RONALD J. ALLEN*

EXPLANATIONISM ALL THE WAY DOWN

ABSTRACT

The probabilistic account of juridical proof meets insurmountable problems. A better explanation of juridical proof is that it is a form of inference to the best explanation that involves the comparative plausibility of the parties' stories. In addition, discrete evidentiary matters such as relevance and probative value are also best understood as involving inference to the best explanation rather than being probabilistic.

The nature of juridical proof is a critical question for the field of evidence. Perhaps it is also an interesting question for epistemology, as this symposium suggests. For obvious reasons, one possibility that has been explored is that juridical proof is, or at least is best understood as, probabilistic. The law assigns decision rules to essentially all cases, and those seem easily understood as probabilistic. The preponderance of the evidence easily bears the interpretation of requiring proof of elements to exceed .5; proof beyond reasonable doubt plausibly means a very high probability of guilt, .95 perhaps; and intermediate (and lower) burdens of persuasion may entail intermediate (or lower) probability levels. Clear and convincing is perhaps somewhere between .5 and "very high," and probable cause is somewhere between 0.0 and .5. The equally obvious probabilistic interpretation to give to these numbers is that they are relative frequencies, as propensity and classical accounts are plainly inapposite. That raises an immediate difficulty because virtually never is the data presented at trial in relative frequency formats, and even when it is (DNA evidence, good statistical evidence of disparate treatment), it must be combined with evidence that is not ("The defendant raped me", "I was treated in a way that people with different skin color were not"). The solution to this problem appears equally obvious: subjective Bayesianism. One can translate impressions about evidence into subjective beliefs and then compute posterior probabilities in the light of new evidence. This not only maintains consistency among belief states, but has the added advantage of seeming to approximate what trials seem to be about, which is updating beliefs in the light of new evidence.

Yet another advantage of this approach is that it provides an answer to other important questions such as the meaning of relevance, prejudice, and probative

 DOI: 10.3366/E1742360008000427

https://doi.org/10.3366/E1742360008000427 Published online by Cambridge University Press

value. "Relevance" means a likelihood ratio of anything other than 1 : 1; "prejudice" means that the evidence is likely to affect the rationality of appraising the likelihood ratio; and "probative value" means how far from a 1 : 1 ratio the likelihood ratio is.

Although the model of juridical proof adumbrated above has obvious appeal, I believe it is wrong all the way down, from the most general questions of the basic structure of proof at trial to the most detailed question of the probative value of discrete pieces of evidence. Juridical proof is not probabilistically based but instead is explanation based; rather than probabilism all the way down, it is explanations all the way down. To some extent, these two theories complement rather than compete with each other, but where they point in different directions the legal system is compatible with explanationism but not probabilism, as I will explain in this brief essay.

A preliminary word about the nature of this project. My effort is to understand the nature of juridical proof and give a true and accurate description of it. The project is thus purely positive rather than normative, although obviously states of the world may have normative implications. Nonetheless, I am not arguing that some description of the state of the world should be adopted for any reason other than that it appears to be the best explanation we can currently give of our observations. In light of that description, others may laud or condemn the state of affairs, but that is not my concern. Neither am I engaging in traditional philosophical conceptual work. It is quite obvious, however naive, what "true" and "truth" mean, for example – they mean whether the facts found at trial correspond with the state the world was actually in at the pertinent time. Was the light green or red when the car entered the intersection? Similarly, I accept as more than adequate a conventional understanding of "explanation." I will return to the meaning of "best", but only because it is there that the intersection between probabilism and explanationism lies.

Two other preliminary remarks are in order to situate the inquiry at a conference like this, with its mix of different disciplines. The first is the somewhat counter-intuitive point that factual accuracy is more fundamental to the American legal system than rights and obligations, because, conversely, they are parasitic on factual accuracy. Take a simple example, existence of property. What does it mean to own something – like the clothes that each of you is wearing? The conventional answer is that you have the right to possess, alienate, control, and consume that physical item, but consider what happens when someone else demands the possession of one of your personal items. The legal system's response is to make available a fact finder to whom can be presented evidence of sale, gift, creation, or discovery to establish that the universe was in a certain state at a certain time. The contours of that state determine who has the right to possess, consume, or dispose of the item in question.

