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 systematic philosophical defense of legal probabilism, a research program whose aim is to harness the powers of probability to analyze, model and improve the evaluation of evidence and the process of decision-making in trial proceedings. The book is informed by three questions. (1) Can the evidence presented at trial be weighed using probability theory? (2) Can legal decision-making and standards of proof such as 'proof beyond a reasonable doubt' be defined using the language of probability? (3) Does the deployment of probability theory in assessing evidence improve the accuracy and fairness of trial decisions? We argue that the answer to these questions is, by and large, affirmative.

Over the last fifty years, these questions have been debated in the literature in philosophy, law, forensic science and artificial intelligence. Despite considerable progress, 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. Trials are not gambles. Two rival theories hold prominence. The first is the "story model" (or its close cousin, "relative plausibility"). According to this theory, judges and jurors make sense of the evidence presented at trial by constructing a story or account of what happened and how the evidence got there. The story that best explains the evidence should prevail. Argumentation theory is another rival. It emphasizes the fact that trials are adversarial. The arguments by one side are scrutinized, attacked, challenged by the other side. The side with the strongest argument should prevail. Both the story model and argumentation theory provide important insights. 'Probability on Trial' has a conciliatory aim. By taking advantage of recent developments in artificial intelligence, in particular Bayesian networks, it shows that probability theory can factor in the insights from its rival theories and make them more rigorous and precise. Hence, the subtitle reads 'Making Sense of Arguments and Stories.'

This project is an exercise in philosophical imagination. What would trial proceedings look like if probability played a more prominent role? Would they be more accurate? Would they be more fair? We think they would be, and we rely on a number of computer simulations written in **R** to test this hypothesis. But even though the argument is to a large extent revisionist of current trial practice, it also pays close attention to it. 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 racial and socio-economic 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, the most well-developed framework to tame uncertainty. 'Probability on Trial' examines whether, and if so how, probability can indeed be the right guide for trial proceedings.

We wish to highlight the following contributions of the book:

- A probabilistic account, based on Bayesian networks, 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 response to standard philosophical objections leveled against legal probabilism, particularly, the paradoxes of naked statistical evidence and the difficulty with conjunction;
- A theory of the standard of proof in four parts: high probability, coherence, resilience, completeness;
- A simulation-based argument that tests the proposed theory in light of accuracy and fairness criteria.

'Probability on Trial' is aimed at philosophers with an interest in legal epistemology and epistemology more generally. 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

The book 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 has proven 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 the book 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 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 or explanations constituted by several sub-propositions. This level of 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 Bayes' theorem which many legal probabilists have found useful: the likelihood ratio. Bayes' theorem requires one to assess the prior probabilities of the hypothesis of interest. But assessing prior probabilities is notoriously difficult. The likelihood ratio, instead, offers a way to evaluate evidence without prior probabilities. 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 likelihood ratios face difficulties in modeling the notion of legally relevance in certain cases.

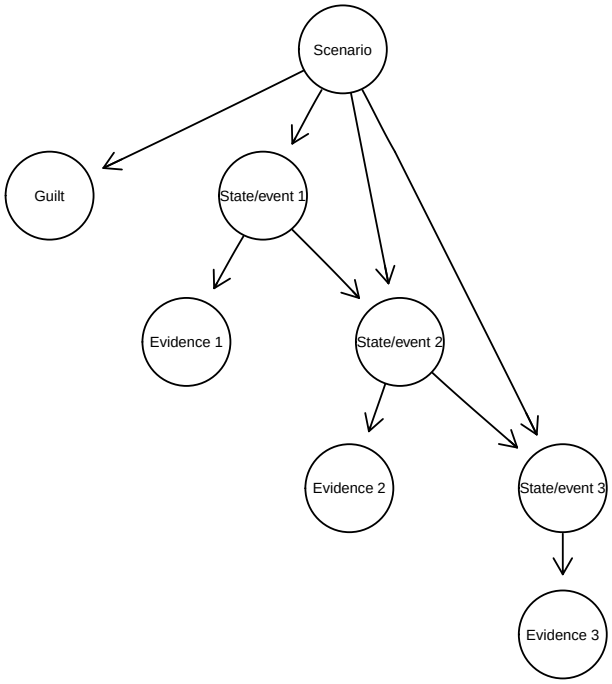
Chapters 3 and 4 show that we need to move beyond simple legal probabilism. Chapter 5 shows that likelihood ratios, while useful in many ways, are still an unsatisfactory approach overall. The journey toward a more sophisticated legal probabilism is accomplished in Part III, Chapters 6 through 10.

