

# Probability on Trial

## Making Sense of Arguments and Stories

Marcello Di Bello and Rafal Urbaniak

## 1 The Book

### 1.1 Brief Description

‘Probability on Trial’ is the first book-length philosophical examination of legal probabilism, a research program whose aim is to harness the powers of probability to guide the evaluation of evidence at trial, model the process of decision-making, and promote accurate and fair trial decisions. The primary motivation for this project comes from the unfortunate fact that the trial system makes mistakes. Innocents are convicted. Perpetrators are acquitted. And certain group demographics tend to be disproportionately the victims of mistaken trial decisions. How can the trial system best guard against these errors and distribute them fairly when they occur? Many have sought guidance in probability theory, starting with the pioneering work of Bernoulli, Laplace, Condorcet, and D’Alambert. This led to a research program that now involves legal scholars, forensic scientists, psychologists, statisticians, computer scientists and a handful of philosophers.

Despite considerable progress, however, legal probabilism faces robust skepticism by legal scholars and philosophers. Many point out that judges and jurors do not concern themselves with the probability that the defendant is guilty, nor should they. It seems irresponsible, even unjust, to decide about a person’s life with numbers. Many scholars subscribe to an alternative theory under the name ‘relative plausibility,’ closely related to the ‘story model’ of jury decision-making. According to these theories, judges and jurors make sense of the evidence presented at trial by constructing a story or account of what happened. The story that best explains the evidence should prevail. Argumentation theory has also been applied to the analysis of how evidence is evaluated at trial. This argumentation-centered approach emphasizes the adversarial structure of legal trials: the arguments by one side are scrutinized, attacked, challenged by the other side.

Relative plausibility, the story model and argumentation theory provide important insights. ‘Probability on Trial’ has a conciliatory aim. It takes advantage of recent developments in artificial intelligence, in particular Bayesian networks, and shows that probability theory can make the insights from these other theories more rigorous and precise. This can serve to develop a more defensible version of legal probabilism. Hence, the subtitle: ‘Making Sense of Arguments and Stories.’

The main contributions of ‘Probability on Trial’ are as follows:

- A probabilistic account of concepts from the story model (coherence, completeness, specificity) and argumentation theory (rebutting, undercutting);
- A solution to the problem of ‘unanticipated possibilities’ via dynamic refinements of a Bayesian network or comparisons of multiple networks;
- A multi-pronged and yet still probabilistic theory of the standard of proof, involving not only posterior probability, but also factors such as coherence, resilience, and completeness;
- Simulation-based arguments that test the proposed theory in light of accuracy and fairness criteria;
- A response to standard philosophical objections leveled against legal probabilism.

‘Probability on Trial’ is aimed at philosophers with an interest in legal epistemology and epistemology more generally. Many of the difficulties of legal probabilism resemble difficulties faced by Bayesianism in epistemology. The book will also be of interest to legal scholars who have championed applications of probability theory to evidence law as well as scholars who have resisted this trend. Another target audience includes computer scientists and psychologists interested in studying evidential reasoning and decision-making under uncertainty. Besides contributing to the literature about legal probabilism, the book aims to introduce unfamiliar readers to the rich interdisciplinary debate on the topic, often scattered throughout journals and books in philosophy, law, computer science, forensic science and psychology. ‘Probability on Trial’ can be a resource for legal practitioners and reformers who aim to improve the accuracy of trial decisions and promote a fairer justice system. Readers can follow different tracks through the book, depending on their interests and inclinations. Some chapters present original research. Others are introductory, suitable for an advanced undergraduate course.

### 1.2 Outline

‘Probability on Trial’ comprises five parts. Part I covers the current state of legal probabilism and the challenges it faces. Part II describes the basic formal tools, Bayes’ theorem and the likelihood ratio, and their limitations. Part III develops a more sophisticated theory using Bayesian networks. Part IV turns to decisions and standards of proof. Part V focuses on accuracy and fairness.

## **Part I - Legal Probabilism and Its Foes**

The first part of the book will instill interest in legal probabilism among unfamiliar readers and refresh seasoned readers about the main points of contention. **Chapter 1** describes legal probabilism in its infancy. In the early days, calculations were carried out by hand. Bayes' theorem was used for weighing evidence and assigning probabilities to hypotheses. Rules of decision were identified with simple probability thresholds fixed via the maximization of expected utility. This repertoire of ideas proved useful in several ways, especially in the assessment of explicitly quantitative evidence such as DNA matches and expert testimony. At the same time, it has faced robust opposition. **Chapter 2** describes the fierce scholarly debates that accompanied the emergence of legal probabilism in the 20th century. A host of conceptual difficulties emerged: the conjunction problem, the problem of priors, and the paradoxes of naked statistical evidence. These difficulties are well-known. Others are less familiar: the problem of complexity, soft variables, unanticipated possibilities, and corroboration.