This is a general and ubiquitous point. All rights and obligations are meaningless without accurate fact finding. Whether the issue is the age of adulthood, the right to an abortion, the various powers of government, or your right to possess, consume,

https://doi.org/10.3366/E1742360008000427 Published online by Cambridge University Press

and dispose of your clothes, it is the conditioning of rights and obligations on facts that gives them substance. This is the feature that most distinguishes liberal democracies and market economies from autocratic states and centralized economies, and the consequences are obvious.

In finding facts, juridical fact finders are supposed to rely on the evidence produced at trial, but "evidence" cannot be restricted to testimony, exhibits, and demeanor presented at trial. Exhibits and demeanor must be described by propositions to fit into deliberation, which is done by the fact finder. More deeply, the meaning of language, rules of logic, formations of likelihood ratios, or judgments of plausibility come from elsewhere than the "evidence" presented at trial. The law reflects this in the explicit jury instructions to rely on common sense, which means one's background, and to deliberate, which means share one's background.

My final preliminary remark is that, while facts are important, they are not the only thing, for they come with a price attached. Decisions have to be made by somebody – either the parties or the government – as to how much to invest in investigation and trial. Parties in civil cases bear most of the cost of a trial and its attendant circumstances, but there is some government subsidy; in criminal cases, the costs are almost exclusively borne by the government. In any event, cost is plainly a variable that compromises the legal system's pursuit of fact accuracy.

I now move to some of the difficulties of giving juridical proof a conventional probabilistic interpretation. These difficulties open up space for an alternative explanation, that juridical proof is an example of inference to the best explanation. The conventional probabilistic understanding of a preponderance of the evidence requires the fact finder to compare the probability of each of the elements to the probability of its negation and to decide for the plaintiff only if each element is more probably true than false. Because the probability of an element being either true or false exhausts the possibilities, this equates with a requirement that the plaintiff prove each element by more than a .5 probability. To see the difficulties with this, note first that if one of the elements of a cause of action did not occur, a verdict for the plaintiff would be in error. In a torts case, if there was either no duty or no harm, yet plaintiff won a verdict, that would be an error. The specter of costs now raises its head. Errors in fact finding are inevitable, and the objective largely is to distribute them equally over the sets of plaintiffs and defendants, or alternatively reduce the sum of errors. I say "largely" because the legal system actually must take account not only of errors, but also of correct decisions, and, probably more importantly still, of the relationship between the matrix of decisions at trials and incentive effects created by those trials rules (Allen & Laudan 2009).

For now I will ignore the complexities just noted and focus on the traditional conception of proof rules as primarily concerned with error allocation. On that assumption, the three conventional beliefs noted in the previous paragraph are logically inconsistent. If the probability of each of two independent elements of a cause of action, such as breach of duty and causation in tort litigation, is .6 and they

https://doi.org/10.3366/E1742360008000427 Published online by Cambridge University Press

are independent, the probability of their both being true is .6 × .6 = .36; thus, the probability of the defendant not having negligently harmed the plaintiff is 1.0 – .36 = .64. If findings are made for plaintiffs in such circumstances, errors should favor plaintiffs over defendants at a ratio of close to 2 to 1.

In addition to illogic, the probabilistic interpretation leads to bizarre results. Assume in Case 1, another torts case, that breach of duty is proven to .9 and causation to .4. The verdict would be for the defendant, since one of the elements – causation – is not proven by a preponderance of the evidence. Compare that to Case 2, in which both elements are proven to .6. In Case 2, the verdict would be for the plaintiff even though the probability of the defendant having negligently harmed the plaintiff is identical (.9 × .4 = .6 × .6 = .36). Or consider Case 3, in which breach of duty is proven to .9 and causation to .5, still resulting in a verdict for the defendant. But the probability of the defendant having negligently harmed the plaintiff is .9 × .5 = .45. This would be a defendant's case, yet the probability of negligent harm (.45) exceeds the probability in Case 2 (.36), which was a plaintiff's verdict. To be sure, elements of a cause of action will not typically be independent, but that simply means that the conjunction problem will be lessened by the extent of the dependency. Criminal cases have these same attributes, obviously. If a crime has four elements, each is proven to .9, and they are independent, the probability that the defendant is guilty is only .66 (.9 × .9 × .9 × .9 = .66).

