Evidential Reasoning

Marcello Di Bello and Bart Verheij

When a suspect appears in front of a criminal court, there is a high probability that he will be found guilty. In the USA, statistics for recent years show that the conviction rate in federal courts is roughly 90%, and in Japan reaches as high a rate as 99%.[1] In the UK, the numbers are slightly lower, with a conviction rate of roughly 80%, while in the Netherlands the conviction rate is around 90%.[2] This does not mean that the fact finders deciding about the facts of a case have an easy job. Whether laypeople, such as jury members selected from the general public, or professionals, often experienced judges having completed postgraduate education, all face the difficulties associated with handling the evidence that is presented in court. What to do with conflicting testimonies? Does an established DNA match outweigh the testimony that the suspect was not seen at the crime scene? How to coherently interpret a large body of evidence? When is there enough evidence to convict?

The primary aim of this chapter is to explain the nature of evidential reasoning, the characteristic difficulties encountered, and the tools to address these difficulties.

[1] On the conviction rate in US federal courts, see the statistical reports of the Offices of the United States Attorneys, available at www.justice.gov/usao/resources/annual-statistical-reports. Most of these convictions are guilty pleas, not convictions after trial. On Japan's conviction rate, see *White Paper on Crime 2014*, Part 2, Chap. 3, Sect. 1, available at http://hakusyo1.moj.go.jp/en/63/nfm/mokuji.html.

[2] On the UK conviction rate, see *Criminal Justice Statistics–March 2014*, available at www.gov.uk/government/statistics. As in the US case, the rate include mostly guilty pleas. For the Netherlands, see CBS, the Dutch central bureau of statistics, publishing its data at www.cbs.nl.

M. Di Bello (✉)
Lehman College - City University of New York, Bronx, USA
e-mail: marcello.dibello@lehman.cuny.edu

B. Verheij
Faculty of Science and Engineering, University of Groningen, Groningen, The Netherlands
e-mail: bart.verheij@rug.nl

G. Bongiovanni et al. (eds.), *Handbook of Legal Reasoning and Argumentation*,
https://doi.org/10.1007/978-90-481-9452-0_16

Our focus is on evidential reasoning in criminal cases. There is an extensive scholarly literature on these topics, and it is a secondary aim of the chapter to provide readers the means to find their way in historical and ongoing debates.

This chapter does not aim to offer legal practitioners, lawyers, judges and expert witnesses, practical tools that can be immediately used to better litigate a case. But we hope that practitioners interested in the theoretical underpinnings of evidential reasoning in court will benefit from reading this chapter. And, more generally, philosophers, legal scholars, statisticians, logicians, and those scholars and practitioners interested in the theoretical aspects of reasoning with evidence will hopefully find this chapter an interesting resource and point of departure for further thinking on the matter.

1 Setting the Stage

We set the stage by using two important and often encountered kinds of evidence as an illustration: eyewitness testimony and DNA profiling. These two kinds of evidence will be used to establish a list of central questions that structure the exposition that follows.

1.1 Eyewitness Testimony

Eyewitness testimony has always been a central source of information in criminal proceedings. It typically takes the form of oral statements by the witness in court, in response to questions by the prosecution, the defense, the court, and sometimes, albeit rarely, the jury. Eyewitness testimony can also come in the form of reports of oral examinations written in the pre-court stages of the criminal investigation.

Eyewitness testimony can provide information about what happened on the scene of the crime. Here is an example.

> Q: Can you describe what happened that day?
>
> A: I was in the park and suddenly heard a lot of noise, close by. I saw two men quarreling, shouting. Suddenly one of them pulled a gun, and I heard a shot. The other man fell to the ground. The shooter looked around, looked me in the eye, and then started to run.
>
> Q: Can you describe the shooter?
>
> A: He was a young men, in his twenties, I think. Tall, blonde, with a white skin, and unusually blue eyes. He looked unhealthy, with bad teeth, like a drug addict. He was wearing a perfectly ironed shirt, which surprised me.

The information contained in the testimony can be more or less detailed, and on its basis, the fact finders can form a hypothesis about what happened. Still, it remains a hypothesis. There are many reasons why the hypothetical events reconstructed on the basis of the testimony might not be true. Typical reasons against the truth of the

events reported by an eyewitness include that the witness wrongly interpreted what she saw, that time distorted her memories, or that the witness is lying.

1.2 DNA Evidence

DNA evidence has become very common in criminal cases. Perpetrators sometimes leave traces of themselves and their actions, such as pieces of hair, skin tissues, drops of blood, or other bodily fluids. By using forensic DNA technology, a genetic profile associated with the crime traces can be created, and if this profile matches with an individual's profile, this establishes a link, at least *prima facie*, between the matching individual and the crime.

What is a DNA profile? A DNA profile is determined by analyzing a number of specific locations, the so-called *loci*, of a DNA molecule. Different countries use different sets of *core loci* for their DNA profiles. For instance, the CODIS system in the USA uses 20 core loci.[3] At each specific locus, a different *allele* might occur. A core locus that is often used, called CSF1PO, has up to 16 allele types, depending on how often the molecular sequence AGAT is repeated at that location.[4] A DNA profile, then, consists in a list of allele types for a certain number of select core loci.

The evidential relevance of a DNA profile stems from the fact that although most of the structure of the DNA molecule is shared among all human beings (more than 99%), the select core loci used to construct DNA profiles are highly specific. To be sure, DNA profiles need not be unique, but their proportional frequency in a reference population is expected to be very low. So, how is this low number arrived at?

Many countries have created extensive reference databases that contain millions of DNA profiles, and these are used to assess the rarity of a profile. This is a two-step process. First, the number of occurrences of each allele at each core locus in the reference database is counted. This gives a measure of the proportional frequency of each allele at each core locus in the population. Second, the measured proportional frequencies for the alleles at the core loci are multiplied. This allows us to assess the overall proportional frequency of the DNA profile or to use a more common terminology, the *Random Match Probability*. More recently, the terminology *Conditional Genotype Probability* has also been introduced. The sets of core loci have been chosen such that Random Match Probabilities, or Conditional Genotype Probabilities, are typically small, for instance, in the order of 1 in 50 billion, amply exceeding the number of people on our planet.

There is discussion about whether we can reasonably affirm such low numbers and if it makes sense to report them. A key assumption underlying the model—used when multiplying the measured proportional frequencies of specific alleles—is that there are no dependencies among the alleles at different loci in the population considered. This assumption does not always hold, for instance, in a population with

[3] See www.fbi.gov/services/laboratory/biometric-analysis/codis.

[4] See www.cstl.nist.gov/strbase/str_CSF1PO.htm.

family relations. Scientists have also established certain dependencies among the profiles within ethnic groups. Testing the independence assumption can be hard and would require the assessment of more profiles than reasonably possible.

With this background in place, suppose now that a trace of blood—which allegedly came from the perpetrator—is found at the crime scene and that the DNA profile created from the trace matches the DNA profile of a suspect. The match lends support to the hypothesis that the suspect—as opposed to an unknown individual—is the source of the blood trace, and the Random Match Probability associated with the profile provides a measure of the evidential strength of the match. Importantly, the hypothesis that a DNA match can support is rather circumscribed. It is limited to the suspect being the *source* of the trace and should not be confused with the hypothesis that the suspect is *factually guilty*, at least absent other information about how the trace got there.[5] Further, given the declared match, the hypothesis itself that the suspect is the source need not be true. We should always be wary of possible laboratory errors and false positive matches. And even if no laboratory error occurs, the suspect and the perpetrator, while different individuals, might share the same DNA profile, either because they are identical twins or because, though unrelated, they happen to share the same profile by sheer coincidence.[6]

1.3 Central Questions

Using the two kinds of evidence as an illustration, we now provide a list of central questions about evidential reasoning in the law. These questions will structure the discussion that follows:

Question 1: How should we understand conflicts between pieces of evidence? Legal disputes often occur because the evidence provides conflicting perspectives on the crime. For instance, a witness claims that the criminal has blond hair, but the suspect whose DNA matches the traces at the crime scene has dark hair. How should we understand conflicts between pieces of evidence? What are the different ways in which such conflicts arise?

Question 2: How should we handle the strength of the evidence? Some evidence is stronger than other evidence. This is most obvious in the case of DNA evidence, where DNA profiles are associated with different Random Match Probabilities or Conditional Genotype Probabilities. But also some eyewitness testimonies are stronger than others. For instance, the description of a criminal by a witness who could only view the crime scene in bad lighting conditions is of lesser value. How to address the strength of evidence?

[5] We use the terms "factually guilty" or simply "guilty" to express factual guilt, which is not the same as the legal verdict of guilt.

[6] At a rate of a dozen or more twin births per 1000 live births, identical twins are not that rare. Source https://en.wikipedia.org/wiki/Twin#Statistics.

Question 3: How should we coherently interpret the available evidence? A DNA match can support the claim that the suspect is the source, and a witness can add information about how the crime was committed. In general, there is a lot of evidence that needs to be coherently combined in order to make sense of what has happened. How do we combine all information in a coherent whole?

Question 4: How should we decide about the facts given the evidence? When are we done? After a careful and exhaustive investigation in the pretrial and trial phases of the criminal proceedings, the question arises of when a decision can be made and what that decision is. When is the burden of proof met? What is "proof beyond a reasonable doubt"?

The plan is as follows. In the next section (Sect. 2), we discuss three normative frameworks that can help us understand how to correctly handle the evidence. In the remaining sections, we discuss the four questions we set out above in light of the three frameworks and their distinctive features (Sects. 3, 4, 5 and 6).

2 Three Normative Frameworks

In this section, we discuss in broad outline three normative frameworks for the assessment of the evidence presented in a case: arguments, probabilities, and scenarios. Those frameworks constitute systematic and well-regulated methods for examining, analyzing, and weighing the evidence. In this section, we shall only briefly emphasize their distinctive theoretical strengths. Arguments can naturally capture the dialogical dimension, by modeling relations of support and attack. These are the issues raised by the first question above about conflicting pieces of evidence. Probabilities are better suited to quantify the value of the evidence. This is the issue raised by the second question about different pieces of evidence having different strengths. Finally, scenarios are best in offering a coherent and holistic interpretation of large bodies of evidence. These are the issues raised by the third question above about combing different pieces of evidence in a coherent whole. Neither framework, by itself, is well suited to address the fourth question about reaching a decision on the basis of the evidence. As we shall see toward the end, each framework will have to be supplemented with a decision-theoretic component.

2.1 Arguments

The first normative framework that we discuss uses arguments as its primary tool. Arguments are best analyzed in a dialogical setting, for they contain reasons that *support* or *attack* a certain conclusion of interest. For instance, when a witness reports that she saw the suspect at the crime scene, this evidence constitutes a reason for the conclusion that the suspect was, in fact, at the crime scene. But if the DNA profile found at the crime scene does not match the suspect's DNA profile, this constitutes

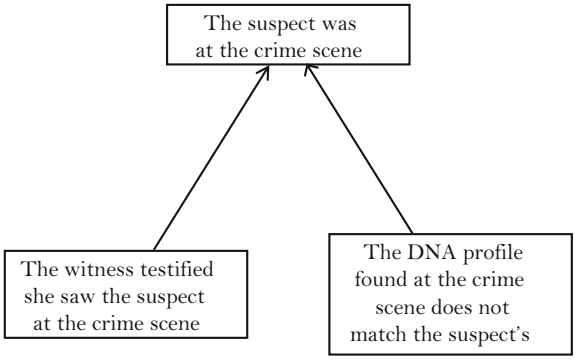

Fig. 1 Arguments with supporting and attacking reasons

a reason attacking the conclusion. An argument with a supporting and an attacking reason is represented in Fig. 1.

The analysis of the structure of arguments goes back to the early twentieth century when John Henry Wigmore (1913) developed his famous evidence charts. The work by Anderson et al. (2005) continued from Wigmore's insights. Independently, and not focusing on evidence in criminal cases, the structure of arguments for and against conclusions was formalized and studied computationally by the philosopher John Pollock (1987, 1995). Pollock's work stimulated an extensive literature on the formal and computational study of arguments (van Eemeren et al. 2014a).

2.2 Probabilities

The second normative framework uses probabilities as its primary tool. In handling evidence in court, a crucial question from the probabilistic perspective is, how probable is a certain hypothesis H given a body of evidence E? This is the *conditional probability* of H given E, or in symbols, $\Pr(H|E)$. Another crucial question is, how does the probability of H change in light of evidence E? This *probability change* is expressed by the difference between the so-called posterior probability $\Pr(H|E)$ and prior probability $\Pr(H)$. Both questions can be addressed with Bayes' theorem:

$$\Pr(H|E) = \frac{\Pr(E|H)}{\Pr(E)} \cdot \Pr(H).$$

This formula—which can be easily proven from the probability axioms—shows how the posterior probability $\Pr(H|E)$ of hypothesis H given evidence E can be computed by the prior probability $\Pr(H)$ and the factor $\Pr(E|H)/\Pr(E)$.[7]

The interest in probabilistic calculations as a tool for the good handling of the evidence has recently been stimulated by the statistics related to DNA profiling and by some infamous miscarriages of justice that involved statistics, in particular the Lucia de Berk and Sally Clark cases (Dawid et al. 2011; Fenton 2011; Schneps and Colmez 2013). The interest is not new (Finkelstein and Fairley 1970; Tillers 2011) and can in fact be traced back to early developments of probability theory (Bernoulli 1713; Laplace 1814) and forensic science in the late nineteenth century (Taroni et al. 1998). To what extent probabilistic calculations have a place in courts has always been, and remains, the subject of debate (Fenton et al. 2016; Tribe 1971).

2.3 Scenarios

Finally, the third normative framework centers around scenario analysis. In a scenario, a coherent account of what may have happened in a case is made explicit. Scenario analysis proves helpful when considering a complex case and its evidence. For instance, the following brief scenario can help to make sense of a murder case:

> The robber killed the victim when caught during a robbery but lost a handkerchief.

This scenario can make sense of a number of facts, for example, that no one in the victim's circle of acquaintances is a possible suspect; that there are signs someone broke into the victim's apartment; and that a handkerchief was found on the floor although it does not belong to the victim. Such a unifying explanation in the form of a scenario can be regarded as a sense-making tool for handling cases with a large dossier.

Legal psychology has contributed to our knowledge about the role of scenarios in handling the evidence (Bennett and Feldman 1981; Pennington and Hastie 1993b). Scenario analysis is also connected with inference to the best explanation (Pardo and Allen 2008). Scenarios, however, can be misleading. Experiments have shown that a false scenario told in a sensible chronological order can be more persuasive than a true scenario whose events are told in a random order. Still, the legal psychologists Wagenaar et al. (1993) have emphasized the usefulness of scenario analysis for the rational handling of the evidence. In their work, they use scenario analysis for debunking dubious case decisions.

---

[7] Bayes' theorem can be derived using the definition of conditional probability. We have $\Pr(E|H) = \Pr(H \wedge E)/\Pr(H)$. Here, we use logical conjunction $\wedge$ to write the combined event H and E. Since $\Pr(H \wedge E) = \Pr(E|H) \cdot \Pr(H)$, it follows that $\Pr(H|E) = \Pr(H \wedge E)/\Pr(E) = \Pr(E|H) \cdot \Pr(H)/\Pr(E)$, proving Bayes' theorem. Note that the theorem holds generally for probability functions and does not assume a temporal ordering of taking evidence into account, as instead suggested by the terminology of "prior" and "posterior" probability. This terminology is standard in the context of Bayesian updating.