The first two chapters provide the motivation for a deeper examination of legal probabilism and the development of its more sophisticated version. The remaining parts of the book cover two distinct topics: evidence assessment (Part II and Part III, Chapters 3 through 10) and decision-making (Part IV and Part V, Chapters 11 through 17). This distinction reflects the fact that legal probabilism is both a theory of evidence assessment (or evidence evaluation, evidence weighing) as well as a theory of decision-making. These two topics are intertwined, but are best kept separate for analytical clarity.

## **Part II - Evidence Assessment First Pass**

The second part of 'Probability on Trial' discusses in great detail Bayes' theorem and likelihood ratios, two formal tools essential for legal probabilism. The objective is to understand how these tools help to evaluate evidence and what their shortcomings may be.

**Chapter 3** begins with Bayes' theorem and surveys many of its applications, for example, as a tool to avoid reasoning fallacies such as the prosecutor's fallacy and the base rate fallacy. At the same time, its applications are also limited. As discussed in **Chapter 4**, court cases often require fact-finders—experienced judges or lay jurors—to weigh several pieces of evidence, sometimes conflicting and susceptible to different interpretations. The hypotheses that the fact-finders are asked to evaluate in light of the evidence are structured stories constituted by several sub-propositions. This complexity can hardly be modeled by successive discrete applications of Bayes' theorem. A more sophisticated machinery for evidence assessment is needed.

**Chapter 5** describes a formal tool distinct from (and arguably, simpler than) Bayes' theorem which many legal probabilists have found useful: the likelihood ratio. We illustrate how it can be used to evaluate both quantitative and non-quantitative evidence, focusing in particular on DNA match evidence and eyewitness testimony. Despite its versatility, however, the likelihood ratio should be deployed with care. It can be hard to interpret in practice and only provides a local assessment of the evidence. A further problem, as critics have alleged, is that the likelihood ratio faces difficulties in modeling the notion of legal relevance.

Chapters 3 and 4 show that we need to move beyond simple legal probabilism. Chapter 5 shows that likelihood ratios, while useful in many ways, are still an unsatisfactory approach overall. The journey toward a more sophisticated legal probabilism is accomplished in Part III, Chapters 6 through 10.

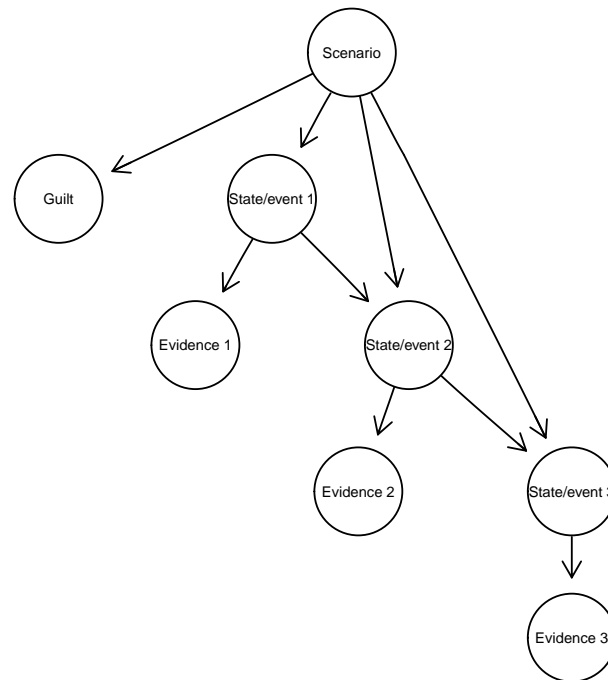
## **Part III - Evidence Assessment More and Better**

The third part of 'Probability on Trial' is informed by two observations about the evaluation of trial evidence. First, the process is holistic. Judges and jurors make sense of the totality of the evidence not in a piecemeal manner, but rather, by constructing a unifying story or explanation. Many have criticized this story model because of its potential for bias. Judges and jurors may discount important evidence just because it does not fit with their preferred story. Arguably, the best way to address this worry is to deploy probability theory and make the process of story construction more rigorous and transparent. The second aspect of how evidence is evaluated in trial proceedings is its adversarial structure. Each side has the opportunity to scrutinize via cross-examination the evidence presented by the other side. The defense may respond by attacking the supporting evidence, the internal consistency of the opponent's account, or by offering an alternative account.

Our working hypothesis is that Bayesian networks constitute the formal machinery necessary for developing a more sophisticated legal probabilism that is able to model these phenomena. The coherence of a story as well as conflicts between pieces of evidence can be modeled by corresponding properties of, operations on, and relations between Bayesian networks.

