

Craig Callen on the Burden of Proof

Dale A. Nance*

I was delighted to accept Michael Risinger's invitation to comment on a paper by the late Craig Callen, finally published with the proceedings of this symposium.¹ His draft, dated March 21, 2008, is clearly unfinished. So I will write at a relatively high level of generality about Craig's contribution in this final work, concentrating on what I take to be his important insights. Before doing so, I should mention that, over the years, I had numerous rewarding conversations with Craig, usually at conferences we were both attending. I am sure I am not alone in my gratitude for his constant reminders of the importance of the cognitive limitations of adjudicative factfinders. Unsurprisingly, these reminders play a role in his parting intellectual message.

Craig's most important thesis in this paper is that we should be skeptical of solely comparative theories of the burden of proof because they do not capture an important aspect of proof requirements associated with the reluctance of the factfinder to award a verdict in the face of unacceptable "gaps" in the evidence. The context in which he makes this claim is an assessment of the theory of proof burdens for civil cases articulated by Michael Pardo and Ron Allen in their 2008 paper entitled, "Juridical Proof and the Best Explanation."² This context requires Craig to make some additional observations about problems with the particular comparative theory advanced by Professors Michael Pardo and Ron Allen. Finally, Craig observes how standards of directed verdicts and summary judgments reflect judicial supervision of jury verdicts involving a non-comparative element.

I will address only in passing the second of these topics, namely, that "inference to the best explanation," as a theory of the standard of proof, is unworkable in its present form. That theory says that the task of the factfinder is to identify, with the help of the parties' stories (or theories of the case), the most plausible hypothesis explaining the evidence and give

* Albert J. Weatherhead III & Richard W. Weatherhead Professor of Law, Case Western Reserve University.

¹ Craig R. Callen, *Spotting a Preponderance of the Evidence in the Wild: Inference to the Best Explanation and Sufficiency of the Evidence*, 48 SETON HALL L. REV. 1517 (2018).

² Michael S. Pardo & Ronald J. Allen, *Juridical Proof and the Best Explanation*, 27 LAW & PHIL. 223 (2008).

verdict for the plaintiff, if and only if, that most plausible explanation satisfies or instantiates the elements of the plaintiff's cause of action.³ I have addressed this theory elsewhere,⁴ and in the end I think it is largely a distraction from Craig's main thesis, which applies to all comparative theories of the standards of proof and, in fact, to many that are not comparative. But to explain this point, it is useful to be a bit clearer than most previous discussions have been about the meaning of a "comparative" standard of proof, for this terminology appears regularly in the literature without being clearly defined.

By one conception, a comparative standard of proof is any standard of proof that requires (only) a comparison of some quantity or quality derived from the evidence and allocated to the plaintiff to a quantity or quality similarly derived and allocated to the defendant. By this conception, the conventional probabilistic theory of the civil standard of proof, that a verdict should be awarded the plaintiff, if and only if, the probability that the true facts instantiate the plaintiff's claimed cause of action is true exceeds 0.5, seems non-comparative. I say it "seems" non-comparative because that standard is equivalent to the theory that a plaintiff's verdict is appropriate if and only if the probability that the facts instantiate the plaintiff's claimed cause of action is greater than the probability that they instantiate the defendant's claim that the plaintiff's claim is false. Stated this way, the conventional probabilistic theory is also comparative in this sense. Indeed, the probabilistic articulation of other standards of proof are also comparative in this sense, although they must introduce a number, as in the assessment that "instantiation of the plaintiff's cause of action is at least three time more probable than instantiation of the defendant's denial of that cause of action."⁵

This is close to what Craig had in mind by "comparative," but it is not quite the same. Yet another conception, which is what I think Craig had in mind, is that a standard of proof is comparative when it involves (only) the comparison of some quantity or quality derived from the evidence and allocated to the plaintiff's specific theory of the case with a similar quantity or quality derived and allocated to the defendant's specific theory of the case. In this sense, the conventional probabilistic theory is not comparative, whereas Craig seems to regard Pardo and Allen's theory as comparative.

³ Here and throughout, I simplify the discussion by assuming a single asserted cause of action with no formal affirmative defenses.

