TOPICS
TOPICS IN COGNITIVE SCIENCE

Topics in Cognitive Science 12 (2020) 1053–1067

ISSN:1756-8765 online
DOI: 10.1111/tops.12519

This article is part of the topic "Models of Rational Proof in Criminal Law," Henry Prakken, Floris Bex, and Anne Ruth Mackor (Topic Editors). For a full listing of topic papers, see http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1756-8765/earlyview

Editors' Review and Introduction: Models of Rational Proof in Criminal Law

Henry Prakken,[a,b] Floris Bex,[a,c] Anne Ruth Mackor[b]

[a]*Department of Information and Computing Sciences, Utrecht University*
[b]*Faculty of Law, University of Groningen*
[c]*Tilburg Institute for Law, Technology and Society, University of Tilburg*

Received 5 December 2019; received in revised form 3 September 2020; accepted 3 September 2020

Abstract

Decisions concerning proof of facts in criminal law must be rational because of what is at stake, but the decision-making process must also be cognitively feasible because of cognitive limitations, and it must obey the relevant legal–procedural constraints. In this topic three approaches to rational reasoning about evidence in criminal law are compared in light of these demands: arguments, probabilities, and scenarios. This is done in six case studies in which different authors analyze a manslaughter case from different theoretical perspectives, plus four commentaries on these case studies. The aim of this topic is to obtain more insight into how the different approaches can be applied in a legal context. This will advance the discussion on rational reasoning about evidence in law and will contribute more widely to cognitive science on a number of topics, including the value of probabilistic accounts of cognition and the problem of dealing with cognitive biases in reasoning under uncertainty in practical contexts.

Keywords: Rational proof; Criminal law; Arguments; Scenarios; Probabilities; Case study

Correspondence should be sent to Henry Prakken, Department of Information and Computing Sciences, Utrecht University, P.O. Box 80089, 3508 TB Utrecht, The Netherlands. E-mail: h.prakken@uu.nl

1. Motivation

In the context of criminal law, decisions concerning legal proof of facts must be rational because of what is at stake, but the decision-making process must also be cognitively feasible because of the decision-makers' cognitive limitations, and it must obey the relevant legal–procedural constraints. The aim of this topic is to compare three approaches focusing, respectively, on arguments, probabilities, and scenarios, in light of these demands. All three approaches acknowledge that evidence cannot provide watertight support for a factual claim, as there is always room for doubt. Probabilistic approaches account for this by applying Bayesian probability theory, argumentative approaches emphasize the comparison of supporting and attacking arguments, and scenario approaches advocate the comparison of alternative scenarios about what may have happened.

In this topic (combinations of) the three approaches are compared in six case studies in which different authors analyze a manslaughter case from different theoretical perspectives. In addition, the topic contains four comments on specific case studies. The aim of centering the topic on the study of a particular case is to obtain more insight into how the different approaches can be applied in a legal context. Thus we hope to advance the discussion on theories of rational reasoning about evidence in legal cases. In addition, the topic aims to more widely contribute to cognitive science on a number of topics. It relates to the recent interest in probabilistic models as alternatives to deductive models of cognition (Chater, Tenenbaum, & Yuille, 2006) and more specifically to attempts to develop a Bayesian theory of argumentation (Hahn & Hornikx, 2016; Hahn & Oaksford, 2007; Harris, Hahn, Madsen, & Hsu, 2016). It also relates to the literature on cognitive biases in reasoning (Tversky & Kahneman, 1974), since it is known that reasoning under uncertainty is especially prone to such biases. The issue of the cognitive feasibility of the various approaches relates to the study of "smart heuristics" for decision-making (Gigerenzer & Todd, 1999). Among other things, the question arises whether argumentation or scenario approaches are cognitively feasible alternatives to strict Bayesian thinking.

2. Theories of rational criminal proof

The study of principles of rational legal proof arguably started with the ground-breaking work of the American legal scholar John Henry Wigmore (Wigmore, 1931). While his contemporaries exclusively focused on doctrinal analysis of evidence law, Wigmore wanted to develop a "science" of legal proof. His "scientific" approach was further developed since the 1970s by several Anglo-American legal scholars called the new evidence scholars (Lempert, 1986), using a range of methods from various disciplines. In the 1980s and 1990s, psychologists became interested in the topic and the increasing importance of forensic evidence (e.g., DNA and fingerprint databases) made statisticians and forensic scientists more interested, advocating mainly probabilistic approaches. Finally, since 2000, researchers from artificial intelligence and law have applied various AI models to legal proof.

The various approaches to legal proof thus each have their roots in a particular discipline—artificial intelligence (AI), mathematics, philosophy, psychology, legal theory—and hence the literature on the topic is very diverse and interdisciplinary, touching on a wide range of subjects such as the relations between different approaches, their philosophical or mathematical foundations, their cognitive feasibility, and their legal plausibility. Three main approaches can be distinguished: arguments, probabilities, and scenarios (Verheij et al., 2016). All three approaches acknowledge that evidence cannot provide watertight support for a factual claim, as there is always room for doubt.