**Chapter 6** offers a crash course on Bayesian networks with a focus on the assessment of legal evidence. They are graph-like structures that embody probability distributions and model complex structures of evidence. In the last decade, the literature in artificial intelligence and forensic science has made significant progress in modeling holistic notions such as the coherence of a story and argument-based notions such as

conflicts between pieces of evidence. In particular, Charlotte Vlek, together with Bart Verheij and Henry Prakken, proposed to model the coherence of a story by adding a node in the Bayesian network, call it a ‘story node.’ The story node has other nodes as its children corresponding to the events that make up the story. In turn, these events are linked to their supporting evidence. An example of a Bayesian network with a story node (or scenario node to use Vlek’s terminology) is depicted below:



Since the story node unifies the parts of a story, changes in the probabilities of these parts can model the notion of coherence. The stronger the (positive) probabilistic dependence between the parts, the more coherent the story. To model conflicts of evidence, a Bayesian network can be built that comprises two competing stories, each supported by their own evidence. The network would specify that these stories are incompatible and cannot be true concurrently. Another approach to model conflicting evidence and competing stories was developed by Norman Fenton and his research group. Separate stories are represented by separate Bayesian networks, and Bayesian model comparison is then used for assessing the comparative evidential support of the competing stories.

**Chapter 7** assesses Vlek’s story node approach as an account of coherence. The critical argument is followed by a positive proposal. Adding a story node by fiat—without any good reason for supposing that the different parts of the story are connected other than being part of one story—introduces unnecessary probabilistic dependencies between the elements of a story. In addition, the story node approach is simplistic as an account of coherence and fails to engage with the rich philosophical literature on the topic. After the critical argument, the chapter articulates a more adequate probabilistic account of coherence. Using Bayesian networks, we show that a formal notion of ‘structured coherence’ that reflects background causal knowledge addresses the objections against probabilistic accounts of coherence in the philosophical literature. We also show that structured coherence can model related ideas, such as the specificity and completeness of a story.

**Chapter 8** focuses on conflicts between pieces of evidence. Neither Vlek’s story node approach nor Fenton’s model comparison approach adequately capture how pieces of evidence and competing stories may conflict with one another. The complex adversarial dialectic that takes place in a trial cannot be modeled by simply averaging different Bayesian networks (Fenton) or postulating relationships of incompatibility between different story nodes (Vlek). We need an account of more fine-grained notions, such as undercutting and rebutting evidence, and more generally we need an account of how cross-examination operates at trial. What cross-examination often accomplishes is not so much the creation of an alternative story, but the reinterpretation of an existing story by supplying additional information. This process of re-interpretation can be represented formally as the refinement of an existing Bayesian network. Conflicts between pieces of evidence such as undercutting and rebutting can be modeled by drawing additional arrows and nodes between existing nodes. Unanticipated possibilities and new alternative scenarios that may arise in the adversarial confrontation can be modeled by comparisons across multiple Bayesian networks.

The reverse of the phenomenon of conflicting evidence is that of convergence, in particular, the fact that one piece of evidence corroborates another. Corroboration has been the focus of extensive scholarly

debate often independently of the debates within legal probabilism. **Chapter 9** surveys the literature on corroboration and the main difficulties that have been leveled against probabilistic accounts. The chapter then formalizes a notion of corroboration using Bayesian networks that overcomes these difficulties.

**Chapter 10** draws some morals. The more sophisticated version of legal probabilism developed in Chapters 6 through 9—call it *legal probabilism 1.02*—can be challenged because of its questionable empirical adequacy since judges and jurors hardly follow probability theory. The claims of ‘Probability on Trial’ are theoretical and philosophical in nature. They rest on idealizations of the practice of trial proceedings. Nevertheless, we emphasize that legal probabilism 1.02 rivals in explanatory power its main competitors: argumentation theory and relative plausibility. The former is well suited to model conflicts between evidence, but cannot easily model the fact that evidence may conflict more or less strongly with other evidence. Legal probabilism 1.02 offers an account of evidential support, conflict and convergence that captures how these relations come in degrees of strength. The other competing theory, relative plausibility, is often criticized because the defense need not present a full-fledged alternative story. Particularly in criminal trials, it seems enough for the defense to weaken the prosecutor’s story and bring it below a threshold of acceptability. Legal probabilism 1.02 is flexible enough to model conflicts without the need to construct a full-fledged alternative (as critics of relative plausibility might prefer).

In addition, legal probabilism 1.02 offers a richer epistemological theory beyond what critics have recognized. Susan Haack, for example, criticized legal probabilism for its monodimensional account of evidential support. This is true of simple legal probabilism in which evidential support is modeled by the posterior probability of a hypothesis given the evidence or the likelihood ratio. In legal probabilism 1.02, however, evidential support also depends on the degree of specificity, completeness, and coherence of the story put forward, and the extent to which the supporting evidence withstands objections. These features can serve to formulate decision criteria that are not unidimensional, as discussed in the next part of the book.