3 Conflicting Evidence

In many situations, it is clear what the facts are. In a standard case of tax evasion, for example, it will be easy to establish whether you filed for taxes on time and whether your employer paid you 100,000 dollars in 2015. Only in special circumstances, such as administrative errors, there will be something to dispute here. But cases that are litigated in court are typically more complicated. Disputes emerge because the two parties—who then become the defense and the prosecution in a criminal trial—introduce evidence that supports conflicting reconstructions of the facts. In this section, we illustrate how each of the three frameworks can represent and model conflicts between different pieces of evidence.

3.1 Arguments

In the argument-based framework, conflicting evidence is analyzed in terms of reasons for and against a certain conclusion. Consider a criminal case where a witness testifies she saw the suspect at the crime scene. The witness testimony constitutes a reason supporting the conclusion that the suspect indeed was at the crime scene. This can be understood as an argument *from* "a witness testified she saw the suspect at the crime scene" *to* "the suspect was in fact at the crime scene." This argument consists of three parts: the conclusion; the reason (also called the premise); and the connection between the reason and the conclusion. In what follows, we describe three ways this argument can be attacked and three symmetric ways the same argument can be further supported by additional reasons.

Three kinds of attack can be distinguished: rebutting, undercutting, and undermining. Consider the argument that the suspect was at the crime scene because the witness reports that she saw the suspect at the crime scene. First, the conclusion can be attacked. For example, suppose DNA testing shows that the suspect does not genetically match with the traces found at the crime scene. Such an attacking reason is called a *rebutting attack*. It supports the opposite conclusion, namely that the suspect was *not* at the crime scene. Second, the reason itself can be attacked. For instance, if the witness never actually testified that she saw the suspect at the crime scene, this attacks the existence of the supporting reason itself. This kind of attack is referred to as *undermining attack*. Third, the connection between the reason and the conclusion can be attacked. The fact that the lighting conditions were bad when the witness saw the crime is an example of such an attack, referred to as an *undercutting attack*. In contrast with a rebutting attack, an undercutting attack provides no support for the opposite conclusion. In the example, if the lighting conditions were bad, there would be no reason explicitly supporting that the suspect was not at the crime scene. The three examples of the different kinds of attack are shown in Fig. 2.

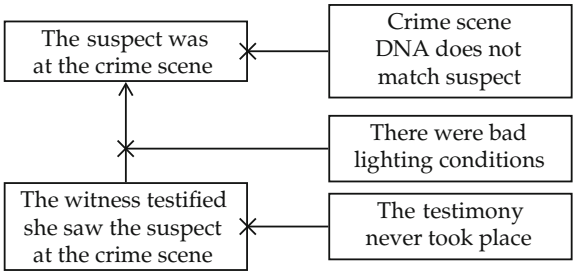

Fig. 2 Three kinds of attack

Three kinds of support can be distinguished: multiple, subordinated, and coordinated. Just as attacking reasons can target the conclusion of an argument, its supporting reason, or the connection between the two, additional reasons can provide further support for each of these parts. Additional reasons can be seen as responses to attacking reasons or as reasons strengthening an existing argument.

Consider, once again, the argument that the suspect was at the crime scene because the witness reports that she saw the suspect at the crime scene. First, the conclusion can be further supported, for example, by a second witness testimony. If a conclusion is supported by more than one reason, this is referred to as *multiple support*. Second, the reason itself can be supported, for example, by a video recording of the witness testimony itself. Support of the reason itself is called *subordinating support*. Finally, the connection between the reason and the conclusion can be further supported, for example, by another testimony that the witness has always been trustworthy and reliable. Support for the connection between the reason and the conclusion does not have a standard name, but is closely related to a third named kind of support: *coordinated support*. In coordinated support, the support for the conclusion consists of at least two supporting reasons which, in their conjunctive combination, provide support for the conclusion. Coordinated support is distinguished from multiple support because in the latter each supporting reason provides support for the conclusion by itself.

Figure 3 shows the three kinds of (further) support. Multiple and subordinated support are graphically visualized with an arrow, whereas coordinated support is shown with a line. An arrow indicates the support of the connection between reason and conclusion.

Arguments can involve complex structures of supporting and attacking reasons. So far we have looked at an elementary argument, consisting of a reason and a conclusion, along with three types of attacking reasons and three types of symmetric supporting reasons. But an argument can also be more complex; for example, it can contain *chains of reasons*.

Consider, once again, the example of a witness who reports that she saw the suspect at the crime scene. The witness testimony constitutes a reason supporting the conclusion that the suspect was at the crime scene, and this conclusion—in turn—functions as a reason that supports the conclusion that the suspect committed the crime. This chain of supporting reasons is graphically depicted in Fig. 4, on the left.

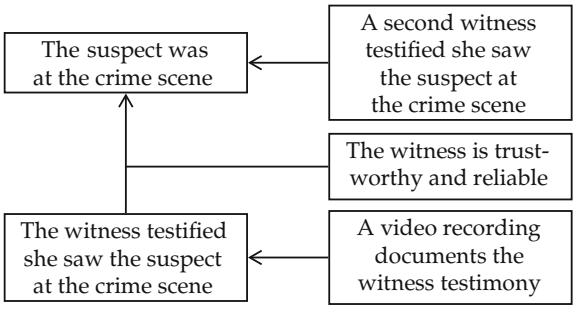

Fig. 3 Three kinds of (further) support

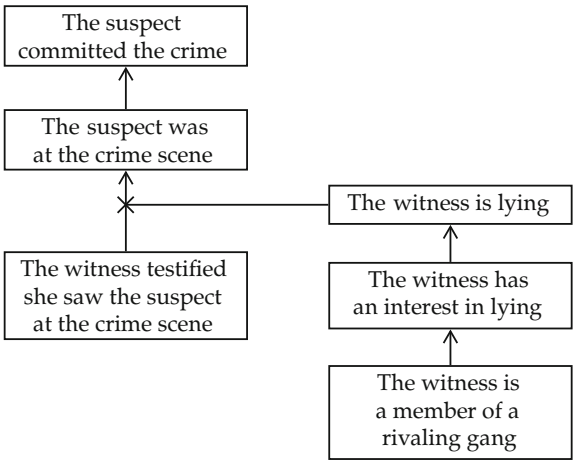

Fig. 4 Supporting and attacking reasons can be chained

Attacking reasons can also be chained. For example, when it is discovered that the witness is a member of a rivaling gang, this constitutes a reason for concluding that the witness has an interest in lying, and further, for concluding that the witness is in fact lying (Fig. 4, on the right). This conclusion attacks—undercuts, to be precise—the connection between the witness testimony and the conclusion the suspect was at the crime scene.

3.2 Scenarios

In the scenario-based framework, conflicts are analyzed by considering different scenarios about what may have happened. While in the previous framework, conflicts were modeled as conflicts between attacking and supporting reasons within arguments, here the perspective is more holistic, and conflicts are modeled as conflicts between scenarios.

There may be conflicting scenarios about what happened. The prosecution and the defense sometimes present different scenarios about what happened. In a murder case, for example, prosecution and defense may put forward the following conflicting scenarios:

S_1: The defendant killed the victim when caught during a robbery.

S_2: The victim's partner killed the victim after a violent fight between the two.

The two scenarios conflict insofar as they offer incompatible reconstructions of the killing and point to two different perpetrators.

At trial, however, while the prosecutor is expected to identify the perpetrator, the defense is not expected to identify another perpetrator. Two scenarios, then, can be conflicting even though they do not each point to a different perpetrator, such as the following:

S_1: The defendant killed the victim when caught during a robbery.

S_3: The defendant was at home with his wife.

Scenarios S_1 and S_3 are still clearly in conflict because they cannot be both true. Still, scenario S_3 does not say who killed the victim or how the crime occurred. It only asserts, in the form of an *alibi*, that the defendant did not do it.

Evidence can be explained by one scenario, but not by another. Conflicts between scenarios can also exist in relation to the evidence, for example, when one scenario can explain a piece of evidence but the other cannot. Two senses of "explanation" are relevant here. First, a scenario explains the evidence in the sense that it *predicts* the evidence. If the scenario is assumed to be true, the evidence must be (likely to be) there. There is another, albeit closely related, sense of explanation. A scenario explains the evidence in the sense that it exhibits the *causal process* by which the evidence was brought about.

To understand the difference between the two senses of explanation, consider the conflicting scenarios S_1 and S_2, one referring to the robber scenario and the other to the partner scenario described above. Suppose now that laboratory analyses find a genetic match between the DNA profile of a tissue trace found under the victim's fingernails and her partner, and it is clear that the skin tissue could not have gotten there unless there was a violent fight between the two. Scenario S_1, the robber scenario, cannot explain the presence of the trace matching the victim's partner. Scenario S_2, the partner scenario, can explain the presence of the matching trace. The explanation is that the victim's partner is the source of the trace, which was deposited during the violent fight between the two. The scenario can predict the presence of the trace, in the sense that if the scenario is assumed to be true, the matching trace must be there or likely to be there. The scenario also exhibits the causal process that brought about the trace, namely the violent fight, altercation and physical contact between the two.

However, suppose another piece of evidence is that the victim's house was in fact robbed in concomitance with the victim's death. Scenario S_1 can explain this evidence, both in terms of prediction and in terms of causal process. By contrast,

scenario S_2 cannot offer the same explanation. All in all, scenarios S_1 and S_2 are not only inconsistent on their face and they also diverge in terms of the evidence that they can or cannot explain.

Scenarios can be contradicted by evidence. So far we considered scenarios that are inconsistent with one another because they cannot be both true, and also scenarios that diverge in terms of the evidence they can or cannot explain. There is another type of conflict worth discussing. This takes the form of a quasi-inconsistency between scenarios and evidence. The quasi-inconsistency occurs when the evidence taken at face value—typically testimonial, not physical evidence—asserts that such-and-such an event occurred, while the scenario denies precisely that.

Suppose a video recording shows the defendant breaking into the victim's house, and upon being discovered, killing the victim and later stealing the jewelry. This evidence contradicts scenario S_2 in which the victim's partner is the killer. More precisely, insofar as the evidence is taken at face value—that is, the video is taken to be truthful—scenario S_2 is inconsistent with the evidence, while scenario S_1 is consistent.

3.3 Probabilities

In the probability-based framework, conflicts are modeled as conflicts between pieces of evidence which support or attack a certain hypothesis, where "support" and "attack" are described in probabilistic terms.

Support can be characterized as "probability increase" or "positive likelihood ratio." A piece of evidence E supports an hypothesis H whenever E raises the probability of H, or in symbols, $\Pr(H|E) > \Pr(H)$. For example, a witness testifies that she saw the defendant around the crime scene at the time of the crime. The testimony supports the hypothesis that the defendant is factually guilty. This can be described probabilistically, as follows:

$$\Pr(\mathit{guilt}|\mathit{testimony}) > \Pr(\mathit{guilt}).$$

There is another characterization of evidential support. Instead of comparing the initial probability $\Pr(H)$ and the probability $\Pr(H|E)$ of the hypothesis given the evidence, a so-called likelihood ratio of the form $\Pr(E|H)/\Pr(E|\neg H)$ can also be used. On this account, E supports H whenever the likelihood ratio $\Pr(E|H)/\Pr(E|\neg H)$ is greater than one. This means that the presence of the evidence is regarded as more probable if the hypothesis is true than if the hypothesis is false. Given the example considered earlier, we have:

$$\frac{\Pr(\mathit{testimony}|\mathit{guilt})}{\Pr(\mathit{testimony}|\neg\mathit{guilt})} > 1.$$

These two characterizations of evidential support—in terms of probability increase and positive likelihood ratio—are in fact equivalent. For the following statement holds[8]:

$$\Pr(H|E) > P(H) \text{ iff } \frac{\Pr(E|H)}{\Pr(E|\neg H)} > 1.$$

The equivalence, however, only holds if the two hypotheses being compared in the likelihood ratio are one the negation of the other, such as *guilt* and ¬*guilt*.

Attack can be characterized as "probability decrease" or "negative likelihood ratio." By contrast, a piece of evidence E attacks a hypothesis H whenever E lowers the probability of H, or in symbols, $\Pr(H|E) < \Pr(H)$. For example, if a DNA test shows no match between the traces found at the crime scene and the defendant, this evidence attacks the hypothesis that the defendant is factually guilty. Probabilistically,

$$\Pr(\textit{guilt}|\textit{no DNA match}) < \Pr(\textit{guilt}).$$

Similarly, a piece of evidence E attacks a hypothesis H whenever the likelihood ratio is lower than one. This means that the presence of the evidence is less probable if the hypothesis is true than if the hypothesis is false. For the example considered earlier, we have:

$$\frac{\Pr(\textit{no DNA match}|\textit{guilt})}{\Pr(\textit{no DNA match}|\neg \textit{guilt})} < 1.$$

Just as the two characterizations of evidential support are equivalent, so are the two characterizations of evidential attack, that is:

$$\Pr(H|E) < \Pr(H) \text{ iff} \frac{\Pr(E|H)}{\Pr(E|\neg H)} < 1.$$

The equivalence holds because the two hypotheses being compared in the likelihood ratio are the negation of each other.

---

[8]To see why, note that

$$\frac{\Pr(H|E)}{\Pr(\neg H|E)} = \frac{\Pr(E|H)}{\Pr(E|\neg H)} \cdot \frac{\Pr(H)}{\Pr(\neg H)},$$

which implies

$$\frac{\Pr(E|H)}{\Pr(E|\neg H)} > 1 \text{ iff } \frac{\Pr(H|E)}{\Pr(\neg H|E)} > \frac{\Pr(H)}{\Pr(\neg H)}.$$

To prove the left-right direction of the equivalence in the text, if $\Pr(H|E) > P(H)$, then $1 - \Pr(H|E) < 1 - \Pr(H)$. This means that $\frac{\Pr(H|E)}{1-\Pr(H|E)} > \frac{\Pr(H)}{1-\Pr(H)}$, and thus $\frac{\Pr(H|E)}{\Pr(\neg H|E)} > \frac{\Pr(H)}{\Pr(\neg H)}$. So, by the equivalence above, $\frac{\Pr(E|H)}{\Pr(E|\neg H)} > 1$. For the other direction, if $\frac{\Pr(E|H)}{\Pr(E|\neg H)} > 1$, then $\frac{\Pr(H|E)}{\Pr(\neg H|E)} > \frac{\Pr(H)}{\Pr(\neg H)}$, again by the equivalence above. The latter is the same as $\frac{\Pr(H|E)}{1-\Pr(H|E)} > \frac{\Pr(H)}{1-\Pr(H)}$. To establish $\Pr(H|E) > \Pr(H)$, suppose for contradiction that $\Pr(H|E) \leq \Pr(H)$, which implies $1 - \Pr(H|E) \geq 1 - \Pr(H)$. This means that $\frac{\Pr(H|E)}{1-\Pr(H|E)} \leq \frac{\Pr(H)}{1-\Pr(H)}$. This contradicts $\frac{\Pr(H|E)}{1-\Pr(H|E)} > \frac{\Pr(H)}{1-\Pr(H)}$, and thus $\Pr(H|E) > \Pr(H)$.

