Rethinking legal probabilism

Rafał Urbaniak

Contents

1 Scientific goal

As many miscarriages of justice indicate, scientific evidence is easily misinterpreted in court. This happens partially due to miscommunication between the parties involved, partially due to the usual probabilistic fallacies, but also because incorporating scientific evidence in the context of a whole case can be really hard. While probabilistic tools for piecemeal evaluation of scientific evidence and spotting probablistic fallacies in legal contexts are quite well developed, the construction of a more general probabilistic model of incorporating such evidence in a wider context of a whole case, probabilistic explication of and theorizing about evidence evaluation and legal decision standards, remain a challenge. Legal probabilism (LP), for our purpose, is the view that this challenge can and should be met. This project intends to contribute to further development of this enterprise in a philosophically motivated manner.

> say sth about replicability crisis in forensic sciences at some point?

The assessment of evidence in the court of law can be viewed from at least three perspectives: as an interplay of arguments, as an assessment of probabilities involved, or as an interaction of competing narrations. Each perspective presents an account of legal reasoning (Di Bello & Verheij, 2018; van Eemeren & Verheij, 2017). Individually, each of these strains has been investigated. The probablistic approach is the most developed one but LP is still underdeveloped —to a large extent this is so in light of various lines of criticism developed by the representatives of the other strains.

The **goal of this project** is to contribute to the **development of legal probabilism** by formulating its variant that accomodates important **insights provided by its critics**. A crucial point of criticism is that the fact-finding process should be conceptualized as **a competition of narrations**. I plan to develop methods that allow the probabilist to take this perspective, and explain how such methods allow the legal probabilist to address various objections present in the literature. The key idea is that once **narrations are represented as bayesian networks, various criteria on, features of and operations on narrations can be explicated in terms of corresponding properties of and operations on bayesian networks**. Further, the hypothesis is that such an improved framework will facilitate addressing key objections raised against LP.

The conceptual developments will be accompanied by technical accounts. **R** code capturing the technical features developed will be made available to the reader. Thus, the output will be a **unifying extended probabilistic model embracing key aspects of the narrative and argumentative approaches, susceptible to AI implementation.** The methods employed include: Bayesian statistical methods (including Bayesian approach to higher-order probability), imprecise probabilities, and Bayesian networks.

2 Significance

(state of the art, justification for tackling a specific scientific problem, justification for the pioneering nature of the project, the impact of the project results on the development of the research field and scientific discipline);

2.1 State of the art

One of the functions of the trial is to resolve disputes about questions of facts. Did the defendant rob the bank? Who is the father of the child? Did this drug cause birth defects? To answer such questions, the litigants will present evidence of different kinds: eyewitness testimonies, DNA matches, epidemiological studies, etc. The evidence presented will often be in conflict with other evidence. In a bank robbery case, for example, the prosecution may present eyewitness testimony that the defendant was seen driving a truck near the bank a few minutes after the robbery took place. The defense may respond that no traces were found at the crime scene that would match the defendant. The fact-finders, judges or lay jurors, should address these conflicts by assessing and weighing the evidence, and on that basis reach a final decision. This is a difficult task. The evidence presented at trial can be complex and open to multiple interpretations, and even when it is assessed carefully, it may still lead to an incorrect verdict. How should judges and jurors respond to this uncertainty?

From among the three perspectives mentioned in the beginning, the probablistic approach will be my point of departure, for the following key reasons:

- The project is to be informed by and reflect on the actual practice of legal evidence evaluation, and much of scientific evidence in such contexts has probabilistic form.
- Probabilistic tools are fairly well-developed both for applications and within formal epistemology, reaching a state of fruition which should inspire deeper reflection.
- Statistical computing tools for such methods are available, which makes programming development and preliminary computational and data-driven evaluation of the ideas to be defended a viable enterprise.

Accoringly, the view in focus of this research is legal probabilism (LP)—an ongoing research program that comprises a variety of claims about evidence assessment and decision-making at trial. At its simplest, it comprises two core tenets: first, that the evidence presented at trial can be assessed, weighed and combined by means of probability theory; and second, that legal decision rules, such as proof beyond a reasonable doubt in criminal cases, can be explicated in probabilistic terms.