#### **Part IV - Decision-making**

The fourth part of ‘Probability on Trial’ examines trial decision-making, specifically, to what extent standards of proof such as ‘preponderance of the evidence’ and ‘proof beyond a reasonable doubt’ can be understood through the lenses of probability theory.

**Chapter 11** scrutinizes different strategies for theorizing about standards of proof using probabilistic language. There has been a spur of research arguing that legal probabilism is unfit to model standards of proof. But this research often holds a narrow view of legal probabilism. The most natural decision criterion is a probability threshold whose stringency is determined by expected utility maximization. Unfortunately, this account falls prey to well-known objections, most notably, the puzzles of naked statistical evidence and the difficulty with conjunction (discussed in greater detail in later chapters). Alternative accounts, in particular, the comparative strategy (Cheng) and the likelihood ratio strategy (Dawid, Kaplow, Sullivan) suffer from other complications. A more promising approach is multidimensional. That is, the standard of proof is a function of several criteria: the probability of liability, the specificity and coherence of the accusatory story, the comprehensiveness of the supporting evidence and its ability to withstand objections. These criteria—probability, specificity, completeness and resilience—can be modeled using Bayesian networks. So this multidimensional account lies within the confines of legal probabilism, though not the narrow version its critics have in mind. The following chapters illustrate the theoretical payoffs of this approach.

**Chapter 12** tackles the puzzles of naked statistical evidence, a hotly debated topic for over forty years. Ronald Allen and Lewis Ross have persuasively shown that these puzzles are theoretical constructions with little to no significance for legal practice. We still devote an entire chapter to them because of the enormous scholarly attention they have garnered in the recent philosophical literature. Our proposed resolution of these puzzles rests on two premises. First, an accusation of liability should be substantiated by a well-specified account—or story, narrative—whose moving parts are each supported by adequate evidence. The second premise is that the supporting evidence should be ‘causally grounded.’ This grounding contributes to a well-specified story and is achieved during cross-examination by eliciting additional information about the relation between the evidence and the alleged facts—for example, information about the visibility conditions; the academic credentials of an expert witness; the chain of custody of a document. In cases of naked statistical evidence, the probability of liability is usually quite high, but the accompanying narrative is suspiciously under-specified and thus fails to causally ground the evidence.

**Chapter 13** tackles another central problem for legal probabilism, the difficulty with conjunction. We show that previous attempts in the literature on legal probabilism, though promising and instructive, are ultimately unsatisfactory. This verdict applies to the likelihood ratio approach (Sullivan, Dawid and others) as well as the comparative strategy (Cheng). Our proposal follows the holistic approach by Allen and Pardo. Legal probabilism can address the difficulty with conjunction so long as it can account for holistic

notions such as coherence and specificity. Once the prosecution has established a well-specified accusatory narrative—and its case withstands criticism—each element of the accusation is established to the required standard if and only if the story as a whole is established to the required standard.

**Chapter 14** compares our probabilistic account of decision-making and standards of proof to other, non-probabilistic accounts in the literature. We advance two main points of criticism. First, other accounts are not necessarily incompatible with legal probabilism 1.02 which may in fact provide a more rigorous way to express their insights. This point applies to foundeherentism (Hack), normic support (Smith), argumentation theory (Sartor and Prakken), the story model (Pennington and Hastie; van Koppen and Mackor) and to some extent relative plausibility (Allen and Pardo). The second point we make is that other accounts, in particular knowledge accounts of legal proof, borrow ideas from contemporary analytic epistemology and apply them to the legal domain. These accounts can be criticized for failing to take seriously into account the values and objectives that are integral to decision-making in the justice system.

We might ourselves be accused of unquestionably taking probability theory as a paradigm of rationality for legal-decision making. Admittedly, there is no empirical evidence that a probabilistic turn would improve trial decisions, if only because what it would mean to ‘improve’ them is unclear. A set of criteria are needed for assessing the desired improvements. ‘Probability on Trial’ tackles this normative question in the final part and shows how probability theory can be of service here.

## **Part V - Accuracy and Fairness**

What values and objectives should trial decisions seek to realize? What norms and criteria should guide trial decisions so that they can further these values and objectives? The fourth part of the book addresses these questions by focusing on accuracy and fairness.

As a preliminary step, **Chapter 15** surveys different values and objectives that may inform the design of the trial system. In response to criminal and civil wrongs, we assume that the state has, in principle, the authority to impose punishments and decide about monetary compensations. Given this assumption, we survey the most common values and objectives that trial decisions should conform with: they should be accurate; they should be fair; they should be accompanied by a justification that is public and subject to scrutiny; they should be reversible under appeal; they should contribute to further social cohesion, compliance and deterrence; they should be humane and respectful of people’s dignity. We also explore to what extent these values and objectives may be in tension with one another and to what extent they can be subject to restrictions. The subsequent chapters focus on accuracy and fairness only. We focus on them, not because other values and objectives are not important, but because we believe they are foundational for the others.