The conflict between two pieces of evidence can be described probabilistically. Two pieces of evidence come into conflict with one another insofar as one supports a hypothesis and the other attacks the same hypothesis. The conflict can be described probabilistically, in that one piece of evidence increases the probability of the hypothesis, while the other decreases it, or equivalently, the likelihood ratio is positive (for one piece of evidence) and negative (for the other).

For example, the testimony that the defendant was around the crime scene conflicts with the lack of a DNA match. Probabilistically, the testimony increases the probability of the defendant's guilt (or equivalently, the likelihood ratio is greater than one), while the lack of a DNA match decreases the probability of the same hypothesis (or equivalently, the likelihood ratio is lower than one).

4 Evidential Value

The evidence in a criminal case has different levels of evidential value: Some evidence is strong, other not so much. How is evidential value handled in each of the three normative frameworks? That is the topic of this section.

4.1 Probability

In the probabilistic framework, evidential value is quantified numerically using various concepts based on the probability calculus, that is, probabilistic difference, likelihood ratio, and conditional probability on the evidence.

The incremental evidential value is measured by probabilistic change. The incremental value of evidence for, or against, a hypothesis can be quantified probabilistically in various ways. One approach considers the difference between the probability of the hypothesis with and without the evidence, that is, $\Pr(H|E) - \Pr(H)$. The larger the positive difference, the higher the value of the evidence for the hypothesis. An alternative approach is given by the likelihood ratio $\Pr(E|H)/\Pr(E|\neg H)$. For any value greater than one, the higher the likelihood ratio, the higher the value of the evidence for the hypothesis. By contrast, a negative difference $\Pr(H|E) - \Pr(H)$ and a likelihood ratio lower than one quantify the value of the evidence *against* a hypothesis. The larger the negative difference and the lower the likelihood ratio (for any value below one), the higher the value of the evidence against the hypothesis.

Note that these two approaches parallel the two characterizations of evidential support and attack in the previous section, as probability increase/decrease and positive/negative likelihood ratio. While these notions were only qualitative, probability increases/decreases and likelihood ratios, as measures of evidential value, express quantities.

The overall evidential value is measured by the overall conditional probability. In contrast with the incremental evidential value of evidence that is measured by

a probabilistic difference or likelihood ratio, the overall evidential value of the full body of evidence is measured by the conditional probability of the hypothesis given the evidence. The higher, or lower, the probability $\Pr(H|E)$, the higher the overall value of the evidence for, or against, the hypothesis. If there are different pieces of evidence $E_1, \ldots, E_k$, the overall evidential value of the evidence is measured as $\Pr(H|E_1, \ldots, E_k)$.

Overall and incremental evidential value should not be confused. To illustrate, suppose we have strong evidence E_1 for the hypothesis H that a suspect was at the crime scene, for instance, security camera footage in which the suspect is easily recognizable. In this case, the overall evidential value $\Pr(H|E_1)$ of the evidence is high. If this is the only evidence, then also the incremental evidential value is high: Before the evidence is considered, the hypothesis is not strongly supported, i.e., $\Pr(H)$ is low, whereas after the evidence is considered, the hypothesis is strongly supported, i.e., $\Pr(H|E_1)$ is high. In this case, the overall and incremental evidential value of E_1 are both high. But suppose a witness testifies that the defendant was not at the crime scene (evidence E_2), but as it turns out, the witness is unreliable as a known accomplice of the suspect. Consider now the overall evidential value $\Pr(H|E_1, E_2)$ of the two pieces of evidence together. This will not have changed much when compared to $\Pr(H|E_1)$. As a result, the incremental evidential value of E_2 is low, while still the overall evidential value $\Pr(H|E_1, E_2)$ is high, even though E_2 did not contribute much.

The difference between overall and incremental evidential value can be especially confusing when there is a single piece of evidence. Consider the hypothesis $\neg H$ that the suspect was not at the crime scene and the evidence E_2, the testimony of the unreliable witness. Now, if $\Pr(\neg H)$ is high, then $\Pr(\neg H|E_2)$ will be equally high because E_2 has no incremental value. Uncritically interpreted, the high value of $\Pr(\neg H|E_2)$ suggests that the testimony of the unreliable witness has a high evidential value. But incrementally E_2 did not change much. The hypothesis $\neg H$ is, in totality, still strongly supported after the incrementally weak evidence E_2, since the hypothesis was already strongly supported before that evidence.

The use of evidence with high incremental evidential value has complications. As an illustration, we discuss the likelihood ratio of a DNA match. When introduced in court, a DNA match comes with a so-called Random Match Probability or, as of late, with its Conditional Genotype Probability. This (roughly) is the probability that the DNA of a random person, who had nothing to do with the crime, would match. Let us denote this probability by γ.

With some simplifications (on these later), the evidential value of the DNA match M in favor of the hypothesis that the suspect is the source of the crime traces, abbreviated S, is as follows

$$\frac{\Pr(M|S)}{\Pr(M|\neg S)} = \frac{1}{\gamma}.$$

The numerator $\Pr(M|S)$ equals 1 because we assume that if the defendant is the source of the crime traces, the laboratory test will report a match. As for the denominator,

assuming $\Pr(M|\neg S) = \gamma$ is plausible because the probability that a match would be reported if the defendant was *not* the source is roughly the same as the chance that someone who had no contact with the victim would randomly match. For example, if γ is 1 in 200 million, the likelihood ratio would be

$$\frac{\Pr(M|S)}{\Pr(M|\neg S)} = \frac{1}{\frac{1}{200\text{ million}}} = 200\text{ million}.$$

Since the likelihood ratio in question is a high number, the DNA match has strong evidential value in favor of the hypothesis that the suspect is the source. More generally, a low random match or genotype probability γ corresponds to a match with a rare profile and thus has a high evidential value.

Still, even with a low γ one should beware of the complications when using a DNA match in a criminal case. Consider the following non-equivalent hypotheses:

1. The *lab reports* that the defendant's genetic profile matches with the crime traces;
2. The defendant's genetic profile *truly matches* with the crime traces;
3. The defendant is the *source* of the traces;
4. The defendant *visited* the crime scene; and
5. The defendant is *factually guilty*.

The inferential path from "reported match" to "guilt," passing through the intermediate steps "true match," "source" and "visit," is a long one, and each step comes with sources of error that may undermine the inference along the way.

First, the inference from "reported match" to "true match" depends on whether or not the laboratory made a mistake. A key source of laboratory mistakes is human errors, much more frequent than DNA profiles. Second, the inference from "true match" to "source" can go wrong in several ways. For one, someone who is entirely unrelated with the crime could be, by sheer coincidence, a true match. This could happen, although the chance of this happening remains typically low as measured by the Random Match Probability or Conditional Genotype Probability. For another, a suspect who is not the source of the crime traces could still match because of close family relations with the actual perpetrator. Think of a perpetrator who has a genetically identical twin. Third, the inference from "source" to "visiting the crime scene" is not infallible. In particular, the traces can have been accidentally transferred to the crime scene or have been planted there. Fourth, the inference from "visiting the crime scene" to "factual guilt" can go wrong in many ways, because having visited a crime scene is not the same as having committed the crime.

4.2 Arguments

The evidential value of arguments can be analyzed in terms of the strength of the reasons they are built from, but also by asking critical questions about the reasons of the argument, its conclusion, and the connection between reasons and conclusion.

The reasons used can be conclusive or defeasible. A reason is conclusive when, given the reason, its conclusion is guaranteed. The main type of conclusive reason corresponds to deductive, logically valid reasoning. An example of a conclusive reason occurs in the logically valid argument from the reasons "John is shot" and "If someone is shot, he dies" to the conclusion "John dies." Its logical validity is connected to the underlying logical structure of the argument: from "A" and "A implies B" conclude "B."

Many reasons are not conclusive, but defeasible. There are circumstances in which the conclusion does not follow, although the reason obtains. The reason "The witness reports to have seen the suspect at the crime scene" supports the conclusion "The suspect was at the crime scene" but does not guarantee that conclusion, because the witness could have made a mistake. A defeasible reason can provide *prima facie* justification for a conclusion, which might later be withdrawn in light of countervailing reasons. Reasons that occur in so-called *abductive arguments* are also defeasible, where abductive arguments can be thought of as providing an explanation. For example, from "John's DNA matches the crime trace" conclude "John left the trace." The fact that John left the trace is put forward as an explanation for the fact that John's DNA matches the trace. Abductive arguments are typically defeasible because there often are alternative explanations. Someone with the same genetic profile as John might have left the trace.

Arguments can be evaluated by asking critical questions. Consider again the one-step argument from the reason "The witness reports that she saw the suspect at the crime scene" to the conclusion "The suspect was at the crime scene." Critical questions can be asked about the argument. They include, for example, whether there are reasons to doubt the suspect was at the crime scene, such as an alibi; whether there are reasons to doubt that the witness testimony supports the conclusion that the suspect was at the crime scene, for instance, the witness is lying; and whether there are reasons to doubt the existence of the witness testimony, such as a fraudulent report. The first of these questions is directed at the argument's conclusion, the second at the argument step from reason to conclusion, and the third at the argument's reason. These different kinds of critical questions are connected to the three kinds of argument attack discussed in Sect. 3.1 (see in particular Fig. 2, page 455).

But what do critical questions do? How do they help us assess the strength of arguments? Suppose that initially it is believed that the suspect was at the crime scene because of the witness testimony. A positive answer to any of the critical questions mentioned above will weaken the support for the conclusion that the suspect was at the crime scene, perhaps up to the point of making it no longer believable.

It can be subject to debate whether a reason supports or attacks a conclusion. Whether a reason supports a conclusion depends on an underlying general rule. For instance, the argument from a witness testimony (the reason) to the suspect's being at the crime scene (the conclusion) rests on the general rule that what witnesses say can generally be believed. Following Toulmin (1958)'s terminology, such general rules making explicit how to get from the reason to the conclusion are referred to as *warrants*. Support for a warrant is called the backing of the warrant.

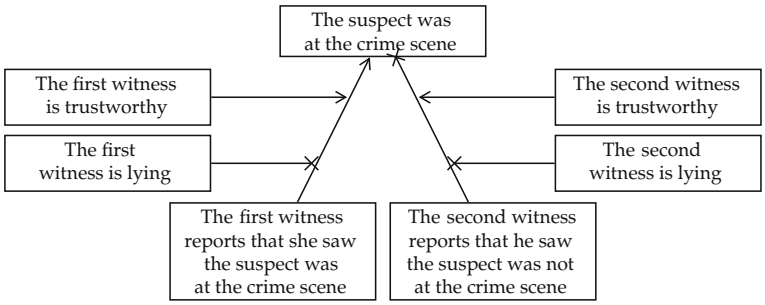

Fig. 5 Arguments about whether a reason is supporting or attacking

More generally, a reason can either support or attack a conclusion, so the relation between reason and conclusion can be a supporting relation or an attacking relation. These supporting or attacking relations can, in turn, be themselves supported or attacked. This gives rise to four different combinations: support of a supporting relation; support of an attacking relation; attack of a supporting relation; and attack of an attacking relation. In Fig. 5, these situations are illustrated by two opposite witness testimonies.

4.3 Scenarios

The evidential value of a scenario depends on how well it matches up with the evidence. This matching up can be understood in three ways: the scenario's plausibility and logical consistency; its power to explain the evidence; its consistency with the evidence. We examine each in turn.

Scenarios can be plausible and logically consistent. Plausibility measures how well a scenario matches up with our background assumptions and knowledge of the world. At least, a scenario should not violate the laws of nature or commonsense. If a scenario asserts that the same individual was in two different locations at the same time, or moved from one location to another in too short amount of time, the scenario would lack plausibility. The scenario "an alien did it" lacks plausibility because it describes something that rarely happens. Lack of plausibility can become so pronounced that it amounts to a lack of *logical consistency*, for example, claiming that the defendant had and did *not* have a motive for killing the victim.

Recall now the two conflicting scenarios we considered earlier:

S_1: The defendant killed the victim when caught during a robbery.

S_2: The victim's partner killed the victim after a violent fight between the two.

Which one is the most plausible? Statistics suggest that people are less often killed by strangers than by people they know. If so, scenario S_2 would be initially more plausible. However, suppose we acquired more background information about the

relationship between the victim and her partner, and it turned out their relationship was peaceful. In light of this new information, scenario S_2 will appear less plausible than S_1. It does not happen often that anger and violence manifest themselves unannounced, while it is natural that a robber, once he is discovered and has no alternative, will resort to violence.

In assessing plausibility, the evidence with which the scenario is expected to match up is not the evidence specific to the case, but rather, background information about the world. Plausibility has something to do with what we might call *normality*, that is, with what happens most of the time. It is true, however, that criminal cases are often about odd coincidences, unexpected and improbable events. Plausibility only measures the persuasiveness or credibility of a scenario *prior to* considering any more specific evidence about the crime. An initially plausible scenario may turn out to be weakly supported in light of more evidence presented about the crime.

The more evidence a scenario can explain, the better. When a case comprises several items of evidence, the more items of evidence a scenario can accommodate, preferably from both the prosecutor and the defense, the better the scenario. This depends on the scenario's explanatory power and consistency with the evidence.

Consider a case in which two items of evidence must be explained. The first is the presence of fingerprint traces at the crime scene, traces whose presence is consistent with just innocent contact. The second is that the fingerprints match with the defendant. Scenarios S_1 and S_2—the robber scenario and the victim's partner scenario, respectively—both explain the presence of fingerprint traces at the crime. They were left either by the robber, if S_1 is true, or by the victim's partner, if S_2 is true. Still, only S_1 can explain the fact that the traces match with the defendant (who is the alleged robber).

But suppose that in order to defend S_2—the victim's partner scenario—a new detail is added to the story: the victim's partner, right after killing the victim and with the intent to mislead the investigators, implanted fingerprint traces that match the defendant. This new scenario, however implausible, can explain both items of evidence: the presence of the fingerprint and the fact that they match the defendants. As far as explanatory power goes, scenarios S_1 and S_2, when properly supplemented, are now on a par with one another. Still, further evidence may distinguish the two. For example, if a witness testified she saw the defendant/robber walk toward the location of the crime immediately before the crime was committed, scenario S_2 cannot easily explain the testimony, even when supplemented with additional information. By contrast, S_1 can easily explain the testimony. Absent other evidence, scenario S_1 explains more evidence than the competing scenario S_2, in both its original and updated versions.

The more pieces of evidence a scenario is consistent with, the better. Besides plausibility and explanatory power, we can evaluate a scenario by checking whether it is consistent with the evidence presented in a case. The more pieces of evidence the scenario is consisted with, the better.

We can define consistency as lack of inconsistency between the evidence (taken at face value) and the scenario. For example, if a witness testifies that the defendant was

at home with his girlfriend at 6 PM, while according to the scenario proposed by the prosecutor, the defendant was at the crime scene at 6 PM, the two are inconsistent. Here, we are dealing we what we earlier called quasi-inconsistency, in the sense that insofar as the evidence is taken at face value—that is, the witness is taken to be truthful—the scenario is inconsistent with the evidence. An inconsistency in this sense between the evidence and a proposed scenario need not be damning for the scenario. It might, in fact, turn out that the witness was untruthful or simply confused about the timing. If so, the evidence will be discarded, not the scenario.