⁴ See DALE A. NANCE, *THE BURDENS OF PROOF: DISCRIMINATORY POWER, WEIGHT OF EVIDENCE, AND TENACITY OF BELIEF* 73–84 (2016).

⁵ For example, if the criterion for decision is that the probability of the claim being true must exceed 0.75, then this is equivalent to saying that the probability that the claim is true must be greater than 3 times the probability that the claim is false. Thus, under yet another conception of "comparative" standards—viz., that the standard does not incorporate any quantification—any use of a number—these higher standards are non-comparative.

2018] *CRAIG CALLEN ON THE BURDEN OF PROOF* 1571

True, Professor Allen's original relative plausibility theory, which would prescribe a verdict for the plaintiff, if and only if, the plaintiff's theory of the case is more plausible than the defendant's, is comparative in this sense.⁶ But this feature was lost when Allen accepted the importance of recognizing that the factfinder considers, and must be allowed to consider, theories of the case that neither party has asserted.⁷ Given that concession, the Pardo and Allen theory remains comparative in the first sense articulated above, however, because the best available explanation will favor one side and, a fortiori, it must be a better explanation than the best explanation the factfinder can discern that favors the other side. So their theory could be stated as requiring the factfinder to compare the best discernible explanation favoring the plaintiff (whether or not specifically advanced by the plaintiff) to the best discernible explanation favoring the defendant (whether or not specifically advanced by the defendant).

With these clarifications in mind, consider Craig's theory articulating the burden of persuasion in a civil case:

[T]he preponderance of the evidence standard for proof of element *E* requires that the fact finder conclude that (i) *E* is more likely than not-*E*, given the evidence in the record and the stories or explanations the fact finder is considering, and (ii) the possible disutility of finding *E* in the absence of further information or explanation does not warrant reliance on the default rule by finding for the party that does not have the burden.⁸

The first prong of this test reflects Craig's rejection of the relative plausibility idea, which he articulates in the context of an example given by Pardo and Allen:

Hypothesizing a situation in which the trier believes the probability of the plaintiff's case is .4 and the defendant has two

⁶ See Ronald J. Allen, *A Reconceptualization of Civil Trials*, 66 B.U.L. REV. 401 (1986).

⁷ See NANCE, *supra* note 4, at 78–79.

⁸ Callen, *supra* note 1, at 1564 (citing Craig R. Callen, *Kicking Rocks with Dr. Johnson: A Comment on Professor Allen's Theory*, 13 CARDOZO L. REV. 423, 431 (1991)). For trials in which more than one element of the cause of action is litigated, it is not clear in Craig's draft whether he would require anything more than a set of such tests, one for each litigated element of the cause of action. In the earlier article that he cites, it is somewhat clearer that he does not contemplate a test that would require satisfying the requirement that the conjunction of the elements be more likely than not. See *Kicking Rocks*, 13 CARDOZO L. REV. at 425–27. This implicates the so-called "conjunction paradox," which evidently Craig was attempting to avoid, but a discussion of that would also distract us from Craig's principal theme, so I will refer hereafter to the first prong of his test as roughly coinciding with the conventional probabilistic test for civil cases, though I recognize that Craig might not agree with that characterization. This difference is not important in what follows.

defenses, which each have a probability of .1, they argue that the trier should find for the plaintiff. [Pardo and Allen's theory] requires that the plaintiff's more probable explanation prevail.⁹

Craig criticizes Pardo and Allen's approach because it "seems to implicitly assume that the party with the better story is more likely to be right than its opponent, and so should win the verdict, despite the uncertainty."¹⁰