**Chapter 16** examines what it means for trial decisions to be accurate and how accuracy can be promoted. Accuracy can be understood as predictive or diagnostic. The former is the probability that a defendant who is found liable at trial (or found not liable) is actually liable (or actually not liable); the latter is the probability that a defendant who is actually liable (or actually not liable) is found liable at trial (or found not liable). The two notions are related but not equivalent. We run several computer simulations whose aim is to compare the performance of possible models of the standard of proof. For example, one simulation compares a model that simply requires that the defendant’s liability be established with a high probability and a model that requires, in addition to high probability, that the story presented by the prosecution or the plaintiff be reasonably specific. The first model prevails in terms of predictive accuracy provided the probability is calibrated, while the second prevails in terms of diagnostic accuracy. However, the probability assessment in the first model is mis-calibrated more often than in the second model, and when mis-calibration occurs, predictive accuracy also suffers. Thus, from the point of view of accuracy, understanding the standard of proof as the combination of ‘high probability’ and ‘narrative specificity’ is a more rational model of decision-making.

**Chapter 17** turns to the fairness of trial decisions. In a substantive, non-formalistic sense, fairness requires that defendants not be subject to excessive burdens, and that these burdens be roughly equally distributed. We examine whether admitting certain types of evidence makes trial decisions unfair, and to what extent trial decisions are inevitably unfair because of structural inequalities in society. We argue that admitting certain forms of evidence, specifically profile evidence and group-based generalizations, tends to burden certain groups of defendants more heavily than others, thus deteriorating the fairness of trial decisions. We also compare, via a computer simulation, different models of the standard of proof and examine which model is conducive to fairer decisions. We show that a standard of proof that incorporates multiple criteria—high probability of liability, coherence and specificity, resistance to objections and comprehensiveness of the evidence—allocates burdens and benefits more fairly across defendants than a standard of proof that simply requires that the probability of liability be above a fixed threshold.

Thus, the multidimensional account of the standard of proof in ‘Probability on Trial’ fares well in light of accuracy and fairness. No such detailed comparison would have been possible without relying on the

rigorous language of probability theory. Even those who resist a direct application of probability in evidence evaluation and decision-making at trial will concede that it plays an important analytic role. It is needed to define the meaning of ‘accuracy’ and ‘fairness,’ and then test competing models against these criteria.

**Chapter 18** draws some conclusions and points to open problems. We emphasize that our aim in ‘Probability on Trial’ is to show that legal probabilism is a richer and more flexible theory than it might seem at first. It is a theory that is able to incorporate insights from other, possibly rival theories. Legal probabilism 1.02 can satisfactorily address long-standing conceptual difficulties: the problem of naked statistical evidence, the difficulty with conjunction, corroboration and cross-examination. We also made some progress on other conceptual difficulties, in particular, the problem of unanticipated possibilities. We underscore how, besides furthering the academic debate, ‘Probability on Trial’ can be used by legal practitioners and reformers who aim to strengthen the protection for defendants against wrongful convictions, improve the accuracy of trial decisions and promote a fairer justice system. We sketch how our theoretical findings can be implemented in the trial setting, but leave a more precise articulation of the challenges of implementation for future work.

