Law, Probability and Risk Advance Access published August 27, 2014

 doi:10.1093/lpr/mgu011

To catch a thief with and without numbers: arguments, scenarios and probabilities in evidential reasoning

Bart Verheij[†]

CodeX, Stanford University, Stanford, USA; Artificial Intelligence, University of Groningen, Groningen, The Netherlands

[Received on 14 March 2014; revised on 1 May 2014; accepted on 5 August 2014]

Mistakes in evidential reasoning can have severe consequences. Especially, errors in the use of statistics have led to serious miscarriages of justice. Fact-finders and forensic experts make errors in reasoning and fail to communicate effectively. As tools to prevent mistakes, three kinds of methods are available. Argumentative methods analyse the arguments and counterarguments that are presented in court. Narrative methods consider the construction and comparison of scenarios of what may have happened. Probabilistic methods show the connections between the probability of hypothetical events and the evidence. Each of the kinds of methods has provided useful normative maxims for good evidential reasoning. Argumentative and narrative methods are especially helpful for the analysis of qualitative information, but do not come with a formal theory that is as well-established as probability theory. In probabilistic methods, the emphasis is on numeric information, so much so that a standard criticism is that these methods require more numbers than are available. This article offers an integrating perspective on evidential reasoning, combining the strengths of each of the kinds of methods: the adversarial setting of arguments pro and con, the globally coherent perspective provided by scenarios, and the gradual uncertainty of probabilities. In the integrating perspective, arguments and scenarios are interpreted in the quantitative setting of standard probability theory. In this way, the integrated perspective provides a normative framework that bridges the communicative gap between fact-finders and forensic experts. Both qualitative and quantitative information can be used safely, focusing on what is relevant.

Keywords: evidential reasoning; argumentation; narratives; probabilities.

1. Introduction

Human reasoning is prone to error (e.g. Kahneman, 2011); evidential reasoning is not special in this respect. However, since errors in evidential reasoning can have very serious consequences, the prevention of errors has high priority. One source of errors is probabilistic reasoning, which for instance played a significant role in the high profile miscarriage of justice cases of Sally Clark (in the UK) and Lucia de Berk (in the Netherlands) (Buchanan, 2007; Derksen and Meijsing, 2009; Schneps and Colmez, 2013). In both these cases, it became clear after the trial that the computation of the numbers used was wrong, and that, even if the numbers had been computed correctly, the wrong conclusions were based on them.[1]

[†]Email: verheij@stanford.edu, b.verheij@ai.rug.nl

[1] In the Sally Clark case, an expert stated in court that the probability that in one family two babies die in their beds without a medical explanation is 1 in 73,000,000. The number's computation indefensibly assumed independence between the deaths. In the Lucia de Berk case, an expert reported that the probability that a nurse accidentally encountered as many unexpected deaths

These and other miscarriages of justice have refueled the long-standing debate about which methods of evidence-based fact finding can help prevent errors. Recent discussions are influenced by the successful role of DNA evidence in criminal trial.[2] The associated probabilistic model of DNA profile matching has proven to be a valuable tool to measure how distinguishing a profile match is.[3] Since DNA evidence has a strong scientific background—statistics and probability theory—it becomes pertinent to consider whether all kinds of evidence can be given such a solid underpinning. An experienced expert in the field reports that judges, prosecutors and attorneys now seem to expect precise numbers with all kinds of evidence, e.g. a likelihood ratio for a positive identification in a line-up (van Koppen, 2013). This 'DNA effect' has refueled the debate about the use of probabilities in court. The debate is long-standing (Finkelstein and Fairley, 1970; Tribe, 1971)[4], but every now and then gains momentum.[5]

The discussion is an example of what might be called the *qualitative–quantitative tension*: Some common methods are qualitative, in the sense that they do not involve numbers. Examples of such methods are argumentative and narrative analysis. Other methods are quantitative, in the sense that they do use numbers. Probabilistic tools are of this kind. Qualitative and quantitative methods are sufficiently different in nature to have led to different bodies of expertise and best practices. For instance, in criminal investigation and decision-making, some professionals, in particular lawyers, most often use the qualitative methods of argumentation and narrative analysis, while other professionals, in particular forensic scientists, focus on quantitative methods founded in probability theory. The result of the qualitative–quantitative tension is a communicative gap that has proven hard to overcome (see, e.g. Broeders, 2009, for a discussion of communicating statistical information in court).

One would hope that the scholarly literature can help out. It turns out that in the literature three theoretical approaches to evidential reasoning are presented: argumentative, narrative and probabilistic (Kaptein *et al.*, 2009; Dawid *et al.*, 2011; Anderson *et al.*, 2005).

Argumentation methods focus on arguments for and against what has happened in a criminal case, using reasons grounded in the available evidence. The argumentation diagramming techniques developed by John Henry Wigmore (1931) at the start of the 20th century are precursors of today's software for the diagramming, analysis and construction of arguments (Kirschner *et al.*, 2003; Verheij, 2005). Also, the New Evidence Scholarship (Anderson *et al.*, 2005) focused on the refinement and extension of argumentation techniques (but see Kadane and Schum, 1996 for a perspective including probabilistic tools). The formal modelling of argumentation is an active area of research (e.g. Pollock, 1995; Dung, 1995; Prakken, 2010).

In narrative methods, plausible scenarios are constructed as alternative hypotheses about what has happened, and evaluated and compared on the basis of the evidence. The role of scenarios in evidential

and resuscitations during her shifts as Lucia de Berk was 1 in 342,000,000. The number's computation used an erroneous method for combining numbers for three different wards that Lucia de Berk had been working for. It has been argued that, in both cases, the number was wrongly taken to imply a low chance of innocence, cf. the Lucia de Berk case where the statistical expert notoriously claimed 'That cannot be by chance' (in Dutch: 'Dat kan geen toeval zijn').

[2] See Thompson (2013) for a discussion of DNA evidence, including possible pitfalls.

[3] The web site http://dna-view.com/ developed by the forensic mathematician Charles Brenner provides helpful resources.

[4] Tillers (2011) reports on this 1970s instance of the debate, commenting that it became 'generally unproductive and sterile years ago'.

[5] For a recent installment of the debate, see the extensive discussion following the judgment of the UK Court of Appeal (R v T [2010] EWCA Crim 2439) in the journals *Science and Justice* (Volumes 51 and 52, 2011/2012) and *Law, Probability and Risk* (with a special issue in Volume 11, 2012).

TABLE 1 *Characteristics of the three normative frameworks considered*

	Arguments	Scenarios	Probabilities
Adversarial setting	+	±	−
Global coherence	−	+	−
Uncertainty in gradations	±	−	+
Standard formalization	±	−	+

reasoning has been studied in legal psychology. Pennington and Hastie (1993) showed the persuasive effect of good story telling—associated with the risk that well-told stories may be believed more easily than well-supported stories—while Wagenaar *et al*. (1993) have used scenario methods as a rationality tool. Recently, the formal and computational connections between argumentation and scenario methods have been studied (Bex, 2011; Bex *et al*., 2010). In probabilistic approaches, the probabilities of hypothetical events are calculated, compared and updated in the light of new evidence (see Aitken and Taroni, 2004 for a textbook treatment of the subject matter). Probabilistic methods have been used to explain fallacious reasoning, such as the prosecutor's and defense attorney's fallacy (Thompson and Schumann, 1987). The formal and computational techniques associated with Bayesian Networks (Jensen and Nielsen, 2007) have also been applied to reasoning with evidence (Taroni *et al*., 2006; Hepler *et al*., 2007; Fenton *et al*., 2013; Kadane and Schum, 1996; Vlek *et al*., 2013, Timmer *et al*., 2013). Some contemporary scholars have defended the primacy of probabilistic methods (Fenton, 2011; Fenton and Neil, 2013), while others doubt that they can play a helpful role at all (e.g. van Koppen, 2012, 2013, favouring narrative tools).