In the Middle Ages, before the advent of probability theory, there existed an informal mathematics of legal evidence (Wigmore, 1901). Formalistic procedures fixed the number of witnesses required to establish a claim. Lawyers would list ways in which items of evidence could be added or subtracted to weaken or strengthen one's case. This formalistic system fell into disrepute as the Enlightenment principle of 'free proof' gained wide acceptance (Damaška, 1995). Concurrently, the development of probability theory brought forth a new approach to weighing evidence and making decisions under uncertainty. The early theorists of probability in the 17th and 18th century were as much interested in games of chance as they were interested in the uncertainty of trial decisions (Daston, 1988; Franklin, 2001; Hacking, 1975). Bernoulli (1713) was one of the first to formulate probabilistic rules for combining different pieces of evidence in legal cases and assessing to what extent they supported a claim of interest. He was also one of the first to suggest that decision rules at trial could be understood as probability thresholds.

Bernoulli's prescient insights attained greater popularity in the 20th century amidst the law and economics movement (Becker, 1968; Calabresi, 1961; Posner, 1973). In a seminal article, Finkelstein & Fairley (1970) gave one of the first systematic analyses of how probability theory, and Bayes' theorem in particular, can help to weigh evidence at trial. Lempert (1977) was one of the first to rely on probability theory, specifically likelihood ratios, for assessing the relevance of evidence. Such contributions fueled what has been called the New Evidence Scholarship, a rigorous way of studying the process of legal proof at trial (Lempert, 1986).

In response to such developments, Tribe (1971) attacked what he called 'trial by mathematics'. His critique ranged from listing well-known cases of misuse or probabilities in legal contexts and practical difficulties in assessing the probability of someone's criminal or civil liability to the dehumanization of trial decisions. After Tribe, many have criticized legal probabilism on a variety of grounds, both theoretical and practical, arguing that probabilistic models are either inadequate or unhelpful (Brilmayer, 1986; Dant, 1988, pp. @Allen1986A–Reconceptuali; Underwood, 1977, p. cohen86).

After the discovery of DNA fingerprinting in the eighties, many legal probabilists focused on how probability theory could be used to quantify the strength of a DNA match under various circumstances (Kaye, 1986, 2010;

Koehler, 1996; National Research Council, 1992; Robertson & Vignaux, 1995).

Some legal scholars and practitioners have voiced their support for legal probabilism explicitly (Tillers & Gottfried, 2007). Yet skepticism about mathematical and quantitative models of legal evidence is still widespread among prominent legal scholars and practitioners (see, for example, Allen & Pardo, 2007). Even among legal probabilists, few would think it possible to quantify precisely the probability of someone's guilt or civil liability. In response, Taroni, Biedermann, Bozza, Garbolino, & Aitken (2014) suggest that the probabilistic formalism 'should primarily be considered as an aid to structure and guide one's inferences under uncertainty, rather than a way to reach precise numerical assessments' (p. xv).

Perhaps the most difficult callenge challenge for legal probabilism—at least, one that has galvanized philosophical attention in recent years—comes from the paradoxes of legal proof or puzzles of naked statistical evidence. In a number of seminal papers, (Cohen, 1981; Nesson, 1979; Thomson, 1986) formulated scenarios in which, even if the probability of guilt or civil liability, based on the available evidence, is particularly high, a verdict against the defendant seems unwarranted.

Imagine you are a trier of fact in a legal proceeding in which the defendant's guilt is identified as equivalent to a certain factual statement G and that somehow you succeeded in properly evaluating $P(G|E)$—-the probability of G given the total evidence presented to you, E One question that arises in such a situation is: when should you decide against the defendant? when is the evidence good enough? What we are after here is a condition Ψ, formulated in (primarily) probabilistic terms, such that the trier of fact, at least ideally, should accept any relevant claim A (including G) just in case $\Psi(A, E)$. One straightforward attempt might be to say: convict if $P(G|E)$ is above a certain threshold, otherwise acquit.