Part III - Evidence Assessment More and Better

The third part of the book is motivated 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. We show that 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 between evidence nodes and hypothesis nodes. Unanticipated possibilities and alternative scenarios that may arise in the adversarial process can be modeled by comparisons across multiple Bayesian networks.

The reverse of the phenomenon of conflicting evidence is that of converging evidence, 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 proposed probabilistic accounts. The chapter then formalizes a notion of corroboration using Bayesian networks that overcomes most of the difficulties of existing accounts.

Chapter 10 draws some general 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 emphasis that legal probabilism 1.02 outperforms, in several respects, its main two rivals: argumentation theory and relative plausibility. Argumentation theory 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. Unlike argumentation theory, 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 competing stories (in agreement with relative plausibility) or 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 thresholds. Trial decisions are discussed more extensively in the next part of the book.

Part IV - Decision-making

The third part of the book 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. 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. We show that the version formulated in the previous chapters of the book provides a more adequate framework.

Chapter 11 examines different strategies for theorizing about standards of proof using probabilistic language. We begin with the most natural decision criterion, a probability threhsold whose stringency is determined by expected utility maximization. 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). We then turn to alternative accounts, in particular, the comparative strategy (Cheng) and the likelihood ratio strategy (Dawid, Kaplow, Sullivan). Finally, we present our own proposal. 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. We emphasize that these criteria—probability, specificity, etc.—can be modeled using Bayesian networks. So our proposal lies within the confines of legal probabilism, though not the narrow version its critics have in mind. In the following chapters, we illustrate the theoretical payoffs that come from endorsing our proposal.

Chapter 12 tackles the puzzles of naked statistical evidence. This is a topic of enormous scholarly attention in the recent philosophical literature. Our solution 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. In cases of naked statistical evidence, the probability of liability is high, but the specificity of the accompanying narrative is suspiciously low. The second premise is that the supporting evidence should typically 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 credential of an expert witness; the chain of custody of a document.

Chapter 13 tackles another central problem for legal probabilism, the difficulty with conjunction. We first provide a detailed argument for why previous attempts in the literature on legal probabilism have failed, focusing on the likelihood ratio approach (Sullivan, Dawid and others), and the comparative strategy (Cheng). Our proposal follows the holistic approach by Allen and Pardo. We show how 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 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) and to some extent relative plausibility (Allen and Pardo). The second criticism we make is that other accounts are engaged with what we might call 'epistemology fetishism'—that is, they borrow ideas in contemporary analytic epistemology and force them onto legal-decision making. This criticism applies in particular to knowledge accounts of legal proof. To be sure, we might ourselves be accused of 'probability fetishism'—that is, unquestionably taking probability theory as a paradigm of rationality and force it onto legal-decision making. Admittedly, there is no empirical evidence that a probabilistic turn in legal decision-making would improve trial decisions, if only because what it would mean to 'improve' trial decisions is unclear. A set of criteria are needed for assessing the desired improvements. We tackle this normative question in the final part of the book and show 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. We first distinguish accuracy in the single instance and accuracy in the long run. Accuracy in the single instance consists in the conviction of a defendant that is factually guilty and the acquittal of a defendant that is factually innocent. A similar definition applies to civil liability. Accuracy in the long run, instead, can be understood as predictive or diagnostic. The former is the probability that, if the defendant is found (not) liable at trial, the defendant is actually (not) liable; the latter is the probability that if the defendant is actually (not) liable, the defendant is (not) be found liable at trial. The two notions are related but not equivalent. Distinguishing them has implications for how we should understand standards of proof. 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 narrative 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 risk of being mis-calibrated more often than in the second model, and when mis-calibration occurs, predictive accuracy will also suffer. Thus, from the point of view of accuracy, understanding the standard of proof as the combination of 'high probability' and 'narrative specificity' (as we have argued in earlier chapters) is a more rational model of decision-making.