Craig's reference to "uncertainty" suggests a different criticism of Pardo and Allen's relative plausibility theory than the one that I have articulated, and the difference is illuminating. To explain the difference, I make two simplifying assumptions that, I think, do not change the basic analysis. First, I assume that the two defensive theories (which apparently are not true affirmative defenses on which the defendant bears the burden of persuasion) to which Pardo and Allen refer are mutually exclusive (i.e., they cannot both be true), so the probability of at least one of them being true is 0.2. Second, I assume that the missing 0.4 probability is attributed to a single explanation or theory of the case. Now my criticism of Pardo and Allen's theory was that one can check this unendorsed alternative explanation against the elements of the claim and determine whether it instantiates the plaintiff's claim or not. If it favors the plaintiff, then the plaintiff should win both under the conventional probabilistic criterion, which would add this 0.4 probability to the 0.4 probability already attributed to plaintiff's case, and under the theory advanced by Pardo and Allen. But if it favors the defendant, then the defendant should win, contrary to Pardo and Allen's theory, because the probability that the plaintiff's claim is true is 0.4, while the probability that it is false is 0.6. This result can be avoided by allowing the factfinder to aggregate the probability attributable to unendorsed explanations with that of those that are advanced by the parties, an aggregation that Pardo and Allen explicitly concede to be appropriate under their theory, giving the same results as under the conventional probabilistic criterion. In other words, insofar as it articulates a decision criterion, as distinct from an inferential methodology, Pardo and Allen's theory collapses into the conventional probabilistic criterion.¹¹

Craig, however, has interpreted the example offered by Pardo and Allen in a different way. Although Pardo and Allen do not say this explicitly (and, in fact, there is little in their article to suggest this is what they were thinking), Craig interprets the unendorsed explanation as *unknown*, as one of "an unknown nature having a probability of .4 in the mathematical model."¹²

⁹ Callen, *supra* note 1, at 1561 (citing Pardo & Allen, *supra* note 2, at 256).

¹⁰ Callen, *supra* note 1, at 1562.

¹¹ See NANCE, *supra* note 4, at 80–81.

¹² Callen, *supra* note 1, at 1562.

Because it is of an unknown nature, one cannot check it against the elements of the claim to determine whether it favors the plaintiff or the defendant, so one cannot aggregate it with the explanations offered by the plaintiff or those offered by the defendant. Craig then proceeds to examine the implications of such “uncertainty.” Craig identifies the source of this uncertainty as arising from incompleteness in the evidence, which can leave gaps in the factfinder’s understanding of what happened, and it is to this problem that the second prong of his test is directed.

There is much to be said for this understanding of what Pardo and Allen were trying to illustrate with their suggestive example. At first glance, however, it would seem that, on this interpretation, what they suggested is simply incoherent, because the probability that the plaintiff’s cause of action is instantiated is not the complement of the probability that it is not, which violates the conventional axioms of probability. Nevertheless, the point of their illustration can be saved by Craig’s interpretation if we re-characterize their example in terms of Shafer-style belief functions instead of probabilities.¹³ It then makes perfect sense to speak of a belief that the claim is true of 0.4 and a belief that it is false of 0.2, with 0.4 being *uncommitted evidential support*. Indeed, in his book articulating the theory of belief functions, Glenn Shafer explicitly connects this withholding of support to the problem of evidential incompleteness.¹⁴

If this reconstruction of the example is used, the dilemma becomes how to make a decision in such an epistemic context. I have addressed this elsewhere, canvassing several solutions suggested in the literature, and in my opinion the most sensible way to use belief functions to make decisions in the presence of uncommitted support is to ignore the uncommitted support (precisely because it refers to an inability to discriminate between the contending hypotheses) and to use the ratio of beliefs (which are based on committed support) as the odds ratios for the propositions in question.¹⁵ In other words, to use Pardo and Allen’s example, if the belief that the plaintiff’s claim is true is 0.4, while the belief that the plaintiff’s claim is false is 0.2, their ratio, 2:1 in favor of the plaintiff, is the only meaningful odds on the plaintiff’s claim—the equivalent of a probability that the plaintiff’s claim is true of 0.66—which yields a verdict for plaintiff just as Pardo and Allen suggest.

¹³ See GLENN SHAFER, *A MATHEMATICAL THEORY OF EVIDENCE* (1976). To be clear, Pardo and Allen do not use the concept of belief functions in their paper, nor does anything about their presentation suggest that they would be amenable to such formalism, but their theory can be rendered more plausible in this manner.

¹⁴ See, e.g., *id.* at 5–6, 22–25, 38.

¹⁵ See NANCE, *supra* note 4, at 167–75.