Table 1 summarizes some key characteristics of the three normative frameworks encountered. The *adversarial setting* of evidential reasoning is most clearly represented in the normative framework of arguments, where the pros and cons of different positions are structured and evaluated (indicated by a+ in the table). In the scenarios framework, the possibility of considering different scenarios also points to the adversarial setting in which the parties involved will present competing views of what has happened, but not as directly as in an arguments setting (hence the ± in the table). The characteristic strong point of a scenarios framework is its emphasis on the *global coherence* of the evidence and case as a whole, by analysing and evaluating the scenarios of what can have happened. The ensuing holistic, synthetic point of view helpfully balances the atomistic, reductionistic style of both argumentative and probabilistic models. The probabilistic framework gives tools to analyse and evaluate *uncertainty in gradations*. Although variations and non-standard models do exist, probabilistic approaches come with a *standard formalization* that is widely accepted, applied and well-understood (Hájek, 2011). For argumentative methods, the situation is very different. There exist several formal and computational models (Dung, 1995; Besnard and Hunter, 2009; Prakken, 2010; Bench-Capon, 2003; Verheij, 2003), but there is as yet much ongoing work on possible variations and connections between approaches, in particular also with classical logic and probability theory.[6] The normative theory of scenario models has only been formalized to a limited extent. State-of-the-art proposals make a connection to argumentative modelling (Bex, 2011) or probabilistic approaches (Shen *et al*., 2006).

The assumption of this article is that the three kinds of methods are closely connected, and that an integrated perspective is possible. The aim of this article is therefore to present a perspective on

[6] See in particular the proceedings of the 'Computational Modeling of Argument' conference series (COMMA), biennially held since 2006.

evidential reasoning in which argumentative, narrative and probabilistic methods go side by side, emphasizing the strengths of each. Recent research combining arguments and scenarios (Bex *et al.*, 2010), scenarios and probabilities (Vlek *et al.*, 2013), and arguments and probabilities (Timmer *et al.*, 2013) has been an inspiration.[7]

The structure of the article is as follows. In the next section, an example robbery case is presented that will be used throughout the article as an illustration (Section 2). Then the integrated perspective on arguments, scenarios and probabilities is presented, starting in Section 3 with the formal background. In Section 4, the integrating perspective is presented, focusing on the main ideas. In Section 5, the robbery example is formalized. In Section 6, the integrated perspective on arguments, scenarios and probabilities is revisited, emphasizing the formal ideas. In Section 7, related research is discussed, highlighting relations with the present proposal.[8]

2. Example: robbery

Assume that the police receive a phone call about a single-person robbery at a jewelry shop. By their knowledge of the local criminal milieu, the police hypothesize that one of the usual suspects, say 8 in total, may be the robber. In order to know more about what happened, the surveillance video is checked. It is found that the robber has a specific kind of tattoo, associated with membership of a certain gang. Two of the usual suspects are known members of this gang. Upon interrogation, it becomes clear that the first of these two does not have the tattoo, but the second suspect does. When the police show the video, the gang member with the tattoo decides to cooperate hoping for a lower punishment, and confesses. He reveals the location of the stolen jewelry, which is subsequently recovered. Since the suspect showed such specific knowledge of the crime, no doubt is left that he is the robber.

Let us see how the evidence and hypotheses about what has happened in this robbery develop. A graphical illustration is shown in Figure 1.

After the phone call (evidence E_1), the police assume eight possible hypotheses about what has happened: one of the usual suspects $1, \ldots, 8$ has committed the robbery ($H_1, \ldots, H_8$). These hypotheses are shown as the row of eight boxes at the top of the figure. The hypotheses are based on the knowledge that robberies are typically performed by one of the usual suspects.

The surveillance camera gives additional evidence (E_2). The state of the hypotheses given the combined evidence $E_1 \wedge E_2$ is shown on the second line in the figure. Two things have changed. First, the evidence has shown that the robber has a specific tattoo (T). If one of the suspects is the robber, he will have such a tattoo. The evidence has given an expectation that would hold in any of the hypotheses, if the hypothesis were true. The expectation T is indicated on the right hand side of the figure.

The second change concerns the relative probability of the hypotheses. Since suspects 3 and 7 are the known gang members, the police consider it more probable that one of them is the robber than any of the other suspects. In the figure, this is suggested by reducing the size of the other boxes.

[7] See the website http://www.ai.rug.nl/~verheij/nwofs/ for further information about the research project ‘Designing and Understanding Forensic Bayesian Networks with Arguments and Scenarios’, funded in the NWO Forensic Science program.

[8] Initial ideas were presented at the 2013 Workshop on Formal Argument and Evidential Inference (FAREI 2013), organized by Giovanni Sartor, Scott Brewer, and Gustavo Ribeiro, at the 14th International Conference on AI and Law (ICAIL 2013) in Rome.

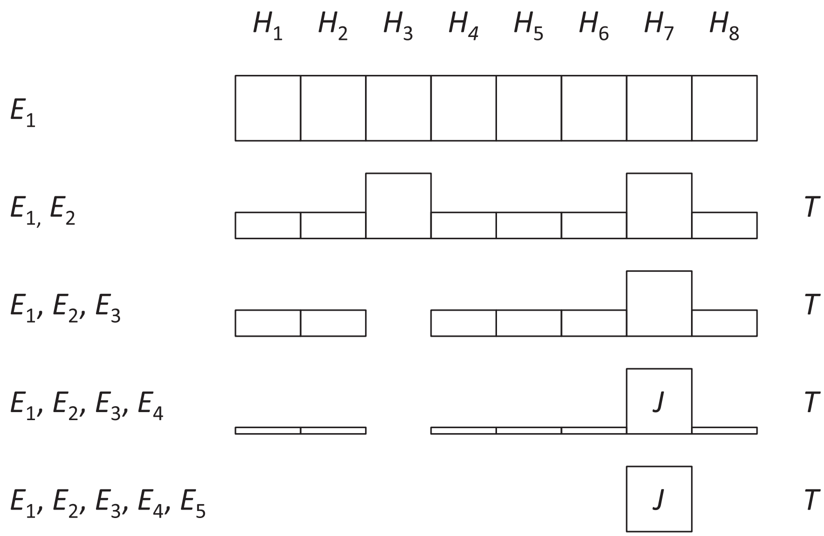

FIG. 1. Development of the evidence and hypotheses in the robbery example.

When suspect 3 is interrogated, it is found that he does not have the tattoo (evidence E_3), contradicting the expectation T. As a result, the box of H_3 has been removed in the figure.

