section will briefly outline several previously identified burden input possibilities, and then survey some of the particular features that might make these competing candidates more or less attractive. As we shall see, the particular features incorporated into a model will influence its suitability as a psychological, doctrinal, or normative account, and it is doubtful that a model that is optimized for one purpose can serve the others equally well.

2.1 Probability of Charged Conduct Versus Other Grounds

One common view is that burdens of proof might operate based on assessed *probabilities* of liability or guilt.[4] Using this approach, if a jury finds that the probability that the defendant committed the charged wrong against the plaintiff is greater than 50%, then the preponderance standard has been satisfied. Similarly, we might say, for the sake of exposition, that a probability greater than 66% meets the "clear and convincing evidence" threshold, while a probability above 90% meets the "beyond a reasonable doubt" standard, while acknowledging that some might prefer slightly different numerical thresholds. This approach is often linked to Bayesian decision theory, a mathematical framework that offers decision-makers normative guidance as to which choices they should make if they seek to maximize their future utility under conditions of uncertainty.[5] It has also been used within economic analyses that aim to identify which litigation outcomes maximize social welfare,[6] and some have incorporated it within psychological models of fact-finding as well.[7]

Against this baseline, scholars have proposed many alternative approaches. Some rivals agree that the thresholds should be quantitative in nature but disagree with using *probability* estimates. Edward Cheng and Sean Sullivan have separately argued in favor of a *likelihood ratio* account.[8] This differs from the probability approach in two key ways. First, rather than relying on the isolated probability of the defendant's culpability, the likelihood ratio approach focuses on the comparative strength of each side's case. Second, this approach measures case strength by assessing the likelihood of observing the given body of evidence *assuming* that either side's proposed factual account was true. The preponderance standard would be met whenever the plaintiff's account makes the evidence more likely than the defendant's account, while heightened burdens would require higher ratios.

---

[4] *See generally* Friedman 2019.
[5] *See, e.g.*, Kaye 1982; Kaplan 1968.
[6] *See, e.g.*, Posner 1999.
[7] *See* Devine 2012, 22-26.
[8] Sullivan 2019, 26-29; Cheng 2013, 1268; *see also* Dahlman & Kolflaath, Chapter 6, this volume.

A related approach seeks to ground verdict decisions on *degrees of committed belief in wrongdoing*. Drawing on the theory of belief functions, Kevin Clermont has urged that we should model fact-finders as having some degree of committed belief in wrongdoing, some degree of committed belief in innocence, and some degree of uncommitted belief (which might later be committed to either side should new evidence arise that is relevant to the case).[9] Like the likelihood theory, this account is essentially comparative. But degrees of belief do not depend strictly on the parties' proposed case theories, and the mathematics of combining them does not align with ordinary probabilistic reasoning.[10]

Another kind of theory focuses on the *weight* of accumulated evidence, either on its own or in addition to the probability of wrongdoing. Evidential weight refers to the quantity of evidence that has been collected on a question, relative to the *ideal* amount of evidence that might be considered in resolving the same question.[11] L. Jonathan Cohen, for instance, argued that some burden thresholds should depend on the comparative quantity of evidential weight offered by each party.[12] Neil Cohen offered a blended theory, in which not just the probability of wrongdoing, but also its confidence interval, would need to exceed the proof threshold to support a verdict decision.[13] Neither of these authors, however, explained how one might apply these ideas in deciding a run-of-the-mill court case based upon conflicting witness testimonies.[14]

A fifth group have argued against quantification entirely. One recent approach, championed prominently by Ronald Allen and Michael Pardo, has drawn on the concept of inference to the best explanation. Advocates of this approach maintain that rather than determine the probability of any facts, jurors engage in a comparative evaluation of the *strength of each party's explanation of the evidence*.[15] Such explanations can consist of simple stories about how the litigated events might have actually transpired, but they can also take more complex forms. The strength of an explanation is evaluated in a holistic manner based on factors such as consistency with the evidence, internal consistency, simplicity, and coherence with background beliefs about the world.[16]

A final group of theories has arisen in the context of psychological accounts of jury decision-making, which attempt to elucidate the underlying mental mechanisms at play in juridical inference. Some of these accounts have modelled jurors' assessments of case strength in mathematical terms, although not all assume that the numbers in question are derived in a mathematically ideal fashion.[17] By contrast, Nancy Pennington and Reid Hastie's *story model* analyzed the process in terms of a choice between salient *stories* about the litigated events. Fact-finders, in their view, use the evidence to construct coherent narratives about the disputed events in a case, and then award victory to the party who has the "best" story based on factors such as

---

[9] Clermont 2018, 1068-69.
[10] *See id.* at 1068-69, 1071-75. *See also* Mark Spottswood, *Paradoxes of Proof*, this volume at Chapter 7.
[11] Nance 2016, 130; Nance, Chapter 4, this volume.
[12] Cohen 1977, 252-56 (focusing on the preponderance standard).
[13] Cohen 1985, 404-09.
[14] Nance 2016, 138.
[15] Pardo & Allen 2008, 233-42; Allen & Pardo, Chapter 2, this volume.
[16] *Id.* at 230.
[17] *See* Devine 2012, 22-26 (2012).