But, if a scenario is inconsistent with several pieces of evidence, this becomes an increasingly powerful challenge against the scenario. For example, if the timing provided by the scenario is not only inconsistent with the first witness testimony but also with the testimony of a pizza delivery man, who claims to have delivered a pizza to the house of the defendant's girlfriend, around 6 PM, and remembers having received money from the defendant, then the prosecutor's scenario is further undermined. In short, the more pieces of evidence inconsistent with the scenario, the more powerful the challenges against the scenario. This conclusion can also be stated more positively. The more pieces of evidence consistent with the scenario, the higher the evidential value of the scenario.

5 Coherently Interpreting the Evidence

The dossiers of criminal cases can be large, and the coherent interpretation of the evidence in such a dossier can be daunting, whichever normative framework is used. For each framework, we discuss how the coherent interpretation of the evidence can be addressed.

5.1 Scenarios

Scenarios can provide coherent interpretations that make sense of the evidence. We examine three dimensions along which competing scenarios, considered holistically in their entirety, can be assessed: coherence; completeness; and explanatory power.

Scenarios are coherent clusters of events, ordered in time and with causal relations. Earlier we encountered an elementary murder case scenario; namely, a robber kills the victim when caught during a robbery (S_1). This scenario can be analyzed as having a specific temporal structure: First, the robber enters the victim's house (H_1); then, the victim accidentally encounters the robber (H_2); and finally, the robber kills the victim (H_3).

Some of the events in this temporally ordered scenario are also causally connected. The accidental encounter is the cause that triggers a reaction in the robber who then kills the victim. Causal relations among the different parts, or episodes, in a scenario are important to evaluate what we might call the *coherence* of a scenario.

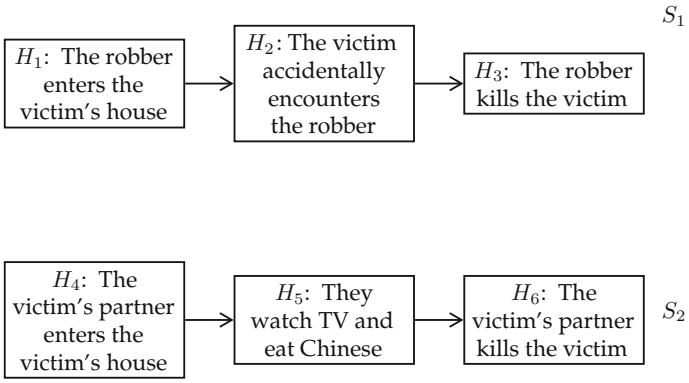

Fig. 6 Scenarios and their structure. The second scenario lacks in causal structure

Compare the robbery scenario S_1 with the partner scenario S_2 we countered earlier. Suppose this scenario is articulated more in detail as follows: The victim and her partner were watching a show on TV and eating Chinese takeout, when the partner killed the victim. This scenario has a clear temporal structure: The victim's partner arrives at the victim's home (H_4); then, they watch TV while eating Chinese takeout (H_5); finally, the victim's partner kills the victim (H_6). Still, something is missing here, that is, the causal link between "watching TV" and "killing." Why would peacefully watching TV suddenly turn into fatal violence? In comparison, the first scenario scores better in terms of causal structure. The first scenario is more coherent than the second (Fig. 6).

Scenarios can be more or less complete. Another criterion to evaluate scenarios is their *completeness*. Since scenarios are discursive arrangements of events, ordered according to temporal and causal relations, they may contain gaps in time, space, and causality. A scenario may not describe the defendant's whereabouts between 4 and 6 PM, while it describes, rather precisely, what the defendant did at 7 PM, immediately before the killing took place. The temporal gap between 4 and 6 PM makes it less complete than a scenario which describes the defendant's whereabouts between 4 and 7 PM without gaps. Yet, this might not be the notion of completeness that is important here to evaluate scenarios.

The law is not very specific in this respect. Besides defining the crime and requiring that both *mens rea*—the intention to do harm—and *actus reus*—the occurrence of the physical harm—be established, the law does not say how detailed the prosecutor's reconstruction of the crime should be. So, how is completeness a criterion to evaluate a scenario?

Some suggest that scenarios must follow certain patterns, schematic structures, or scripts. For example, in most violent crimes, we can identify an initial moment of conflict. This triggers a psychological reaction that gives rise to the formation of an intention, which, in turn, later results in the violent act. On this account, a scenario is complete whenever it has *all of its parts*, at least given an appropriate scenario script

or schematic structure. Scenario S_2, in this sense, is incomplete because it does say why and how the victim's partner formed the intention to kill the victim nor does it describe any initial moment of conflict. Scenario S_1 does not say, exactly, why the robber killed the victim. But the reason can be easily inferred. Presumably, the robber formed the intention to kill the victim when he was caught by surprise and saw no better alternative.

Weaker scenarios can be better supported by the evidence. The coherence and completeness of a scenario play a role in its evaluation. However, a weaker scenario in terms of coherence and completeness may be the best explanation of the evidence. Earlier we saw that the robbery scenario was more coherent than the scenario in which the victim's friend kills the victim while eating Chinese takeout in front of the TV. But now suppose that the pieces of evidence are as follows: The investigators find Chinese takeout in the victim's house (E_1); the saliva on one fork matches with the victim's friend DNA (E_2); there are no signs of forced entry into the victim's house (E_3). While the robbery scenario was more coherent, the Chinese takeout scenario explains the three items of evidence. In fact, the robbery scenario cannot explain any of them. So, a scenario might be superior to another on one dimension, for example, the robbery scenario is more coherent than the Chinese takeout scenario, but inferior on another dimension, for example, the robbery scenario has less explanatory power than the Chinese takeout scenario.

5.2 Arguments

An analysis of a case in terms of arguments can become complex. When Wigmore (1913) developed his charting method for analyzing the evidence in a criminal case, he was well aware of this complexity. Figure 7 provides a Wigmore diagram of the murder case *Commonwealth v. Umilian* (1901). The diagram for this relatively simple case already contains some two dozen nodes. Diagrams for more complex cases contain many more nodes.

Here, we describe three sources of complexity in the analysis of cases from the argumentation perspective: arguments and subarguments; attacks, counterattacks, and chains of attacks; and conflicts between reasons and their resolution.

The evaluation of an argument can depend on its subarguments. The structure of a complex argument influences its evaluation, and in particular, the subarguments of a larger argument determine the evaluation of the whole. For example, consider the argument in Fig. 4 (page 456). This can be analyzed as consisting of two subarguments. The first is that the witness testimony supports the intermediate conclusion that the suspect was at the crime scene. This intermediate conclusion, in turn, supports the conclusion that the suspect committed the crime, and this is the second subargument. If the first subargument is successfully attacked by a counterargument alleging that the witness is lying, the subargument supporting the intermediate conclusion that the suspect was at the crime scene breaks down and its conclusion no

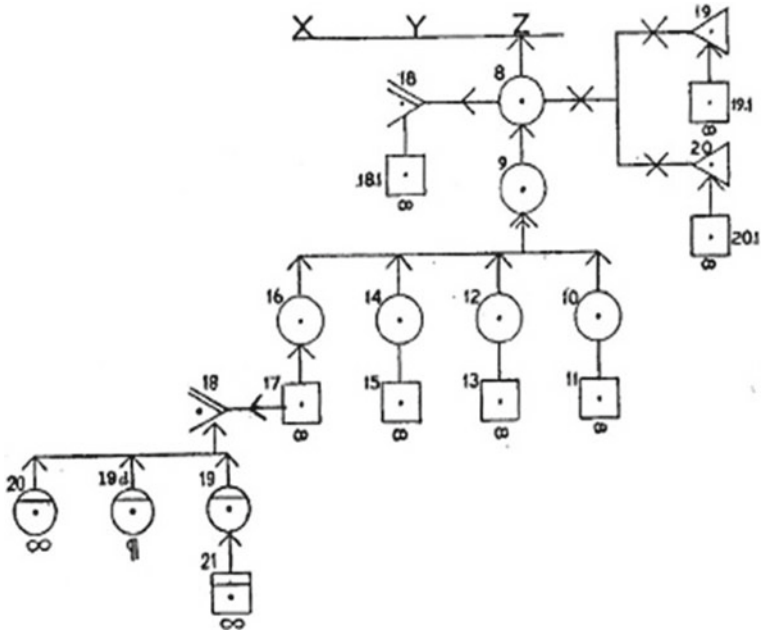

Fig. 7 A Wigmore chart

longer follows. Since the subargument does not successfully support the intermediate conclusion, also the larger argument for the final conclusion does not successfully support its conclusion.

The evaluation of an argument can depend on chains of attacks. Besides the modular relationship between arguments and subarguments, a further source of complexity in the analysis of arguments is that attacks can be chained. An attack against an argument can be countered by a further attack. When an attack is countered by a further attack, the original argument can be reinstated, in the sense that it again successfully supports its conclusion. Fig. 8 shows an example. A first witness, witness A, testifies that the suspect was at the crime scene. This testimony successfully supports the conclusion that the suspect was at the crime scene. However, suppose a second witness, witness B, claims that A is lying. The claim by witness B attacks the original argument, and thus, the conclusion that the suspect was at the crime scene is no longer supported. But if a third witness, witness C, claims that B is lying, the attack by B against A is countered. Witness B is no longer believable, so there is no reason to conclude that A is lying. As a result, A's testimony can again support the conclusion that the suspect was at the crime scene, and the original argument is thus reinstated.

Conflicts between reasons can be addressed by exceptions, preferences, and weighing. Another source of complexity in the analysis of arguments is conflicts between reasons. Reasons may support a certain conclusion or oppose it. When the same conclusion is supported by a reason (or set of reasons) and opposed by another

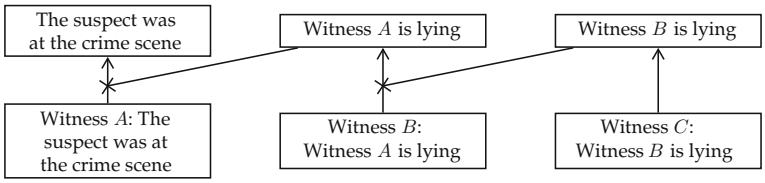

Fig. 8 Reinstatement

reason (or set of reasons), different reasons come into conflict. Since a conclusion cannot be both supported and opposed, the question arises of how to address and resolve conflicts between reasons. We distinguish three ways in which conflicts between reasons can be addressed.

First, if a reason supports a conclusion and another reason opposes it, the conflict between them can be resolved if an exception exists which excludes one of the reasons. For instance, suppose two witnesses, *A* and *B*, make conflicting statements about whether the suspect was at the crime scene. If there is evidence that one witness is lying, the conflict is resolved in favor of the witness against whom there is no evidence of lying. If there is evidence that witness *B* is lying, the conflict is resolved in favor of *A*'s testimony (see the top of Fig. 9). Evidence that *B* is lying undercuts the connection between *B*'s testimony and its conclusion (cf. Sect. 3.1). This, by contrast, leaves intact the connection between *A*'s testimony and its conclusion. The conflict is thus resolved in favor of *A*'s testimony.

In the second way of addressing a conflict between reasons, there is again a reason for a conclusion and a reason against the same conclusion. This time the resolution of the conflict occurs not by excluding one of the two reasons, but rather, by giving preference to one reason over the other. If two witnesses give conflicting testimonies about whether the suspect was at the crime scene, the conflict will remain unresolved insofar as the two reasons oppose one another and are assigned equal weight. Taking into account further information that can justify preferring one reason over the other will resolve the conflict. A reason can be preferred over another, for instance, when it is stronger. A preference (indicated by the $>$-sign in Fig. 9) can be justified if one witnesses is shown to be more reliable than the other. In this case, the conclusion that follows from the testimony by the more reliable witness should be drawn, while the conclusion that follows from the other, weaker testimony should not be drawn. This resolves the conflict.

The third way of addressing conflicts between reasons involves more than two conflicting reasons. For instance, there can be more than two witnesses, offering conflicting testimonies (Fig. 9, bottom). Resolving such conflicts can be thought of as weighing the reasons involved, where the weighing of reasons is done by generalizing a preference ordering of reasons to an ordering of sets of conflicting reasons.

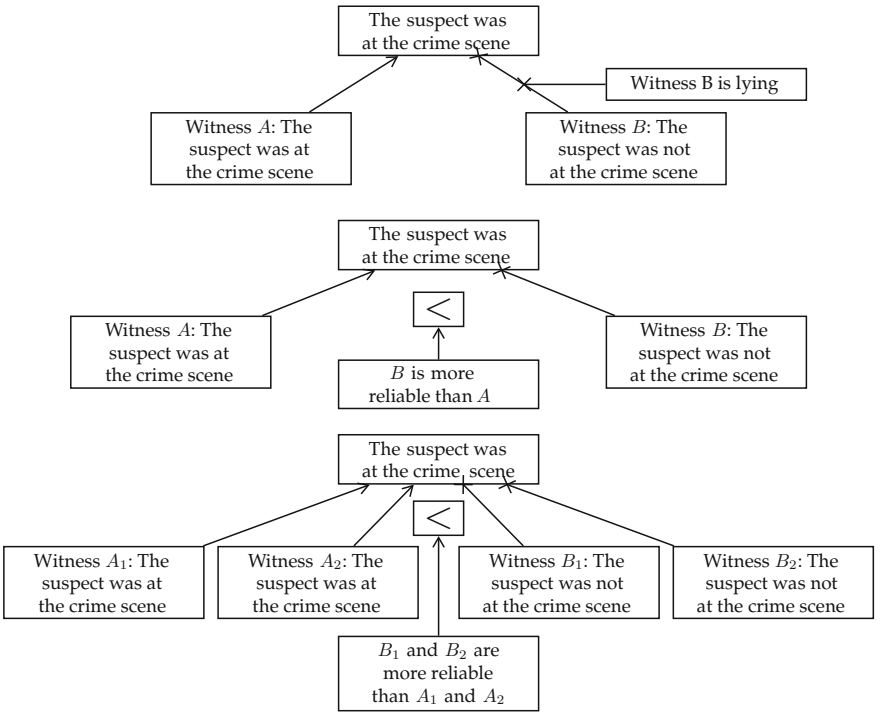

Fig. 9 Three kinds of addressing conflicts of reasons. The sign $<$ indicates a preference ordering between one reason (or set of reasons) and another

5.3 Probability

The probability calculus provides formal rules for the coherent interpretation of the evidence. We begin by discussing an elementary application of Bayes' theorem when only one piece of evidence is involved, and then turn to an analysis that involves more than one piece of evidence. We briefly discuss Bayesian networks, formal and computational tools for handling complex bodies of evidence in the probabilistic setting.