Suspect 7 has the tattoo and confesses to the crime (evidence E_4). As a result, the police consider the probability that someone else than suspect 7 is the robber to have again become smaller. This is indicated by reducing the size of the boxes for these hypotheses. In his confession, suspect 7 reveals the location of the jewelry (J). If he actually is the robber, the location of the jewelry is expected to be correct. This expectation holds only for H_7 so has been indicated in the figure in the appropriate box.

The jewelry is found at the revealed location (E_5), removing reasons to doubt that suspect 7 is the robber. The robber is considered to have been identified.[9] In the figure, only the box for hypothesis H_7 has remained. Given all evidence $E_1, \ldots, E_5$, hypothesis H_7 is believed 'beyond a reasonable doubt'. The hypothesis matches the expectations of the tattoo T and the location of the jewelry J. The overview in Figure 1 can be regarded as a graphical model of the development of the evidence, the hypotheses and the expectations.

The example illustrates how evidential reasoning under uncertainty involves iterated observation, with the aim of identifying what has happened. In the example, initially, there are several possibilities of what can have happened. At this initial stage, none is identified as the actual course of affairs. By ever looking better, it is finally established that suspect 7 is the robber.

3. Formal background: classical logic and probability theory

The formal background used in this article does not go beyond introductory textbooks on logic and probability theory. We do so on purpose, in an attempt to show what can be done using standard, widely known, and well-tested tools, without going to specific techniques developed for specific settings (of which there are many available). We use classical logic and probability theory.

[9] See Sections 6 and 8 for a discussion of the hints of doubt that always remain.

We assume a logical language L with two binary connectives $\wedge$ and $\vee$, expressing conjunction and disjunction, respectively, and a unary connective $\neg$, expressing (classical) negation. The language is constructed inductively from a non-empty set of propositional Boolean variables P, with an associated classical deductive consequence relation $\vDash$. Sentences $\varphi \in L$ and $\psi \in L$ are said to be logically incompatible when $\vDash \neg(\varphi \wedge \psi)$. A sentence $\varphi \in L$ is a logical truth when $\vDash \varphi$.

Probability functions are real-valued functions governed by the standard Kolmogorov axioms (Hájek, 2011):

(1) $\mathrm{p}(\varphi) \geq 0$ for all $\varphi \in L$.
(2) If $\varphi \in L$ is a logical truth, then $\mathrm{p}(\varphi) = 1$.
(3) $\mathrm{p}(\varphi \vee \psi) = \mathrm{p}(\varphi) + \mathrm{p}(\psi)$ for all $\varphi \in L$ and $\psi \in L$ such that φ and ψ are logically incompatible.[10]

We recall the definition of conditional probabilities: for sentences φ and $\psi \in L$, with $\mathrm{p}(\varphi) > 0$, define $\mathrm{p}(\psi \mid \varphi) := \mathrm{p}(\varphi \wedge \psi)/\mathrm{p}(\varphi)$.

4. Integrating arguments, scenarios and probabilities

Very briefly, the integrated perspective of arguments, scenarios and probabilities can be summarized as follows:

> The available evidence supports one or more scenarios of what has happened. As such, the evidence provides a supporting argument for a scenario. The evidence can give different degrees of support to different scenarios. Degrees of support can range from possibility to certainty, and can sometimes be numerically estimated, by numbers that behave like conditional probabilities. Scenarios can be more or less specific about what has happened. Incompatible scenarios can express different versions of what has happened. Additional evidence can change the support of a scenario, even exclude it. Scenarios give rise to expectations of what follows from their truth. Expectations are supported to different degrees, given the scenario they are based on. By collecting further evidence, it can happen that no reason for doubting a scenario remains. Then the opposite of the scenario is not considered a possibility anymore.

Figure 2 summarizes the perspective. The place of arguments and scenarios is explicitly shown, but the term 'probability' does not appear. Instead, we have used the term 'strength', as preferred term for the measure of how strongly or weakly an argument supports its conclusion. Argument strength is modelled by numbers that formally behave like probabilities. Also one may wonder how counter-arguments fit in the picture, as the emphasis in the figure is on supporting arguments. Without going into detail now (for more, see below; Sections 5 and 6), a counterargument is here treated as additional information that reduces the strength of an argument.

Let us flesh out this summary by formalizing the robbery example.

5. Formalization of the robbery example

In this section, we go over the robbery example again, showing its formalization, using the tools of standard probability and its underlying classical logic. While developing the formalization, terms and tools associated with arguments, scenarios and probabilities are used side by side.

[10] In this article, we will not discuss countably infinite additivity.

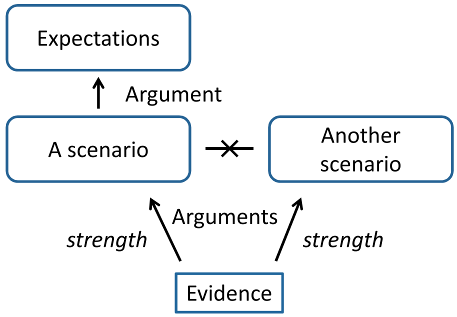

FIG. 2. The evidence supports one or more scenarios.

The example is formalized using a logical language built from the variables already introduced:

- E_1: the phone call;
- E_2: the evidence provided by the surveillance camera;
- E_3: the interrogation of suspect 3;
- E_4: the interrogation of suspect 7;
- E_5: the recovery of the jewelry;
- $H_1, \ldots, H_8$: the hypotheses that usual suspect 1, 2, 3, 4, 5, 6, 7, or 8 has committed the robbery;
- T: the suspect has a tattoo;
- J: the correctness of the location of the jewelry.

The example starts with the phone call about the robbery, E_1. On the basis of this evidence, the police consider eight possible hypotheses $H_1, \ldots, H_8$, each hypothesis representing that one of the usual suspects has committed the robbery. Figure 3 illustrates how the evidence is an argument supporting each hypothesis. Clearly, the argument from the evidence to each separate hypothesis is not very strong, but each is worthy of further scrutiny. The inference to each of the hypotheses is presumptive: the police consider each hypothesis a possibility, not a certainty. The inferences are also defeasible: further evidence may prune some of the presumptive possibilities, hopefully identifying the actual robber.

The evidence supports each hypothesis, but not simultaneously. In the figure, it is shown that each hypothesis attacks the other hypotheses: as there is only one robber, the truth of the hypothesis that one is the robber implies the falsity of the other hypotheses. In the figure, this has been indicated by the use of a diamond-pointed arrow.

Each hypothesis can be regarded as an unspecific mini-scenario, expressing who committed the robbery, without going into detail. The evidence matches each scenario, but no two scenarios simultaneously.

In the language of probability theory, the possibility of each hypothesis given the evidence can be expressed by saying that the conditional probability of the hypothesis given the evidence is positive:

$$\mathrm{p}(H_i \mid E_1) > 0, \text{ for each } i. \tag{1}$$

Here the index i ranges from 1 to 8, as will be silently assumed in the following. Using the language of argument strength, we can say that the argument from the evidence to a hypothesis has positive

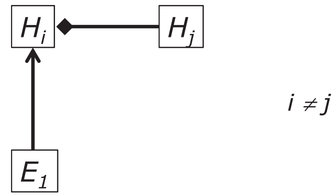

FIG. 3. Evidence 1: a phone call about a robbery.

strength. As there is no information about how strong exactly, or even approximately, the only thing to be said about the strength is that it is positive. The police have no belief about how big or small the probability for each of the suspects is, but each hypothesis is a possibility, expressed by stipulating a positive probability.