evidential coverage, consistency, plausibility, and completeness.[18] More recently, I argued that juridical decisions arise out of an interaction between implicit cognition, which operates in a non-comparative, associative manner, and explicit reasoning, which often involves the articulation and comparison of semantic stories.[19]

2.2 Subjective vs. Objective

One source of disagreement among rival theorists has been whether the input to the burden function should be understood as a *subjective* quality within the fact-finder's mind versus an *objective* feature of the underlying case. Most obviously, the psychological models are clearly designed to explain how subjective aspects of reasoning lead to particular outcomes in cases. Moreover, many adherents to the probabilistic conception of burdens take a Bayesian perspective, interpreting the underlying probabilities as subjective.[20] (In practice, we must make a distinction between truly subjective accounts, which recommend outcomes based upon a fact-finder's *actual* credence concerning guilt or liability, versus epistemic approaches, which also advise her what credence she *ought* to hold, given the evidence presented to her and her prior beliefs.)[21] The belief function theory likewise draws on an essentially subjective quality, by relying on a fact-finder's degree of committed belief. And the likelihood ratio approach can also be framed in subjective terms.[22]

But some theorists maintain that it is undesirable to incorporate subjective facts into a model of proof burdens. Allen and Pardo, for instance, have maintained that credences have "no necessary relationship to advancing accurate outcomes." Moreover, they worry that premising trial outcomes on subjective beliefs conflicts with rules that allow judges to override jury verdicts when there is only one "reasonable" outcome in a case.[23] Thus, the relative plausibility theory explicates burdens as functions that depend on what the *true* balance is between competing explanations, rather than a particular fact-finder's *beliefs* regarding that balance. Some other theorists take a similar approach.[24]

Ultimately, the desirability of basing burdens on credences versus objective features of a case will depend on our goals. At one extreme, it is simply odd to complain when *psychological* accounts draw upon subjective facts, as the goal of such theories is to link decisions to underlying mental processes. At the other extreme, reliance on a fact-finder's actual degrees of belief will limit the reach of normative theorizing. After all, such a theory ought to be able to tell us how each case should be properly decided.[25] This would, ideally, include the ability to tell us when a fact-finder's beliefs or preferences regarding case outcomes are poorly justified or biased. But if those beliefs are also the ground of a normative theory, that theory could then lose the power to help us identify some sources of error.

---

[18] Pennington & Hastie 1991, 527-28.
[19] *See* Spottswood 2013, 171-93 (2013).
[20] *E.g.*, Friedman 1996, 276-78.
[21] *See* Nance 2016, 42-49.
[22] Cheng 2013, 1266-68.
[23] Allen & Pardo 2019, 12.
[24] *E.g.*, Sullivan 2019, 35.
[25] *Cf.* Laudan 2006, 79-81.

Harder questions arise when we focus on the goal of describing burden of proof doctrines. On the one hand, judges are permitted to set aside some jury results on the basis that they are unreasonable or against the clear weight of the evidence,[26] which might suggest that there is something other than the jurors' subjective mental states that can serve as a yardstick. On the other hand, some jury instructions do seem to invite a decision based upon the juror's own views and beliefs concerning the evidence.[27] And on reflection, it is far from clear what *non-subjective* referents real-world decision-makers have access to that could ground burdens. Consider relative judgments of explanatory plausibility, for instance. One can tell a juror to decide based on which explanation is *truly* best, but in applying this instruction they will inevitably draw on their own beliefs regarding the strength of each explanation. A reviewing court, if they are to reverse a verdict, will likewise draw on their own judgments. It is almost unimaginable that any of them would say, "*Personally* I feel that the plaintiff has the stronger explanation, but in *truth* the defendant's explanation is better, so I will vote against liability." Thus, defining burdens in putatively objective terms may obscure how they must operate in practice.

2.3 Single Quantity vs. Comparative

Another key question is whether the input to a burden decision is an entity that can be meaningfully described in isolation, versus one that can only be described in comparison with something else. The classical probabilistic approach, for instance, would say that the preponderance standard is met when the jury has a confidence level in liability that is greater than 50%. Competing offers of proof and arguments from the parties might adjust this level upwards or downwards, but in the end the outcome depends upon a single numerical quantity. Some psychological models also incorporate features that require a factual account to be sufficiently strong *in isolation* to become intuitively appealing, whether or not a stronger competitor has been offered.[28]

By contrast, other approaches take an explicitly comparative approach. The likelihood ratio approach, for example, awards victory to whichever party's case makes the evidence more probable, without demanding that either party's case be more likely than not to be true.[29] In other words, the approach *ranks* the strength of each case in an ordinal fashion, without demanding that case strength be assessed relative to a fixed, cardinal scale. Similarly, the story model suggests that fact-finders choose among competing stories, rather than assess the individual likelihood of any one story's truth.[30] In addition to these, the belief function approach[31] and the relative plausibility approach[32] also take a comparative approach.

Once again, the strength or weakness of each approach can only be considered in relation to the various uses we might have for describing burdens. At the psychological level, I have argued

---

[26] *See* Byrd v. Blue Ridge Rural Elec. Co-op., Inc., 356 U.S. 525, 540 (1958); Brady v. S. Ry. Co., 320 U.S. 476, 479–80 (1943).

[27] *See, e.g.*, Committee on Pattern Jury Instructions for the Seventh Circuit 2017, § 1.27; In re Winship, 397 U.S. 358, 371 (1970) (Harlan, J., concurring).