The likelihood ratio formula shows how to find the posterior odds given the evidence. The odds of a hypothesis H are given by the ratio $\Pr(H)/\Pr(\neg H)$ of the probability of the hypothesis and the probability of its negation. The odds $\Pr(H)/\Pr(\neg H)$ of the hypothesis, unconditioned on the evidence, are called the *prior odds* of the hypothesis, and the odds $\Pr(H|E)/\Pr(\neg H|E)$ of the hypothesis, conditioned on evidence E, are called the *posterior odds*. The latter can be found by multiplying the prior odds with the likelihood ratio[9]:

[9]To derive the likelihood ratio formula, one first applies Bayes' theorem to both H and $\neg H$. We get $\Pr(H|E) = \Pr(E|H) \cdot \Pr(H)/\Pr(E)$ and $\Pr(\neg H|E) = \Pr(E|\neg H) \cdot \Pr(\neg H)/\Pr(E)$. Using these,

$$\frac{\Pr(H|E)}{\Pr(\neg H|E)} = \frac{\Pr(E|H)}{\Pr(E|\neg H)} \cdot \frac{\Pr(H)}{\Pr(\neg H)}.$$

This formula shows that the (incremental) evidential value of the evidence for a hypothesis, expressed by the likelihood ratio $\frac{\Pr(E|H)}{\Pr(E|\neg H)}$, does not by itself give the posterior odds. The prior odds are needed as well. If the posterior odds $\frac{\Pr(E|H)}{\Pr(E|\neg H)}$ are known, the *posterior probability* $\Pr(H|E)$ can be derived by applying the following formula[10]:

$$\Pr(H|E) = \frac{\frac{\Pr(H|E)}{\Pr(\neg H|E)}}{1 + \frac{\Pr(H|E)}{\Pr(\neg H|E)}}.$$

Consider an example. The incremental evidential value of a DNA match, call it M, relative to the hypothesis that the defendant is factually guilty, call it G, is given by the likelihood ratio $\frac{\Pr(M|G)}{\Pr(M|\neg G)}$. Suppose this ratio is assigned a numerical value, as follows:

$$\frac{\Pr(M|G)}{\Pr(M|\neg G)} = \frac{1}{\frac{1}{2{,}000{,}000}} = 2{,}000{,}000.$$

Suppose, also, that the prior odds are as follows:

$$\frac{\Pr(G)}{\Pr(\neg G)} = \frac{\frac{1}{200{,}000}}{\frac{199{,}999}{200{,}000}} \approx \frac{1}{200{,}000}.$$

The posterior odds of the hypothesis given the match $\frac{\Pr(G|M)}{\Pr(\neg G|M)}$ are therefore as follows:

$$\frac{\Pr(G|M)}{\Pr(\neg G|M)} = \frac{\Pr(M|G)}{\Pr(M|\neg G)} \cdot \frac{\Pr(G)}{\Pr(\neg G)} \approx 2{,}000{,}000 \cdot \frac{1}{200{,}000} = 20.$$

So the poster probability of the hypothesis is as follows:

$$\Pr(G|M) = \frac{\frac{\Pr(G|M)}{\Pr(\neg G|M)}}{1 + \frac{\Pr(G|M)}{\Pr(\neg G|M)}} \approx \frac{20}{1+20} \approx 95\%.$$

A generalization of the formula shows how to handle more pieces of evidence. So far we have considered only one piece of evidence. In a straightforward generalization

---

we find:

$$\frac{\Pr(H|E)}{\Pr(\neg H|E)} = \frac{\Pr(E|H) \cdot \Pr(H)/\Pr(E)}{\Pr(E|\neg H) \cdot \Pr(\neg H)/\Pr(E)} = \frac{\Pr(E|H) \cdot \Pr(H)}{\Pr(E|\neg H) \cdot \Pr(\neg H)},$$

proving the likelihood ratio formula.

[10] $\Pr(H|E) = \frac{\Pr(H|E)}{\Pr(H|E)+\Pr(\neg H|E)} = \frac{\frac{\Pr(H|E)}{\Pr(\neg H|E)}}{\frac{\Pr(H|E)+\Pr(\neg H|E)}{\Pr(\neg H|E)}} = \frac{\frac{\Pr(H|E)}{\Pr(\neg H|E)}}{\frac{\Pr(H|E)}{\Pr(\neg H|E)}+1}.$

of the formula for two pieces of evidence E_1 and E_2, the likelihood ratios of the individual pieces of evidence are multiplied, as follows:

$$\frac{\Pr(H|E_1 \wedge E_2)}{\Pr(\neg H|E_1 \wedge E_2)} = \frac{\Pr(E_2|H)}{\Pr(E_2|\neg H)} \cdot \frac{\Pr(E_1|H)}{\Pr(E_1|\neg H)} \cdot \frac{\Pr(H)}{\Pr(\neg H)}.$$

However, this generalization only holds provided that the two pieces of evidence are independent conditional on the hypothesis, that is, $\Pr(E_2|H) = P(E_2|H \wedge E_1)$.

To illustrate, consider now two pieces of evidence: a DNA match and a witness testimony. The DNA match, call it M, holds between the crime traces and the defendant, and the witness, call it W, in her testimony asserts that the defendant was seen at the crime scene. Both pieces of evidence, intuitively, support the hypothesis G that the defendant is factually guilty. To assign an explicit numerical value, assume the DNA match has a likelihood ratio $\frac{\Pr(M|G)}{\Pr(M|\neg G)}$ of 2 million, and the witness testimony a likelihood ratio $\frac{\Pr(W|G)}{\Pr(W|\neg G)}$ of 1,000. These numbers are purely illustrative, but are needed to perform the probabilistic calculations. (Of course, there remains the question of how the numbers can be obtained and whether the numbers needed to carry out the calculations are available in the first place. This is a topic of debate.) Finally, assume the two pieces are independent conditional on the hypothesis G, that is, $\Pr(W|G) = \Pr(W|G \wedge M)$.

The combined (incremental) evidential value of the two pieces of evidence is given by multiplying the two likelihood ratios, that is, $2{,}000{,}000 \times 1{,}000 = 2{,}000{,}000{,}000$, which is a higher value than the two pieces considered independently. If the prior odds $\frac{\Pr(G)}{\Pr(\neg G)}$ are roughly $\frac{1}{200{,}000}$ as before, the posterior odds are therefore as follows:

$$\frac{\Pr(G|M \wedge W)}{\Pr(\neg G|M \wedge W)} = \frac{\Pr(M|G)}{\Pr(M|\neg G)} \cdot \frac{\Pr(W|G)}{\Pr(W|\neg G)} \cdot \frac{\Pr(G)}{\Pr(\neg G)} \approx 2{,}000{,}000 \cdot 1{,}000 \cdot \frac{1}{200{,}000} = 20{,}000.$$

So the posterior probability of the hypothesis given the two pieces of evidence is as follows:

$$\Pr(G|M \wedge W) \approx \frac{20{,}000}{1 + 20{,}000} \approx 99\%.$$

Compare this probability with $\Pr(G|M)$, which was 95%, a lower value. The probability calculus can offer a numerical representation of the intuitive fact that two pieces of evidence, taken together, have a higher (overall) evidential value than one piece alone.

If the two pieces of evidence are not independent, the likelihood ratio formula for two pieces of evidence takes the following, more general form:

$$\frac{\Pr(H|E_1 \wedge E_2)}{\Pr(\neg H|E_1 \wedge E_2)} = \frac{\Pr(E_2|H \wedge E_1)}{\Pr(E_2|\neg H \wedge E_1)} \cdot \frac{\Pr(E_1|H)}{\Pr(E_1|\neg H)} \cdot \frac{\Pr(H)}{\Pr(\neg H)}.$$

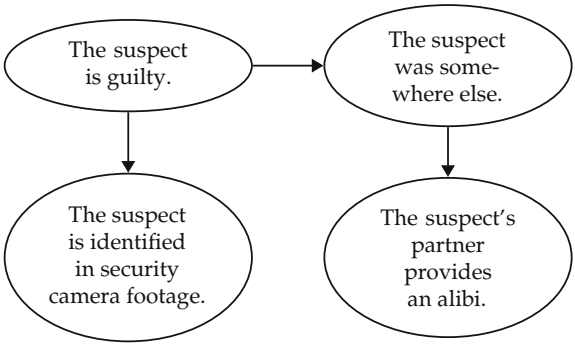

Fig. 10 An example of a Bayesian network: directed acyclic graph

The first generalization follows from the second, assuming independence between the two pieces of evidence conditional on the hypothesis of interest. The first generalization does not always hold because the evidential value of a piece of evidence, as measured by the likelihood ratio, can change in the face of other evidence.

More complex analytic tools can be used, in particular Bayesian networks. Probabilistic analyses become more complex when more elements are involved, and calculations quickly become unmanageable without appropriate modeling tools. We discuss Bayesian networks, a prominent example of a modeling tool that allows for complex probabilistic representations and calculations. A great strength of Bayesian networks is that, once the network structure is in place and the probabilities are assigned, many computer programs exist that can perform all calculations automatically.

A Bayesian network has a graphical and a numeric part. The graphical part is a directed, acyclic graph of the relevant probabilistic variables. Each node in the graph represents a variable—a hypothesis or a piece of evidence—that can take the value "true" or "false." Consider, for instance, the Bayesian network in Fig. 10. At the top, two hypotheses are shown. One is that the suspect is factually guilty (abbreviated, "Guilt"), and the other hypothesis is that the suspect was somewhere else when the crime was committed (abbreviated, "Elsewhere"). At the bottom, two pieces of evidence are shown: security camera footage that incriminates the suspect (abbreviated, "Camera"), and the testimony of the suspect's partner who provides an alibi (abbreviated, "Partner"). The arrows connecting the nodes indicate dependency relations. The "Elsewhere" variable has the "Guilt" variable as a parent, and each evidence node has a hypothesis as parent.

Some might wonder how a Bayesian network should be constructed. This is not a trivial task, and there can be different Bayesian network models of the same case. For example, some might prefer to the Bayesian network in Fig. 10 a network whose nodes are arranged in a slightly different way, or using different variables. The construction of a Bayesian network for the purpose of evidence assessment is ultimately an art.

Table 1 An example of a Bayesian network: conditional probability tables

Guilt	Pr(Guilt)
False	0.999
True	0.001

Guilt	Elsewhere	Pr(Elsewhere|Guilt)
False	False	0
False	True	1
True	False	1
True	True	0

Guilt	Camera	Pr(Camera|Guilt)
False	False	0.99
False	True	0.01
True	False	0.3
True	True	0.7

Elsewhere	Partner	Pr(Partner|Elsewhere)
False	False	0.9
False	True	0.1
True	False	0
True	True	1

Turning now to the numeric part of a Bayesian network, this consists of conditional probability tables. In these tables, the conditional probabilities of each variable are specified, conditioned on the different values of the parent variables. Table 1, for example, contains four tables of conditional probabilities, one per variable. Consider first the probability assignments in the top two tables. The prior probability of factual guilt is set to an arbitrary low value, 0.1%, or 1 in a 1000 (top left table). Since the two hypotheses—"guilt" and "elsewhere"—are assumed to exclude one another, it can never occur that they are both true or both false (top right table). Consider now the two bottom tables. The bottom left table shows how the camera identification depends on the suspect's factual guilt. If the suspect is factually guilty, there is a 70% probability that the suspect is identified in the security camera footage. If the suspect is not factually guilty, this probability is much lower, 1%. The likelihood ratio $\Pr(\text{Camera}|\text{Guilt}) / \Pr(\text{Camera}|\neg\text{Guilt})$ expressing the strength of the camera footage evidence relative to hypotheses "guilt" and "not-guilt" is therefore 70. The bottom right table shows how the partner's alibi testimony depends on whether the suspect was in fact elsewhere when the crime was committed. If the "Elsewhere" hypothesis is true, the partner will surely provide an alibi in favor of the defendant. If the "Elsewhere" hypothesis is false, the table specifies that there is a 10% chance that the suspect's partner will still provide an alibi in favor of the defendant. The likelihood ratio $\Pr(\text{Partner}|\text{Elsewhere}) / \Pr(\text{Partner}|\neg\text{Elsewhere})$ expressing the strength of the evidence (in this case, the partner's testimony providing an alibi) relative to hypotheses "Elsewhere" and "not-Elsewhere" is therefore 10.

The question now is how to combine the two pieces of evidence, the incriminating camera footage and the exculpatory partner's alibi testimony, and thus how to assess the probability of the defendant's guilt. Multiplying the likelihood ratios for each piece of evidence will not do. Bayesian networks can help here. For illustrative purposes, we write down the somewhat tedious calculations, although there is software that can do them automatically. Now, if the goal is to calculate $\Pr(\text{Guilt}|\text{Partner} \wedge \text{Camera})$—that is, the conditional probability of guilt given both the camera footage

and the partner's alibi testimony—we need to calculate both Pr(Partner ∧ Camera) and Pr(Guilt ∧ Partner ∧ Camera).

Pr(Partner ∧ Camera) can be broken down as the sum of four probabilities. That is,

$$\begin{aligned}\Pr(\text{Partner} \wedge \text{Camera}) = \Pr(&\text{Partner} \wedge \text{Camera} \wedge \text{Elsewhere} \wedge \text{Guilt}) \\ + \ &\Pr(\text{Partner} \wedge \text{Camera} \wedge \text{Elsewhere} \wedge \neg\text{Guilt}) \\ + \ &\Pr(\text{Partner} \wedge \text{Camera} \wedge \neg\text{Elsewhere} \wedge \text{Guilt}) \\ + \ &\Pr(\text{Partner} \wedge \text{Camera} \wedge \neg\text{Elsewhere} \wedge \neg\text{Guilt})\end{aligned}$$

Let us calculate each in turn. First, we have:

$$\begin{aligned}&\Pr(\text{Partner} \wedge \text{Camera} \wedge \text{Elsewhere} \wedge \text{Guilt}) \\ &= \Pr(\text{Partner}|\text{Camera} \wedge \text{Elsewhere} \wedge \text{Guilt})\ \Pr(\text{Camera}|\text{Elsewhere} \wedge \text{Guilt})\ \Pr(\text{Elsewhere}|\text{Guilt})\ \Pr(\text{Guilt}) \\ &= \Pr(\text{Partner}|\text{Elsewhere})\ \Pr(\text{Camera}|\text{Guilt})\ \Pr(\text{Elsewhere}|\text{Guilt})\ \Pr(\text{Guilt}) \\ &= 1 \times 0.7 \times 0 \times 0.001 = 0\end{aligned}$$

Note that we relied on the equalities

$$\begin{aligned}&\Pr(\text{Partner}|\text{Camera} \wedge \text{Elsewhere} \wedge \text{Guilt}) = \Pr(\text{Partner} \wedge \text{Elsewhere}); \text{ and} \\ &\Pr(\text{Camera}|\text{Elsewhere} \wedge \text{Guilt}) = \Pr(\text{Camera} \wedge \text{Guilt}).\end{aligned}$$

They hold for the Bayesian network because the node representing the partner's testimony has only the "Elsewhere" variable node as a parent, and the "Camera" node has only the "Guilt" node as a parent. Given the Bayesian network, the "Partner" variable does not depend on the "Camera" or "Guilt" variable, and the "Camera" variable does not depend on the "Elsewhere" variable.