Note how the formalization refines the representation in Figure 1: there the boxes of equal sizes suggest that the probabilities are equal, whereas only their positive size matters for now.

The mutual incompatibility of the hypotheses is expressed by saying that the probability of a combination of a pair of hypotheses is 0, given the evidence:

$$p(H_i \wedge H_j \mid E_1) = 0, \text{ for each } i \text{ and } j \text{ with } i \neq j. \tag{2}$$

The surveillance camera E_2 reveals the tattoo T. So, given the evidence, for each of the hypotheses it is expected that the suspect has a tattoo if the hypothesis is true. The associated argument structure is illustrated in Figure 4. By the (conjunctively) combined evidence $E_1 \wedge E_2$ it is supported that the truth of a hypothesis H_i implies that the suspect has a tattoo T.

The example shows how evidence can lead to expectations about what is true. *If* a particular suspect is the robber, he will have the tattoo as seen on the surveillance camera. For most suspects this will be a counterfactual expectation, since most will not have the specific tattoo.

The expectation can probabilistically be expressed as follows:

$$p(T \mid H_i \wedge E_1 \wedge E_2) = 1, \text{ for each } i. \tag{3}$$

In other words, assuming the evidence and the truth of a hypothesis, the suspect has, with certainty, the tattoo.

By the evidence E_2, the police also apply their knowledge about the local crime milieu: suspects 3 and 7 are known gang members. As a result, they now consider their guilt more probable than that of the other usual suspects. Probabilistically this can be expressed as an inequality:

$$p(H_i \mid E_1 \wedge E_2) > p(H_j \mid E_1 \wedge E_2), \text{ for } i = 3 \text{ or } 7, \text{ and each } j \neq 3 \text{ and } \neq 7. \tag{4}$$

In the graphical representation of the associated argument structure (Figure 5), arrows of differing thickness are used to suggest different strengths.

When suspect 3 is interrogated it turns out that he does not have the tattoo (evidence E_3). So if he is the robber (H_3), the evidence ($E_1 \wedge E_2 \wedge E_3$) implies that he has no tattoo ($\neg T$). This expectation is graphically illustrated in Figure 6.

But recall that we also have the expectation that, if he is the robber, he has the tattoo, on the basis of the surveillance camera evidence. Informally, we conclude that he cannot be the robber.

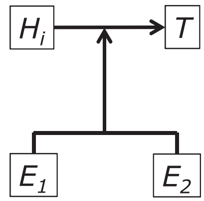

FIG. 4. Evidence 2(a): the surveillance camera.

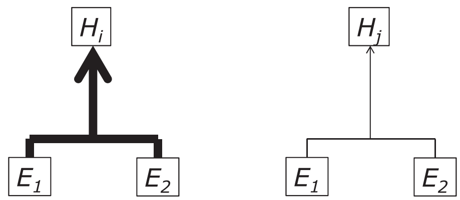

FIG. 5. Evidence 2 (b): the surveillance camera.

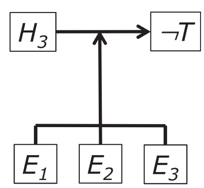

FIG. 6. Evidence 3: the interrogation of suspect 3.

This reasoning has a counterpart in the given formalization. First, we need the representation of the expectation that suspect 3 does not have the tattoo. As a conditional probability, this is expressed like this:

$$\mathrm{p}(\neg T \mid H_3 \wedge E_1 \wedge E_2 \wedge E_3) = 1 \tag{5}$$

If we combine this with the certain expectation that the robber has the tattoo ($\mathrm{p}(T \mid H_3 \wedge E_1 \wedge E_2) = 1$), it follows using formal reasoning[11] that the evidence and hypothesis 3 are incompatible:

$$\mathrm{p}(H_3 \wedge E_1 \wedge E_2 \wedge E_3) = 0 \tag{6}$$

In other words, given the evidence, suspect 3 cannot be the robber. Using the terminology of scenarios: the scenario that he is the robber does not match the evidence.

When suspect 7 confesses (evidence E_4), the probability of the guilt of the other suspects, when compared to that of suspect 7, becomes much lower than before. The confession does not

[11] By $\mathrm{p}(T \mid H_3 \wedge E_1 \wedge E_2) = 1$, it follows that $\mathrm{p}(T \mid H_3 \wedge E_1 \wedge E_2 \wedge E_3) = 1$, unless $\mathrm{p}(H_3 \wedge E_1 \wedge E_2 \wedge E_3) = 0$. But we also have $\mathrm{p}(\neg T \mid H_3 \wedge E_1 \wedge E_2 \wedge E_3) = 1$, contradicting $\mathrm{p}(T \mid H_3 \wedge E_1 \wedge E_2 \wedge E_3) = 1$. So it follows that $\mathrm{p}(H_3 \wedge E_1 \wedge E_2 \wedge E_3) = 0$.

remove all doubt, as it happens more often than one would naively expect that a suspect falsely confesses. Still, suspect 7 being the robber has become much more probable, when compared to the guilt of the others.

Probabilistically, such a change of comparative probability can be expressed by the use of an inequality of ratios of conditional probabilities:

$$\frac{\mathrm{p}(H_i \mid E_1 \wedge \ldots \wedge E_4)}{\mathrm{p}(H_7 \mid E_1 \wedge \ldots \wedge E_4)} < \frac{\mathrm{p}(H_i \mid E_1 \wedge E_2 \wedge E_3)}{\mathrm{p}(H_7 \mid E_1 \wedge E_2 \wedge E_3)}, \text{ for each } i \text{ except 3 and 7.} \tag{7}$$

During his interrogation, suspect 7 reveals the location of the stolen jewelry (J), so if suspect 7 is indeed the robber, we expect that location to be correct. The expectation can be graphically depicted in analogy with our earlier figures concerning the expectation of the tattoo. This time it is a specific expectation, that only is connected to the hypothesis that suspect 7 is the suspect. Probabilistically, we write:

$$\mathrm{p}(J \mid H_7 \wedge E_1 \wedge \ldots \wedge E_4) = 1 \tag{8}$$

The jewelry is indeed found at the mentioned location (evidence E_5). The evidence is so specific that it is considered a certainty that suspect 7 is the robber:

$$\mathrm{p}(H_7 \mid E_1 \wedge \ldots \wedge E_5) = 1 \tag{9}$$

Note that this certainty does not follow from the other elements of the theory about the case. It is an expression of certainty, as modelled in the theory about the case.

As the expectations associated with the truth of the hypothesis were certain, they follow (using formal reasoning):

$$\mathrm{p}(H_7 \wedge T \wedge J \mid E_1 \wedge \ldots \wedge E_5) = 1 \tag{10}$$

Figure 7 illustrates the development of the argument against suspect 7. Initially, there is a weak presumption for his guilt, and there is very un specific information. Finally, the theory about the case expresses certainty about what has happened, as the information uncovered in the evidence is so specific that no reasonable doubt remains. Also, the circumstances of the scenario of what has happened have become more specific, now that they include the tattoo and the location of the jewelry.