Chapter 17 turns to the fairness of trial decisions. What does it means for trial decisions to be fair? In a formalistic sense, fairness merely requires that every rule be applied to all defendants in the same way.

In a substantive sense, it requires, that trial defendants not be subject to excessive burdens, and also, that burdens (and benefits) be roughly equally distributed across different defendants. Call the latter substantive comparative fairness. It can be understood as the requirement that all defendants be exposed to roughly the same risks of mistaken conviction. Equipped with this notion, 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 conceptual, analytic role. It is needed to define what we mean by accuracy and fairness, and then test competing models of the standard of proof against these normative criteria.

Chapter 18 draws some conclusions and points to open problems for legal probabilism. We emphasize that our aim in the book is to show that legal probabilism is a richer and more flexible theory that 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 criminal 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
- 1 The emergence of legal probabilism
  - 1.1 Famous cases
  - 1.2 Probabilistic evidence
  - 1.3 Trial by mathematics
  - 1.4 Some history
- 2 A skeptical perspective
  - 2.1 The difficulty about conjunction
  - 2.2 The problem of priors
  - 2.3 Naked statistical evidence
  - 2.4 Complexity
  - 2.5 Soft variables
  - 2.6 Corroboration
  - 2.7 The reference class problem
  - 2.8 Unanticipated possibilities
  - 2.9 Non-probabilistic theories

II Evidence assessment (first pass)
- 3 Bayes' theorem and the usual fallacies
  - 3.1 Assuming independence
  - 3.2 The prosecutor's fallacy
  - 3.3 Base rate fallacy
  - 3.4 Defense attorney's fallacy
  - 3.5 Uniqueness fallacy
  - 3.6 Case studies
- 4 Complications and caveats
  - 4.1 Complex hypotheses, complex evidence
  - 4.2 Source, activity and offense level hypotheses
  - 4.3 Where do the numbers come from?
  - 4.4 Modeling corroboration
  - 4.5 Stories, explanations and coherence
- 5 Likelihood ratios and relevance
  - 5.1 Likelihood ratio is better than Bayes factor
  - 5.2 Match evidence and error probabilities
  - 5.3 Eyewitness identification and likelihood ratio
  - 5.4 Hypothesis choice
  - 5.5 Relevance and the small-town murder
  - 5.6 Appendix: Confirmation measures

III Evidence assessment (more and better)
- 6 Bayesian Networks
  - 6.1 Bayesian networks to the rescue
  - 6.2 Legal evidence idioms
  - 6.3 Scenario idioms
  - 6.4 Modeling relevance
  - 6.5 Case study: Sally Clark
  - 6.6 DNA evidence
- 7 Coherence
  - 7.1 Existing probabilistic coherence measures
  - 7.2 An array of counterexamples
  - 7.3 Structured coherence with Bayesian networks
  - 7.4 Specificity
  - 7.5 Completeness
  - 7.6 Application to legal cases
- 8 Conflicts
  - 8.1 Argumentation theory
  - 8.2 Undercutting and rebutting
  - 8.3 Cross-examination in Bayesian networks
  - 8.4 Unanticipated possibilities
  - 8.5 Refining and comparing networks