Probabilistic approaches account for this by applying probability theory (Dawid, 2005). Probabilistic approaches have been central to much of the research on proof, particularly when the probabilities can be determined based on "hard" data (e.g., DNA and fingerprint databases). Often a Bayesian approach is taken (Fenton & Berger, 2016). In its simplest form, so-called Bayes' Rule is applied, according to which the conditional probability of the evidence given the hypotheses (the likelihood ratio) and the prior probability of the hypotheses (i.e., their probability before taking any evidence into account) are used to compute the posterior probability of the hypotheses given the available evidence. A limitation of this approach is that it is only feasible if the various pieces of evidence can be assumed to be statistically independent, an assumption which in practice is often not warranted. Accordingly, there is increasing use of Bayesian networks (de Zoete, Sjerps, Lagnado, & Fenton, 2015; Fenton, Lagnado, & Neil, 2013; Fenton & Neil, 2011; Hepler, Dawid, & Leucari, 2007; Vlek, Prakken, Renooij, & Verheij, 2014), which overcome this limitation by using graph theory to graphically represent statistical independence relations. Bayesian networks are state of the art in AI, and many software packages exist for creating Bayesian networks and computing with them.

However, probabilistic approaches are by no means uncontroversial. A theoretical objection, stemming from the field of epistemology, is that criminal proof is about justified belief, or warrant, and that probabilities are not the same as degrees of (justified) belief (Cohen, 1977; Pollock, 2007b). For example, in probability theory, it is hard to distinguish between ignorance and uncertainty: Assigning a probability of 50% to a statement says much more than "I don't know." Moreover, it has been argued that the "weight" of evidence (in that the probability of a hypothesis given a large body of evidence is better justified than the same probability given just a small subset of the evidence) cannot easily be represented in a probabilistic way. Alternative numerical accounts of degrees of (justified) belief have been proposed, such as the theory of belief functions (Shafer, 1976) and possibility theory (Dubois & Prade, 1988). However, these theories have never become popular in the study of criminal proof, possibly since they share with probability theory some other problems. A psychological objection to probability theory is that for nonspecialists Bayesian thinking is very difficult and counterintuitive (Tversky & Kahneman, 1974). Thus, probability theory seems to impose a standard of rationality that cannot be attained in practice, so that their application would lead to more, instead of fewer, errors (Lempert, 1986). Related psychological concerns are that using probability theory may "dwarf" softer types of evidence in that only easily quantifiable forms of evidence will be considered, or that the mathematical form of a testimony

gives the misleading impression of objectivity, or that the presentation of small numbers like random-match probabilities can easily mislead judges or jurors (Tribe, 1971). Finally, a practical objection against probability theory (and numerical alternatives) is that in legal cases the required numbers are usually not available, either because there are no data or because experts or judges are unable or reluctant to provide estimates of probabilities. This leads to the practical problem of how fact finders can make probability estimates in a reliable way. Below we will see that the alternative approaches strongly rely on commonsense knowledge in the form of generalizations. However, it is not straightforward to turn such qualitative generalizations into quantitative conditional probabilities.

A second approach takes *arguments*, more specifically, series of inferences from evidence to conclusions, as the main concept. This approach goes back to Wigmore's charting method for legal proof (Wigmore, 1931). The idea of this method is that making the various inferences in an argument explicit allows one to identify sources of doubt in these arguments. Wigmore's charting method was modernized in the 1980s and 1990s by the new evidence scholars (Anderson & Twining, 1991; Kadane & Schum, 1996; Schum & Tillers, 1991), who, among other things, paid special attention to the dual nature of generalizations: They are the glue that hold evidential arguments together, but they are also important sources of doubt in these arguments, since they can have exceptions or be based on subjective judgment or even prejudice (Twining, 1999). Therefore, comparing conflicting arguments is as important as constructing arguments. The argument-based approach has been formalized in terms of AI frameworks for argumentation (Bex, Prakken, Reed, & Walton, 2003), which include arguments and counterarguments in a logical model of argument-based attack and defeat. The basic ideas behind the argumentation approach are widespread and widely accepted, and they have been used in diverse areas such as communication theory, logic, philosophy, and law (van Eemeren et al., 2014). Furthermore, lawyers and judges are familiar with the idea of argumentation. Thus, argumentation seems to be a more natural approach to rational proof than probability theory. However, a problem is that argumentative reasoning does not explicitly allow for the construction and comparison of alternative scenarios as a way of maintaining overview of a mass of evidence.