The difficulty about conjunction arises because intuitively there should be no difference betweenthe trier's acceptance of A and B separately, and her acceptance of their conjunction, $A \wedge B$, that is, that $\Psi(A, E)$ and $\Psi(B, E)$ just in case $\Psi(A \, wedge B, E)$. If $\Psi(H, E)$ just the threshold criterion, requiring that $P(()H|E)$ be sufficiently high, Ψ in general fails to satisfy this equivalence.

A variant of the gatecrasher paradox goes as follows. suppose our guilt threshold is high, say at 0.99. Consider the situationin which 1000 fans enter a football stadium, and 991 of them avoidpaying for their tickets. A random spectator is tried for not paying. The probability that the spectator under trial did not pay exceeds 0.99. Yet,intuitively, a spectator cannot be considered liable on the sole basis of the number of people who did and did not pay.

Arguably, these scenarios underscore a theoretical difficulty with probabilistic accounts of legal standards of proof. Many articles have been written on the topic, initially by legal scholars. In the last decade, philosophers have also joined the debate, while interest among legal scholars has waned.[1] Other conceptual challenges for legal probabilism include the so-called problem of conjunction and the reference class problem.

One important difficulty is that there are various thought experiments in which the probability of guilt is very high and yet conviction or finding of liability is intuitively unjustified—these are known as proof paradoxes (Cohen, 1977; Redmayne, 2008).

At least *prima facie*, then, it seems that some conditions other than high posterior probability of liability have to be satisfied for the decision to penalize (or find liable) to be justified. Accordingly, various informal notions have been claimed to be essential for a proper explication of judiciary decision standards (Haack, 2014; Wells, 1992). For instance, evidence is claimed to be insufficient for conviction if it is not *sensitive* to the issue at hand: if it remained the same even if the accused was innocent (Enoch & Fisher, 2015). Or, to look at another approach, evidence is claimed to be insufficient for conviction if it doesn't *normally support* it: if—given the same evidence—no explanation would be needed even if the accused was innocent (Smith, 2017). A legal probabilist needs either to show that these notions are unnecessary or inadequate for the purpose at hand, or to indicate how they can be explicated in probabilistic terms.

More recently, alternative frameworks for modeling evidential reasoning and decision-making at trial have been proposed. They are based on inference to the best explanation (Allen, 2010; Pardo & Allen, 2008), narratives and stories (Allen, 1986, 2010; Allen & Leiter, 2001; Clermont, 2015; Pardo, 2018; Pennington & Hastie, 1991a), and argumentation theory (Bex, 2011; Gordon, Prakken, & Walton, 2007; Walton, 2002). Those who favor a conciliatory stance have combined legal probabilism with other frameworks, offering preliminary sketches of hybrid theories (Verheij, 2014, p. @urbaniak2018narration).

Another important point of criticism of LP is that legal proceedings are back-and-forth between opposing parties in which cross-examination is of crucial importance, reasoning goes not only evidence-to-hypothesis, but

---

[1]For critical surveys see (Gardiner, 2018; Redmayne, 2008; pardo2019; Urbaniak, 2019).

also hypotheses-to-evidence (Allen & Pardo, 2007; Wells, 1992) in a way that seems analogous to inference to the best explanation (Dant, 1988), which notoriously is claimed to not be susceptible to probabilistic analysis (Lipton, 2004). An informal philosophical account inspired by such considerations—The **No Plausible Alternative Story (NPAS)** theory (Allen, 2010)—is that the courtroom is a confrontation of competing narrations (Ho, 2008; Wagenaar, Van Koppen, & Crombag, 1993) offered by the sides, and the narrative to be selected should be the most plausible one. The view is conceptually plausible (Di Bello, 2013), and finds support in psychological evidence (Pennington & Hastie, 1991b, 1992).

It would be a great advantage of legal probabilism if it could model phenomena captured by the narrative approach. But how is the legal probabilist to make sense of them? From her perspective, the key disadvantage of NPAS is that it abandons the rich toolbox of probabilistic methods and takes the key notion of plausibility to be a primitive notion which should be understood only intuitively.

Pioneering nature of the project

Recent work in Artificial Intelligence made it possible to use probability theory—in the form of Bayesian networks—to weigh and assess complex bodies of evidence consisting of multiple components.

