

# RETHINKING LEGAL PROBABILISM

Rafał Urbaniak

## 1. Scientific goal

The **goal of this project** is to **develop of a probabilistic modelling method of the interaction of various items of evidence and hypotheses and the resulting decisions in the court of law**. In light of the current criticism of the probabilistic approach, such a model should (1) be **sensitive to the argumentative structure** involved, and (2) capture the idea that in a legal context we are dealing with a **class of competing narrations**. At this point in the literature, it is not clear how this is to be achieved within the probabilistic approach.

The key idea is to represent **narrations as bayesian networks**. Then, the argumentative structure becomes clear, and various features requested by the critics can be explicated in terms of corresponding properties of, operations on, and relations between bayesian networks.

The output will be a **unifying extended probabilistic model embracing key aspects of the narrative and argumentative approaches, with implementation in the programming language R**. What the project will uniquely bring to the table is joining the familiarity with epistemological debates, familiarity with the details of evidence assessment in legal cases and technical skill to programmatically implement, simulate and test various theoretical moves.

## 2. Significance

### 2.1. State of the art

**Legal probabilism (LP)** comprises two core tenets: (1) the evidence presented at trial can be assessed, weighed and combined by means of probability theory; and (2) legal decision rules, such as proof beyond a reasonable doubt in criminal cases, can be explicated in probabilistic terms. LP gained popularity amidst the law and economics movement (Becker, 1968; Calabresi, 1961; Finkelstein & Fairley, 1970; Lempert, 1977, 1986; Posner, 1973).

These developments faced **criticism** (Allen, 1986; Brilmayer, 1986; Cohen, 1986; Dant, 1988; Tribe, 1971; Underwood, 1977). The negative trend has been somewhat mitigated by the discovery of DNA fingerprinting in the eighties and progress in forensic science in general, with the increasing role of quantitative evidence in the court of law (Kaye, 1986, 2010; Koehler, 1996; National Research Council, 1992; Robertson & Vignaux, 1995).

Skepticism about wider mathematical and quantitative models of legal evidence is still widespread among prominent legal scholars and practitioners Allen & Pardo (2007), partially in light of **conceptual difficulties** extensively discussed in the literature, which arise when one wants to formulate a probabilistic decision criterion for the court of law. For instance, the threshold view has it that sufficiently high probability of guilt is sufficient for conviction (DeKay, 1996; Kaye, 1979; Laudan, 2006). The threshold view is blocked by the so-called paradoxes of legal proof or puzzles of naked statistical evidence (Cohen, 1981; Nesson, 1979; Thomson, 1986). Another conceptual problem is the so-called difficulty about conjunction. It arises, because intuitively there should be no difference between the trier's acceptance of  $A$  and  $B$  separately, and her acceptance of their conjunction,  $A \wedge B$ . On the threshold view, such a difference might arise.

How to define the **standards of proof**, or whether they should be even defined in the first place, remains contentious (Diamond, 1990; Gardiner, 2018; Horowitz & Kirkpatrick, 1996; Laudan, 2006; Newman, 1993; Redmayne, 2008; Walen, 2015). Various informal notions have been claimed to be essential for a proper explication of judiciary decision standards (Enoch & Fisher, 2015; Haack, 2014; Smith, 2017; Wells, 1992). 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.

**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; Hastie, 2019; Ho, 2019; Nance, 2019; Pardo & Allen, 2008; Schwartz & Sober, 2019), 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). Preliminary sketches of hybrid theories (Urbaniak, 2018a; Verheij, 2014) are also available.

The main point of criticism is that legal proceedings are back-and-forth between opposing parties in which cross-examination and the argumentative structure are of crucial importance, reasoning goes not only evidence-to-hypothesis, but 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 LP if it could model such phenomena and by doing so resolve the already mentioned conceptual challenges.