Precisely this is a strong point of the third approach to evidence and proof: the *scenario* approach (sometimes also called the story-based or narrative approach), which states that fact finders should construct and compare alternative explanations of the evidence. Empirical research (Pennington & Hastie, 1993) suggests that when dealing with a mass of evidence, it is much easier for people to construct scenarios that explain the evidence instead of building the arguments based on evidence as expressed by, for example, Wigmore charts. Wagenaar, van Koppen, and Crombag (1993) gave a normative twist to Pennington and Hastie's story model in their "anchored narratives" theory, based on the premise that fact finding in criminal procedures should take the form of constructing and comparing alternative scenarios about what may have happened. The scenario that best explain the evidence should, if it does so to a sufficient degree, be accepted as true. In their theory, generalizations have an equally central role as in argumentation approaches in that scenarios should be anchored in common sense generalizations, which should be critically examined on their acceptability. This, together with the requirement to compare

alternative scenarios on their explanatory power, should mitigate the psychological danger that plausible and persuasive but false scenarios are accepted.

The scenario approach is arguably the most cognitively feasible in that it best describes how people actually reason when confronted with a lot of evidence. However, the normative and formal aspects of the scenario approach, both of which should be clearly defined in any rational theory of proof, are the least developed of all three main approaches. One link with normative theories of evidential reasoning is through the requirement to test which scenario best explains the evidence. In this sense, scenario approaches are closer to probabilistic than to argumentation-based approaches, since this comparison is similar to the likelihood ratio of Bayes' rule. Alternatively, scenario approaches can be seen as applications of philosophical theories of inference to the best (Lipton, 2004). Several AI researchers have formalized aspects of the scenario approach with logical AI formalisms for abduction, which essentially is inference to the best explanation (Bex, 2011; Josephson, 2001; Poole, 2001). An explanation-based approach to legal proof has also been advocated by the American legal scholars Pardo and Allen (2008), although not with a focus on scenarios. Finally, Thagard (2004) has modeled the scenario approach with his connectionist computational theory of explanatory coherence, claiming that it avoids the number problem that is inherent in using Bayesian networks (see also Amaya, 2009). An interesting question is whether these various approaches are pragmatic approximations of Bayesian probability theory or whether they model something intrinsically different than probability (Lipton, 2004).

The last decade has seen quite a few attempts to compare and integrate the different approaches. Bex (2011) has combined a scenario and argumentative approach in a formal "hybrid theory." The connection between ideas from argumentation and Bayesian networks has been studied by, among others, Kadane and Schum (1996). The connection between scenarios and Bayesian networks has also received attention recently (Vlek, Prakken, Renooij, & Verheij, 2016). Moreover, van Koppen (2011) has proposed a rational foundation of his scenario approach in terms of Popper's philosophy of science. Finally, several studies address the issue of comprehensibility of Bayesian networks. Work on using so-called idioms (Fenton et al., 2013) or templates for crime scenarios (Vlek et al., 2014) aims to support the construction of Bayesian networks, while Timmer, Meyer, Prakken, Renooij, and Verheij (2017) study the extraction of arguments from Bayesian networks in order to explain these networks to judges or prosecutors.

3. Debating models of rational proof—The Simonshaven case study

In 2015, we organized a 3-day workshop at the Zentrum für Interdisciplinäre Forschung (Centre for Interdisciplinary Research) in Bielefeld, Germany, and invited legal scholars and practitioners, computer scientists, philosophers, and cognitive scientists. Follow-up workshops were organized at the Isaac Newton Institute for Mathematics in Cambridge, England, in 2016 and at Queen Mary University London, England, in 2017. In Bielefeld, the idea arose to model a real legal case using the various approaches, in order to compare

their strong and weak points. For this purpose, a Dutch manslaughter case (the "Simonshaven" case) was chosen. The various analyses were made on the basis of English translations of the Court ruling in the first instance, the prosecutor's and defense's arguments on appeal and the court of appeal's decision (available as Appendix S1 to this topic). During the workshop in Cambridge, initial draft analyses of the case were discussed and new versions were discussed at the workshop in London. This topic presents the final versions of the case analyses.

The Simonshaven case concerns the death of a woman called Jenny Lourens,[1] who was violently killed in a recreational area near the village of Simonshaven, close to Rotterdam, while walking there with her husband, Ed Lourens. The couple were known to have had marital problems and were living apart at the time of the events. The trial court convicted Ed Lourens of manslaughter on his wife by intentionally hitting and/or kicking her head and strangling her. It did so on the basis of circumstantial evidence and several witness testimonies about Ed's earlier verbal and nonverbal aggressive behavior and his relationship with Jenny. Ed maintained his innocence throughout the case and claimed that someone had suddenly jumped out of the bushes and had knocked him unconscious; when he woke up, he found his wife lying next to him, dead. After having made this statement as a witness, Ed became a suspect in this case and from then on he appealed to his right to remain silent. Some time after the conviction by the lower court, new evidence came up that another man had attacked and killed three other women in the same area. A map of the area with several crosses near the murder scenes was found at the killer's home and at some of these locations a pit was found. An accomplice stated that the killer used these pits as hiding places. New investigations were carried out, but no evidence linking the killer to Jenny's death was found. Ed appealed against his conviction but was convicted again.