Initial philosophical analysis of the approach has been performed, (Di Bello, 2013) pioneering a probabilistic understanding of narrations.

add more about Marcello

2.2 Choice of problem

2.3 Pioneering nature & impact

3 Concept and work plan

(general work plan, specific research goals, results of preliminary research, risk analysis);

4 Research methodology

(underlying scientific methodology, methods, techniques and research tools, methods of results analysis, equipment and devices to be used in research);

References

Allen, R. J. (1986). A reconceptualization of civil trials. *Boston University Law Review*, 66, 401–437.

Allen, R. J. (2010). No plausible alternative to a plausible story of guilt as the rule of decision in criminal cases. In J. Cruz & L. Laudan (Eds.), *Prueba y esandares de prueba en el derecho*. Instituto de Investigaciones Filosoficas-UNAM.

Allen, R. J., & Leiter, B. (2001). Naturalized epistemology and the law of evidence. *Virginia Law Review*, 87(8), 1491–1550. JSTOR.

Allen, R., & Pardo, M. (2007). The problematic value of mathematical models of evidence. *The Journal of Legal Studies*, 36(1), 107–140. JSTOR.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy*, 76, 169–217. Springer.

Bernoulli, J. (1713). *Ars conjectandi*.

Bex, F. J. (2011). *Arguments, stories and criminal evidence: A formal hybrid theory*. Law and philosophy library. Dordrecht ; New York: Springer.

Brilmayer, L. (1986). Second-order evidence and bayesian logic. *Boston University Law Review*, 66, 673–691.

Calabresi, G. (1961). Some thoughts on risk distribution and the law of torts. *Yale Law Journal*, 70, 499–553.

Clermont, K. M. (2015). Trial by Traditional Probability, Relative Plausibility, or Belief Function? *Case Western Reserve Law Review*, 66(2), 353–391.

Cohen, J. (1977). *The probable and the provable*. Oxford University Press.

Cohen, J. L. (1981). Subjective probability and the paradox of the Gatecrasher. *Arizona State Law Journal*, 627–634.

Damaška, M. R. (1995). Free proof and its detractors. *The American Journal of Comparative Law*, 43(3), 343–357.

Dant, M. (1988). Gambling on the truth: The use of purely statistical evidence as a basis for civil liability. *Columbia Journal of Law and Social Problems*, *22*, 31–70. HeinOnline.

Daston, L. (1988). *Classical probability in the enlightenment*. Princeton University Press.

Di Bello, M. (2013). *Statistics and probability in criminal trials* (PhD thesis). University of Stanford.

Di Bello, M., & Verheij, B. (2018). Evidential reasoning. In *Handbook of legal reasoning and argumentation* (pp. 447–493). Springer.

Enoch, D., & Fisher, T. (2015). Sense and sensitivity: Epistemic and instrumental approaches to statistical evidence. *Stan. L. Rev.*, *67*, 557–611. HeinOnline.

Finkelstein, M. O., & Fairley, W. B. (1970). A Bayesian approach to identification evidence. *Harvard Law Review*, *83*(3), 489–517.

Franklin, J. (2001). *The science of conjecture: Evidence and probability before pascal*. John Hopkins University Press.

Gardiner, G. (2018). Legal burdens of proof and statistical evidence. In D. Coady & J. Chase (Eds.), *Routledge handbook of applied epistemology*. Routledge.

Gordon, T. F., Prakken, H., & Walton, D. (2007). The Carneades model of argument and burden of proof. *Artificial Intelligence*, *171*(10-15), 875–896.

Haack, S. (2014). Legal probabilism: An epistemological dissent. In *Haack2014-HAAEMS* (pp. 47–77).

Hacking, I. (1975). *The emergence of probability: A philosophical study of early ideas about probability, induction and statistical inference*. Cambridge University Press.

Ho, H. L. (2008). *A philosophy of evidence law: Justice in the search for truth*. Oxford University Press.

Kaye, D. H. (1986). The admissibility of "probability evidence" in criminal trials—part I. *Jurimetrics Journal*, 343–346.

Kaye, D. H. (2010). *The double helix and the law of evidence*. Harvard University Press.