[28] *E.g.*, Spottswood 2013, 197-99.

[29] *E.g.,* Cheng 2013, 1265 & n. 17.

[30] Pennington & Hastie 1991, 527-529.

[31] Clermont 2018, 1079.

[32] Allen & Pardo 2019, 15.

that both singular and comparative processes are at work in real-life juridical reasoning. First, while fact-finders listen passively to the evidence, their minds are actively organizing it into an associative model that connects people, places, events, and moral categories into a single, unitary framework. Later, when the case has finished and the fact-finders are deliberating about how to decide, they will often approach their task by explicitly comparing each party's account of the case to the body of evidence and the relevant legal rules.

At the level of doctrine, the singular approach is more defensible than the comparative one. To be sure, some language in conventional jury instructions can be read in a comparative fashion. References to the "preponderance" or the "greater weight" of the evidence, for instance, do seem to ask the jury to prefer whichever party's case is stronger. Other language, however, fits more neatly within a singular account. Whether a case is "clear and convincing," or whether a doubt is strong enough to be "reasonable," do not seem to be inherently comparative questions. And some cases have made it explicit that even the preponderance standard is not meant to be purely comparative. On multiple occasions, the Supreme Court has stated that the preponderance standard is not met simply because the jury finds the plaintiff's account more convincing than the defendant's account. As the Court has put it, "it is not enough to *dis*believe the [defendant]; the factfinder must *believe* the plaintiff's explanation."[33]

Courts, moreover, enforce a number of related doctrines that are at odds with a comparative approach. Plaintiffs and prosecutors can have their cases dismissed by a court if they are too weak to permit a reasonable jury to find in their favor, and such rulings can occur *before* the defendant has offered any counter-evidence.[34] Likewise, criminal defendants sometimes adopt a strategy of identifying internal weaknesses within the prosecutor's theories and evidence, without offering any counter-narrative. The comparative account might suggest that such a strategy requires a verdict for the prosecutor, given the absence of a competing story or explanation, but no such doctrine exists. Thus, although comparative *reasoning* is certainly an accepted way for juries to analyze whether a party who bears a burden of proof has satisfied it,[35] a fully comparative interpretation of the burden *itself* is harder to defend.

Lastly, there is the question of whether burdens *should* be comparative as a normative matter. There are two possibilities. One is that a comparative approach considers *all* the relevant possibilities on each side of the case, such that a plaintiff receives credit for each account that implies liability and the defense receives credit for each account that negates liability. Some theories, such as the explanatory account, permit a fact-finder to proceed in this way. Taking this approach, the outcome of a comparative analysis should closely approximate the probability that the plaintiff's case is true, possibly implying that the gap between normative explanationism and normative Bayesianism may be smaller in practice than it might appear at first glance.[36]

---

[33] Reeves v. Sanderson Plumbing Prods., 530 U.S. 133, 146 (2000) (quoting from St. Mary's Honor Ctr. v. Hicks, 509 U.S. 502, 519 (1993).
[34] Lempert 1986, 471-72.
[35] *Cf.* Anderson v Griffin, 397 F.3d 515, 521–22 (7th Cir. 2005) (noting that one way of raising the estimated probability of an event's occurrence is to rule out competing possibilities).
[36] *See* Lipton 2004, 107-117; Nance 2016, 80-81.

By contrast, other comparative models do not permit the aggregation of inconsistent stories as a means of supporting a case. Edward Cheng's likelihood ratio model, for instance, takes as an input the likelihood of each party's strongest *internally consistent* account. He explicates his approach using an automobile tort hypothetical. The plaintiff, in this example, claims that the crash occurred because the defendant was speeding and that she suffered a neck injury as a result. The defendant argues that he was not speeding and that in any event the neck injury was a pre-existing condition. Cheng's approach compares the likelihood of the plaintiff's account (speeding & caused neck injury) with the likelihood of whichever *single story* offered by the defendant is most probable.[37] Imagine, for instance, that the overall likelihoods are as follows:

	Separate Probability	Aggregated Probabilities
Speeding and New Injury	33%	33% Chance Across All Plaintiff Stories
Not Speeding and New Injury	22%	66% Chance Across All Defense Stories
Speeding and Pre-Existing Injury	22%	
Not Speeding and Pre-Existing Injury	22%	

With a pattern like this, Cheng's theory would conclude that the plaintiff should prevail, because no single defense story is more likely than 33%. This might be useful in a psychological model, if we think jurors are irrationally reluctant to give due credit to arguments in the alternative.[38] It *may* also have merit as a doctrinal account, given that courts sometimes instruct jurors in ways that diverge from what is mathematically ideal.[39] But as a normative guide to inference, it has grave deficiencies. Put simply, when parties argue in the alternative, the overall likelihood that they are right on one of their points is always *greater* than the likelihood that their strongest argument is correct. Thus, precluding aggregation will, in general, cause us to give too little credit to accounts that are disjunctive in form, and thereby create an excessive risk that we will erroneously find parties liable. We would do better either by relying on the strength of the plaintiff's claim alone or by comparing it to the *aggregated* probability that one or more of the defendant's accounts is true.