A few things are noteworthy about the Dutch criminal-law system. First, as a civil law system, it is inquisitorial in nature, which means that courts actively investigate the facts of the case. In this respect, civil law systems differ from most common law systems, which are adversarial in nature. Another difference with most common law systems is that Dutch criminal law does not have a jury system; cases are wholly decided by professional judges. Finally, again unlike many common law systems, Dutch criminal law has very few rules of evidence. Therefore, courts are relatively free in their decisions about the relevance and weight of evidence and about the proof of facts.

We next discuss the nature of the case studies and this topic. Anderson and Twining (1991, pp. 120–121) argue that anyone who analyzes a legal case on its evidential aspects must declare his or her standpoint by answering three questions: Who am I? In what stage of what process am I? What am I trying to do? We will answer these questions for all authors of the case studies.

3.1. Who are we?

We are all academics interested in theories of how to rationally prove facts in a criminal court case. We have had academic training in law (Dahlman, Van Koppen, Mackor, and Prakken), philosophy (Dahlman, Mackor, and Prakken), psychology (Van Koppen and

Lagnado), mathematics (Fenton, Neil, and Verheij), computer science (Yet), or AI (Bex). Some of us have experience as an expert witness in court (Dahlman, Fenton, Van Koppen, and Prakken) or in assisting parties involved in a court case (Dahlman and Fenton). Several of us have given tutorials to judges on rational criminal proof (Dahlman, Fenton, Van Koppen, Mackor, and Prakken). One of us has been a jury member (Fenton). None of the other authors has ever been a crime investigator or a fact finder in court procedures.

3.2. In what stage of the process are we?

We are at a stage at which the formal court proceedings have closed. In carrying out our case studies, we were not under any real time or institutional pressure of the kind experienced by crime investigators or fact finders in criminal-legal procedures. We hypothetically put ourselves in the position of the fact finders at the trial and offer an analysis of their verdict and the evidence from different theoretical perspectives.

3.3. What are we trying to do?

We are analyzing the quality of the reasoning of the court of appeal in the Simonshaven case when justifying its decision to convict. As such, we are not investigating how the court should think and reason but how it should *justify* its decision. We do this in terms of a particular theory of rational proof. Some of these theories are formal in that they use mathematical tools; other theories are informal. Most of us apply a theory which we believe to be appropriate for its purpose. An exception is Henry Prakken's argumentation-based analysis, which was made to have an analysis of this kind for the purpose of comparison. The main aim of our case studies is not to find out the truth about the case. This would require access to the complete case files, which we did not have. Moreover, it would require expertise about various aspects of the case (such as ballistic evidence, DNA evidence, and reliability of eye witnesses), which we (with some exceptions on particular points) do not have. Accordingly, our selections of scenarios or generalizations, and our probability or preference estimates have no objective status, while the same holds for the conclusions of our case studies about the question whether the court's conviction of the suspect was rationally justified. Instead, the main aim of the case studies is to obtain insight into the theoretical and practical strengths and weaknesses of our respective methods when applied as a reasoning model in criminal court cases. Since we are post hoc analyzing the quality of the court's reasoning, our insights are primarily on the merits of the methods for *analyzing* evidential reasoning and at best only indirectly on the merits of the methods for *performing* evidential reasoning as a fact finder.

The authors of the case studies were asked to explicitly answer the following questions:

- To what extent is the analysis objective and to what extent is it based on subjective beliefs, assumptions, and choices?
- How natural is the analysis from a cognitive and legal point of view?

- Did your analysis identify errors or biases in the reasoning of the judge, prosecutor, or defense?
- Does your analysis respect the legal constraints, such as the burden and standard of proof and the right to remain silent?

Of course, the case studies do not provide final answers to these questions, not in the least given their just-mentioned limitations. Nevertheless, we hope that our topic will significantly advance the ongoing debate, since so far there have been no systematic comparisons of the three approaches in an actual substantial case.

4. The papers and commentaries in this topic

Henry Prakken gives an argumentation-based analysis of the case. He uses formal tools developed in AI, namely, the *ASPIC*$^+$ framework for specifying arguments and their relations, and a version of the theory of abstract argumentation frameworks for evaluating the arguments. The analysis is graphically displayed in a way that is similar to Wigmore charts. Prakken regards evidential argumentation as the construction and attack of "trees of inference" from evidence to conclusions by applying generalizations to evidence or intermediate conclusions. The outcome of his analysis is not a conclusion on whether the accused was guilty but a list of relevant issues on which a (possibly reasoned) choice for one of two conflicting arguments influences such a conclusion. He argues that the main benefits of the argumentation approach are that it clearly shows how evidence and hypotheses relate and what are the points of disagreement in a case. He acknowledges that weak points are the lack of overview over a case and the lack of a systematic account of degrees of uncertainty. Finally, he claims that his analysis is not far from how the court justified its decision but that he filled in many reasoning steps that the court left implicit, and in these steps the court may have reasoned in a scenario-based or even Bayesian way but left this implicit in its justification.