6. Integrating arguments, scenarios and probabilities, revisited

In Section 4, we gave an informal overview of the integrated perspective, illustrated in Figure 2. We revisit the integrated perspective, reviewing the formal ideas presented.

- The evidence, scenarios and expectations are expressed in the logical language underlying a probability function. We used *standard probability theory and its underlying classical logical language*. For our purposes, a propositional language was sufficient.
- Information about the evidence, scenarios and expectations is expressed in terms of the properties of a probability function. The properties of the probability function provide a *theory* about the evidence, scenarios and expectations in the case.[12] The theory of our example case is gradually

[12] These properties are expressed in a meta-language about probability functions, in which it is possible to evaluate and compare probabilities. See Halpern (2003) for the logical study of the formal properties of probability functions, and other uncertainty formalisms.

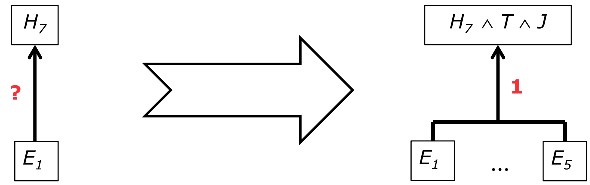

FIG. 7. Evidence 5: from a weak presumption to certainty.

constructed and is expressed by the equations (1) through (10) in Section 5. As in general, many specific numbers will not be available, the probability function used will in general be only partially specified by the theory.

- Given a theory about a case, a logical sentence expresses a *possible* state of affairs when it has positive probability (and *impossible* otherwise), and a *certain* state of affairs when its probability is equal to 1 (*uncertain* otherwise).[13]
- Evidence, scenarios and expectations are expressed using *structured sentences*, typically conjunctions of propositional variables and their negations. In particular, the overall evidence was expressed as the conjunction of the available individual evidence (e.g. $E_1 \wedge E_2 \wedge E_3 \wedge E_4 \wedge E_5$ at the end), and a scenario as the conjunction of individual events (in the example $H_7 \wedge T \wedge J$).
- The *strength of an argument* from premises to conclusions is modeled as the conditional probability of the conclusions given the premises, where premises and conclusions are conjunctively combined. The strength of an argument may or may not follow from a theory about the case. Typically, many strengths will not be modelled in the theory, as in the example where only a few of the actual probabilities of the function p were made specific.
- Given a theory, an argument is *presumptive* when it has positive strength, according to the theory, and *certain* when it has strength 1. Certain arguments are almost monotonic: When an argument from φ to ψ is certain, an argument with more specific premises, say from $\varphi \wedge \varphi'$ to ψ, is also certain, except when $\varphi \wedge \varphi' \wedge \psi$ is not possible. The exception occurs when there is a local inconsistency.
- An uncertain presumptive argument leaves *reason for doubt*, in the sense that the opposite conclusion is also supported. This follows from the following property:

$$\text{If } \mathrm{p}(\psi \mid \varphi) < 1, \text{ then } \mathrm{p}(\neg\psi \mid \varphi) > 0.$$

As a result, a presumptive argument that is not certain (i.e. an argument with positive strength smaller than 1), has an associated presumptive argument for the opposite conclusion. The complementary argument of an uncertain presumptive argument can be regarded as the expression of the doubt associated with the argument. Note that a weak argument (with a strength close to 0), has an associated complementary argument that is strong (close to 1). One may wonder whether this means that for instance a weak alibi implies a strong argument for guilt. This is not the case. A weak alibi provides a weak argument for innocence on the basis of the alibi. There is then a strong

[13] Probability functions serve as the models of a theory about the evidence, scenarios and expectations. So, actually, a sentence expresses a possible state of affairs according to a theory when it has positive probability *in all models* of the theory; and similarly for impossibility, certainty and uncertainty. Cf. the previous footnote that a theory in general does not specify a unique probability function.

argument against innocence on the basis of the alibi. There can be other innocent scenarios and other guilty scenarios, some more strongly, and some more weakly supported.

- Given a presumptive argument from φ to ψ and a presumptive argument from ψ to χ, it is natural to consider chaining—or, if you like, rule application. However, the associated 'argument' from φ to $\psi \wedge \chi$ need not be presumptive, as it is possible that $\varphi \wedge \psi \wedge \chi$ is not possible (has probability zero). If it is presumptive, the strength of the argument will not be larger than that of the argument from φ to ψ. When given presumptive arguments from φ to ψ and from $\varphi \wedge \psi$ to χ, the argument from φ to $\psi \wedge \chi$ that results from chaining is also presumptive. Its strength is the product of the strengths of the two underlying arguments.
- There can be *arguments for incompatible positions*, in particular for *incompatible scenarios*. In the example, we saw presumptive arguments for the guilt of each of eight usual suspects. The hypotheses about these suspects' guilt were mutually incompatible. The conclusions of presumptive arguments cannot be conjunctively combined. For instance, E_1 is a presumptive argument for the guilt of suspect 1 expressed as H_1 and of suspect 2, H_2, but E_1 is not a presumptive argument for $H_1 \wedge H_2$.
- The *coherence of a scenario* is expressed by a positive probability, meaning that the scenario is possible. A higher probability expresses a more plausible scenario. There is a formal connection between the coherence of scenarios and the existence of presumptive arguments. On the one hand, there is a presumptive argument from premises φ to conclusions ψ if and only if the conditional probability of ψ given φ is positive. On the other hand, a scenario $\varphi \wedge \psi$ is coherent if and only if the probability of $\varphi \wedge \psi$ is positive. But $\mathrm{p}(\psi \mid \varphi) > 0$ is equivalent to $\mathrm{p}(\varphi \wedge \psi) > 0$. So a scenario $\varphi \wedge \psi$ is coherent if and only if the argument from premises φ to conclusions ψ is presumptive. And since the coherence of a scenario does not depend on the order of its elements, it is possible to switch the premises and the conclusions, without losing the presumptiveness of the argument. Of course the strength of the argument (if known) would change by such a switch in accordance with Bayes' rule.
- Additional information can *change the strength* of an argument. Adding a premise can make the argument stronger or weaker. For instance, by the gradually collected additional evidence, the initial presumptive argument for the guilt of usual suspect 2 on the basis of E_1 is first weakened by evidence E_2 (although the precise values are not made explicit), and ends with strength 0. The argument for the guilt of usual suspect 7 is strengthened by E_2 and ends with strength 1. Note that the gradually constructed theory has not involved a change of the properties of the probability function; only gradually more information about the function p has been incorporated in the theory.
- By gradually collecting further evidence, possible *scenarios can be excluded*. Exclusion of a scenario is formally expressed by a probability of 0 of the combination of the scenario and the evidence. An example is equation (6) of the theory of the robbery case. The *expectations* about possible new evidence based on possible scenarios can guide the investigative process. We have seen the role of the tattoo in the robbery case.
- A scenario is certain when the negation of the scenario is not possible, according to the theory about the case and the evidence. The negation of the scenario encompasses competing scenarios of what can have happened. In this situation, which will in practice not always be achieved, the *case is solved*, in the sense that the evidence has identified the scenario of what has happened, given the theory and the observed evidence. Possible remaining doubt is about the theory and the observation of the evidence, for instance whether the theory about the case has been carefully and openly constructed, considering all information as it is provided by the two sides.