One potential logical solution to all this is to conceive of proof as requiring the conjunction to exceed .5 in civil cases and some high number, maybe .95, in criminal cases. However, the law nowhere requires this, and it simply changes the weirdness around (Allen & Jehl 2003, 893). An example from criminal law emphasizes the point. Murder is typically defined as having three elements: voluntary act, causation, intentionality. Theft has a bunch: the taking and carrying away of the personal property of another with the intent to deprive permanently. Now, compare proof of the mental element in murder to that of theft. On average, intent to deprive in theft cases would have to be proved to a much higher probability than intent to kill in murder cases in order that the conjunction exceed whatever the standard is. That is bizarre in itself, and it is even more bizarre to think that either common law judges or legislators took into account these effects in articulating the contours of liability.

The difficulties press even more deeply. The actual effect of decisions under any legal decision rule is an empirical rather than an analytical question. Take a simple example. Employing a .5 decision rule will minimize errors only under certain stringent assumptions about the distribution of cases going to trial and accuracy of assessments of the probabilities of liability. For example, if the two subsets of cases – those in which plaintiffs and those in which defendants deserve to win – are of roughly equal size, the probability of liability is normally distributed, and fact finders find those probabilities reliably, then things may work out as planned. One can hypothesize other situations where the decision rules may work out. However, there are an infinite number of possibilities that will result in the legal decision rule

https://doi.org/10.3366/E1742360008000427 Published online by Cambridge University Press

badly misallocating errors. What if, for example, there are no representatives of one or the other subset? What if fact finders are systematically bad at appraising the probability of liability? The point is obvious. At trial, decision rules are not in an algorithmic relationship to actual results (Allen 2003).

Now add back in the implications of dependence. One might think that the possibility of dependence relationships might ameliorate some of these problems, but in fact it exacerbates them. We now run up against Simpson's paradox. The consequences of dependence relationships is completely unpredictable. By confirming each of two conjunctions, for example, you can disconfirm their conjunction, and so on (Allen 1994).

I could continue – for example, the entire probabilistic program with regard to juridical proof suffers from a deep reference class problem (Allen & Pardo 2007) – but I will make just one last point. Much of the debate about the nature of juridical proof proceeds as though it were clear what the terms of the debate refer to, but that is false. Take the simple statement that the plaintiff must prove an element to greater than .5. What, exactly, does that mean? Does it mean that the plaintiff must prove that, of all the ways the universe might have been the day in question, half plus one favor liability? That is, in fact, the implications of a standard relative frequency account, but it also demonstrates that the standard relative frequency account must be wrong in the juridical context. No one could establish all the ways in which the universe might have been at sometime in the past, let alone that half of them would favor liability. Or consider the plight of the prosecutor in a criminal case. Can a defendant defend by simply walking into court with the residential phone book of the city where the crime occurred and demand that each of the residents be shown not to be the perpetrator? Of course not, but that is equivalent to saying that the standard conventional probabilistic approach must not capture the essence of juridical proof.

To some extent, these difficulties are why legal theorists began to explore subjective Bayesianism as a possible explanation or interpretation of juridical proof, but the difficulties here are immense as well. By far the most important is precisely the subjectivity of subjective Bayesianism. It provides a means of ordering one's beliefs, but those beliefs can be in literally any relationship at all to reality. Moreover, there is, again, literally no evidence in the legal system that fact finders are supposed to optimize their utilities; quite the contrary, they are told to behave as the law requires, including adopting the law's utilities (Allen, Kuhns, Swift & Schwartz 2006, 61).

There are in addition technical problems. Subjective Bayesianism requires the formulation of a probability space of mutually exclusive propositions that sum to 1.0, a recognition of all logical truths in the domain, and updating in light of new evidence. None of this occurs at trial. Before deliberation, fact finders only have a vague idea as to what the parties' actual claims will be. If there is a jury, the jury has an even vaguer understanding of the law prior to being instructed. In any event, by the time the initial probability space is formed, all the trial evidence has already

https://doi.org/10.3366/E1742360008000427 Published online by Cambridge University Press

been taken into account. And of course hovering over all of this is the intractable problem of computational complexity (Allen 1997).

Alvin Goldman has tried to bring some order to all of this through an interesting theorem that demonstrates that under certain assumptions there is a positive probability of evidence increasing the probability of a truthful verdict, but in fact his theorem highlights the incompatibility of juridical proof and Bayesianism (Goldman 1992, 245–52). Similar to the convergence to truth theorems, Goldman's proof requires objectively true likelihood ratios, but, outside of the rare occurrence of something like indisputable DNA,[1] that condition never obtains at trial. The likelihood ratios that fact finders would create are just as subjective as their initial priors, and there is no means of adjudicating their differences. One juror might think sweating means lying and another that it is a sign of a witness striving to remember the truth, and so on.