Norman Fenton et al. present a detailed Bayesian network analysis of the case, using their previously developed set of building blocks ("idioms") for legal applications of Bayesian networks. They claim that the relatively short time it took the authors to build their analysis indicates that their idioms approach can facilitate the construction of Bayesian networks, which are often claimed to be hard to build. Their network yields a posterior probability of guilt of 74%. The authors acknowledge that this outcome is relative to the subjective nature of their judgments but they claim that their sensitivity analysis (which investigates the effect of variations in their probability estimates) increases the confidence in this outcome, since it shows that their model is reasonably robust to variations and that the posterior probability of guilt generally remains below the 95% which they claim to be the threshold for conviction. The authors also claim that the use of their idioms approach further reduces the subjectivity of their design decisions, as does their new opportunity-based method for estimating the prior probability of guilt (which is often claimed to be a hard problem). The authors also acknowledge some weaknesses of their

model, such as that they left some evidence out. They further acknowledge that their Bayesian network model may not be as easily accessible as scenario- or argumentation-based modelings but they still claim that a Bayesian approach models the standard approach to legal proof, which is to continually revise beliefs under new evidence. Another claimed advantage of a Bayesian approach is that it provides an explicit probability of guilt, which they believe is the ultimate requirement in reaching a verdict in legal cases.

Christian Dahlman analyzes the case with a generalized version of Bayes' rule that can handle dependencies. Sometimes when Dahlman regards pieces of evidence E and E' as dependent, he estimates the likelihood ratio of E relative to the guilt and innocence hypotheses *plus* E'. And sometimes he groups several pieces of evidence that he regards as dependent on each other together in a bundle of evidence and then treats the bundle as one piece of evidence. This approach is mathematically correct but does not allow a very precise analysis of the interdependencies. Thus Dahlman sacrifices some expressiveness compared to Bayesian networks in order to gain on cognitive feasibility.

Dahlman claims that the use of this method can help a fact finder avoid three kinds of bias in probabilistic reasoning—*false dichotomy* (assuming that the two compared hypotheses exhaust the space of possibilities), *dependence neglect* (neglecting statistical dependencies between pieces of evidence), and *miss rate neglect* (disregarding the probability of a piece of evidence given the innocence hypothesis). He claims to have identified occurrences of all three kinds of bias in the (trial and appellate) courts' decisions. The outcome of Dahlman's case study is that the posterior probability of the guilt hypotheses fails to satisfy the standard of proof beyond reasonable doubt. Dahlman acknowledges that this is relative to his own subjective estimates. Finally, he acknowledges the danger that a mathematical analysis may give a false impression of objectivity to fact finders, but he claims (like Fenton et al. in their case study with Bayesian networks) that an advantage of the mathematical form of his analysis is that it forces him to make his assumptions explicit, which can then be scrutinized.

Peter Van Koppen and *Anne Ruth Mackor* apply their version of the scenario approach to the case. First they further explicate and expand this approach in several ways. They explicitly link it to inference to the best explanation and to theories of explanatory coherence, and they expand the approach in order to accommodate the difference between explaining known facts and predicting novel facts. This distinction is important in the Simonshaven case, since the defendant stated from the beginning that a madman who jumped from the bushes had killed his wife. Whereas this claim seemed very implausible at first, evidence about a madman jumping out of the bushes in other cases shed a new light on his claim. Van Koppen and Mackor claim that a strong point of their approach is that it is feasible from a cognitive point of view since it stays close to descriptive theories of evidential reasoning. They acknowledge that their approach can be further developed and be made more precise, but they want to keep it informal, so that legal professionals can apply it.

Floris Bex analyzes the case in an informal version of his hybrid theory. In this theory, (alternative) hypothetical scenarios, Bex calls them "stories," are constructed to explain

important observations in the case, similar to the scenario approach. However, a key difference with that approach is that arguments based on evidence are used to reason about these alternative stories: Stories can be anchored in evidence by supporting them with arguments, or stories can be attacked by arguments based on evidence. Stories and arguments are thus used in a dialectical process, where critical questions (e.g., "is the story plausible?"; "is there an alternative explanation?"; "are there arguments that support or attack the story?") lead to (alternative) stories and (counter-) arguments. Bex starts by presenting one possible story, and then proceeds to ask and answer such questions, ending up with three alternative stories and more than 50 arguments. The case study shows that the story supporting conviction is better than the two stories supporting acquittal and, perhaps more importantly, that Ed never answered some of the critical questions brought against him in the case. Bex argues that the main strength of the hybrid theory is that it provides the best of both worlds: Cognitively feasible story-based reasoning is combined with more detailed rational argument-based reasoning. A weak point of the hybrid theory is that, like the argument-based approach, it does not provide a systematic account of uncertainty.