2.4 Quantified vs. Unquantified

The last question is whether the criteria upon which burdens operate should be quantified or not. Many approaches, including the probabilistic frameworks, the likelihood ratio theories, the belief function account, and some psychological models, agree that burdens depend on something that can be described in numerical terms. Other approaches, including the relative plausibility account and the story model, avoid such quantification. As we shall see, the desirability of using

---

[37] Cheng 2013, 1263-65.

[38] *See* Spottswood 2016, 281-82.

[39] *See id.* at 267, 282. On the other hand, some legal rules allow parties to make arguments in the alternative. *See, e.g.,* FED. R. CIV. P. 8(d)(3) (providing that "[a] party may state as many separate claims or defenses as it has, regardless of consistency").

numerical quanties as inputs will depend on whether we are offering a psychological, doctrinal, or normative account of burdens.

At the psychological level, it is doubtful that many judges or jurors pause in their deliberations to attach explicit numbers that represent their belief in the strength of a case.[40] Thus, a quantified model will fail to capture important aspects of the internal experience of fact-finding. Of course, there may still be uses for quantified *descriptions* of the process, purely for analytic purposes. For instance, if we wish to assess whether jurors' votes line up with a quantified *normative* model of inference, it may help to elicit a quantified statement regarding their confidence in guilt, even if that would not normally be a part of the decisional process, and use that to construct a mathematicised account of their actual decisions.[41] But if we are trying to get closer to the root causes of how cases are typically decided, modelling burdens in a way that depends on quantified levels of confidence in wrongdoing may obscure as much as it clarifies.

Doctrinally, some courts describe the preponderance standard as requiring the jury to think it more than 50% likely that the plaintiff's claim is correct. [42] However, when we shift our focus to the burden of proof beyond a reasonable doubt, we find that judges display an open hostility to the notion of quantification.[43] So as before, although there may be utility to modelling these doctrines mathematically for some analytic purposes, such simplifications will fail to capture important features of existing law.

Finally, scholars have fiercely debated whether a normative model of burdens should incorporate quantification. Views on this topic will necessarily depend, at least in part, on broader notions concerning the inferential process by which jurors should decide cases. If one thinks that sequential probabilistic adjustments in response to evidence will help fact-finders reach more accurate verdicts,[44] then it might make sense to retain the quantified output of such a process as the basis for decision. Similarly, some normative proposals suggest that it would be desirable to vary sanctions fluidly in response to varying likelihoods of wrongful conduct, and thus require the input side of the equation to be a sliding scale.[45] If we think such proposals are attractive, representing the input side of the equation as a continuously variable numeric quantity will ease its implementation. More pragmatically, quantification might yield special benefits when judges or jurors must determine how to incorporate explicitly statistical evidence into their decisions.[46]

Others have worried that the seeming precision of a quantified burden presents a false hope. Most evidence does *not* present itself in a quantified form, and if one considers the vast multitude of sequential inferences needed to decide even a simple case, one must quickly conclude that any attempt to give numerical assignments of strength to each item of evidence must involve a great deal of rough guesswork.[47] Moreover, if we wish to constrain our estimates of evidence strength through reliance on real-world frequency data, we will find that there is no

---

[40] *See* Devine 2012, 23.
[41] E.g., Koehler, Chia, & Linsdey 1995.
[42] *See, e.g.*, Brown v. Bowen, 847 F.2d 342, 345 (7th Cir. 1988).
[43] *See, e.g.*, McCullough v. State, 657 P.2d 1157, 1159 (Nev. 1983).
[44] *Cf.* Spottswood 2019, 80.
[45] *See infra* at Part III.
[46] *Cf.* National Research Council 1996, 200-02.
[47] *See* Allen 1996, 269-70.

data to be had on some questions, while others force us to make potentially arbitrary choices among competing reference classes.[48] Ergo, if we employ burden rules that specify their input functions quantitatively, in practice most fact-finders will derive the required numbers from vague intuitive impressions rather than strict mathematical analysis.

But this observation, while important, does not foreclose further normative discussion. Even if jurors produce numerical estimates intuitively, perhaps asking for a number will induce them to consider potential sources of error more carefully before deciding. Alternatively, perhaps combining each juror's probabilistic judgments using averaging would return a more trustworthy result than all-or-nothing voting with respect to the overall outcome.[49] Or maybe unquantified burdens are interpreted in inconsistent or unpredictable ways by decision-makers, and quantifying them might help us better implement our goals regarding distribution of errors and equality of treatment.[50] Although these ideas are speculative, they do suggest that there could be benefits to quantifying burden inputs even if the numbers are generated by holistic intuition rather than mechanistic analysis.

We have seen that the differing approaches to modelling burdens can be classified along several interesting dimensions. Some operate on subjective states of mind, while others focus on objective case features. Some depend upon a single factor, while others draw on a comparison between alternatives. And some rely upon a factor that can be expressed numerically, while others eschew quantification. As the foregoing analysis has shown, the treatment of these details will affect the degree of fit a model has at the psychological, doctrinal, and normative levels of analysis. Moreover, the approach that works best at one level can be suboptimal for the others.

3. Transforming Inputs into Outputs

Next, let us consider how potential burdens transform representations of case strength into varying levels of sanctions. Doctrinally, the most common approach is a discontinuous, "all or nothing" rule, which awards zero sanctions up to a specified level of confidence and full sanctions above it. But interesting variations exist, including some that have been tried by courts in isolated areas, and others that have been studied theoretically but never implemented. The variations include rules that vary the threshold from case to case, rules that permit intermediate steps between zero sanctions and full sanctions, and rules that vary the sanction size continuously in response to variations in a fact-finder's confidence in liability or guilt.