The idea that **Bayesian networks** can be used for probabilistic reasoning in legal fact-finding started gaining traction in late eighties and early nineties (Edwards, 1991). A Bayesian network comprises two components: a directed acyclic graph of relations of dependence (represented by arrows) between variables (represented by nodes) and conditional probability tables corresponding to these relations. Fairly simple graphical patterns (called *idioms*) often appear while modeling the relationships between evidence and hypotheses. Complex graphical models can be constructed by combining these basic patterns in a modular way (Bovens & Hartmann, 2004; Fenton & Neil, 2018a; Fenton, Neil, & Lagnado, 2013; Friedman, 1974; Hepler, Dawid, & Leucari, 2007; Neil, Fenton, & Nielson, 2000; Taroni, Biedermann, Bozza, Garbolino, & Aitken, 2014).

Some attempts have been made to use Bayesian networks to weigh and assess complex bodies of evidence consisting of multiple components. Kadane & Schum (2011) made one of the first attempts to model an **entire criminal case**, Sacco & Vanzetti from 1920, using probabilistic graphs. Similarly, Fenton & Neil (2018b) constructed a Bayesian network for the famous Sally Clark case. The literature contains also a general approach to the use of BNs for modeling whole cases. The main idea is that once all the pieces of evidence and claims are represented as nodes, one should use the **scenario idiom** to model complex hypotheses, consisting of a sequence of events organized in space and time: a scenario (Vlek, Prakken, Renooij, & Verheij, 2014).<sup>1</sup> A BN that uses the scenario idiom would consist of the following components: first, nodes for the states and events in the scenario, with each node linked to the supporting evidence; second, a separate scenario node that has all states and events as its children; finally, a node corresponding to the ultimate hypothesis as a child of the scenario node. The scenario node in a sense unifies the different events and states. Because of this unifying role, increasing the probability of one part of the scenario will also increase the probability of the other parts. This is intended to capture the idea that the different components of a scenario form an interconnected sequence of events.

This strategy is supposed to make sense of the notion of the **coherence of a scenario** as different from its probability given the evidence. On this approach (Vlek, 2016; Vlek, Prakken, Renooij, & Bart Verheij, 2015; Vlek, Prakken, Renooij, & Verheij, 2013; Vlek et al., 2014), coherence is identified with the prior probability of the scenario node. The approach formally represents reasoning with **multiple scenarios**. Given a class of narrations, all the nodes used in some of the separate BNs are to be used to build one large BN, and separate scenario nodes are to be added to it, so that one BN supposedly represents multiple scenarios at once.

An **alternative approach** to representation of and reasoning with multiple scenarios has been developed by Neil, Fenton, Lagnado, & Gill (2019). They criticize (Urbaniak, 2018a) by pointing out that it makes no connection to BNs and so it “fails to offer a convincing and operational means to structure and compare competing narratives.” This is a fair assessment of the limits of what I have achieved so far. They represent separate narrations in terms of separate BNs, and deploy Bayesian model comparison and averaging (linear pooling) as a tool for reasoning with multiple scenarios.

## 2.2. Pioneering nature of the project

Here are the key reasons why I am convinced **the scenario node approach is not satisfactory**. (A) Adding a parent node by *fiat* without any good reasons to think the nodes are connected other than being a part of a single story, introduces probabilistic dependencies between the elements of a narration. (B) Another problem results from the identification of prior probability with coherence. This does not add up intuitively. (C) In general, the legal probabilistic approach to coherence is very simple and fails to engage with rich philosophical literature exactly on this topic (Douven & Meijs, 2007; Fitelson, 2003a, 2003b; Glass, 2002; Meijs & Douven, 2007; Olsson, 2001; Shogenji, 1999).<sup>2</sup> (D) The merging procedure with scenario nodes assumes that for the nodes that are common to the networks to be merged, both the directions of the arrows in the DAGs and the conditional

<sup>1</sup> A discussion of modelling crime scenarios by means of graphical devices mixed with probabilities can be also found in the work of Shen, Keppens, Aitken, Schafer, & Lee (2007), Bex (2011), Bex (2015), Dawid & Mortera (2018) and Verheij (2017). See also the survey by Di Bello & Verheij (2018).