Finally, *Bart Verheij* analyzes the case with his own recently developed case model formalism, which was designed to connect arguments, scenarios, and probabilities in one approach, and to bridge quantitative and qualitative analytic styles. Case models represent what can be the case in a set of logical formulas, each representing a case. Scenarios can be expressed as cases conjoining evidence and hypotheses. The cases in a case model are ordered by a preference ordering, which can but need not express a probability distribution. Arguments are a pair (premises, conclusion) and are evaluated given a case model (unlike the formalisms used by Prakken, which evaluate arguments in light of their interaction with conflicting arguments). An argument is *coherent* if there is a case in which both its premises and conclusion holds. It is *presumptively valid* if its conclusion holds in the maximally preferred cases of those in which its premises hold. And it is *conclusive* if its conclusion holds in all cases of those in which its premises hold. Verheij has similar definitions for when a scenario is coherent, plausible, or beyond a reasonable doubt.

The outcome of Verheij's analysis is that the guilty-as-charged scenario is beyond reasonable doubt, but Verheij notes that his aim was not to give his own analysis of the case but to reconstruct the court of appeal's reasoning (which led to conviction) in his formalism. Verheij claims that this modeling is natural in that it models the gradual development and evaluation of scenarios in light of evidence. Verheij also claims that since the case models have a probabilistic interpretation, the use of his formalism prevents probabilistic reasoning errors, while yet there is no need to express probabilities on all elements of his model (thus addressing the number problem). Verheij acknowledges that his formalism pays less attention to the internal structure of arguments and scenarios than alternative argumentation and scenario approaches.

Our topic also contains four commentaries. *Ronald Meester* comments on the papers of Fenton et al. and Dahlman, which both take a Bayesian approach. Meester elaborates on what he thinks are limitations of Bayesian approaches in the context of criminal proof. *Marcello di Bello* comments on Prakken's argumentative and Van Koppen and Mackor's

scenario approach. He argues that both these approaches face difficulties since they lack an account of plausibility, in particular of degrees of support. *Frank Zenker* makes similar observations in his comment on Bex's hybrid combination of these approaches. Zenker also comments on Verheij's approach, claiming that it promises to bridge qualitative and quantitative accounts but is not easy to use for those not trained in logic. Finally, *Paul Roberts* suggests that the main use of Van Koppen and Mackor's scenario approach may lie not in trial stages but in the investigative and preparatory stages of criminal adjudication, especially in jury-based systems. About Dahlman's Bayesian approach, he acknowledges its usefulness in exposing probabilistic reasoning fallacies but warns that it is far removed from how fact finders in criminal cases actually reason.

5. Concluding remarks

In our concluding remarks, we first summarize the authors' answers to the four questions each paper was meant to address.

To what extent is the analysis objective and to what extent is it based on subjective beliefs, assumptions, and choices? Prakken and Dahlman acknowledge the subjectivity of their modelings, while Bex and Verheij say they have attempted to reconstruct the appeal court's reasoning. Van Koppen and Mackor claim that their reliance on commonsense knowledge, which can be incorrect, makes their modeling neither fully subjective nor fully objective. Fenton et al. also claim some degree of objectivity for their modeling, based on the outcome of their sensitivity analysis and their use of idioms and the opportunity-prior method. The latter method is an interesting result of our topic, separately published as Fenton, Lagnado, Dahlman, and Neil (2017). Interestingly, Dahlman, who is a co-author of the latter paper, did not use the method in his Bayesian analysis. If he had done so, his posterior probability might well have been above his threshold for proof beyond reasonable doubt.

How natural is the analysis from a cognitive and legal point of view? Bex and Van Koppen & Mackor claim that the centrality of scenario construction makes their respective methods practically applicable, although Van Koppen and Mackor note that sometimes direct reasoning from evidence to hypotheses (as allowed by the argumentation part of Bex' hybrid approach) may be more natural than the explanatory style of reasoning in a purely story-based approach. Prakken recommends empirical research on this issue in the context of witness testimony. Both Fenton et al. and Dahlman acknowledge that their Bayesian approaches may be difficult to apply in practice, but Dahlman claims that applying his version of the method is worth the effort while Fenton et al. claim that their approach still models the general process of evidential legal reasoning. Verheij makes a similar claim and also claims that his integration of arguments, scenarios, and probabilities balances the strengths of each. Finally, five of the six case studies apply formal models. While formality may increase precision, it clearly reduces practical usability since legal practitioners are not used to using formal methods. For this reason, two of these five case studies (Bex and Prakken) make use of visualization techniques to make their analyses more understandable, while Fenton et al. claim that the "powerful" visual appeal of

Bayesian networks makes them potentially natural in legal contexts. This highlights the issue of whether the use of visualization techniques, possibly supported by software, can improve legal evidential reasoning (Tillers, 2007).