Second, we have:

$$\begin{aligned}&\Pr(\text{Partner} \wedge \text{Camera} \wedge \text{Elsewhere} \wedge \neg\ \text{Guilt}) \\ &= \Pr(\text{Partner}|\text{Camera} \wedge \text{Elsewhere} \wedge \neg\text{Guilt})\ \Pr(\text{Camera}|\text{Elsewhere} \wedge \neg\text{Guilt})\ \Pr(\text{Elsewhere}|\neg\text{Guilt})\ \Pr(\neg\text{Guilt}) \\ &= \Pr(\text{Partner}|\text{Elsewhere})\ \Pr(\text{Camera}|\neg\text{Guilt})\ \Pr(\text{Elsewhere}\ |\neg\text{Guilt})\ \Pr(\neg\text{Guilt}) \\ &= 1 \times 0.01 \times 1 \times 0.999 = 0.00999\end{aligned}$$

Note that we relied on the equalities

$$\begin{aligned}&\Pr(\text{Partner}|\text{Camera} \wedge \text{Elsewhere} \wedge \neg\text{Guilt}) = \Pr(\text{Partner}|\text{Elsewhere}); \text{ and} \\ &\Pr(\text{Camera}|\text{Elsewhere} \wedge \neg\text{Guilt}) = \Pr(\text{Camera}|\neg\text{Guilt}).\end{aligned}$$

They hold, as before, because of the independence relations among variables in the Bayesian network.

Third, we have:

$$\Pr(\text{Partner} \wedge \text{Camera} \wedge \neg\text{Elsewhere} \wedge \text{Guilt})$$

$= \Pr(\text{Partner}|\text{Camera} \wedge \neg\text{Elsewhere} \wedge \text{Guilt})\ \Pr(\text{Camera}|\neg\text{Elsewhere} \wedge \text{Guilt})\ \Pr(\neg\text{Elsewhere}|\text{Guilt})\ \Pr(\text{Guilt})$
$= \Pr(\text{Partner}|\neg\text{Elsewhere})\ \Pr(\text{Camera}|\text{Guilt})\ \Pr(\neg\text{Elsewhere}|\text{Guilt})\ \Pr(\text{Guilt})$
$= 0.1 \times 0.7 \times 1 \times 0.001 = 0.00007$

Once again, we relied on independence relations. Fourth, since "Guilt" and "Elsewhere" cannot both be true, we have:

$$\Pr(\text{Partner} \wedge \text{Camera} \wedge \neg\text{Elsewhere} \wedge \neg\text{Guilt}) = 0$$

By adding these four probabilities, we find:

$$\Pr(\text{Partner} \wedge \text{Camera}) = 0 + 0.00999 + 0.00007 + 0 = 0.01006$$

Finally, we can compute the guilt probability based on the camera footage and the partner's testimony, as follows:

$$\begin{aligned}\Pr(\text{Guilt} \mid \text{Partner} \wedge \text{Camera}) &= \Pr(\neg\text{Elsewhere} \wedge \text{Guilt} \mid \text{Partner} \wedge \text{Camera}) \\ &= \frac{\Pr(\text{Partner} \wedge \text{Camera} \wedge \neg\text{Elsewhere} \wedge \text{Guilt})}{\Pr(\text{Partner} \wedge \text{Camera})} \\ &= 0.00007/0.01006 \approx 0.7\%\end{aligned}$$

The 0.7% probability is low, but recall that the prior guilt probability was lower, 0.1%. Also note that Pr(Guilt | Camera) equals (roughly) 6.5%.[11] This is a low probability, but still greater than 0.7 or 0.1%. This shows that the camera evidence by itself has incriminating evidential value, as it raises the guilt probability from 0.1 to 6.5%. However, when the partner's alibi testimony is added as evidence, the support for the guilt hypothesis is weakened to 0.7%.

Finally, a note about complexity. For probability functions of many variables that have many dependencies, a Bayesian network representation can be significantly more compact than a general full probability distribution over the variables. Our example has four variables, in general requiring 15 numbers to specify the full distribution. Given the independencies in the network, we only need 7 (note that half of the 14 numbers in Table 1 are superfluous as they follow from the other half).

6 Reasoning and Decision Making

So far we have focused on how the evidence can be evaluated and combined. But once the evidence has been introduced at trial, examined and cross-examined, it comes a time when the fact finders, either a trained judge or a group of lay jurors, must reason from the evidence, reach a conclusion, and decide whether to convict or acquit the

[11] Since $\frac{\Pr(\text{Guilt})}{\Pr(\neg\text{Guilt})} = \frac{0.001}{0.999}$ and $\frac{\Pr(\text{Camera}|\text{Guilt})}{\Pr(\text{Camera}|\neg\text{Guilt})} = 70$, by Bayes' theorem, $\frac{\Pr(\text{Guilt}|\text{Camera})}{\Pr(\neg\text{Guilt}|\text{Camera})} = \frac{0.001}{0.999} \times 70 \approx 0.07$ and thus $\Pr(\text{Guilt}|\text{Camera}) \approx \frac{0.07}{1+0.07} \sim 6.5\%$.

defendant. The decision criterion is defined by law and consists of a standard of proof, sometimes also called burden of persuasion. If the decision makers are persuaded of the defendant's guilt beyond a reasonable doubt, they should convict, or else they should acquit.

Paraphrases of the formulation "proof beyond a reasonable doubt" abound in the case law. Yet, it is unclear whether they improve our understanding. The US Supreme Court might have been right when, in Holland v. United States, 348 U.S. 121 (1954), it wrote that that "attempts to explain the term 'reasonable doubt' do not result in making it any clearer" (140). The three frameworks we considered—probability, arguments, and scenarios—can be supplemented by a decision-theoretic layer and then used to characterize the standard of proof, although they are not immune from shortcomings, as we shall soon see.

6.1 Probability

In a probabilistic treatment, reasoning and decision making are analyzed using the probability calculus combined with elements of decision theory.

The guilt probability is assessed by weighing the evidence with the probability calculus. On the probabilistic framework, the goal is to assess the probability of the defendant's guilt based on all the available evidence. The assessment begins with a relatively low value for the guilt probability, prior to considering any evidence. After all, absent any incriminating evidence, the prior probability that an individual committed a crime should be rather low. As more evidence is presented, the guilt probability moves upward or downward depending on whether the evidence is incriminating or exculpatory. When all the evidence is considered, a final guilt probability value is reached. This forms the basis for the decision to convict or acquit.

The value of the guilt probability is arrived at by applying Bayes' theorem a repeated number of times and by plugging the values of the probabilities that are needed. Sometimes, these probabilities can be based on numerical data about population proportions, as in the case of DNA evidence, but often data are not available. For example, $\Pr(G)$ is required to calculate $\Pr(G|E)$ using Bayes' theorem, where $\Pr(G)$ is the probability of the defendant's guilt regardless of the evidence presented at trial. What probability value should be assigned to $\Pr(G)$? It is subject to debate how to assess this probability, and even whether it makes sense to assign a number to the prior probability of guilt in the first place.

The decision criterion is a guilt probability threshold. In probabilistic terms, proof of guilt beyond a reasonable doubt means that the defendant's *probability of guilt*, given the evidence presented at trial, meets a threshold, say >99 or $>99.9\%$. A numerical value for the threshold can be identified using expected utility theory. Let $c(CI)$ be the cost of convicting an innocent and $c(AG)$ the cost of acquitting a guilty defendant. For a conviction to be justified, the expected cost of convicting an innocent must be lower than the expected cost of acquitting an innocent, that is,

$$\Pr(G|E) \cdot c(AG) > [1 - \Pr(G|E)] \cdot c(CI).$$

The inequality holds just in case

$$\frac{\Pr(G|E)}{1 - \Pr(G|E)} > \frac{c(CI)}{c(AG)}.$$

Suppose $\frac{c(CI)}{c(AG)} = \frac{99}{1}$, as might be more appropriate in a criminal case in which the conviction of an innocent defendant is regarded as far worse than the acquittal of a guilty defendant. Then, the inequality holds only if $\Pr(G)$ meets the threshold 99%. More complicated models are also possible, but the basic idea is that the probability required for a conviction is a function of weighing the costs that would result from an erroneous decision.

It is not obvious how to assess all the required probabilities. The characterization of a decision criterion in terms of a probabilistic threshold is elegant, but its application in practice can be problematic. If a probabilistic threshold is understood as a criterion which the decision makers should mechanically apply whenever they confront the decision to convict or acquit, two difficulties arise. The first difficulty is that assigning a probability value to guilt itself might not be feasible. As seen earlier, the starting probability $\Pr(G)$ cannot be easily determined, and even if it could, other probabilities might be hard to assess. One solution here is that instead of aiming for a unique guilt probability, we can simply aim for an interval of admissible probabilities given the evidence. More generally, the assessment of the probability of guilt can be viewed as an idealized process, a regulative ideal which can improve the precision of legal reasoning.

Another problem with the probabilistic characterization is that it does not take into account the so-called weight of the evidence, that is, whether the evidence presented at trial contains all the evidence in the case or just a partial subset of the evidence. The guilt probability will vary dramatically depending on the evidence that is used to assess it. It is tempting to suggest that the guilt probability must be based on a body of evidence that is complete, or at least as complete as reasonably possible. And yet, it is unclear how to characterize this notion. No body of evidence is, strictly speaking, complete because new evidence could always be discovered and added.

6.2 Arguments

In an argumentative treatment, reasoning and decision making are analyzed in terms of the arguments that are collected.

Supporting and attacking reasons are collected and weighed. In a court of law, the prosecutor puts forward a conclusion and offers supporting reasons. The opposing side responds by offering attacking reasons. The dialectical process can be complex. As seen earlier, there are different attacking reasons: undermining, undercutting, and

rebutting. The process is complex also because it can be iterated. A conclusion can be attacked by an attacking reason, and the latter in turn can be attacked and so on. When the dialectical process reaches an equilibrium point and the opposing parties have nothing more to contribute, the status of a claim and its supporting reasons can be assessed.

On the argument-based framework, the goal is to consider all the available reasons, by representing them in a comprehensive argumentation graph that keeps track of the relations of support and attack. The two competing theories of the cases, the prosecutor's and the defense's theory, will each be supported by a set of reasons. The argument framework, through the aid of argument graphs, allows us to compare the relative strength of the reasons in favor of one side of the case or the other. This comparison of the two sides forms the basis for the trial decision.

Defeating attacking arguments is the criterion for meeting the standard of proof. In order to establish the defendant's guilt beyond a reasonable doubt, all the attacks against the conclusion that the defendant is guilty must be defeated. Now, whether an attack is defeated is not always an all or nothing affair. It is often a matter of degrees. If the reasons for guilt are slightly stronger than all their attacks, this would not be enough yet. To meet the demands of the standard of proof beyond a reasonable doubt, the supporting reasons must be significantly stronger than all their attacks. On the other hand, defeating all the attacks with absolute certainty would be too much to expect. So, more realistically, all attacks must be defeated in an almost definitive way. Perhaps, we need to reintroduce some threshold, even though not in an explicitly probabilistic or numerical way.

It is not obvious when to stop collecting supporting and attacking reasons. The argumentation framework is rather realistic. The idea that meeting the standard of proof requires to answer all attacks against the conclusion that the defendant is guilty is natural enough. A problem is that if the opposing party puts forward no attacks, meeting the standard of proof would be effortless. A possible response here is that the attacks must be all the attacks which a reasonable objector could in principle put forward, not just the attacks that in fact are put forward. But who is this "reasonable objector"?

Another problem consists in identifying the threshold. While the probability-based account can identify a specific probability threshold, at least in theory, by applying the principle of expected utility theory, the argumentation-based framework cannot. How could expected utility theory be applied to the argument framework, as well?

6.3 Scenarios

In a scenario treatment, reasoning and decision making are analyzed by comparing the different scenarios.

Competing scenarios are collected and compared. On the scenario framework, the two parties will put forward competing scenarios, at least two or possibly more

than two. This is partly problematic because in a criminal case, the defense does not have the burden of proof. So it might well be that the defense puts forward a scenario that weakens the prosecutor's scenario, but that is not a scenario that proves innocence. Be that as it may, the various competing scenarios will be evaluated along the different criteria we identified, such as consistency with the evidence, explanatory power, plausibility, coherence. The question arises, which scenario should be selected among the competitors?

The best explanatory scenario is the rule of decision. We can picture the process of evaluation of the competing scenarios as a process of elimination. At the beginning, several scenarios are viable, but as more evidence is considered and the scrutiny of each scenario continues, fewer scenarios will survive. The goal would be to select one scenario, or at least a limited set of scenarios, so that the answer to the question "guilty or not?" would be univocal. On this picture, a scenario meets the demands of the standard of proof whenever it is the *only* scenario left.

But, once again, we confront a recurrent problem. The selection of a scenario is not always an all-or-nothing affair. The term "abduction" or the expression "inference to the best explanation" is sometimes used in this context. The basic idea is that, when confronted with two or more competing scenarios, the best explanation must be chosen. The notion of "best explanation" here is wide-ranging. It includes criteria such as consistency with the evidence, explanatory power (predictive power and causal fit), plausibility, completeness, coherence (temporal and causal structure). Other criteria might also play a role, such as the simplicity of the scenario. The best explanation is the scenario that fares best on some combination of these criteria. In this background, the decision rule would stipulate that the best explanatory scenario should be selected.

It is not obvious how to identify the scenarios and how to compare them. The process of scenario analysis and selection resembles how jurors reason in trial proceedings, whereas—in contrast— it is hard to relate probability to judicial proceedings: Jurors do not naturally quantify guilt, and it can be difficult to quantify it even if we wanted to. Still, a problem with the scenario approach is that the method by which scenarios are identified and selected is not entirely transparent. When are all relevant scenarios identified? Should all scenarios mentioned in trial be taken into account, even when they seem far-fetched? Also, the different criteria, such as consistency, explanatory power, coherence, can pull the decision makers in opposite directions. For example, a scenario might be better in terms of explanatory power, while another might be more plausible. What to do, then? Perhaps a criterion for selecting the best scenario would ultimately be a qualitative version of selecting the most probable scenario, connecting a scenario-based approach with a probabilistic perspective.

7 Summary and Conclusion

We have discussed evidential reasoning in the law. We started out by discussing two common forms of evidence, eyewitness testimonies and DNA matches. We then distinguished three normative frameworks for theorizing about evidential reasoning: one focusing on the arguments for and against the positions taken; the second using probabilities to assess the evidential value of the evidence; and the third considering the scenarios that best explain the evidence. We then discussed four main themes: conflicting evidence; evidential value; the coherent interpretation of the evidence; and reasoning and decision making. For each theme, we discussed how they can be addressed in each of the three frameworks. We now summarize our discussion for each theme, using the highlighted phrases in the preceding sections.

7.1 Conflicting Evidence

Arguments Three kinds of attack can be distinguished: rebutting, undercutting, and undermining. Three kinds of support can be distinguished: multiple, subordinated, and coordinated. Arguments can involve complex structures of supporting and attacking reasons.
Scenarios There may be conflicting scenarios about what happened. Evidence can be explained by one scenario, but not by another. Scenarios can be contradicted by evidence.
Probabilities Support can be characterized as "probability increase" or "positive likelihood ratio." Attack can be characterized as "probability decrease" or "negative likelihood ratio." The conflict between two pieces of evidence can be described probabilistically.

7.2 Evidential Value

Probabilities The incremental evidential value is measured by probabilistic change. The overall evidential value is measured by the overall conditional probability. The use of evidence with high incremental evidential value has complications.
Arguments The reasons used can be conclusive or defeasible. Arguments can be evaluated by asking critical questions. It can be subject to debate whether a reason supports or attacks a conclusion.
Scenarios Scenarios can be plausible and logically consistent. The more evidence a scenario can explain, the better. The more pieces of evidence a scenario is consistent with, the better.