3.1 Potential Approaches

Most commonly, burdens of proof take a simple form, generating zero sanctions below some input threshold and full sanctions above it. In ordinary civil cases, for instance, the plaintiff gets nothing if she proves her case by less than a preponderance of the evidence, and she gets full

---

[48] *See* Pardo & Allen 2008, 259-60.

[49] Imagine two cases that must be decided by 6-member juries. In the first, three jurors agree that liability is 95% likely, while the others think it is 45% likely. In the second, all six jurors agree that liability is 55% likely. The former case has a higher average confidence level (70% in the first versus 55% in the second), but if we lump each judgment into a dichotomous decision on liability, we will obscure that fact.

[50] *See* Kagehiro 1990, 196-98.

damages once that point has been exceeded. In criminal cases, the defendant is subject to no punishment unless the prosecutor can prove her case beyond a reasonable doubt; past that point, he is subject to full punishment (often within a range defined by statute). Affirmative defenses, meanwhile, must be proven to some fixed threshold by a defendant in order to negate these consequences. And in all of these scenarios, the party who "bears" the burden of proof must convince the fact-finder that the applicable threshold has been exceeded, with ties going to their opponent.[51]

If one choses this *single-stepped* rule, one must also decide whether its threshold will be fixed or variable. The clearest example of a fixed threshold rule is the civil preponderance standard, which places the tipping point fairly clearly at the 50% mark, for all fact-finders and case types. But we can also employ a threshold that *moves*, so that some cases require more proof than others to justify sanctioning the defendant. And indeed, our system already employs a variable threshold in one sense: we apply a heightened burden in criminal cases, a more balanced burden in ordinary civil cases, and an intermediate burden in a subset of civil cases in which wrongful ascriptions of liability are thought to be especially costly.

There may also be subtler ways that our proof thresholds vary in practice. At the psychological level, fact-finders who are given vague burden definitions might demand stronger proof when the costs of error seem high, while permitting weaker cases to suffice when the cost of making a mistake seems less grave. For instance, consider the notion of proof "beyond a reasonable doubt." Multiple mock-trial experiments have shown that both judges and jury-eligible laypeople would ask for higher levels of proof before convicting defendants when the cases involve more serious crimes or longer sentences.[52] It has also been suggested that juries who are tasked with deciding whether to sentence convicted defendants to death are often reluctant to do so unless they think the guilt has been proven to a *higher* level than the beyond-a-reasonable-doubt standard would require on its own.[53] Some courts have agreed that this is appropriate, permitting advocates to openly advocate for a life sentence on the basis of such residual doubts, while other courts disagree.[54]

Multi-stepped approaches add one or more intermediate steps between zero and full sanctions. Historically, continental courts employed a system where "full proof" enabled stricter consequences than "half proof," for instance.[55] Likewise, modern Scotland employs a three-tiered verdict system in criminal cases, permitting verdicts of not guilty, not proven, and guilty.[56] The former system imposed reduced punishment at the intermediate proof level, while the only impact of Scotland's intermediate verdict is potential stigma due to the lack of an innocence finding.

Some courts have gone further, replacing single- or multi-stepped thresholds with a smooth transition across the full range of case strength. In toxic tort cases, plaintiffs sometimes have

---

[51] *See* Nance 2016, 3-4.
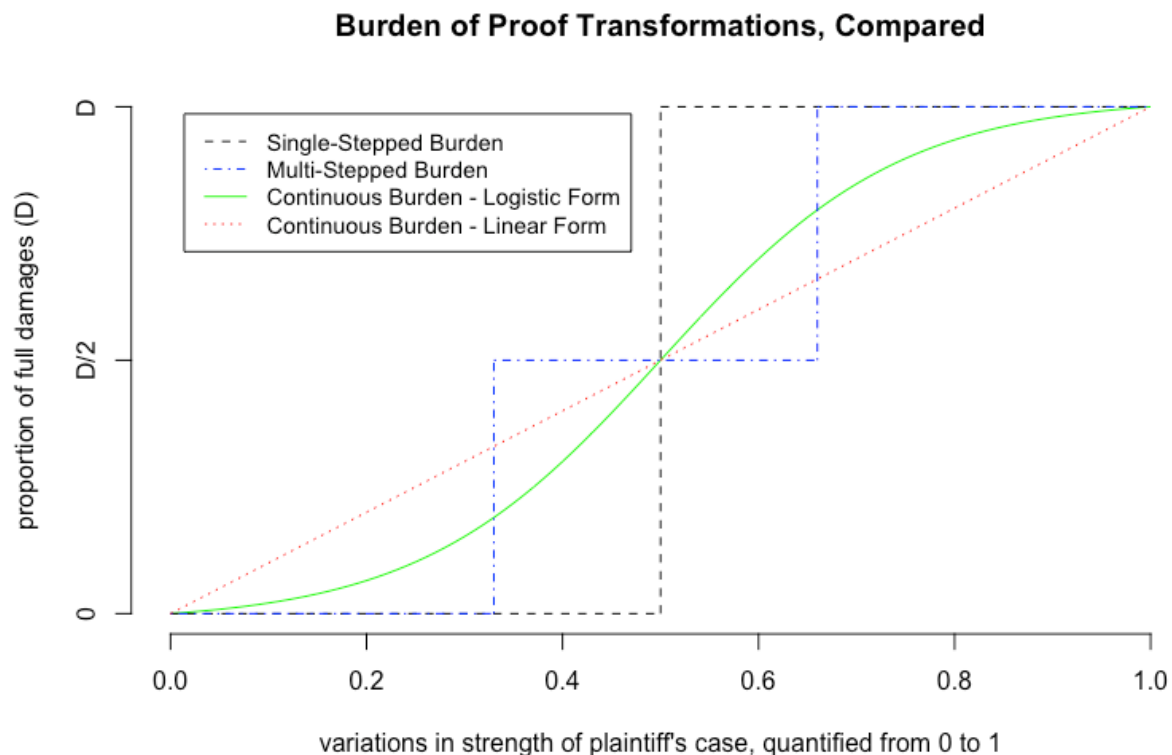[52] *See* Stoffelmayr & Diamond 2000, 780-81.
[53] *See* Antkowiak 2005, 582-83.
[54] Fisher 2012, 840-41 (surveying cases).
[55] Picinali 2018, 556.
[56] *Id.*