<sup>2</sup> This approach also fails to include a long list of counterexamples to the existing proposals and desiderata that a probabilistic coherence measure should satisfy (Akiba, 2000; Bovens & Hartmann, 2004; Crupi, Tentori, & Gonzalez, 2007; Koscholke, 2016; Merricks, 1995; Schippers & Koscholke, 2019; Shogenji, 1999, 2001, 2006; Siebel, 2004, 2006).

probability tables are the same across different narrations. This is unrealistic.

Here are the key **limitations of the model selection and averaging approach**. (E) The assumption of equal priors is highly debatable. This approach would render prior probabilities quite sensitive to the choice of hypotheses and thus potentially arbitrary. Also, this approach seems particularly unsuitable for criminal cases. If the only two hypotheses/models on the table ultimately say “the defendant is guilty” and “the defendant is innocent”, the prior probability of each would be 50%. But defendants in criminal cases, however, should be presumed innocent until proven guilty. A 50% prior probability of guilt seems excessive.<sup>3</sup> Some take the presumption to mean that the prior probability of guilt should be set to a small value (Friedman, 2000; Friedman, Allen, Balding, Donnelly, & Kaye, 1995), but it is not clear whether this interpretation can be justified on epistemological or decision-theoretic grounds. (F) More recent models rely on relevant background information about people’s opportunities to commit crimes (Fenton, Lagnado, Dahlman, & Neil, 2019). But even if these models are successful in giving well-informed assessments of prior probabilities any evidence-based assessment of prior probabilities, they are likely to violate existing normative requirements of the trial system (Dahlman, 2017; Engel, 2012; Schweizer, 2013). People who belong to certain demographic groups will be regarded as having a higher prior probability of committing a wrong than others, and this outcome can be seen as unfair (Di Bello & O’Neil, 2020). (G) Model selection based on likelihood (given equal priors) or posterior model probabilities in general (if priors are not assumed to be equal) boils down to a variant of the threshold view, and so all the difficulties with the threshold view apply. (H) There is a rich literature on the difficulties that linear pooling runs into (see the surveys in Dietrich & List, 2016; Franz Dietrich & List, 2017a, 2017b). One problem is that the method satisfies the unanimity assumption: whenever all models share a degree of belief in a claim, this is exactly the output degree for that belief. Another problem is that linear pooling does not preserve probabilistic independence (List & Pettit, 2011): even if all models agree that certain nodes are independent, they might end up being dependent in the output. There is also a variety of impossibility theorems in the neighborhood (Gallow, 2018).

**The key steps in my proposal are as follows:**

**Representation.** Use BNs taken separately without scenario nodes to represent various narrations. without assuming the conditional probability tables or directions of edges are the same across the BNs, adding another layer of information, dividing nodes into evidence and narration nodes.

**Dynamic BNs.** Averaging does not seem to be the right way to model cross-examination. To take the argumentative approach seriously and to be able to model relations such as “undercutting” and “rebutting” I will consider another dimension: BNs changing through time in light of other BNs.

**BN-based coherence.** I have formulated a BN-based coherence measure that is not a function of a probabilistic measure and a set of propositions alone, because it is also sensitive to the selection and direction of arrows in a Bayesian Network representing an agent’s credal state. Now, it needs to be deployed (implemented in **R** for BNs) and properly tested on real-life cases discussed in the LP literature.

**Divide and conquer.** I propose that ensemble methods should be deployed for multiple narration variants available from one side (as in when, say, the prosecution story comes with uncertainty about the direction of an arrow or about a particular probability table), but selection methods should be used when final decision is to be made between narrations proposed by the opposing sides.

**Ensemble methods.** One question that arises is whether the general concerns about linear pooling arise for such limited applications. If not, the remaining concern is what priors should be used. In light of the controversial nature of equal priors, I plan to study the consequences of rescaling coherence scores (already mentioned) to constitute model priors. The idea is that given that narrations are to be developed by the sides themselves, taking coherence of their narration as determining the prior might be more fair than using equal priors or relying on geographical or population statistics. If yes, perhaps some other methods boiling down to a variant of sensitivity analysis can be deployed: look at all BNs corresponding to some variant of the narration of one of the sides, find the strongest and the weakest one, and these give you a range of possible outcomes.