7.3 Coherently Interpreting the Evidence

Scenarios Scenarios are coherent clusters of events, ordered in time and with causal relations. Scenarios can be more or less complete. Weaker scenarios can be better supported by the evidence.
Arguments The evaluation of an argument can depend on its subarguments. The evaluation of an argument can depend on chains of attacks. Conflicts between reasons can be addressed by exceptions, preferences, and weighing.
Probabilities The likelihood ratio formula shows how to find the posterior odds given the evidence. A generalization of the formula shows how to handle more pieces of evidence. More complex analytic tools can be used, in particular Bayesian networks.

7.4 Reasoning and Decision Making

Probabilities The guilt probability is assessed by weighing the evidence with the probability calculus. The decision criterion is a guilt probability threshold. It is not obvious how to assess all the required probabilities.
Arguments Supporting and attacking reasons are collected and weighed. Defeating attacking arguments is the criterion for meeting the standard of proof. It is not obvious when to stop collecting supporting and attacking reasons.
Scenarios Competing scenarios are collected and compared. The best explanatory scenario is the rule of decision. It is not obvious how to identify the scenarios and how to compare them.

With the thematic discussion of the three normative frameworks, we have aimed to show how each framework contributes to the understanding of conflicting evidence, evidential value, the coherent interpretation of the evidence, and reasoning and decision making. In our perspective, there is no need to choose between the frameworks, since each adds to the normative analysis of evidential reasoning. At the same time, there is room for further studies of how the three normative frameworks relate to one another and how they can be integrated into a unified normative perspective on evidential reasoning.

8 Further Readings

For the interested readers, we now provide some reading suggestions, thematically organized following the headings of the previous sections. The full list of references is provided at the end.

8.1 Setting the Stage

Eyewitness Testimony: Eyewitness misidentification as one of the main causes of mistaken convictions (www.innocenceproject.org). Manipulations of the memory of witnesses (Loftus 1996). Selective attention in perception (Simons and Chabris 1999). Detecting lies and the psychology of lying (Vrij 2008). Holistic versus piecemeal face recognition (Tanaka and Farah 1993). Orientation of faces in identification tasks and the Tatcher illusion (Thomson 1980). Improving the probative value of eyewitness evidence (Wells et al. 2006). On a corroboration requirement for convictions solely based on eyewitness evidence (Thompson 2008; Crump 2009).
DNA Evidence: Basics of DNA evidence (Hicks et al. 2016; Wasserman 2008; Kaye and Sensabaugh 2000). DNA matches as a matter of degrees (Kaye 1993). Confusions between the source hypothesis and the guilt hypothesis and other exaggerations in the presentation of DNA evidence (Koehler 1993). Laboratory errors (Thompson et al. 2003). On whether DNA profiles are unique (Balding 1999; Weir 2007; Koehler and Saks 2010; Kaye 2013). History of the use of DNA evidence in court and the debate on the independence assumption Kaye (2010).

8.2 Three Normative Frameworks

The three frameworks for modeling evidential reasoning (Anderson et al. 2005; Kaptein et al. 2009; Dawid et al. 2011).

Arguments: Wigmore charts (Wigmore 1913). The New Evidence Scholarship (Anderson et al. 2005). Formal and computational study of arguments (Pollock 1987, 1995). Informal and formal argumentation theory (van Eemeren et al. 2014b).

Probabilities: Bayes' theorem (Swinburne 2002). Bayesian epistemology and updating (Bovens and Hartmann 2003a). Evidence and probabilities in the law (Dawid 2002; Schum 1994; Schum and Starace 2001; Mortera and Dawid 2007). Statistics in the law (Finkelstein and Levin 2001; Fenton 2011; Gastwirth 2012). Miscarriages of justice involving statistics (Dawid et al. 2011; Schneps and Colmez 2013). Debate on whether probabilistic calculations have a place in courts (Finkelstein and Fairley 1970; Tribe 1971; Fenton et al. 2016), and more recently, the 2012 special issue of *Law, Probability and Risk*; Vol. 11, No. 4.

Scenarios: Scenarios in evidential reasoning (Bennett and Feldman 1981; Pennington and Hastie 1993a, b). Scenarios and miscarriages of justice (Wagenaar et al. 1993). Inference to the best explanation (Pardo and Allen 2008). Hypothetical explanations of the evidence (Thagard 1989).

Combined Approaches: Combining arguments and scenarios (Bex et al. 2010; Bex 2011). Bayesian networks for evidential reasoning (Hepler et al. 2007; Fenton et al. 2013; Taroni et al. 2014). Combining arguments, scenarios and probabilities (Vlek et al. 2014, 2016; Timmer et al. 2017; Verheij et al. 2016; Verheij 2014, 2017).

8.3 Conflicting Evidence

Arguments: Argument structure and diagrams (Wigmore 1913; Toulmin 1958; Freeman 1991). Defeasible reasoning and nonmonotonic logic (Pollock 1987; Gabbay et al. 1994). Rebutting and undercutting attack (Pollock 1987, 1995). Undermining attack (Bondarenko et al. 1997). Formal evaluation of defeasible arguments (Pollock 1987, 1995; Dung 1995; Prakken 2010). Argumentative dialogue (Toulmin 1958; Walton and Krabbe 1995; Prakken 1997; Hage 2000). Argument diagramming and evaluation software (Pollock 1995; Reed and Rowe 2004; Kirschner et al. 2003; van Gelder 2003; Verheij 2005; Gordon et al. 2007).

Scenarios: Scenarios in evidential reasoning (Bennett and Feldman 1981; Pennington and Hastie 1993a, b). Scenarios and miscarriages of justice (Wagenaar et al. 1993). Inference to the best explanation (Pardo and Allen 2008). Hypothetical explanations of the evidence (Thagard 1989).

Probabilities: On confirmation theory and accounts of evidential support (Carnap 1950; Fitelson 1999; Skyrms 2000; Hacking 2001; Bovens and Hartmann 2003a; Crupi 2015). Probabilistic accounts of evidential support in the law (Lempert 1977). On whether the likelihood ratio should consider exhaustive hypotheses or not (Fenton et al. 2014; Biedermann et al. 2014).

8.4 Evidential Value

Probabilities: Introductions to using probability for weighing evidence (Finkelstein and Fairley 1970; Dawid 2002; Mortera and Dawid 2007). Critique of the probabilistic approach (Tribe 1971; Cohen 1977; Allen and Pardo 2007). Prosecutor's fallacy (Thompson and Schumann 1987). Introduction to DNA evidence (Wasserman 2008; Kaye and Sensabaugh 2000). Different hypotheses for evaluating DNA evidence (Koehler 1993; Cook et al. 1998; Evett et al. 2000). Probabilistic analyses of DNA evidence (Robertson and Vignaux 1995; Hicks et al. 2016; Balding 2005). Lab errors for DNA evidence (Thompson et al. 2003). Match is not all-or-nothing judgment (Kaye 1993). Uniqueness of DNA profiles (Balding 1999; Kaye 2013; Weir 2007). How DNA evidence can be synthesized and implanted (Frumkin et al. 2009). Cold hit controversy in DNA evidence cases (NRC 1996; Balding and Donnely 1996). Comparison between DNA evidence and fingerprints (Zabell 2005). Probabilistic analyses of eyewitness testimony (Friedman 1987; Schum 1994; Schum and Starace 2001).

Arguments: Nonmonotonic reasoning (Gabbay et al. 1994). Prima facie reasons, undercutting and rebutting defeaters (Pollock 1987, 1995). Warrants and backings (Toulmin 1958). Argument schemes and critical questions (Walton et al. 2008). Formal and computational argumentation (van Eemeren et al. 2014a).

Scenarios: Explanation in the deductive nomological model (Hempel and Oppenheim 1948). Explanation and causality (Salmon 1984). Abduction and inference to the best explanation (Lipton 1991). More the philosophical literature on scientific explanation (Woodward 2014). Two directions of fit (Wells 1992). Hypothetical explanations of the evidence (Thagard 1989). Scenario quality (Pennington and Hastie 1993b; Wagenaar et al. 1993; Bex 2011).

8.5 Coherently Interpreting the Evidence

Scenarios: Explanation and unification in philosophy of science (Friedman 1974). Coherence in epistemology (BonJour 1985). The crossword puzzle analogy for coherently evaluating a mass of evidence (Haack 2008). Explanatory coherence (Thagard 2001). Cognitive role of scripts (Schank and Abelson 1977). The story model (Pennington and Hastie 1993a). Scenarios as scripts (Wagenaar et al. 1993). Scenarios in legal cases (Griffin 2013). Evidence and scenario schemes (Bex 2011; Verheij et al. 2016; Vlek et al. 2014, 2016). Worries about scenarios in law (Velleman 2003). Scenarios shifting the legal perspective (Bex and Verheij 2013).

Arguments: Argument structure and their evaluation (Pollock 1995). Formalizing argumentation (Prakken and Vreeswijk 2002). Evaluating argument attack (Dung 1995). Formal argumentation models (Simari and Loui 1992; Vreeswijk 1997; Prakken 2010; Verheij 2003; Gordon et al. 2007). Informal and formal argumentation theory (van Eemeren et al. 2014b). Accrual of reasons and weighing (Pollock 1995; Hage 1997; Verheij 1996; Prakken 2005).

Probabilities: The conjunction paradox (Cohen 1977) and a response (Dawid 1987). Coherence and probability (Bovens and Hartmann 2003b). Probabilistic analysis of an entire legal case (Kadane and Schum 1996; Vlek et al. 2014, 2016). On the use of probability in law (Fenton 2011). Bayesian networks (Pearl 1988; Darwiche 2009; Jensen and Nielsen 2007; Fenton and Neil 2013). Bayesian networks for evidential reasoning (Taroni et al. 2014; Hepler et al. 2007; Fenton et al. 2013; Vlek et al. 2014, 2016; Timmer et al. 2017). Bayesian networks and causality (Pearl 2000/2009; Dawid 2010). Arguments, scenarios and probabilities (Keppens and Schafer 2006; Keppens 2012; Vlek et al. 2014, 2016; Timmer et al. 2017; Verheij et al. 2016; Verheij 2014, 2017).

8.6 Reasoning and Decision Making

Evidence law manuals (Fisher 2008; Méndez 2008). Criminal Procedure manuals (Allen et al. 2016; Roberts and Zuckerman 2010). On difficulties and confusions while defining proof beyond a reasonable doubt (Laudan 2006). Character evidence and its exclusion (Redmayne 2015).

Probabilities: Probabilistic accounts of the burden of proof (Kaplan 1968; Kaye 1986, 1999; Hamer 2004; Cheng 2013). Critique of probabilistic accounts (Cohen 1977; Nesson 1979; Thomson 1986; Stein 2005; Ho 2008; Pardo and Allen 2008; Haack 2014). On the question whether the threshold should be variable (Kaplow 2012; Picinali 2013). The problem of priors (Finkelstein and Fairley 1970; Friedman 2000). A critique of proof beyond a reasonable doubt as understood in the law (Laudan 2006). History of beyond a reasonable doubt standard (Shapiro 1991; Whitman 2008). Other measures, weight, resiliency and completeness of the evidence (Kaye 1999; Stein 2005; Nance 2016).

Arguments: Evaluating arguments and their attacks (Pollock 1995; Dung 1995). Burden of proof and argumentation (Gordon et al. 2007; Gordon and Walton 2009; Prakken and Sartor 2007, 2009). Weighing reasons (Hage 1997).

Scenarios: Inference to the best explanation (Lipton 1991). Application of inference to the best explanation to legal reasoning (Pardo and Allen 2008). Narrative based account of proof beyond a reasonable doubt (Allen 2010; Allen and Stein 2013).

Acknowledgements This chapter has been developed 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 first author would like to thank Infosys Limited which made possible his stay at the Institute for Advanced Study in Princeton for the academic year 2016–17 during which parts of this chapter were written. The second author would like to thank the Isaac Newton Institute for Mathematical Sciences at the University of Cambridge for its hospitality during the program "Probability and Statistics in Forensic Science" which was supported by EPSRC Grant Number EP/K032208/1. The authors would like to thank Ronald Allen, Alex Biedermann, Christian Dahlman, and Norman Fenton for helpful comments and suggestions on an earlier draft.

References

Allen, R.J. 2010. No plausible alternative to a plausible story of guilt as the rule of decision in criminal cases. In *Prueba y Esandares de Prueba en el Derecho*, ed. J. Cruz, and L. Laudan. Mexico: Instituto de Investigaciones Filosoficas-UNAM.

Allen, R.J., and M.S. Pardo. 2007. The problematic value of mathematical models of evidence. *Journal of Legal Studies* 36 (1): 107–140.

Allen, R.J., and A. Stein. 2013. Evidence, probability and the burden of proof. *Arizona Law Journal* 55: 557–602.

Allen, R.J., W.J. Stuntz, J.L. Hoffmann, D.A. Livingston, A.D. Leipold, and T.L. Meares. 2016. *Comprehensive criminal procedure*, 3rd ed. New York, N.Y.: Wolters Kluwer.

Anderson, T., D. Schum, and W. Twining. 2005. *Analysis of Evidence*, 2nd ed. Cambridge: Cambridge University Press.

Balding, D.J. 1999. When can a DNA profile be regarded as unique? *Science & Justice* 39.

Balding, D.J. 2005. *Weight-of-evidence for forensic DNA profiles*. West Sussex: Wiley.

Balding, D.J., and P. Donnely. 1996. Evaluating DNA profile evidence when the suspect is identified through a database search. *Journal of Forensic Science* 41: 603–607.

Bennett, W.L., and M.S. Feldman. 1981. *Reconstructing reality in the courtroom*. London: Tavistock Feldman.

Bernoulli, J. 1713. *Ars Conjectandi*.