Why does juridical proof seem so distant from the orderliness of other disciplines? I think the answer lies in the nature of law itself, which provides multiple webs of regulation that encompass most, and indeed probably all, human interactions. Each of us is constrained by webs of law emanating from cities, counties, states, the central government, and perhaps other sources, many of which overlap. It is hard to imagine a personal action or interaction that does not plausibly have any relationship to legal regulation. That means the legal environment is itself highly, indeed impossibly, complex, but in addition, a priori, the potential facts and evidence of any particular legal dispute are infinitely varied. Consider just the example used above: "Is the witness telling the truth?" There is an enormously rich and complicated data set pertinent to that question, ranging from formal consistency to psychological sleuthing, but try to put it all into algorithmic form. One startling implication of this is that trying to apply algorithmic approaches to everyday life is *more* complicated than "science," because it ranges over hugely larger swaths of existence. It is thus, in turn, less amenable to controlled study, and less likely to be worth doing in any event because the results are likely to be highly idiosyncratic.[2]

A more promising approach to understanding juridical proof is that it is a form of inference to the best explanation. Conceiving of cases as involving the relative plausibility of the parties' claims (normally provided in story or narrative form) substantially resolves all the paradoxes and difficulties noted above. The conjunction paradox is distributed equally over the parties, for example, and the complexity problem is purely a function of the choices made by the parties. Indeed, this points out a subtle but important clarification of the role of parties. They are sometimes thought of as ambiguity generators, but in reality they are ambiguity discarders. Rather than litigate an infinite number of ways in which the universe might have been, they focus on a few and ask fact finders to decide which is the most plausible.[3] To be sure, "plausible" may mean "more likely" but this does not mean comparative plausibility reduces to probability. Rather, what is "plausible"

https://doi.org/10.3366/E1742360008000427 Published online by Cambridge University Press

is a function of the explanation, its coherence, consistency, coverage, consilience, and how it fits into the background knowledge possessed by the fact finder.

At a very fundamental level, the rules of evidence reflect the comparative plausibility structure and are inconsistent with a probabilistic one. I have already identified above certain respects in which this is true, and there are numerous further examples:

1. The rules of evidence favor admissibility even in the face of legitimate claims of irrelevancy. The standard bearer here is of course the Supreme Court's decision in *Old Chief v. U.S.* (1997, 171), but that simply acknowledged the obvious truth of the narrative structure of proof at trial.
2. This narrative structure is enhanced by liberal admission of evidence, and even the rule once claimed as the embodiment of the exclusionary practices of Anglo-American law – the hearsay rule – has morphed into a rule of admission. (Allen 1992). All statements by parties are admissible, for example, no matter when or under what conditions made, as are all present sense impressions and statements of states of mind or physical condition. Business records and government reports all come in readily, along with 35 or so other categories of admission. If none of the formal exceptions work, the courts may make up ad hoc exceptions to facilitate admission (Federal Rule of Evidence 807). Although there are some technical exclusionary rules, in reality, like hearsay, they often make promises that they do not keep. Another wide ranging example is the character evidence rules which promise exclusion but permit generous admissibility due to provisions such as FRE 404(b).
3. Further facilitating a comparative story model, relevancy is defined at such a low standard that it excludes nothing of any value whatsoever, and the power to exclude relevant evidence is further reduced by requiring its harmful aspects to substantially outweigh its beneficial (FRE 401, 403).
4. In addition, the rules explicitly favor holistic accounts. There are various rule of completeness that facilitate narrative richness such as FRE 106 and 612, the lingering effect of the common law *res gestae* rule, and discrete rules such as the admissibility of motive even where motive – such as in a murder – is literally irrelevant except in extreme situations (self-defense, for example).
5. Jury instructions further embed a comparative plausibility approach through instructions to rely on "common sense." This in essence means judge the stories you are hearing by comparison to what you have learned in your life.
6. Absent from the rules is any effort to structure inference, leaving fact finders to their normal, explanation-based, modes of reasoning.