**Selection criteria.** The so-called New Legal Probabilism (NLP) is an attempt to improve on the underspecificity of NPAS (Di Bello, 2013). While still being at most semi-formal, the approach is more specific about the

---

<sup>3</sup>Some (Williamson, 2010) try to defend a variant of the principle of indifference by reference to informational entropy, and a proposal along this line has been used in practical recommendation by expert committees (ENFSI Expert Working Group Marks Conclusion Scale Committee, 2006). However, this attempt has been sensibly criticized by Biedermann, Taroni, & Garbolino (2007). The question remains, what should the proper application of informational entropy in the context of BN selection and averaging look like, given that information entropy considerations independently are in epistemologically decent standing?

conditions that a successful accusing narration is to satisfy for the conviction beyond reasonable doubt to be justified. Di Bello works out the philosophical details of requirements such as evidential support and completeness, resiliency, and narrativity. It is far from obvious that such conditions are susceptible to a Bayesian networks explication. I plan to rely on a more expressive probabilistic framework (Urbaniak, 2018b, 2018a), that is capable of expressing such features within a formalized higher-order language. The key hypothesis is that they can be recast in terms of properties of BNs and that the existing BN programming tools can be extended to implement testing for these criteria. This will make them susceptible to programmatic implementation and further study by means of computational methods. The hope is that on one hand, they will do better than the existing proposals, and where they fail, further insights can be gained by studying the reasons behind such failures.

### 3. Work plan

Throughout the whole project I plan to cooperate with Marcello Di Bello (Arizona State University). Over the last year we co-authored the Stanford Encyclopedia of Philosophy entry on Legal Probabilism. We decided to continue our fruitful cooperation. For over two months we have been working on a book proposal to be submitted to Oxford University Press exactly on the issues to be studied in this research project. Marcello Di Bello is an excellent philosopher with extensive research experience in the philosophy of legal evidence, and he would bring his expertise to the table when working both on the book and on the papers, whereas I would be focused on the technical aspects and the underlying formal philosophy. During the last year of the research **I plan a six months' stay at Arizona University**, to work in person with Di Bello on finalizing the book that presents the results of the project.

Apart from publications, the results will be presented at various conferences devoted to legal reasoning. These include the yearly conferences of the *International Association for Artificial Intelligence and Law* and of the *Foundation for Legal Knowledge Based Systems (JURIX)*, and more general conferences gathering formal philosophers, so that the research is inspired by interaction not only with legal evidence scholars, computer scientists, but also philosophers. I am also already an invited speaker at the upcoming “Probability and Proof” conference that will be part of an international conference on the philosophy of legal evidence (“The Michele Taruffo Girona Evidence Week”) in Girona (Spain) May 23-27 2022.

Stage 1  
Philosophical & formal  
unification

Obtain a unifying extended probabilistic framework by incorporating further insights from philosophical and psychological accounts of legal narrations, and from the argumentation approach. Defend its philosophical plausibility. (6 months)

A philosophical paper published in an academic journal such as *Synthese*, *Mind* or *Ratio Juris*. Working title: *Why care about narration selection principles?*

Stage 2  
AI implementation

Develop Bayesian Network Methods for the obtained formal framework, so that the insights from the argumentation approach and informal epistemology, mediated through it, can be incorporated in AI tools. (12 months)

A technical paper published in a journal such as *IfCoLog Journal of Logics and their Applications*, *Law, Probability and Risk* or *Artificial Intelligence and Law*. Working title: *Implementation of narration assessment criteria in Bayesian Networks with **R***.

Stage 3  
Case studies

Evaluate the developed framework and AI tools by conducting case studies from its perspective. (6 months)

One paper on how the formal framework handles case studies and one paper on how the developed AI tools handle real-life situations in journals such as *Artificial Intelligence* or *Argument & Computation*. Working titles: *Rethinking the famous BN-modeled cases within the narration framework* and *BNarr, an **R** package to model narrations with Bayesian Networks*.