If we revisit the characteristics of arguments, scenarios and probabilities, as they were summarized in Table 1, we see that the *adversarial setting* of arguments for and against positions has found its place in the integrated perspective by the arguments for incompatible positions. Also it is possible to strengthen and weaken arguments by adding further evidence. The *global coherence* provided by scenario methods has been incorporated in the perspective by considering scenarios as composite sentences, here conjunctions of variables and their negations. *Gradual uncertainty* has been approached in the integrated perspective by the use of probability theory as formal background. Probability theory with its underlying logic serves as the *normative framework* of the integrated perspective.

7. Discussion

Summarizing, the proposed integrated perspective uses probability theory as the normative framework in which reasoning with arguments and with scenarios are regulated. Gradually a theory about the case, the evidence, the hypotheses and the expectations is constructed. The elements of such a theory are expressed as properties of a probability function, which typically remains only partially specified by the theory of a case. The strength of an argument from premises to conclusions is measured by a number that formally behaves like the conditional probability of the conclusions given the premises. Scenarios are modelled as composite sentences expressing the conjunction of the facts and events of the scenario. We now discuss how the present approach compares to other proposals.

Argumentation models have typically been developed in contrast with classical logic and probability theory, whereas we have used standard probability theory with its underlying classical logic. The modelling of defeasible reasoning, in which arguments can become defeated in light of new information, has been a driving force in the study of alternative formalisms. For instance, Pollock (1987, 1995) presents his OSCAR model by carefully explaining his philosophical reasons for deviating from classical logic and probability theory. Some recent proposals aim to connect to classical logic and probability theory (e.g. Besnard and Hunter, 2009; Hunter, 2012).

In the computational study of argumentation (Rahwan and Simari, 2009), much attention has been paid to the formal relations between argument structure and argument evaluation. Several definitions have been proposed for how argument structure determines argument defeat. It has turned out that there is a wealth of reasonable proposals. For instance, Dung's (1995) seminal work has shown how a focus on just argument attack already can give rise to several relevant 'argumentation semantics'. Ensuing research showed that a wealth of further proposals can be made, in particular also when also argument support is included [e.g. Verheij, 2003, where 11 formal options are shown (p. 341), some more reasonable than others]. As a result of this diversity, some proposals are presented as a framework with the argumentation semantics as one of the parameters (e.g. Prakken, 2010). The present proposal takes a sidestep from this debate, by not so much focusing on how argument structure determines argument defeat. Arguments are strong, or even certain, depending on the theory about the case at hand. As a result, one could say that it is the world that determines which arguments are good. The formal semantics (here provided by standard probability theory with its underlying classical logic) merely serves as a regulative mechanism. The treatment of the example showed several examples of how argument structure interacts with argument evaluation, in the present proposal.[14]

[14] The figures have been styled in a way that is compatible with the argument structure diagrams presented in (Verheij, 2005, 2007).

Argumentation models can be regarded as a special example of models of non-monotonic and uncertain reasoning (Gabbay *et al.*, 1994; Halpern, 2003). These have typically focused on deviations from classical logic and probability theory. The present proposal also models defeasible uncertain reasoning, in the specific context of establishing the facts. Uncertainty about conclusions is modeled by the different, incompatible positions that can be supported given certain premises. The degree of uncertainty of an argument from premises to conclusions is measured using a conditional probability. Formally, a distinguishing characteristic of the present proposal is that our presumptive arguments do not validate what is called the (And)-rule in the literature: if the premises support two positions, it is not always the case that the conjunction of the positions is also supported.[15] This happens when the positions are incompatible. The robbery example shows that different incompatible positions, each supported, but not in conjunction, are quite common.

Non-monotonic reasoning has two sides in the present proposal. The first is the non-monotonicity that is associated with local inconsistency. An example is that new evidence can show that an initially supported hypothesis is no longer supported. In the robbery example, we saw how the different hypotheses about the guilt of the usual suspects were gradually weeded out, until there was certainty about the guilt of suspect 7. The inconsistency involved is local in the sense that only the 'space of possibilities' is pruned, while the theory of the case remains consistent. The second side of non-monotonicity is associated with global inconsistency. It can happen that new evidence makes one conclude that the whole theory about a case is inconsistent. For instance, in the robbery example, it is thinkable that after it was concluded that suspect 7 is the robber, new evidence shows that he cannot have committed the crime after all, e.g. since he was undergoing hospital surgery at the time. Since his guilt was modelled as a certainty, a global inconsistency has arisen. In such a situation, the theory itself needs to be changed, and not just the conclusions drawn on the basis of the theory.[16]

New possibilities need to be sought for and then incorporated, which may prove hard to discover. For instance, here it could turn out that suspect 7 has wanted to defend his twin brother. A new possibility must be taken into account, expressed by a new positive probability of a hypothesis.

Scenario modelling is connected to coherence approaches to reasoning. For instance, Thagard's model of explanatory coherence (Thagard, 2001, 2004, an application to evidence) is of this kind. His approach is anti-probabilistic, and uses connections between propositions that are not directional. Interestingly, in the present approach, coherent scenarios correspond to presumptive arguments in two directions, de-emphasizing the directionality of arguments (Section 6). However, since our approach is probabilistic, the strength of presumptive arguments changes with their direction (in accordance with Bayes' rule). Another anti-probabilistic approach is Haack's (2012) foundherentist approach in which a balance is struck between coherentist and foundationalist epistemology. Pardo and Allen propose an inference to the best explanation approach, in which explanations play a role similar to scenarios (Pardo and Allen, 2008).

The modelling of scenarios has been studied probabilistically in the context of Bayesian networks (e.g. Keppens, 2012; Fenton *et al.*, 2013; Vlek *et al.*, 2013, inspired by Hepler *et al.*, 2007). Since a Bayesian network is a representation of a full probability function, a recognized issue is how to acquire all numbers needed. In the present proposal, a theory about a case provides (in general) only partial

[15] The (And)-rule has led to considerable debate in the literature, in particular also in connection with the so-called lottery paradox. Without going into details here, the reader is referred to, e.g. Kyburg (1997); Pollock (1995), and more recently, Hawthorne and Makinson (2007). A formally oriented discussion is in Verheij (2012), making connections to the general properties of non-monotonic inference (Makinson, 1994).

[16] In the formal literature, the distinction is referred to as that between defeasible reasoning and belief revision.

information about a probability function. As a result, only the numeric information that is actually available and meaningful is modelled. The example emphasized the relevant role of qualitative information about probabilities, for instance, the comparison of certain values. Scenarios have been studied in connection with argumentation (Bex, 2011; Bex *et al.*, 2010; Bex and Verheij, 2012). In this work, special attention is paid to the role of commonsense knowledge about the world, in the form of scenario schemes (discussed in relation to Schank's scripts (Schank and Abelson, 1977), causal and evidential generalizations, and argument schemes (see also Walton, 2008). The approach is non-probabilistic, and does not commit to a formal semantics, as its formalization is based on the argumentation framework (Prakken, 2010). In legal psychology, the role of scenarios has been studied by Pennington and Hastie (1993). Their empirical results showed that a position is more easily believed when evidence is presented in the order of a coherent story, than when presented in a random order. Wagenaar *et al.* (1993) presented a normative theory of reasoning with scenarios, Anchored Narratives Theory, using it to analyse dubious cases and miscarriages of justice. For instance, one normative maxim is that alternative scenarios should be considered to avoid tunnel vision. The theory is presented informally, and is contrasted with logical and probabilistic thinking. The present proposal shows how scenario reasoning can be embedded in a logico-probabilistic setting.