## Table of contents

|  |   |
|--|---|
| I Legal probabilism 1.01 and its foes                  | 7.3 Structured coherence with Bayesian networks   |
| 1 The emergence of legal probabilism                   | 7.4 Specificity                                   |
| 1.1 Famous cases                                       | 7.5 Completeness                                  |
| 1.2 Probabilistic evidence                             | 7.6 Application to legal cases                    |
| 1.3 Trial by mathematics                               | 8 Conflicts                                       |
| 1.4 Some history                                       | 8.1 Argumentation theory                          |
| 2 A skeptical perspective                              | 8.2 Undercutting and rebutting                    |
| 2.1 The difficulty about conjunction                   | 8.3 Cross-examination in Bayesian networks        |
| 2.2 The problem of priors                              | 8.4 Unanticipated possibilities                   |
| 2.3 Naked statistical evidence                         | 8.5 Refining and comparing networks               |
| 2.4 Complexity   | 9 Corroboration                                   |
| 2.5 Soft variables                                     | 9.1 Boole’s formula and Cohen’s challenge         |
| 2.6 Corroboration                                      | 9.2 Rise in case of agreement                     |
| 2.7 The reference class problem                        | 9.3 Ekelöf’s corroboration measure                |
| 2.8 Unanticipated possibilities                        | 9.4 Multiple false stories and multiple witnesses |
| 2.9 Non-probabilistic theories                         | 10 Towards legal probabilism 1.02                 |
| II Evidence assessment (first pass)                    | 10.1 Outperforming competing accounts             |
| 3 Bayes’ theorem and the usual fallacies               | 10.2 Empirical adequacy                           |
| 3.1 Assuming independence                              | 10.3 Specificity and coherence                    |
| 3.2 The prosecutor’s fallacy                           | 10.4 Resistance against objections                |
| 3.3 Base rate fallacy                                  | 10.5 Comprehensive evidence                       |
| 3.4 Defense attorney’s fallacy                         | 10.6 Bayesian network implementation              |
| 3.5 Uniqueness fallacy                                 | IV Standards of proof                             |
| 3.6 Case studies                                       | 11 Are standards of proof thresholds?             |
| 4 Complications and caveats                            | 11.1 Legal background                             |
| 4.1 Complex hypotheses, complex evidence               | 11.2 Probabilistic thresholds                     |
| 4.2 Source, activity and offense level hypotheses      | 11.3 Theoretical challenges                       |
| 4.3 Where do the numbers come from?                    | 11.4 The comparative strategy                     |
| 4.4 Modeling corroboration                             | 11.5 The likelihood strategy                      |
| 4.5 Stories, explanations and coherence                | 11.6 Four criteria                                |
| 5 Assessing evidential strength with likelihood ratios | 12 Naked statistical evidence                     |
| 5.1 Likelihood ratio is better than Bayes factor       | 12.1 Forty years of hypotheticals                 |
| 5.2 Match evidence and error probabilities             | 12.2 Specific narratives                          |
| 5.3 Eyewitness identification and likelihood ratio     | 12.3 Cross-examination and causal grounding       |
| 5.4 Hypothesis choice                                  | 12.4 Specificity, causality and Bayesian networks |
| 5.5 Relevance and the small-town murder                | 12.5 Are cold-hit DNA matches naked statistics?   |
| 5.6 Appendix: Confirmation measures                    | 13 The Difficulty with Conjunction                |
| III Evidence assessment (more and better)              | 13.1 Formulating the difficulty                   |
| 6 Bayesian Networks                                    | 13.2 The likelihood strategy                      |
| 6.1 Bayesian networks to the rescue                    | 13.3 Evidential Strength                          |
| 6.2 Legal evidence idioms                              | 13.4 The comparative strategy                     |
| 6.3 Scenario idioms                                    | 13.5 Rejecting the conjunction principle?         |
| 6.4 Modeling relevance                                 | 13.6 The proposal: specificity and unity          |
| 6.5 Case study: Sally Clark                            | 14 Other accounts                                 |
| 6.6 DNA evidence                                       | 14.1 Baconian probability                         |
| 7 Coherence  | 14.2 Sensitivity and safety                       |
| 7.1 Existing probabilistic coherence measures          | 14.3 Normic Support                               |
| 7.2 An array of counterexamples                        | 14.4 Foundherentism                               |

|      |   |      |  |
|------|---|------|--|
| 14.5 | Relevant alternatives                   | 16.3 | Predictive and diagnostic accuracy         |
| 14.6 | Relative plausibility                   | 16.4 | Calibration                                |
| 14.7 | Arguments                               | 16.5 | Accuracy, high probability and specificity |
| 14.8 | Knowledge                               | 17   | Fairness in trial decisions                |
| V    | Accuracy and Fairness                   | 17.1 | Procedural v. substantive fairness         |
| 15   | The objectives of trial decisions       | 17.2 | Absolute v. comparative fairness           |
| 15.1 | Accuracy and fairness                   | 17.3 | Distributing risks                         |
| 15.2 | Protecting defendants                   | 17.4 | Profile evidence and generalizations       |
| 15.3 | Public justification                    | 17.5 | Structural inequalities                    |
| 15.4 | Revision and appeal                     | 17.6 | A fairer standard of proof                 |
| 15.5 | Dispute resolution and public deference | 18   | Conclusions                                |
| 16   | Accuracy and the risk of error          | 18.1 | Assessing the progress made                |
| 16.1 | Single case accuracy                    | 18.2 | Open questions problems                    |
| 16.2 | Expected v. actual errors               | 18.3 | To reformers and practitioners             |