difficulty proving which among multiple manufacturers competing in the relevant market actually sold them the product that caused an injury, as many years can elapsed between exposure and the onset of illness. Starting with the *Sindell* decision, some courts have permitted plaintiffs to recover a *pro rata* share of their full damages based on the proportion of products sold by each company in the relevant market, in order to avoid letting companies with minority market positions escape all liability.[57] Thus, with respect to the question of factual causation of harm, the fixed threshold is replaced by a smoothly continuous scale. Some scholars have suggested that this model be generalized to a fully *continuous* burden.[58] Such burdens abandon thresholds entirely, and instead provide for a steady escalation in sanctions levels across the full range of case strength. Thus, a jury that was 25% confident in liability might award 25% of the damages, while a jury that was 75% confident might award 75%. The most commonly discussed continuous burden takes a linear form, but in recent work I have offered arguments in favor of a logistic transformation between confidence and penalties.[59]

Burden of Proof Transformations, Compared

variations in strength of plaintiff's case, quantified from 0 to 1

Thus, the doctrinal landscape with respect to these issues is fairly clear. As set forth above, the single-stepped approach strongly dominates, and the few areas in which an alternative approach has been used have been clearly delineated. Moreover, although we have some reasons to think burden thresholds might behave fluidly as a psychological matter, courts have rarely encouraged this, except occasionally in the context of capital sentencing.

    3.2 Normative Issues

---

[57] *See, e.g.*, Sindell v. Abott Labs, 26 Cal 3d 588 (1980); In re 'Agent Orange' Prod. Liab. Litig., 597 F. Supp. 740 (E.D.N.Y. 1984).

[58] *See generally* Spottswood 2021a (reviewing literature and advocating for continuous burdens).

[59] *Id.*, Part III; Spottswood 2021b, Part IV.

Despite the doctrinal stability in this area, many scholars have contested which of these approaches function best at a normative level. In the space remaining, I will briefly review some key arguments that have been made in defense of conventional, fluid-threshold, and continuous approaches.

One way of proceeding is to compare the *expected error costs* of differing burdens. Expected errors represent the average amount of error that we should expect to arise in a case based on the decision rule, the damages, and the fact-finder's confidence level when deciding. For instance, assume that a judge finds that Alice's damages are $1,000 and believes there is a 75% chance that Bob caused her damages through tortious conduct. Following the preponderance rule, the judge would order Bob to pay the full $1,000. And because the judge believes there is a 25% chance that this decision would be wrong, his expected amount of error would be $250 (25% of $1,000).

David Kaye compared expected error rates across several different types of cases, using both the conventional fixed-threshold civil burden and the alternative of a continuous burden that is linear in form. The latter rule instructs the jury to determine the size of an award by multiplying the found damages by the probability of liability, so that in the above hypothetical the defendant would pay $750 (75% * $1,000) instead of the full $1,000.   Kaye showed that the linear rule generally produces an elevated expected error rate, as compared with the traditional rule, in cases involving a single defendant. He also showed that, if we assume that the cost of an error in favor of either the plaintiff or the defendant are equally problematic, it will generally minimize expected errors to set the threshold between no damages and full damages at 50%, rather than at any other level.[60]

Other have accepted this argument, while arguing that the preponderance rule has other, offsetting defects. Orloff and Stedlinger argued that we should also aim to reduce the chances that individual parties would be subjected to particularly *large* errors.[61] In the example above, the traditional rule imposes an expected error of $250, and because Alice is receiving full compensation, Bob bears the entire risk. By contrast, the linear rule would impose an expected error of only $187.50 on Bob (representing a 25% chance that he has paid $750 to a person who should receive nothing) with Alice expecting to suffer an error of equal size (representing a 75% chance that she has been under-paid by $250). In general, the linear rule spreads the risk of error more evenly across parties, reducing the risk that any one party will suffer from a particularly large error.[62]

Another benefit of multi-stepped or continuous burdens is that, by reducing the chance that small variations in perceived case strength cause large variances in outcomes, they may also reduce the impact of various biases and other unfair influences on a litigation system.[63] Both common

---

[60] Kaye 1982, 499 n.42. Kaye proved a different result for cases in which multiple defendants might have harmed a plaintiff. *Id*. at 503-08.
[61] Orloff & Stedinger 1983, 1165-68.
[62] *See* Spottswood 2021a, Part III (showing that the use of a logistic function, instead of a linear one, can further reduce the risk of large errors).
[63] *Id.* at Part II.

sense and economic theory[64] suggest that defendants will often wish to settle cases that should be easy wins for a plaintiff, and that plaintiffs will fail to bring, or voluntarily dismiss, cases if they think their chances of success at trial are low. This means that our systems will be biased towards trying fairly hard cases while settling easier ones.