8. Conclusion

In this article, a perspective on reasoning with arguments, scenarios and probabilities has been proposed. It has been applied to the setting of evidential reasoning in order to establish the facts about a crime. Characteristics of the proposal are as follows:

(1) Standard probability theory with its underlying classical logic is used as the formal background. As a result, the normative role of the probabilistic framework is kept. Typical fallacies such as the prosecutor's fallacy are also in the present framework fallacious. Standard formulas such as Bayes' rule and the likelihood ratio formula obtain, and can be used inferentially when the relevant numbers are available.

(2) The information about the evidence, arguments, scenarios and probabilities is encoded in a theory about the case and the relevant knowledge. This information can be quantitative or qualitative in nature. In general, a theory about a case will not describe a full probability function. This addresses the commonly recognized issue that probabilistic approaches seem to require more numbers than are available. In the perspective presented here, a theory about a case models both general and case-specific information about the case.

(3) The strength of an argument from premises to conclusions is measured by a number that behaves formally like a conditional probability. In a theory about a case, not all arguments will actually have a strength. An argument is presumptive, when it has positive strength. It is uncertain, when it has strength lower than 1. A certain argument has strength 1. Then there is no reason for doubt: there is no support for the opposite of the argument's conclusion.

(4) Scenarios are modeled as composite sentences that express the facts and events of the scenario. A scenario is coherent, when it is possible, i.e. has positive probability, according to the theory about the case. The evidence can presumptively support incompatible scenarios. The conjunction of incompatible scenarios is not supported by a presumptive argument, as the conjunction of incompatible scenarios has probability 0. The aim is to establish with certainty a scenario that is exactly as specific as is needed to answer the legally relevant issues about a case.

(5) In the proposal, there is reasonable doubt about a conclusion, when the conclusion is not certain, given the evidence and the theory about the case. When a conclusion is uncertain, the conclusion's negation is still considered possible. Further evidence will have to be sought that excludes the alternative possibility. In this way, an answer is provided to a puzzle in probabilistic reasoning: when is the probability of a conclusion sufficiently high? The present answer is: the theory about the case should make the conclusion about what has happened certain. As such, the theory expresses a commitment to a certain conclusion on the basis of the theory about the case.

(6) When a conclusion is established 'beyond a reasonable doubt', there still is room for doubt: the theory about the case, which includes the observation of the evidence, can be wrong. Hence, possible remaining doubt is about the theory about the case itself. This uncertainty is the burden that criminal investigators and fact-finders (such as juries and judges) will always face. They can only reduce the uneasy feeling of this 'unreasonable doubt' by making sure that the theory and evidence in a case are collected in a process of critical, careful and open-minded scrutiny, and has been performed with appropriate effort.

In sum, a normative framework for evidential reasoning has been given that balances qualitative and quantitative methods. Probability theory and its underlying logic provide the formal background, and are used to regulate reasoning with arguments and scenarios. Numbers can do their work, when relevant and meaningful, and at the same time arguments and scenarios can be used safely.

One would hope that the presented perspective on evidential reasoning can prevent mistakes made by criminal investigators, forensic experts and fact-finders. Reasons for this hope are that the proposal is normative, hence can help prevent errors in reasoning. Also, the proposal integrates arguments, scenarios and probabilities, which can help prevent communication errors between experts in either qualitative or quantitative reasoning methods (such as on the one side juries and judges, and on the other forensic statisticians). Efforts can be steered towards the prevention of avoidable errors of reasoning and communication. It is however a fact of life that criminal investigators, forensic experts and fact-finders face a tough task. The proposal emphasizes what is already known: that the construction of a balanced theory about the case and the evidence is where all efforts should go. Tunnel vision, sloppy reasoning, lack of an open mind, an insufficient investigative effort, an insufficiently critical scrutiny of the theory about the case; all of these are among the pitfalls that we as humans can step into.

But even when no avoidable errors have been made, there is always the possibility that a decision is proven wrong after it has been made. This can occur also in the most carefully made investigations and decisions. In the perspective presented here, such unavoidable, post-decision errors involve new possibilities that have become uncovered, and that were not even considered possible before.

Acknowledgements

The research reported in this article has been performed in the context of the project 'Designing and Understanding Forensic Bayesian Networks with Arguments and Scenarios', funded in the NWO Forensic Science program (http://www.ai.rug.nl/∼verheij/nwofs/). The text was completed in the academic year 2013/2014, while the author was a resident fellow at the CodeX centre for legal informatics at Stanford University, on a sabbatical leave, made possible by the Institute of Artificial Intelligence and Cognitive Engineering (ALICE), University of Groningen. The author has benefited from discussions with many, and in particular with the participants of the mentioned project, and with Marcello Di Bello and his students of probability and law.

References

Aitken, C. G. G. & Taroni, F. (2004). *Statistics and the Evaluation of Evidence for Forensic Scientists, 2nd Edition*. Wiley, Chichester.

Anderson, T., Schum, D. & Twining, W. (2005). *Analysis of Evidence, 2nd Edition*. Cambridge University Press, Cambridge.

Bench-Capon, T. J. M. (2003). Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, **13** (3), 429–448.

Besnard, P. & Hunter, A. (2009). Argumentation based on classical logic. In: *Argumentation in Artificial Intelligence* (I. Rahwan & G. R. Simari eds.). 133–152. Springer, Berlin.

Bex, F. J. (2011). *Arguments, Stories and Criminal Evidence: A Formal Hybrid Theory*. Springer, Berlin.

Bex, F. J., Van Koppen, P. J., Prakken, H. & Verheij, B. (2010). A hybrid formal theory of arguments, stories and criminal evidence. *Artificial Intelligence and Law*, **18**, 1–30.

Bex, F. J. & Verheij, B. (2012). Solving a murder case by asking critical questions: an approach to fact-finding in terms of argumentation and story schemes. *Argumentation*, **26**, 325–353.

Broeders, T. (2009). Decision making in the forensic arena. *Legal Evidence and Proof: Statistics, Stories, Logic*. 71–92. Ashgate, Farnham.

Buchanan, M. (2007). Conviction by numbers. *Nature*, **445** (10), 254–255.

Dawid, P., Twining, W. & Vasiliki, M. *Evidence, Inference and Enquiry*. Oxford University Press, Oxford.

Derksen, T. & Meijsing, M. (2009). The fabrication of facts: the lure of the incredible coincidence. In: *Legal Evidence and Proof: Statistics, Stories, Logic* (H. Kaptein, H. Prakken & B. Verheij eds.). 39–70. Ashgate, Farnham.

Dung, P. M. (1995). On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, **77**, 321–357.

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

Fenton, N. E. (2011). Science and law: Improve statistics in court. *Nature*, **479**, 36–37.