Koehler, J. J. (1996). On conveying the probative value of DNA evidence: Frequencies, likelihood ratios, and error rates. *University of Colorado law Review*, *67*, 859–886.

Lempert, R. O. (1977). Modeling relevance. *Michigan Law Review*, *75*, 1021–1057. JSTOR.

Lempert, R. O. (1986). The new evidence scholarship: Analysing the process of proof. *Boston University Law Review*, *66*, 439–477.

Lipton, P. (2004). *Inference to the best explanation*. Routledge/Taylor; Francis Group.

National Research Council. (1992). *DNA technology in forensic science* [NRC I]. Committee on DNA technology in Forensic Science, National Research Council.

Nesson, C. R. (1979). Reasonable doubt and permissive inferences: The value of complexity. *Harvard Law Review*, *92*(6), 1187–1225.

Pardo, M. S. (2018). Safety vs. Sensitivity: Possible worlds and the law of evidence. *Legal Theory*, *24*(1), 50–75.

Pardo, M. S., & Allen, R. J. (2008). Judicial proof and the best explanation. *Law and Philosophy*, *27*(3), 223–268.

Pennington, N., & Hastie, R. (1991a). A cognitive theory of juror decision making: The story model. *Cardozo Law Review*, *13*, 519–557.

Pennington, N., & Hastie, R. (1991b). A cognitive theory of juror decision making: The story model. *Cardozo Law Review*, *13*, 519–557. HeinOnline.

Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of personality and social psychology*, *62*(2), 189–204. American Psychological Association.

Posner, R. (1973). *The economic analysis of law*. Brown & Company.

Redmayne, M. (2008). Exploring the proof paradoxes. *Legal Theory*, *14*(4), 281–309. Cambridge University Press.

Robertson, B., & Vignaux, G. A. (1995). DNA evidence: Wrong answers or wrong questions? *Genetica*, *96*, 145–152.

Smith, M. (2017). When does evidence suffice for conviction? *Mind*.

Taroni, F., Biedermann, A., Bozza, S., Garbolino, P., & Aitken, C. (2014). *Bayesian networks for probabilistic inference and decision analysis in forensic science* (2nd ed.). John Wiley & Sons.

Thomson, J. J. (1986). Liability and individualized evidence. *Law and Contemporary Problems*, *49*(3), 199–219.

Tillers, P., & Gottfried, J. (2007). Case comment–United States v. Copeland, 369 F. Supp. 2d 275 (E.D.N.Y. 2005): A collateral attack on the legal maxim that proof beyond a reasonable doubt is unquantifiable? *Law, Probability and Risk*, *5*(2), 135–157.

Tribe, L. H. (1971). Trial by mathematics: Precision and ritual in the legal process. *Harvard Law Review*, *84*(6), 1329–1393.

Underwood, B. D. (1977). The thumb on the scale of justice: Burdens of persuasion in criminal cases. *Yale Law Journal*, *86(7)*, 1299–1348.

Urbaniak, R. (2018). Narration in judiciary fact-finding: A probabilistic explication. *Artificial Intelligence and Law*, 1–32.

Urbaniak, R. (2019). Probabilistic legal decision standards still fail. *Journal of Applied Logics*, *6*(5).

van Eemeren, F., & Verheij, B. (2017). Argumentation theory in formal and computational perspective. *IFCoLog Journal of Logics and Their Applications*, *4*(8), 2099–2181.

Verheij, B. (2014). To catch a thief with and without numbers: Arguments, scenarios and probabilities in evidential reasoning. *Law, Probability and Risk*, *13*(3-4), 307–325. Citeseer.

Wagenaar, W., Van Koppen, P., & Crombag, H. (1993). *Anchored narratives: The psychology of criminal evidence*. St Martin's Press.

Walton, D. N. (2002). *Legal argumentation and evidence*. Penn State University Press.

Wells, G. (1992). Naked statistical evidence of liability: Is subjective probability enough? *Journal of Personality and Social Psychology*, *62*(5), 739–752. American Psychological Association.

Wigmore, J. H. (1901). Required numbers of witnesses; a brief history of the numerical system in england. *Harvard Law Review*, *15*(2), 83–108.