Although this presents interesting potential for improving Pardo and Allen's relative plausibility theory, it also explains why I said, at the beginning of this comment, that Craig's focus on Pardo and Allen's particular theory can distract us from Craig's most important thesis, which appears in the second prong of his suggested test for civil cases. Indeed, the most important points in Craig's paper, at least from my perspective, are (a) that he asserts a two-prong burden of proof test that adds a criterion separate from, but not incompatible with, a comparative test like the conventional probabilistic criterion (roughly expressed in his first prong), and (b) that the second prong is focused on the significance of evidential incompleteness. To reiterate, his second prong would require that, in order to satisfy the civil burden of persuasion, the factfinder must conclude that "the possible disutility of finding *E* in the absence of further information or explanations does not warrant reliance on the default rule by finding for the party that does not have the burden." This additional requirement would apply, for example, in the context of the reconstructed example from Pardo and Allen, discussed above. Put in terms of belief function theory, it would pose the question whether too much evidential support has been left uncommitted to permit a verdict in accordance with comparative assessments like Craig's first prong.

Craig's discussion of this second prong emphasizes the inability of ordinary persons, subject to their cognitive resource constraints, to engage in "optimization" (systematically equating marginal benefit with marginal cost) with regard to the assessment of evidential completeness. Instead, he argues that the experience of the factfinder (and the trial judge) will provide heuristics and other decisional shortcuts that permit the decision-maker to gauge the adequacy of the parties' evidence presentations:

[O]ne can make, and ordinary people often do make, judgments about the adequacy of evidence, rather than relying exclusively on the better of two or more stories as [Pardo and Allen] would have it. . . .

In order to make good decisions in light of their bounded rationality, humans need to recognize familiar situations and crucial data, search for new data in ways that reflect both benefits and costs of seeking data, and employ simple decision rules to help them make decisions based on the critical data. . . .

Jurors (and probably judges) will tend to believe that the parties have put forward the best cases for their positions. On the other hand, jurors' and judges' experience may suggest that additional evidence should be available if one party's position is accurate. Experience may even teach that the uncertainty remaining after the parties have presented their explanations (and jurors have tried to formulate some on their own initiative) requires a decision in

2018] CRAIG CALLEN ON THE BURDEN OF PROOF 1575

favor of the default. . . . There is no argument here that jurors who rely on their own cognitive strategies will be infallible, just that judges and jurors will make better decisions if permitted to consider whether the evidence is sufficiently complete to justify a departure from the default rule.¹⁶

Had I the opportunity to discuss Craig's draft with him, I would have questioned his inclusion of the second prong as part of the burden of persuasion, not because I question the idea of such an additional test, but because I question the choice of the factfinder as the appropriate institutional actor to make the assessment of evidential completeness.¹⁷ Despite Craig's references in the above-quoted passage to judges, his statement of the preponderance standard has the factfinder addressing both prongs of his test.¹⁸ Presumably, he thinks that judges are involved as well, but only by way of supervising the limits of reasonableness in the jury's assessment under the second prong, just as conventional articulations of the burden of production have the judge supervising the limits of reasonableness in the jury's assessments under his first prong. To be sure, Craig is not primarily writing here about the respective roles of judge and jury, but the arguments he makes from the practical experience of the decision-maker, of the decision-maker's ability to recognize "familiar situations and crucial data," have implications for the allocation of such roles.

"Gaps" in the evidence are of two, very different kinds. Unavoidable gaps in the evidence, like the witness who likely would exist if the plaintiff's account were true but does not exist, should be properly taken into account by the factfinder. However, this can and should be done in determining whether the *first* prong of Craig's test is satisfied. For example, if the plaintiff claims the defendant slandered him in front of several acquaintances, and the defendant denies that the allegedly slanderous statement was uttered, a juror would rightly be skeptical of the plaintiff's claim if diligent efforts have failed to produce a single witness to the utterance other than the parties. That fact makes the plaintiff's claim inherently less plausible relative to defendant's.

But there is another kind of "gap" problem. When it comes to questions about whether the parties (and their lawyers) have done an appropriately diligent search or whether, having done such a search, one or both parties are

¹⁶ Callen, *supra* note 1, at 1555, 1565–66 (citations omitted).

¹⁷ I did have an enlightening conversation with Craig on much the same issue in relation to a paper I published in 2009 in the journal he edited. See Dale A. Nance, *Evidentiary Foul Play: The Roles of Judge and Jury in Responding to Evidence Tampering*, 7 INT'L. COMMENT. ON EVIDENCE 3 (2009).