Stage 4  
Back to challenges & output

Investigate the extent to which the new framework helps to handle the issues raised in points A-H. (12 months)

Completion of the planned book.

## 4. Methodology

Standard arguments for the legitimacy of Bayesianism<sup>4</sup> deploy usually rather abstract pieces of reasoning to the effect that if one's degrees of beliefs satisfy certain conditions, they also have to satisfy the probabilistic requirements. My approach to thinking about the plausibility of Bayesian epistemology is rather unlike such approaches. Instead, I prefer the *proof-of-the-pudding* methodology. I am convinced that an important part of the philosophical assessment of the Bayesian research program has to do with its achievements or failures in contributing to debates in philosophy which are not themselves debates about the status of Bayesianism itself. In particular, it would be great news if insights from Bayesian epistemology could be used to further development of forensic AI and deepening our understanding of judiciary decision making.

I will be using four methods: (a) informal conceptual analysis; (b) formal conceptual analysis; (c) computational methods (R simulations, etc.); and (d) case studies. I will rethink and model the existing implementations of whole-case-scale-BNs in legal evidence from the perspective of the new framework and reconstruct cases which extensively use probabilistic reasoning, but for which BNs have not been yet proposed. Both the literature already listed and many textbooks on quantitative evidence in forensics are great sources of such cases.

What the project will uniquely bring to the table is joining the familiarity with epistemological debates on the nature of coherence (which legal probabilists like Vlek or Fenton ignore or are unaware of), familiarity with the details of evidence assessment in legal cases (which formal epistemologists such as Fitelson ignore) and technical skill to programmatically implement, simulate and test various theoretical moves.

A larger initiative involving reconstructing various cases using different representation methods and comparing the representations, called *Probability and statistics in forensic science*, took place at the Isaac Newton Institute for Mathematical Sciences. My approach will be in the same vein.

I have extensive experience in analytic philosophy, conceptual analysis, philosophical logic and probabilistic and decision-theoretic methods as deployed in philosophical contexts. I also have teaching and publishing record that involves statistical programming in the **R** language (my sample programming projects can be visited at <https://rfl-urbaniak.github.io/menu/projects.html>), and so am also competent to develop AI implementations and tests of the ideas to be developed.

One research risk is that it will turn out that some of the informal requirements cannot be spelled out in probabilistic terms, or expressed in terms of properties of Bayesian networks. In such an event, I will study the reasons for this negative result. It might happen that there are independent reasons to abandon a given condition, or it might be the case that probabilistic inexplicability of a given condition is an argument against the probabilistic approach. Either way, finding out which option holds and why would also lead to a deeper understanding of the framework and lead to a publication in academic journals. Another research risk is that the case studies will show that other methods are more efficient or transparent. This would itself constitute a result that could be used to further modify the framework so that its best aspects could be preserved, while the disadvantages discovered during case studies avoided.

---

<sup>4</sup>See for example (Earman, 1992; Urbach & Howson, 1993) for an early yet fairly comprehensive survey, or (Pettigrew, 2011) for a discussion of more recent contributions. See also (Bovens & Hartmann, 2004; Bradley, 2015; Swinburne, 2001).