### 1.3 Outstanding Features of the Book

- ‘Probability on Trial’ is the first comprehensive philosophical examination of legal probabilism and how it fares against well-known objections. It is also conciliatory in admitting that many insights delivered by critics are valid and should be part of a more sophisticated version of legal probabilism.
- The book is interdisciplinary. It engages with the literature in philosophy (see, for example, the discussion about coherence and corroboration) as well as literature in artificial intelligence and forensic science (see, for example, the discussion of Vlek’s story node approach and Fenton’s averaging).
- The philosophical argument in defense of legal probabilism is accompanied by mathematical proofs and **R** code. We will also rely on Branden Fitelson’s **PrSAT** which offers a decision procedure for the probability calculus. This underscores the theoretical and computational aspiration of the project.
- ‘Probability on Trial’ is suitable for different audiences with different interests. Instead of reading the entire book, one could follow different tracks. One could read the book to learn about the proof paradoxes (Chapter 2, 11, 12 and 13), Bayesian networks for evidence assessment and decision-making (Chapters 6 through 11), legal probabilism and its difficulties (Chapter 1, 2, 3, 4 and 11), accuracy and fairness of trial decisions (Chapters 14, 16 and 17), etc. We will make sure to describe different tracks tailored to different interests that readers may have.
- The project is divided into different parts, each tackling a distinct domain of analysis. The first half (Part II and Part III) examines how the evidence presented at trial should be assessed, weighed and evaluated using the tools of probability theory. The second half examines decisions at trial. The book offers a probability-based explication of legal standards of proof that is theoretically plausible (Part IV). In addition, it provides a normative justification of the proposed theory based on how standards of proof would perform in terms of accuracy and fairness under competing definitions (Part V).

### 1.4 Apparatus

Besides a customary list of references, ‘Probability on Trial’ will contain multiple graphs and plots, Bayesian networks, and other data visualizations generated via the **R** package `ggplot2` as is standard in publications in statistics and machine learning.

Since computer simulations play an integral part in the argument, the book will be accompanied by supplementary materials available on-line. These materials will contain the **R** source code along with tutorials for readers interested in learning the technical details. Some of the derivations and results will be also supported by **Mathematica** notebooks, which will be freely available.

### 1.5 Competition

Several books in print cover topics at the intersection of evidence, law and probability. They, however, do not have the same aims as our book. They can be sensibly grouped by how our book differs from them.

- Consider first books by legal theories and philosophers, such as Stein (2005), *Foundations of Evidence Law*, Oxford University Press; Ho (2008), *A Philosophy of Evidence Law: Justice in the Search for Truth*, Oxford University Press; Haack (2014), *Evidence Matters: Science, Proof, and Truth in the Law*, Cambridge University Press; Nance (2016), *The Burdens of Proof: Discriminatory Power, Weight of Evidence, and Tenacity of Belief*, Cambridge University Press; Dahlman, Stein, and Tuzet (eds.) (2021), *Philosophical Foundations of Evidence Law*, Oxford University Press. These books offer

original theories of evidence law, the standards of proof and legal decision-making under uncertainty. They address many of the conceptual difficulties that we examine in our book. But they do not develop a probabilistic theory of evidence and decision-making that crucially relies on Bayesian networks. Nor do they combine analytic arguments, mathematical proofs and computer simulations.

- Many books by forensic scientists and computer scientists have championed applications of Bayesian networks to legal evidence evaluation, for example, Taroni, Aitken, Garbolino and Biedermann (2006) *Bayesian Networks and Probabilistic Inference in Forensic Science*, Wiley; Taroni, Bozza, Biedermann, Garbolino and Aitken (2010), *Data Analysis in Forensic Science: A Bayesian Decision Perspective*, Wiley; Fenton and Neil (2012/2018) *Risk Assessment and Decision Analysis with Bayesian Networks*, Chapman and Hall/CRC Press. These books examine how trial evidence can be assessed using Bayesian networks, with a focus on expert evidence. They do not address philosophical questions such as the problem of naked statistical evidence or the difficulty with conjunction. Nor do they aim to offer a theory of the standard of proof that makes connections to the story model, relative plausibility and argumentation theory. Finally, they do not address normative questions about the accuracy and fairness of trial decisions.
- Several books exist about Bayesian networks that are more general in content. Some of them focus on the technical dimension, such as Scutari and Denis (2014), *Bayesian Networks With Examples in R*, Chapman & Hall/CRC; Neapolitan (2004), *Learning Bayesian Networks*, Pearson. Some books use Bayesian networks to examine how our minds rely on evidence to understand the world and solve problems, such as Spirtes, Glymour and Scheines (2000), *Causation, Prediction and Search*, MIT Press; Glymour (2001), *The Mind's Arrows. Bayes Nets and Graphical Causal Models in Psychology*, MIT Press; Lagnado (2021), *Explaining the Evidence: How the Mind Investigates the World*, Cambridge University Press. Lagnado's book discusses legal cases extensively, but—like other books in this category—does not extensively discuss the standard of proof or normative questions about accuracy and fairness.
- There are books that focus exclusively on the evaluation of statistical and probabilistic evidence at trial. They include Robertson and Vignaux (1995), *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*, Wiley [second edition: Robertson, Vignaux and Berger, published in 2016]; Lucy (2005), *Introduction to Statistics for Forensic Scientists*, Wiley; Finkelstein (2010), *Statistics for Lawyers*, Springer; Balding and Steele (2015), *Weight-of-Evidence for Forensic DNA Profiles*, Wiley; Buckleton, Bright and Taylor (eds.) (2016) *Forensic DNA Evidence Interpretation*, CRC Press. These books differ from our own because they are more specific, often focusing on just DNA evidence or select forms of expert evidence. Besides DNA and expert evidence, our book also covers common forms of evidence, such as eyewitness testimony. In addition, these books are written from the perspective of forensic science and do not examine the philosophical issues that are in the focus in our book, and they do not develop a probabilistic model of story coherence.
- Finally, a few books examine the legal and ethical problems that arise from relying on statistical and actuarial evidence at trial and in the criminal justice system more generally. These books include Schauer (2006), *Profiles, Probabilities and Stereotypes*, Belknap Press; Harcourt (2008), *Against Prediction: Profiling, Policing, and Punishing in an Actuarial Age*, University of Chicago Press. Some chapters of our book discuss actuarial and statistical evidence, but this is not our sole focus. In addition, the authors of these books do not rely on legal probabilism as their guiding theoretical framework.