¹⁸ See *supra* text accompanying note 8.

withholding relevant evidence without a good reason—the kind of consideration that often seems to be the driving force behind Craig’s second prong¹⁹—the jury is in a much worse position to appraise the significance of such “gaps” or to provide appropriate signals to the litigants about the results of such appraisals in time for the parties to respond by curing evidential deficiencies. Craig rightly emphasizes that practical decision strategies are developed based on experience, and the juridical actor with the far superior experience to develop strategies addressing litigants’ efforts to find, withhold, destroy, or fabricate evidence is the trial judge. Moreover, the trial judge can take actions—ranging from discovery sanctions, to admissibility rulings, to intimations of impending directed verdicts—that can alert litigants to the deficiencies in time for them to correct them or impose appropriate sanctions if the resulting evidential deficiency is incurable. In particular, in those situations where the *defendant* is best situated (or even uniquely able) to cure the defect in the evidential development of the case, recourse to the default rule available to the jury (namely, a decision against the party bearing the burden of persuasion, i.e., the plaintiff) would be perverse. Accordingly, I have argued that this assessment—the one reflected in Craig’s second prong—should fall within the framework of a broadly conceived, judicially-applied burden of production, rather than the jury-applied burden of persuasion.²⁰

But this, I think, is what might be called a second-order disagreement—important, to be sure, but not the most important.²¹ More important is that Craig correctly identifies the under-appreciated fact that the burden of proof involves two distinct components, one involving a practical assessment of

¹⁹ See Callen, *supra* note 1, at 1555 (discussing parties’ incentives not to fully develop the evidence at trial).

²⁰ This is a major theme in my recent book. See NANCE, *supra* note 4, ch. 4, especially at 227–49. Craig’s limited discussion pertaining to the matter of juridical roles consists of the following, by way of reply to a hypothesized argument that juries have no reason to go beyond a comparative assessment:

That argues that jurors should set aside some of the lessons of their cognitive experience, even though the likelihood that they will rely on those lessons relates to an important justification of the jury—their ability to bring everyday experience to bear on questions of fact.

Callen, *supra* note 1, at 1554. But in jury trials, there *is* a special, one might even say “artificial” structure, an allocation of roles between judge and jury, roles that would, in the course of ordinary practical decision-making, reside in a single, unitary decision-maker. This does indeed sometimes require jurors to set aside their own judgment, as when the question is a question of law.

²¹ Craig is certainly in good company: the majority of scholars who have recognized the distinctive role of evidential completeness have placed primary responsibility for its assessment in the hands of the factfinder. I argue that this is a mistake, though an understandable one. See NANCE, *supra* note 4, at 185–86.

2018] *CRAIG CALLEN ON THE BURDEN OF PROOF* 1577

the strength of parties' cases, and one involving a practical assessment of the adequacy of the parties' collective evidential presentation on the basis of which the factfinder makes the former assessment. This, of course, is only one sense in which the burden of proof has two components: the more conventional sense of that idea is that the burden of proof consists of the burden of persuasion and the burden of production. These two senses (of the twin nature of the burden of proof) are related. It is no surprise that, in articulating examples that illustrate the inability of a purely comparative assessment of the parties' cases fails to account for our legal practices, the standards used for summary judgment and directed verdicts figure prominently in Craig's discussion.²² Judges apply the burden of production imbedded in these doctrines, and they have more institutional competence and authority to assess the completeness of the evidence in the course of doing so than juries have in the course of applying the standard of proof. Not surprisingly, the standard of proof applied by the jury is comparative. This, at any rate, is the topic of a conversation that I wish I could have with Craig today.

I have said enough to illustrate the high regard I have for Craig's work. He set an example for all of us in the seriousness with which he approached the interdisciplinary effort to understand what lawyers can learn from cognitive science. And his generosity in sharing his insights was consistent and ever so collegial. His voice will be greatly missed.

²² See Callen, *supra* note 1, at 1557–60. While I am not fully convinced by Craig's argument in the cited pages—that summary termination standards are incompatible with Pardo and Allen's comparative theory of proof—that is partly because I think their comparative theory does not, in the end, diverge discernibly from the conventional probabilistic theory of the standard of proof.