This tendency can have problematic implications. The strength of the evidence is generally a strong predictor of how juries will decide cases[65], but other factors, such as differences in advocacy quality,[66] in-group/out-group bias,[67] or the impacts of emotionally inflammatory evidence of little true relevance[68] might tip the scales when the evidence is closely balanced. Under the conventional rule, such influences can easily tip cases across the line from no damages to full damages, becoming the source of sizeable errors and unfairness. Rules that involve steps of smaller size, or smooth transitions across the full range of confidence, will not eliminate these sorts of influences, but they will cabin the size of errors that they tend to cause. Thus, under a linear rule, the same movement from 45% to 55% confidence leads to a change of only 10% of the full damages, instead of an all-or-nothing change.

Finally, some have suggested that variable thresholds or continuous approaches can produce a more optimal pattern of deterrence of wrongdoing, as compared with what we see under the traditional rule. Consider cases in which proof will hinge on scientific evidence that makes the likelihood that a particular person was harmed a matter of purely probabilistic knowledge. A drug manufacturer might know that its drug will raise the rate of a certain form of cancer by just 50% above existing levels. If there is no known way to distinguish the other cases from the cancers that are caused by the drug, then any plaintiff could prove, at best, a 33% chance that the drug had caused their particular cancer. This, in turn, can lead to the troubling result that the law gives no deterrence against wrongs of this kind.[69] A continuous burden, by contrast, could give 33% damages to all the persons who ingested the drug and later contracted cancer, which would force the manufacturer to internalize the harm its product is causing to the public.[70]

Nor is this kind of under-deterrence unique to toxic torts. Many potential wrongdoers might be able to predict, at the time that they act, that they are unlikely to receive any punishment due to an anticipated deficit of evidence, and thus fail to be deterred by the threat of sanctions.[71] Some have argued that this type of underdeterrence may be particularly acute in the context of the criminal law, which tends to give quite severe sanctions with relatively low probability. This may arise if people value the first year of a prison sentence far more negatively than additional years far down the road, either due to habituation effects or a tendency to steeply discount future utility versus costs and benefits in the "here and now."[72] If that is right, then the system might

---

[64] *See* Spottswood 2021b, Part II; Priest & Klein 1984, 3.

[65] Devine 2012, 122-23 (reviewing experimental and field evidence).

[66] *See, e.g.*, Wheeler, Cartwright, Kagan & Friedman 1987, 408-09, 440-41 (finding that better counsel provided a small but measurable advantage in State Supreme Court litigation).

[67] Spottswood 2011, 847–48.

[68] Grady, Reiser, Garcia, Koeu, & Scurich 2018, 513-14.

[69] *See* Shavell 1987, 115-17.

[70] *Id.*

[71] *See* Spottswood 2021a, Part I-A.

[72] Fisher 2012, 857-58.

deter far more crime by issuing modest sanctions at lower-than-traditional levels of proof[73]—an approach which can easily be realized using multi-stepped or continuous burden structures.[74] Thus, although much work remains to be done in this area, existing research tends to suggest that single-stepped burdens reduce expected errors. By contrast, other burdens may be advantaged in their ability to spread errors, reduce the impact of some problematic influences on verdicts, and produce more optimal levels of deterrence.

4. Conclusion

Burdens of proof can be understood as functions that map a measure of case strength onto variations in the level of applicable sanctions. On the input side, scholars have hotly debated how case strength should be measured, and there may be differing "best" answers at the psychological, doctrinal, and normative levels of analysis. With respect to the best way of transforming case strength judgments into sanctioning levels, the picture is clearer at the levels of psychology and doctrine, but scholars have hotly debated which approach is most desirable at the normative level. This literature remains quite active, and I expect that the coming years will bring many new and interesting insights, both about how burdens currently operate and how they could be improved.

5. References

Allen, R.J. "Rationality, algorithms and juridical proof: a preliminary inquiry," *International Journal of Evidence & Proof* 1:254-75 (1997).

Allen, R.J. & Pardo, M.S. "Relative plausibility and its critics," *International Journal of Evidence & Proof* 23:5-59 (2019).

Antkowiak, B.A. "Judicial Nullification," *Creighton Law Review* 38:545-610 (2005).

Cheng, E.K. "Reconceptualizing the Burden of Proof," *Yale Law Journal* 122:1254-79 (2013).

Clermont, K.M. "Common Sense on Standards of Proof," *Seton Hall Law Review* 48:1057-80 (2018).

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

Cohen, N.B., "Confidence in Probability: Burdens of Persuasion in a World of Imperfect Knowledge," *New York University Law Review* 60:385-422 (1985).