Did your analysis identify errors or biases in the reasoning of the judge, prosecutor, or defense? Dahlman, Fenton et al., and Verheij claim that their respective methods can avoid probabilistic reasoning errors, but only Dahlman gives examples from this case study. Bex, Van Koppen and Mackor, and Prakken criticize the court's reasoning on various specific points of detail. We note that their respective methods all involve the comparison of conflicting arguments and/or scenarios, which may help avoiding confirmation bias.

Does your analysis respect the legal constraints, such as the burden and standard of proof and the right to remain silent? Finally, all authors claim that their methods and analyses generally respect the legal constraints. In addition, Fenton et al. claim that Bayesian approaches can give a precise interpretation of the beyond-reasonable-doubt proof standard. Van Koppen and Mackor note that their view on confirmed risky predictions may be at odds with the right to remain silent.

How should all these claims be evaluated? We recall that any insights obtained by the case studies are primarily on *analyzing* and only indirectly on *performing* evidential reasoning. Moreover, our case study experiment was not meant to be a controlled experiment, suitable to yield formally valid and reliable conclusions. Instead, in our opinion, its main value lies in how it can advance the continuing debate on models of rational legal proof. Now six detailed case studies are available of a real case, which can be used by other researchers in their research. Moreover, the authors' own comments and those of the four commentators are arguably valuable contributions to the ongoing debate, since these comments are made in the context of a realistic case study experiment.

Acknowledgments

We acknowledge the financial support and hospitality of the Zentrum für Interdisciplinäre Forschung (Centre for Interdisciplinary Research) in Bielefeld, Germany; of the Isaac Newton Institute for Mathematical Sciences, Cambridge, UK (in the context of the program Probability and Statistics in Forensic Science, supported by EPSRC grant no. EP/K032208/1); and of Prof. Norman Fenton's group at the School of Electronic Engineering and Computer Science, Queen Mary University London, which enabled us to organize three successive workshops devoted to the Simonshaven case study. We thank Hylke Jellema and Nadine Smit for translating the Simonshaven case files from Dutch into English and for anonymizing the files.

Note

1. All names in this topic of persons related to the crime are fictional, to ensure anonymity.

References

Amaya, A. (2009). Inference to the best legal explanation. In H. Kaptein, H. Prakken, & B. Verheij (Eds.), *Legal evidence and proof: Statistics, stories, logic* (pp. 135–159). Farnham, UK: Ashgate.

Anderson, T., & Twining, W. (1991). *Analysis of evidence. How to do things with facts based on Wigmore's science of judicial proof*. Boston: Little, Brown and Company.

Bex, F. (2011). *Arguments, stories and criminal evidence: A formal hybrid theory*. Dordrecht, the Netherlands: Springer.

Bex, F., Prakken, H., Reed, C., & Walton, D. (2003). Towards a formal account of reasoning about evidence: Argumentation schemes and generalisations. *Artificial Intelligence and Law*, *12*, 125–165.

Chater, N., Tenenbaum, J., & Yuille, A. (2006). Probabilistic models of cognition: Conceptual foundations. *Trends in Cognitive Sciences*, *10*, 287–291.

Cohen, L. (1977). *The probable and the provable*. Oxford, UK: Clarendon Press.

Dawid, P. (2005). Probability and proof. Online appendix to T. J. Anderson, D. A. Schum and W. L. Twining: Analysis of Evidence. Available at: http://tinyurl.com/tz85o

de Zoete, J., Sjerps, M., Lagnado, D., & Fenton, N. (2015). Modelling crime linkage with Bayesian networks. *Science and Justice*, *55*, 209–217.

Dubois, D., & Prade, H. (1988). *Possibility theory*. New York: Plenum.

Fenton, N., & Berger, D. (2016). Bayes and the law. *Annual Review of Statistics and Its Application*, *3*, 51–77.

Fenton, N., Lagnado, D., Dahlman, C., & Neil, M. (2017). The opportunity prior: A simple and practical solution to the prior probability problem for legal cases. In *Proceedings of the 16th International Conference on Artificial Intelligence and Law* (pp. 69–78). New York: ACM Press.

Fenton, N., Lagnado, D., & Neil, M. (2013). A general structure for legal arguments using Bayesian networks. *Cognitive Science*, *37*, 61–102.

Fenton, N., & Neil, M. (2011). Avoiding legal fallacies in practice using Bayesian networks. *Australian Journal of Legal Philosophy*, *36*, 114–151.

Gigerenzer, G., & Todd, P. (1999). *Simple heuristics that make us smart*. Oxford, UK: Oxford University Press.

Hahn, U., & Hornikx, J. (2016). A normative framework for argument quality: Argumentation schemes with a Bayesian foundation. *Synthese*, *193*, 1833–1873.

Hahn, U., & Oaksford, M. (2007). The rationality of informal argumentation: A Bayesian approach to reasoning fallacies. *Psychological Review*, *114*, 704–732.

Harris, A., Hahn, U., Madsen, J., & Hsu, A. (2016). The appeal to expert opinion: Quantitative support for a Bayesian network approach. *Cognitive Science*, *40*, 1496–1533.