## References

- Akiba, K. (2000). Shogenji's probabilistic measure of coherence is incoherent. *Analysis*, 60(4), 356–359. Oxford University Press (OUP). Retrieved from <https://doi.org/10.1093/analys/60.4.356>
- 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.
- Bex, F. (2015). An integrated theory of causal stories and evidential arguments. In *Proceedings of the 15th International Conference on Artificial Intelligence and Law - ICAIL '15* (pp. 13–22). San Diego, California: ACM Press.
- Bex, F. J. (2011). *Arguments, stories and criminal evidence: A formal hybrid theory*. Law and philosophy library. Dordrecht ; New York: Springer.
- Biedermann, A., Taroni, F., & Garbolino, P. (2007). Equal prior probabilities: Can one do any better? *Forensic Science International*, 172(2-3), 85–93. Elsevier BV. Retrieved from <https://doi.org/10.1016/j.forsciint.2006.12.008>
- Bovens, L., & Hartmann, S. (2004). *Bayesian epistemology*. Oxford University Press.
- Bradley, D. (2015). *A critical introduction to formal epistemology*. Bloomsbury Publishing.
- 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. L. (1981). Subjective probability and the paradox of the Gatecrasher. *Arizona State Law Journal*, 627–634.
- Cohen, J. L. (1986). Twelve questions about Keynes's concept of weight. *British Journal for the Philosophy of Science*, 37(3), 263–278.
- Crupi, V., Tentori, K., & Gonzalez, M. (2007). On Bayesian Measures of Evidential Support: Theoretical and Empirical Issues. *Philosophy of Science*, 74(2), 229–252.
- Dahlman, C. (2017). Determining the base rate for guilt. *Law, Probability and Risk*, 17(1), 15–28.
- 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.
- Dawid, A. P., & Mortera, J. (2018). Graphical models for forensic analysis. In *Handbook of graphical models* (pp. 491–514). CRC Press.
- Dekay, M. L. (1996). The difference between Blackstone-like error ratios and probabilistic standards of proof. *Law and Social Inquiry*, 21, 95–132.
- Di Bello, M. (2013). *Statistics and probability in criminal trials* (PhD thesis). University of Stanford.
- Di Bello, M., & O'Neil, C. (2020). Profile evidence, fairness and the risk of mistaken convictions. *Ethics*, 130(2), 147–178.
- Di Bello, M., & Verheij, B. (2018). Evidential reasoning. In *Handbook of legal reasoning and argumentation* (pp. 447–493). Springer.
- Diamond, H. A. (1990). Reasonable doubt: To define, or not to define. *Columbia Law Review*, 90(6), 1716–1736.
- Dietrich, F., & List, C. (2016). Probabilistic opinion pooling. In A. Hajek & C. Hitchcock (Eds.), *Oxford handbook of philosophy and probability*. Oxford University Press.
- Dietrich, F., & List, C. (2017a). Probabilistic opinion pooling generalized. part one: General agendas. *Social Choice and Welfare*, 48(4), 747–786. Springer.
- Dietrich, F., & List, C. (2017b). Probabilistic opinion pooling generalized. part two: The premise-based approach. *Social Choice and Welfare*, 48(4), 787–814. Springer.
- Douven, I., & Meijs, W. (2007). Measuring coherence. *Synthese*, 156(3), 405–425. Springer Science; Business Media LLC. Retrieved from <https://doi.org/10.1007/s11229-006-9131-z>
- Earman, J. (1992). *Bayes or bust? A critical examination of bayesian confirmation theory*. Cambridge: MIT Press.
- Edwards, W. (1991). Influence diagrams, bayesian imperialism, and the collins case: An appeal to reason. *Cardozo Law Review*, 13, 1025–1074.
- ENFSI Expert Working Group Marks Conclusion Scale Committee. (2006). Conclusion scale for shoeprint and toolmarks examinations. *Journal of Forensic Identification*, 56, 255–280.
- Engel, C. (2012). Neglect the Base Rate: It's the Law! *Preprints of the Max Planck Institute for Research on Collective Goods*, 23.
- Enoch, D., & Fisher, T. (2015). Sense and sensitivity: Epistemic and instrumental approaches to statistical evidence. *Stan. L. Rev.*, 67, 557–611. HeinOnline.
- Fenton, N., & Neil, M. (2018a). *Risk assessment and decision analysis with Bayesian networks*. Chapman; Hall.
- Fenton, N., & Neil, M. (2018b). *Risk assessment and decision analysis with bayesian networks*. Chapman; Hall.
- Fenton, N., Lagnado, D., Dahlman, C., & Neil, M. (2019). The opportunity prior: A proof-based prior for criminal cases. *Law, Probability and Risk*, [online first].
- Fenton, N., Neil, M., & Lagnado, D. A. (2013). A General Structure for Legal Arguments About Evidence Using Bayesian Networks. *Cognitive Science*, 37(1), 61–102.
- Finkelstein, M. O., & Fairley, W. B. (1970). A Bayesian approach to identification evidence. *Harvard Law Review*, 83(3), 489–517.
- Fitelson, B. (2003a). A Probabilistic Theory of Coherence. *Analysis*, 63(3), 194–199.
- Fitelson, B. (2003b). Comments on jim franklin's the representation of context: Ideas from artificial intelligence (or, more remarks on