Committee on Pattern Jury Instructions for the Seventh Circuit (2017). *Federal Civil Jury Instructions for the Seventh Circuit.* World Wide Web: http://www.ca7.uscourts.gov/pattern-jury-instructions/7th_cir_civil_instructions.pdf

---

[73] *See* Nagin 2013, 227-28.

[74] Due to space constraints, some interesting arguments concerning the ideal structure of proof burdens could not be discussed. *See, e.g.*, Kaplow 2012; Picinali 2019.

Devine, D.J. (2012). *Jury Decision Making: The State of the Science*. New York, NY: New York University Press.

Fisher, T. "Conviction without Conviction," *Minnesota Law Review* 96:833-85 (2012).

Friedman, R.D. "Answering the Bayesioskeptical Challenge," *International Journal of Evidence & Proof* 1:276-91 (1996).

Friedman, R.D. "The Persistence of the Probabilistic Perspective," *Seton Hall Law Review* 48:1589-600 (2018).

Grady, R.H., Reiser, L., Garcia, R.J., Koeu, C., & Scurich, N. "Impact of Gruesome Photographic Evidence on Legal Decisions: A Meta-Analysis," *Psychiatry, Psychology and Law* 25:503-21 (2018).

Kagehiro, D.K. "Defining the Standard of Proof in Jury Instructions," *Psychological Science* 1:194-200 (1990).

Kaplan, J. "Decision Theory and the Fact-Finding Process," *Stanford Law Review* 20:1065-92 (1968).

Kaplow, L. "Burden of Proof," *Yale Law Journal* 121:738-859 (2012).

Kaye, D.H. "The Limits of the Preponderance of the Evidence Standard: Justifiably Naked Statistical Evidence and Multiple Causation," *American Bar Foundation Research Journal* 1982:487-516 (1982).

Koehler, J.J., Chia, A., & Lindsey, S. "The Random Match Probability in DNA Evidence: Irrelevant and Prejudicial?," *Jurimetrics* 35:201-19 (1995).

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

Lempert, R.O. "The New Evidence Scholarship: Analyzing the Process of Proof," *Boston University Law Review* 66:439-77 (1986).

Lipton, P. (2d Ed. 2004). *Inference to the Best Explanation*. New York, NY: Routledge.

Nagin, D.S. "Deterrence in the Twenty-First Century," *Crime and Justice* 42:199-263 (2013).

Nance, D.A. (2016). *The Burdens of Proof: Discriminatory Power, Weight of Evidence, and Tenacity of Belief*. Cambridge: Cambridge University Press.

National Research Council (1996). *The Evaluation of Forensic DNA Evidence: An Update*. Washington, D.C.: National Academy Press.

Orloff, N. & Stedinger, J. "A Framework for Evaluating the Preponderance-of-the-Evidence Standard,"

Pardo, M.S. & Allen, R.J. "Juridical Proof and the Best Explanation," *Law and Philosophy* 27: 223-68 (2008).

Pennington, N. & Hastie, R. "A Cognitive Theory of Juror Decision Making: The Story Model," *Cardozo Law Review* 13:519-57 (1991).

Picinali, F. "Do Theories of Punishment Necessarily Deliver a Binary System of Verdicts? An Exploratory Essay," *Criminal Law and Philosophy* 12:555-74 (2018).

Picinali, F. "Justice In-Between: The Decision-Theoretic Case for Many Valued Criminal Verdicts," *available at* https://www.ssrn.com/ (2019).

Posner, R.A. "An Economic Approach to the Law of Evidence," *Stanford Law Review* 51:1477-1546 (1999).

Priest, G.L. & Klein, B. "The Selection of Disputes for Litigation," *Journal of Legal Studies* 13:1-55 (1984).

Shavell, S. (1987). *Economic Analysis of Accident Law*. Cambridge, MA: Harvard University Press.

Spottswood, M. "Live Hearings and Paper Trials," *Florida State University Law Review* 38:827-82 (2011).

Spottswood, M. "The Hidden Structure of Fact-Finding," *Case Western Reserve Law Review* 64:131-200 (2013).

Spottswood, M. "Unraveling the conjunction paradox," *Law, Probability and Risk* 15:259-96 (2016).

Spottswood, M. "On the limitations of a unitary model of the proof process," *International Journal of Evidence & Proof* 23:75-81 (2019).

Spottswood, M. "Continuous Burdens of Proof," *Nevada Law Journal* 21:xx-xx (forthcoming 2021) (a).

Spottswood, M. "Proof Discontinuities and Civil Settlements," *Theoretical Inquiries in Law* 22:xx-xx (forthcoming 2021) (b).

Stoffelmayr, E. & Diamond, S.S. "The conflict between precision and flexibility in explaining "beyond a reasonable doubt," *Psychology, Public Policy, and Law*, 6:769–787 (2000).

Sullivan, S.P. "A Likelihood Story: The Theory of Legal Fact-Finding," *Colorado Law Review* 90:1-66 (2019).

Tuzet, G. "Assessment criteria or standards of proof? An effort in clarification," *Artificial Intelligence and Law* 28:91-109 (2020).

Wheeler, S., Cartwright, B., Kagan, R.A., & Friedman, L.M. "Do the 'Haves' Come Out Ahead? Winning and Losing in State Supreme Courts, 1870-1970," *Law & Society Review* 21:403-45 (1987).