Hepler, A., Dawid, A., & Leucari, V. (2007). Object-oriented graphical representations of complex patterns of evidence. *Law, Probability and Risk*, *6*, 275–293.

Josephson, J. (2001). On the proof dynamics of inference to the best explanation. *Cardozo Law Review*, *22*, 1621–1643.

Kadane, J., & Schum, D. (1996). *A probabilistic analysis of the Sacco and Vanzetti evidence*. New York: John Wiley & Sons.

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

Lipton, P. (2004). *Inference to the best explanation* (2nd ed.). London: Routledge.

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

Pennington, N., & Hastie, R. (1993). The story model for juror decision making. In N. Hastie (Ed.), *Inside the juror, the psychology of juror decision making* (pp. 192–221). Cambridge: Cambridge University Press.

Pollock, J. (2007). Defeasible reasoning. In J. Adler & L. Rips (Eds.), *Reasoning: Studies of human inference and its foundations* (pp. 451–470). Cambridge: Cambridge University Press.

Poole, D. (2001). Logical argumentation, abduction and Bayesian decision theory: A Bayesian approach to logical arguments and its application to legal evidential reasoning. *Cardozo Law Review*, *22*, 1733–1745.
Schum, D., & Tillers, P. (1991). Marshaling evidence for adversary litigation. *Cardozo Law Review*, *13*, 657–704.
Shafer, G. (1976). *A mathematical theory of evidence*. Princeton, NJ: Princeton University Press.
Thagard, P. (2004). Causal inference in legal decision making: Explanatory coherence vs. Bayesian networks. *Applied Artificial Intelligence*, *18*, 231–249.
Tillers, P. (Ed.). (2007). *Law, probability and risk* (vol. 6). Special issue on visualizing evidence and inference in legal settings. Oxford, UK: Oxford University Press.
Timmer, S., Meyer, J.-J.Ch., Prakken, H., Renooij, S., & Verheij, B. (2017). A two-phase method for extracting explanatory arguments from Bayesian networks. *International Journal of Approximate Reasoning*, *80*, 475–494.
Tribe, L. (1971). Trial by mathematics: Precision and ritual in the legal process. *Harvard Law Review*, *84*, 1329–1393.
Tversky, A., & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131.
Twining, W. (1999). Necessary but dangerous? Generalisations and narrative in argumentation about "facts" in criminal process. In M. Malsch & F. Nijboer (Eds.), *Complex cases. Perspectives on the Netherlands criminal justice system* (pp. 69–98). Amsterdam: Thela Thesis.
van Eemeren, F., Garssen, B., Krabbe, E., Henkemans, A. S., Verheij, B., & Wagemans, J. (2014). *Handbook of argumentation theory*. Dordrecht: Springer.
van Koppen, P. (2011). *Overtuigend Bewijs. Indammen van Rechterlijke Dwalingen*. Amsterdam: Nieuw Amsterdam (in Dutch).
Verheij, B., Bex, F., Timmer, S., Vlek, C., Meyer, J.-J., Renooij, S., & Prakken, H. (2016). Arguments, scenarios and probabilities: Connections between three normative frameworks for evidential reasoning. *Law, Probability and Risk*, *15*, 35–70.
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.
Vlek, C., Prakken, H., Renooij, S., & Verheij, B. (2016). A method for explaining Bayesian networks for legal evidence with scenarios. *Artificial Intelligence and Law*, *24*, 285–324.
Wagenaar, W., van Koppen, P., & Crombag, H. (1993). *Anchored narratives. The psychology of criminal evidence*. New York: St. Martin's Press.
Wigmore, J. (1931). *The principles of judicial proof* (2nd ed.). Boston: Little, Brown and Company.

Papers in this Topic

Bex, F. J. (2020). The hybrid theory of stories and arguments applied to the Simonshaven case. *Topics in Cognitive Science*, *12*(4), 1152–1174.
Dahlman, C. (2020). De-biasing legal fact-finders with Bayesian thinking. *Topics in Cognitive Science*, *12*(4), 1115–1131.
Fenton, N., Neil, M., Yet, B., & Lagnado, D. (2020). Analyzing the Simonshaven case using Bayesian networks. *Topics in Cognitive Science*, *12*(4), 1092–1114.
Prakken, H. (2020). An argumentation-based analysis of the Simonshaven case. *Topics in Cognitive Science*, *12*(4), 1068–1091.
Prakken, H., Bex, F. J., & Mackor, A. R. (2020). Editors' review and introduction: Models of rational proof in criminal law. *Topics in Cognitive Science*, *12*(4), 1053–1067.
van Koppen, P. J., & Mackor, A. R. (2020). A scenario approach to the Simonshaven case. *Topics in Cognitive Science*, *12*(4), 1132–1151.

Verheij, B. (2020). Analyzing the Simonshaven case with and without probabilities. *Topics in Cognitive Science*, *12*(4), 1175–1199.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Appendix S1. Case documents in the Simonshaven case.