## 2 Market Considerations

The target audience comprises scholars in various disciplines, primarily philosophers and legal theorists with an interest in epistemology, evidence and decision-making under uncertainty. Computer scientists and psychologists who work on similar topics will also be interested in reading the book.

Selection of chapters from the book are suitable to be used as primary readings in advanced undergraduate courses or graduate seminars with titles such as Legal Probabilism, Legal Epistemology, Probability and the Law, Bayesian Epistemology, Bayesian Networks in Philosophy, Statistics in the Law, Math on Trial. We have ourselves taught similar courses in the past and felt the need of a book such as ours. We believe other instructors felt the same.



## 3 Status of the Work

### 3.1 Timetable

We plan to write 18 chapters. Of the first five chapters—corresponding to Part I and Part II—one chapter on likelihood ratios has already been written (submitted as a sample chapter). The other four will draw from preparatory materials for the entry ‘Legal Probabilism’ which we wrote for the *Stanford Encyclopedia of Philosophy*. These four chapters will take four months in total to complete.

Chapters 6 to 10 of Part III partly build on work that has already been published or near publication. Urbaniak has a paper under submission on coherence and Di Bello has published a piece on cross-examination. However, Part III will require substantive new work to weave together existing materials and formulate new lines of argument. We think that, on average, we will need two months per chapter, so ten months in total.

Chapters 11 to 14 of Part IV will be written at a faster pace. Chapter 13 on the problem with conjunction has already been written (submitted as a sample chapter). Chapter 12 on naked statistical evidence will be a development of a recent paper by Di Bello on the topic. Two other chapters remain, Chapter 11 on different probabilistic theories of the standard of proof and Chapter 14, a comparison between legal probabilism and other accounts in the philosophical literature. We will need six months to complete these four chapters.

Part V consists of four chapters that have not been written. We are currently carrying out the research and the computer simulations. Chapter 16 on accuracy and Chapter 17 on fairness are two of the most innovative. We predict we will spend roughly three months on each, while the other chapters will require two months of work each. So part V will require a total of ten months.

Given this schedule, we predict thirty months of work. We should add six months for proofreading, delays, revisions and running various versions of the chapters by our colleagues. So, our schedule suggests completion in about three years. We aim to complete the manuscript by the end of 2024 or the start of 2025.

### 3.2 Size

Existing chapters are about 20 page long, single-spaced, 14,000 words. We expect some chapters to be shorter, 15 page long, roughly 11,000 words. The entire book should run for about 200,000 words.

Some graphs will be needed. Early chapters won’t need many if any at all. The existing chapters have around 10 each. If this is the number of graph needed in four chapters, will need about 40 graphs.

### 3.3 Class and audience testing

We have already used some of the materials on which this book is based in graduate seminars, for example, in the seminar ‘Bayesian Networks in Philosophy’ that Di Bello taught at Arizona State University in Spring 2021. We plan to use select chapters to teach graduate seminars and advanced undergraduate courses. By the time the book is published, we hope to have student-tested a significant portion of the book (in compliance with copyright restrictions as needed).

To better reach our audience, we plan to present portions of the book at international conferences in the coming years. We were invited speakers at an international conference on Evidence Law in Girona, Spain in Spring 2022. This conference gathers some of the leading scholars in law, philosophy, forensic science and psychology interested in topics at the intersection of law, evidence, probability and decision-making. We have also been invited to other international conferences in legal philosophy in Fall 2022 and Spring 2023.

We do not plan to include any material requiring additional copyright permissions.

## 4 Bibliography

For bibliographic references, please see the list at the end of the entry ‘Legal Probabilism’ which we have recently written for the *Stanford Encyclopedia of Philosophy*.

## 5 Additional documents

Two sample chapters and a technical appendix are attached to this submission, as well as the authors’ CVs.