Fenton, N. E. & Neil, M. D. (2013). *Risk Assessment and Decision Analysis with Bayesian Networks*. CRC Press, Boca Raton.

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

Gabbay, D. M., Hogger, C. J. & Robinson, J. A. *Handbook of Logic in Artificial Intelligence and Logic Programming. Volume 3. Non-monotonic Reasoning and Uncertain Reasoning*. Clarendon Press, Oxford.

Haack, S. (2012). The embedded epistemologist: dispatches from the legal front. *Ratio Juris*, **25**, 206–235.

Hájek, A. (2011). Interpretations of probability. In: *The Stanford Encyclopedia of Philosophy* (E. N. Zalta ed.). Stanford University, Stanford.

Halpern, J. (2003). *Reasoning About Uncertainty*. The MIT Press, Cambridge (MA).

Hawthorne, J. & Makinson, D. (2007). The quantitative/qualitative water-shed for rules of uncertain inference. *Studia Logica*, **86** (2), 247–297.

Hepler, A. B., Dawid, A. P. & Leucari, V. (2007). Object-oriented graphical representations of complex patterns of evidence. *Law, Probability and Risk*, **6** (1–4), 275–293.

Hunter, A. (2012). A probabilistic approach to modelling uncertain logical arguments. *International Journal of Approximate Reasoning*, **54**, 47–81.

Jensen, F. V. & Nielsen, T. D. (2007). *Bayesian Networks and Decision Graphs*. Springer, Berlin.

Kadane, J. B. & Schum, D. A. (1996). *A Probabilistic Analysis of the Sacco and Vanzetti Evidence*. Wiley, Chichester.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Penguin, London.

Kaptein, H., Prakken, H. & Verheij, B. *Legal Evidence and Proof: Statistics, Stories, Logic (Applied Legal Philosophy Series)*. Ashgate, Farnham.

Keppens, J. (2012). Argument diagram extraction from evidential Bayesian networks. *Artificial Intelligence and Law*, **20**, 109–143.

Kirschner, P. A., Shum, S. J. B. & Carr, C. S. (2003). *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*. Springer, Berlin.

Kyburg, H. (1997). The rule of adjunction and reasonable inference. *Journal of Philosophy*, **94**, 109–125.

Makinson, D. (1994). General patterns in nonmonotonic reasoning. In: *Handbook of Logic in Artificial Intelligence and Logic Programming. Volume 3. Nonmonotonic Reasoning and Uncertain Reasoning* (D. M. Gabbay, C. J. Hogger & J. A. Robinson eds.). 35–110. Clarendon Press, Oxford.

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

Pennington, N. & Hastie, R. (1993). Reasoning in explanation-based decision making. *Cognition*, **49** (1–2), 123–163.

Pollock, J. L. (1987). Defeasible reasoning. *Cognitive Science*, **11** (4), 481–518.

Pollock, J. L. (1995). *Cognitive Carpentry: A Blueprint for How to Build a Person*. The MIT Press, Cambridge (MA).

Prakken, H. (2010). An abstract framework for argumentation with structured arguments. *Argument and Computation*, **1** (2), 93–124.

Rahwan, I. & Simari, G. R. *Argumentation in Artificial Intelligence*. Springer, Dordrecht.

Schank, R. & Abelson, R. (1977). *Scripts, Plans, Goals and Understanding, an Inquiry into Human Knowledge Structures*. Lawrence Erlbaum, Hillsdale.

Schneps, L. & Colmez, C. (2013). *Math on Trial: How Numbers Get Used and Abused in the Courtroom*. Basic Books, New York.

Shen, Q., Keppens, J., Aitken, C., Schafer, B. & Lee, M. (2006). A scenariodriven decision support system for serious crime investigation. *Law, Probability and Risk*, **5**, 87–117.

Taroni, F., Aitken, C., Garbolino, P. & Biedermann, A. (2006). *Bayesian Networks and Probabilistic Inference in Forensic Science*. Wiley, Chichester.

Thagard, P. (2001). *Coherence in Thought and Action*. The MIT Press, Cambridge (MA).

Thagard, P. (2004). Causal inference in legal decision making: explanatory coherence vs. Bayesian networks. *Applied Artificial Intelligence*, **18**, 231–249.

Thompson, P. (2013). Forensic DNA evidence. The myth of infallibility. *Genetic Explanations: Sense and Nonsense*. 227–255. Harvard University Press, Cambridge (MA).

Thompson, W. C. & Schumann, E. L. (1987). Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and the defense attorney's fallacy. *Law and Human Behavior*, **11**, 167–187.

Tillers, P. (2011). Trial by mathematics — reconsidered. *Law, Probability and Risk*, **10**, 167–173.

Timmer, S., Meyer, J. J., Prakken, H., Renooij, S. & Verheij, B.. (2013). Inference and attack in Bayesian Networks. In: *25th Benelux Conference on Artificial Intelligence (BNAIC 2013)* (K. Hindriks, M. de Weerdt, B. van Riemsdijk & M. Warnier eds.). 199–206. Delft University, Delft.

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

van Koppen, P. J. (2012). *Overtuigend Bewijs: Indammen van Rechterlijke Dwalingen. (Convincing Evidence: Damming Judicial Errors)*. Nieuw Amsterdam, Amsterdam.

van Koppen P. J. (2013). Spoiled judges and the fluidity of evidence in criminal cases. In: *Fourteenth International Conference on Artificial Intelligence and Law. Proceedings of the Conference*. xxii–xxiv. ACM Press, New York.

Verheij, B. (2003). DefLog: on the logical interpretation of prima facie justified assumptions. *Journal of Logic and Computation*, **13** (3), 319–346.

Verheij, B. (2005). *Virtual Arguments. On the Design of Argument Assistants for Lawyers and Other Arguers*. T.M.C. Asser Press, The Hague.

Verheij, B. (2007). Argumentation support software: boxes-and-arrows and beyond. *Law, Probability and Risk*, **6**, 187–208.

VERHEIJ, B. (2012). Jumping to conclusions. A logico-probabilistic foundation for defeasible rule-based arguments. In: *13th European Conference on Logics in Artificial Intelligence, JELIA 2012. Toulouse, France, September 2012. Proceedings (LNAI 7519)* (L. Farinas del Cerro, A. Herzig & J. Mengin eds.). 411–423. Springer, Berlin.

VLEK, C., PRAKKEN, H., RENOOIJ, S. & VERHEIJ, B. (2013). Modeling crime scenarios in a Bayesian Network. *The 14th International Conference on Artificial Intelligence and Law (ICAIL 2013). Proceedings of the Conference*. 150–159. ACM Press, New York.

WAGENAAR, W. A., VAN KOPPEN, P. J. & CROMBAG, H. F. M. (1993). *Anchored Narratives. The Psychology of Criminal Evidence*. Harvester Wheatsheaf, London.

WALTON, D. N., REED, C. & MACAGNO, F. (2008). *Argumentation Schemes*. Cambridge University Press, Cambridge.

WIGMORE, J. H. (1931). *The Principles of Judicial Proof or the Process of Proof as Given by Logic, Psychology, and General Experience, and Illustrated in Judicial Trials, 2nd Edition*. Little, Brown and Company, Boston (MA).