Explanationism is not just the best explanation of the macro structure of proof; it is also the best explanation of such issues as relevance and weight of the evidence. Indeed, it is difficult to see how to give a sensible probabilistic interpretation to such matters. A likelihood ratio explanation of both has superficial plausibility, but

https://doi.org/10.3366/E1742360008000427 Published online by Cambridge University Press

immediately runs aground of the universal lack of objective relative frequencies. The explanationist account is by contrast quite straightforward. Evidence is relevant if it fits into an explanatory account, and its weight is a function of its significance for that explanatory account as compared to whatever other accounts may be under consideration. In one sense, the most critical limitation of the conventional probabilistic account of the evidentiary process is their modeling juridical proof as involving the probability of the hypothesis given the evidence. A more plausible account that is more consistent with the evidence I have discussed is that juridical proof involves the probability of the evidence given the hypotheses formulated by the parties, with the best, more or less, explanatory account winning out (Allen & Pardo 2008).

There are some philosophical worries that emerge from this account. For example, it is fair that what I have described is not exactly what the philosophers refer to as inference to the best explanation, at least not in criminal cases, where the critical issue is whether there is a plausible story of innocence rather than whether it is the best. Fair enough, but the central point is the explanation-based nature of juridical proof. What is most plausible is another worry, to which I simply recur to common sense. This will not satisfy the philosophical worries (Lipton 2004), but another reminder that my aim is understanding, not conceptual purity, may be in order here. A similar comment is in order with regard to what an explanation is.

Against these worries are obvious advantages of the explanatory account. It explains how cases distribute uncertainty and ambiguity over the parties, supporting the system's norms. It has the fundamental virtue of being consistent with the way people actually reason rather than imposing an odd and alien form of thinking. Indeed, this is surely why it is quite consistent with the great bulk of evidentiary regulation (which is mostly nonregulation). And it captures the role of the parties, which is to determine the range of ambiguity and uncertainty over which to dispute.

REFERENCES

Allen, Ronald J. 1992. "The Hearsay Rule as a Rule of Admission." 76 *Minnesota Law Review* 76: 797–812.

Allen, Ronald J. 1994. "Factual Ambiguity and a Theory of Evidence." *Northwestern University Law Review* 88: 604–40.

Allen, Ronald J. 1997. "Rationality, Algorithms, and Juridical Proof: A Preliminary Inquiry." *International Journal of Evidence and Proof* 1997: 254–75.

Allen, Ronald J. 2003. "The Error of Expected Loss Minimization." *Law, Probability & Risk* 2: 1–7.

Allen, Ronald J. & Sarah A. Jehl. 2003. "Burdens of Persuasion in Civil Cases: Algorithms v. Explanations." *Michigan State Law Review* 2003: 893.

Allen, Ronald J., Richard B. Kuhns, Eleanor Swift & David S. Schwartz. 2006. *Evidence: Text, Problems, and Cases*. New York: Aspen Publishers.

https://doi.org/10.3366/E1742360008000427 Published online by Cambridge University Press

Allen, Ronald J. & Larry Laudan. 2009. Forthcoming. "Deadly Dilemmas." *Texas Tech Law Review*.

Allen, Ronald J. & Michael S. Pardo. 2007. "The Problematic Value of Mathematical Models of Evidence." *Journal of Legal Studies* 36: 107–40.

Allen, Ronald J. & Michael S. Pardo. 2008. "Juridical Proof and the Best Explanation." *Law & Philosophy* 27: 223–68.

Federal Rules of Evidence 2006.

Goldman, Alvin I. 1992. *Liaisons: Philosophy Meets the Cognitive and Social Sciences*. Cambridge, MA: The MIT Press.

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

Lipton, Peter. 2004. *Inference to the Best Explanation*. 2nd ed. New York: Routledge.

Old Chief v. United States, 519 U.S. 171 (1997).

NOTES

* John Henry Wigmore Professor, Northwestern University School of Law. I am indebted to the participants of the conference for helpful comments and invigorating conversation.

1 Rare to the point of extinction. Even in DNA cases, there is always the possibility of lab error or planted evidence.

2 Thus Larry Laudan's call for more science is unlikely to be satisfied (Laudan 2006, 84–8).

3 In criminal cases, this amounts to determining whether there is a plausible story of guilt, and if so, in addition a plausible story of innocence.

Ronald J. Allen is the John Henry Wigmore Professor of Law at Northwestern University. His primary fields are evidence, procedure, constitutional law, and jurisprudence.

https://doi.org/10.3366/E1742360008000427 Published online by Cambridge University Press