- the contextuality of probability). *Law, Probability and Risk*, 2(3), 201–204. Oxford Univ Press.
- Friedman, M. (1974). Explanation and scientific understanding. *Journal of Philosophy*, 71, 5–19.
- Friedman, R. D. (2000). A presumption of innocence, not of even odds. *Stanford Law Review*, 52(4), 873–887.
- Friedman, R. D., Allen, R. J., Balding, D. J., Donnelly, P., & Kaye, D. H. (1995). Probability and proof in *State v. Skipper*: An internet exchange. *Jurimetrics*, 35(3), 277–310.
- Gallow, J. (2018). No one can serve two epistemic masters. *Philosophical Studies*, 175(10), 2389–2398. Springer Verlag.
- Gardiner, G. (2018). Legal burdens of proof and statistical evidence. In D. Coady & J. Chase (Eds.), *Routledge handbook of applied epistemology*. Routledge.
- Glass, D. H. (2002). Coherence, Explanation, and Bayesian Networks. In G. Goos, J. Hartmanis, J. van Leeuwen, M. O'Neill, R. F. E. Sutcliffe, C. Ryan, M. Eaton, et al. (Eds.), *Artificial Intelligence and Cognitive Science* (Vol. 2464, pp. 177–182). Berlin, Heidelberg: Springer Berlin Heidelberg.
- 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).
- Hastie, R. (2019). The case for relative plausibility theory: Promising, but insufficient. *The International Journal of Evidence & Proof*, 23(1-2), 134–140.
- Hepler, A. B., Dawid, A. P., & Leucari, V. (2007). Object-oriented graphical representations of complex patterns of evidence. *Law, Probability and Risk*, 6(1-4), 275–293.
- Ho, H. L. (2008). *A philosophy of evidence law: Justice in the search for truth*. Oxford University Press.
- Ho, H. L. (2019). How plausible is the relative plausibility theory of proof? *The International Journal of Evidence & Proof*, 23(1-2), 191–197.
- Horowitz, I. A., & Kirkpatrick, L. C. (1996). A concept in search of a definition: The effect of reasonable doubt instructions on certainty of guilt standards and jury verdicts. *Law and Human Behaviour*, 20(6), 655–670.
- Kadane, J. B., & Schum, D. A. (2011). *A probabilistic analysis of the sacco and vanzetti evidence*. John Wiley & Sons.
- Kaye, D. H. (1979). The laws of probability and the law of the land. *The University of Chicago Law Review*, 47(1), 34–56.
- 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.
- Koscholke, J. (2016). Evaluating Test Cases for Probabilistic Measures of Coherence. *Erkenntnis*, 81(1), 155–181.
- Laudan, L. (2006). *Truth, error, and criminal law: An essay in legal epistemology*. Cambridge University Press.
- 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.
- List, C., & Pettit, P. (2011). *Group agency: The possibility, design, and status of corporate agents*. Oxford University Press.
- Meijs, W., & Douven, I. (2007). On the alleged impossibility of coherence. *Synthese*, 157(3), 347–360.
- Merricks, T. (1995). Warrant entails truth. *Philosophy and Phenomenological Research*, 55, 841–855.
- Nance, D. A. (2019). The limitations of relative plausibility theory. *The International Journal of Evidence & Proof*, 23(1-2), 154–160.
- National Research Council. (1992). *DNA technology in forensic science* [NRC I]. Committee on DNA technology in Forensic Science, National Research Council.
- Neil, M., Fenton, N., & Nielson, L. (2000). Building large-scale Bayesian Networks. *The Knowledge Engineering Review*, 15(3), 257–284.
- Neil, M., Fenton, N., Lagnado, D., & Gill, R. D. (2019). Modelling competing legal arguments using bayesian model comparison and averaging. *Artificial Intelligence and Law*. Retrieved from <https://doi.org/10.1007/s10506-019-09250-3>
- Nesson, C. R. (1979). Reasonable doubt and permissive inferences: The value of complexity. *Harvard Law Review*, 92(6), 1187–1225.
- Newman, J. O. (1993). Beyond “reasonable doubt”. *New York University Law Review*, 68(5), 979–1002.
- Olsson, E. J. (2001). Why Coherence Is Not Truth-Conducive. *Analysis*, 61(3), 236–241.
- 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.
- Pettigrew, R. (2011). Epistemic utility arguments for probabilism. In *Stanford encyclopedia of philosophy*.
- 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.
- Schippers, M., & Koscholke, J. (2019). A General Framework for Probabilistic Measures of Coherence. *Studia Logica*.
- Schwartz, D. S., & Sober, E. (2019). What is relative plausibility? *The International Journal of Evidence & Proof*, 23(1-2), 198–204.
- Schweizer, M. (2013). The Law Doesn't Say Much About Base Rates. *SSRN Electronic Journal*.
- Shen, Q., Keppens, J., Aitken, C., Schafer, B., & Lee, M. (2007). A scenario-driven decision support system for serious crime