9 Corroboration
9.1 Boole's formula and Cohen's challenge
9.2 Rise in case of agreement
9.3 Ekelöf's corroboration measure
9.4 Multiple false stories and multiple witnesses
10 Towards legal probabilism 1.02
10.1 Outperforming competing accounts
10.2 Empirical adequacy
10.3 Specificity and coherence
10.4 Resistance against objections
10.5 Comprehensive evidence
10.6 Bayesian network implementation
IV Standards of proof
11 Are standards of proof thresholds?
11.1 Legal background
11.2 Probabilistic thresholds
11.3 Theoretical challenges
11.4 The comparative strategy
11.5 The likelihood strategy
11.6 Four criteria
12 Naked statistical evidence
12.1 Forty years of hypothetical
12.2 Specific narratives
12.3 Cross-examination and causal grounding
12.4 Specificity, causality and Bayesian networks
12.5 Are cold-hit DNA matches naked statistics?
13 The Difficulty with Conjunction
13.1 Formulating the difficulty
13.2 The likelihood strategy
13.3 Evidential Strength
13.4 The comparative strategy
13.5 Rejecting the conjunction principle?
13.6 The proposal: specificity and unity
14 Other accounts
14.1 Baconian probability
14.2 Sensitivity abnd safety
14.3 Normic Support
14.4 Foundherentism
14.5 Relevant alternatives
14.6 Relative plausibility
14.7 Arguments
14.8 Knowledge
V Accuracy and Fairness
15 The objectives of trial decisions
15.1 Accuracy and fairness
15.2 Protecting defendants
15.3 Public justification
15.4 Revision and appeal
15.5 Dispute resolution and public deference
16 Accuracy and the risk of error
16.1 Single case accuracy
16.2 Expected v. actual errors
16.3 Predictive and diagnostic accuracy
16.4 Calibration
16.5 Accuracy, high probability and specificity
17 Fairness in trial decisions
17.1 Procedural v. substantive fairness
17.2 Absolute v. comparative fairness
17.3 Distributing risks
17.4 Profile evidence and generalizations
17.5 Structural inequalities
17.6 A fairer standard of proof
18 Conclusions
18.1 Assessing the progress made
18.2 Open questions problems
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 the 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. 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 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 the customary list of references, the book 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 for the simulations 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, and since they can be sensibly grouped with respect to how our book will differ from them, we'll discuss them in batches accordingly.

double check if all references have publishers

- Consider first the key books written 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, from a philosophical perspective, 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. They also focus on rather abstract examples and their use of real-life cases and discussions between the practitioners is rather limited.
- 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 are groundbreaking in that they offer a detailed examination of how expert evidence can be assessed using Bayesian networks. They, however, do not address philosophical problems such as the problem of naked statistical evidence or the difficulty of conjunction, nor do they aim to offer a novel theory of the standard of proof, and they make no connection to the narrative approach or evidence evaluation criteria brought up by the critics of legal probabilism. Unlike our book, they do not address normative questions such as the accuracy or fairness of trial decisions.
- Several books exists 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*, (Neapolitan 2004), *Learning Bayesian Networks* or Fenton, Neli and Martin (2013), *Risk Assessment and Decision Analysis with Bayesian Networks*. Some are more psychological and 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*, Glymour (2001), *The mind's arrows. Bayes Nets and Graphical Causal Models in Psychology*, or Lagnado (2021), *Explaining the Evidence: How the Mond Investigates the World*, Cambridge University Press. These general books lack focus on legal applications (some legal applications are discussed in Fenton, but the coverage is fairly narrow as compared to our intent—the authors focus on covering on presenting a few technical representation methods of various types of evidence).

More needs to be said about Lagnado, as he does focus on legal evidence

- 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 general approach to the probabilistic modelling of narrations.
- Finally, a few books focus on 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

2.1 The Primary Market

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

We plan 16 chapters, two of which already exist, so 14 remain. We predict each chapter will take around 2 months to complete, which results in 28 months of work. We add eight months for proofreading, delays, revisions and running various versions of the chapters by our colleagues, so overall the fairly lenient schedule suggests completion after three years.

With appendices the existing chapters have around 32k words each. We expect the early chapters chapters (including appendices) to be shorter, and the later chapters to be of approximately that length, the whole book being around 350k words.

It is not clear how many graphs will be needed. Early chapters won't need many, the existing chapters have around 10-15 each (including appendices). Supposing this is the number for eight chapters, we seem to be looking at 80-100 graphs and visualizations.

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 the Philosophy that one of us taught at Arizona State 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 intended audience, we plan to present portions of the book at international conferences in the coming years. We have been invited to an international conference in Girona, Spain on Evidence Law 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 do not plan to include any material requiring additional copywright permissions.

4 Sample Chapters

Two sample chapters with their appendices are attached to this submission.

5 Reviews

- Floris Bex
- Alex Biedermann
- Edward Cheng
- Christian Dahlman
- Norman Fenton
- David Lagnado
- Anne Hsu

- Brian Hedden
- Silja Renooij
- Frederick Schauer
- Silvia Bozza
- Bart Verheij

6 Author Background

Two CVs are attached to this submission.