investigation. *Law, Probability and Risk*, 5(2), 87–117.

Shogenji, T. (1999). Is Coherence Truth Conducive? *Analysis*, 59(4), 338–345.

Shogenji, T. (2001). Reply to akiba on the probabilistic measure of coherence. *Analysis*, 61(2), 147–150. Oxford University Press (OUP). Retrieved from <https://doi.org/10.1093/analys/61.2.147>

Shogenji, T. (2006). Why does coherence appear truth-conducive? *Synthese*, 157(3), 361–372. Springer Science; Business Media LLC. Retrieved from <https://doi.org/10.1007/s11229-006-9062-8>

Siebel, M. (2004). On Fitelson's measure of coherence. *Analysis*, 64, 189–190.

Siebel, M. (2006). Against probabilistic measures of coherence. In *Coherence, truth and testimony* (pp. 43–68). Springer.

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

Swinburne, R. (2001). *Epistemic justification*. Oxford University Press.

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.

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.

Urbach, P., & Howson, C. (1993). *Scientific reasoning: The bayesian approach*. Open Court.

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

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

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.

Verheij, B. (2017). Proof with and without probabilities. correct evidential reasoning with presumptive arguments, coherent hypotheses and degrees of uncertainty. *Artificial Intelligence and Law*, 1–28. Springer.

Vlek, C. (2016). *When stories and numbers meet in court: Constructing and explaining bayesian networks for criminal cases with scenarios*. Rijksuniversiteit Groningen.

Vlek, C. S., Prakken, H., Renooij, S., & Bart Verheij. (2015). Representing the quality of crime scenarios in a bayesian network. In A. Rotolo (Ed.), *Legal knowledge and information systems* (pp. 133–140). IOS Press.

Vlek, C., Prakken, H., Renooij, S., & Verheij, B. (2013). Modeling crime scenarios in a bayesian network. In *Proceedings of the fourteenth international conference on artificial intelligence and law* (pp. 150–159). ACM.

Vlek, C., Prakken, H., Renooij, S., & Verheij, B. (2014). Building bayesian networks for legal evidence with narratives: A case study evaluation. *Artificial Intelligence and Law*, 22, 375–421. Springer.

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

Walen, A. (2015). Proof beyond a reasonable doubt: A balanced retributive account. *Louisiana Law Review*, 76(2), 355–446.

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.

Williamson, J. (2010). *In defence of objective bayesianism*. Oxford University Press Oxford.