Bex, F.J. 2011. *Arguments, stories and criminal evidence: A formal hybrid theory*. Berlin: Springer.
Bex, F.J., P.J. van Koppen, H. Prakken, and B. Verheij. 2010. A hybrid formal theory of arguments, stories and criminal evidence. *Artificial Intelligence and Law* 18: 1–30.
Bex, F.J., and B. Verheij. 2013. Legal stories and the process of proof. *Artificial Intelligence and Law* 21 (3): 253–278.
Biedermann, A., T. Hicks, F. Taroni, C. Champod, C. Aitken. On the use of the likelihood ratio for forensic evaluation: Response to Fenton, et al. 2014. *Science and Justice* 54 (4): 316–318.
Bondarenko, A., P.M. Dung, R.A. Kowalski, and F. Toni. 1997. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence* 93: 63–101.
BonJour, L. 1985. *The structure of empirical knowledge*. Cambridge, MA: Harvard University Press.
Bovens, L., and S. Hartmann. 2003a. *Bayesian Epistemology*. Oxford: Oxford University Press.
Bovens, L., and S. Hartmann. 2003b. Solving the riddle of coherence. *Mind* 112: 601–633.
Carnap, R. 1950. *Logical foundations of probability*. Chicago, IL: University of Chicago Press.
Cheng, E. 2013. Reconceptualizing the burden of proof. *Yale Law Journal* 122 (5): 1104–1371.
Cohen, L.J. 1977. *The probable and the provable*. Oxford: Clarendon Press.
Cook, R., I.W. Evett, G. Jackson, P.J. Jones, and J.A. Lambert. 1998. A hierarchy of propositions: Deciding which level to address in casework. *Science and Justice* 38 (4): 231–239.
Crump, D. 2009. Eyewitness corroboration requirements as protections against wrongful conviction: The hidden questions. *Ohio State Journal of Criminal Law* 7 (1): 361–376.
Crupi, V. 2015. Confirmation. In *Stanford encyclopedia of philosophy*, ed. E.N. Zalta. Stanford University.
Darwiche, A. 2009. *Modeling and reasoning with bayesian networks*. Cambridge: Cambridge University Press.
Dawid, A.P. 1987. The difficulty about conjunction. *Journal of the Royal Statistical Society. Series D (The Statistician)* 36(2/3):91–92.
Dawid, A.P. 2002. Bayes's theorem and weighing evidence by juries. In *Bayes's Theorem*, vol. 113, 71–90, Oxford: Oxford University Press.
Dawid, A.P. 2010. Beware of the DAG! In *JMLR workshop and conference proceedings: Volume 6. Causality: Objectives and assessment (NIPS 2008 workshop)*, eds. I. Guyon, D. Janzing, and B. Schölkopf, 59–86. http://www.jmlr.org/
Dawid, A.P., W. Twining, and M. Vasiliki (eds.). 2011. *Evidence, inference and enquiry*. Oxford: Oxford University Press.
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.
Evett, I., G. Jackson, J.A. Lambert, and S. McCrossan. 2000. The impact of the principles of evidence interpretation on the structure and content of statements. *Science and Justice* 40 (4): 233–239.
Fenton, N., D. Berger, D. Lagnado, M. Neil, and A. Hsu. 2014. When "neutral" evidence still has probative value (with implications from the barry george case). *Science and Justice* 54 (4): 274–287.
Fenton, N., M. Neil, and D. Berger. 2016. Bayes and the law. *Annual Review of Statistics and Its Application* 3.
Fenton, N.E. 2011. Science and law: Improve statistics in court. *Nature* 479: 36–37.
Fenton, N.E., and M.D. Neil. 2013. *Risk assessment and decision analysis with Bayesian networks*. Boca Raton, FL: CRC Press.
Fenton, N.E., M.D. Neil, and D.A. Lagnado. 2013. A general structure for legal arguments about evidence using Bayesian Networks. *Cognitive Science* 37: 61–102.
Finkelstein, M.O., and W.B. Fairley. 1970. A Bayesian approach to identification evidence. *Harvard Law Review* 83: 489–517.
Finkelstein, M.O., and B. Levin. 2001. *Statistics for lawyers*. Berlin: Springer.
Fisher, G. 2008. *Evidence*, 2nd ed. New York, N.Y.: Foundation Press.
Fitelson, B. 1999. The plurality of Bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science* 66: 362–378.

Freeman, J.B. 1991. *Dialectics and the macrostructure of arguments. A theory of argument structure*. Berlin: Foris.
Friedman, M. 1974. Explanation and scientific understanding. *Journal of Philosophy* 71: 5–19.
Friedman, R.D. 1987. Route analysis of credibility and hearsay. *The Yale Law Journal* 97 (4): 667–742.
Friedman, R.D. 2000. A presumption of innocence, not of even odds. *Stanford Law Review* 52: 873–887.
Frumkin, D., A. Wasserstrom, A. Davidson, and A. Grafit. 2009. Authentication of forensic DNA samples. *Forensic Science International: Genetics* 4 (2): 95–103.
Gabbay, D.M., C.J. Hogger, and J.A. Robinson (eds.). 1994. *Handbook of logic in artificial intelligence and logic programming. Volume 3. Nonmonotonic reasoning and uncertain reasoning*. Oxford: Clarendon Press.
Gastwirth, J.L. (ed.). 2012. *Statistical Science in the Courtroom*. Berlin: Springer.
Gordon, T.F., H. Prakken, and D.N. Walton. 2007. The Carneades model of argument and burden of proof. *Artificial Intelligence* 171 (10–15): 875–896.
Gordon, T.F., and D.N. Walton. 2009. Proof burdens and standards. In *Argumentation in artificial intelligence*, ed. I. Rahwan, and G.R. Simari, 239–258. Berlin: Springer.
Griffin, L.K. 2013. Narrative, truth, trial. *Georgetown Law Journal* 101: 281–335.
Haack, S. 2008. Warrant, causation, and the atomist of evidence law. *Journal of Social Epistemology* 5: 253–265.
Haack, S. 2014. Legal probabilism: An epistemological dissent. In *Science, proof, and truth in the law*, ed. Evidence Matters, 47–77. Cambridge: Cambridge University Press.
Hacking, I. 2001. *An introduction to probability and inductive logic*. Cambridge: Cambridge University Press.
Hage, J.C. 1997. *Reasoning with rules. An essay on legal reasoning and its underlying logic*. Dordrecht: Kluwer.
Hage, J.C. 2000. Dialectical models in artificial intelligence and law. *Artificial Intelligence and Law* 8: 137–172.
Hamer, D. 2004. Probabilistic standards of proof, their complements and the errors that are expected to flow from them. *University of New England Law Journal* 1 (1): 71–107.
Hempel, C., and P. Oppenheim. 1948. Studies in the logic of explanation. *Philosophy of Science* 15: 135–175.
Hepler, A.B., A.P. Dawid, and V. Leucari. 2007. Object-oriented graphical representations of complex patterns of evidence. *Law, Probability and Risk* 6 (1–4): 275–293.
Hicks, T., J. Buckleton, J.-A. Bright, and D. Taylor. 2016. A framework for interpreting evidence. In *Forensic DNA Evidence Interpretation (second edition)*, ed. J. Buckleton, J.-A. Bright, and D. Taylor. Boca Raton, FL: CRC Press.
Ho, H.L. 2008. *Philosophy of evidence law*. Oxford: Oxford University Press.
Jensen, F.V., and T.D. Nielsen. 2007. *Bayesian networks and decision graphs*. Berlin: Springer.
Kadane, J.B., and D.A. Schum. 1996. *A probabilistic analysis of the Sacco and Vanzetti evidence*. Chichester: Wiley.
Kaplan, J. 1968. Decision theory and the fact-finding process. *Stanford Law Review* 20: 1065–1092.
Kaplow, L. 2012. Burden of proof. *Yale Law Journal* 121 (4): 738–1013.
Kaptein, H., H. Prakken, and B. Verheij (eds.). 2009. *Legal evidence and proof: statistics, stories, logic (Applied legal philosophy series)*. Farnham: Ashgate.
Kaye, D.H. 1986. Do we need a calculus of weight to understand proof beyond a reasonable doubt? *Boston University Law Review* 66: 657–672.
Kaye, D.H. 1993. DNA evidence: Probability, population genetics and the courts. *Harvard Journal of Law and Technology* 7: 101–172.
Kaye, D.H. 1999. Clarifying the burden of persuasion: What Bayesian rules do and not do. *International Commentary on Evidence* 3: 1–28.
Kaye, D.H. 2010. *The double helix and the law of evidence*. Cambridge, Mass.: Harvard University Press.

Kaye, D.H. 2013. Beyond uniqueness: the birthday paradox, source attribution and individualization in forensic science. *Law, Probability and Risk* 12 (1): 3–11.
Kaye, D.H., and G.F. Sensabaugh. 2000. Reference guide on DNA evidence. In *Reference manual on scientific evidence*, 2nd ed., 576–585. Washington, D.C.: Federal Judicial Center.
Keppens, J. 2012. Argument diagram extraction from evidential Bayesian networks. *Artificial Intelligence and Law* 20: 109–143.
Keppens, J., and B. Schafer. 2006. Knowledge based crime scenario modelling. *Expert Systems with Applications* 30 (2): 203–222.
Kirschner, P.A., S.J.B. Shum, and C.S. Carr. 2003. *Visualizing argumentation: Software tools for collaborative and educational sense-making*. Berlin: Springer.
Koehler, J.J. 1993. Error and exaggeration in the presentation of DNA evidence in trial. *Jurimetrics Journal* 34: 21–39.
Koehler, J.J., and M.J. Saks. 2010. Individualization claims in forensic science: Still unwarranted. *Brooklyn Law Review* 75 (4): 1187–1208.
Laplace, P.-S. 1814. *Essai Philosophique sur les Probabilités*.
Laudan, L. 2006. *Truth, error, and criminal law: An essay in legal epistemology*. Cambridge: Cambridge University Press.
Lempert, R.O. 1977. Modeling relevance. *Michigan Law Review* 75 (5/6): 1021–1057.
Lipton, P. 1991. *Inference to the best explanation*. New York, N.Y.: Routledge.
Loftus, E.F. 1996. *Eyewitness testimony (revised edition)*. Cambridge, MA: Harvard University Press.
Méndez, M.A. 2008. *Evidence: The California code and the Federal rules*, 4th ed. Eagan, MN: Thomson West.
Mortera, J., and P. Dawid. 2007. Probability and evidence. In *Handbook of probability theory*, ed. T. Rudas. Los Angeles, CA: Sage.
Nance, D.A. 2016. *The burdens of proof: Discriminatory power, weight of evidence, and tenacity of belief*. Cambridge: Cambridge University Press.
Nesson, C.R. 1979. Reasonable doubt and permissive inferences: The value of complexity. *Harvard Law Review* 92 (6): 1187–1225.
NRC. 1996. *The evaluation of forensic DNA evidence*. Washington, D.C.: National Academy Press.
Pardo, M.S., and R.J. Allen. 2008. Juridical proof and the best explanation. *Law and Philosophy* 27: 223–268.
Pearl, J. 1988. *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Mateo, CA: Morgan Kaufmann.
Pearl, J. 2000/2009. *Causality: Models, reasoning and inference*, 2nd ed. Cambridge: Cambridge University Press.
Pennington, N., and R. Hastie. 1993a. *Inside the juror*, chap. The story model for juror decision making, 192–221. Cambridge: Cambridge University Press.
Pennington, N., and R. Hastie. 1993b. Reasoning in explanation-based decision making. *Cognition* 49 (1–2): 123–163.
Picinali, F. 2013. Two meanings of "reasonableness": Dispelling the "floating" reasonable doubt. *Modern Law Review* 76 (5): 845–875.
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*. Cambridge, MA: The MIT Press.
Prakken, H. 1997. *Logical tools for modelling legal argument. A study of defeasible reasoning in law*. Dordrecht: Kluwer.
Prakken, H. 2005. A study of accrual of arguments, with applications to evidential reasoning. In *Proceedings of the tenth international conference on artificial intelligence and law*, 85–94, New York (New York): ACM Press.
Prakken, H. 2010. An abstract framework for argumentation with structured arguments. *Argument and Computation* 1 (2): 93–124.

Prakken, H., and G. Sartor. 2007. Formalising arguments about the burden of persuasion. In *Proceedings of the 11th international conference on artificial intelligence and law*, 97–106, New York, N.Y.: ACM Press.

Prakken, H., and G. Sartor. 2009. A logical analysis of burdens of proof. In *Legal evidence and proof: Statistics, stories, logic*, chap. 9, ed. H. Kaptein, H. Prakken, and B. Verheij, 223–253, Farnham: Ashgate.

Prakken, H., and G.A.W. Vreeswijk. 2002. Logics for defeasible argumentation. In *Handbook of philosophical logic*, vol. 4, 2nd ed. D.M. Gabbay, and F. Guenthner, 218–319. Dordrecht: Kluwer Academic Publishers.

Redmayne, M. 2015. *Character evidence in the criminal trial*. Oxford: Oxford University Press.

Reed, C., and G. Rowe. 2004. Araucaria: Software for argument analysis, diagramming and representation. *International Journal of AI Tools* 14 (3–4): 961–980.

Robertson, B., and G.A. Vignaux. 1995. DNA evidence: Wrong answers or wrong questions? *Genetica* 96: 145–152.

Roberts, P., and A. Zuckerman. 2010. *Criminal evidence*, 2nd ed. Oxford: Oxford University Press.

Salmon, W. 1984. *Scientific explanation and the causal structure of the world*. Princeton, N.J.: Princeton University Press.

Schank, R., and R. Abelson. 1977. *Scripts, plans, goals and understanding, an inquiry into human knowledge structures*. Hillsdale: Lawrence Erlbaum.

Schneps, L., and C. Colmez. 2013. *Math on trial: How numbers get used and abused in the courtroom*. New York, N.Y.: Basic Books.

Schum, D.A. 1994. *The evidential foundations of probabilistic reasoning*. New York, N.Y.: Wiley.

Schum, D.A., and S. Starace. 2001. *The evidential foundations of probabilistic reasoning*. Evanston, Il.: Northwestern University Press.

Shapiro, B. 1991. *Beyond reasonable doubt and probable cause: Historical perspectives on the Anglo-American law of evidence*. Oakland, Calif.: University of California Press.

Simari, G.R., and R.P. Loui. 1992. A mathematical treatment of defeasible reasoning and its applications. *Artificial Intelligence* 53: 125–157.

Simons, D.J., and C.F. Chabris. 1999. Gorillas in our minds: Sustained inattention blindness for dynamic events. *Perception* 28: 1059–1074.

Skyrms, B. 2000. *Choice and chance: An introduction to inductive logic*, 4th ed. Belmont, CA: Wadsworth.

Stein, A. 2005. *Foundations of evidence law*. Oxford: Oxford University Press.

Swinburne, R. (ed.). 2002. *Bayes's theorem*. Oxford: Oxford University Press.

Tanaka, J.W., and M.J. Farah. 1993. Parts and whole in face recognition. *The Quarterly Journal of Experimental Psychology* 46A (3): 225–245.

Taroni, F., A. Biedermann, S. Bozza, P. Garbolino, and C. Aitken. 2014. *Statistics in practice. In Bayesian networks for probabilistic inference and decision analysis in forensic science*, 2nd ed. Chichester: Wiley.

Taroni, F., C. Champod, and P. Margot. 1998. Forerunners of Bayesianism in early forensic science. *Jurimetrics* 38: 183–200.

Thagard, P. 1989. Explanatory coherence. *Behavioral and Brain Sciences* 12: 435–502.

Thagard, P. 2001. *Coherence in thought and action*. Cambridge, MA: The MIT Press.

Thompson, S.G. 2008. Beyond a reasonable doubt? reconsidering uncorroborated eyewitness identification testimony. *UC Davis Law Review* 41: 1487–1545.

Thompson, W.C., and E.L. Schumann. 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.

Thompson, W.C., F. Taroni, and C.G.G. Aitken. 2003. How the probability of a false positive affects the value of DNA evidence. *Journal of Forensic Science* 48: 47–54.

Thomson, J.J. 1986. Liability and individualized evidence. *Law and Contemporary Problems* 49 (3): 199–219.

Thomson, P. 1980. Margaret Thatcher: A new illusion. *Perception* 9 (4): 483–484.

Tillers, P. 2011. Trial by mathematics-reconsidered. *Law, Probability and Risk* 10: 167–173.
Timmer, S.T., J.J. Meyer, H. Prakken, S. Renooij, and B. Verheij. 2017. A two-phase method for extracting explanatory arguments from Bayesian networks. *International Journal of Approximate Reasoning* 80: 475–494.
Toulmin, S.E. 1958. *The uses of argument*. Cambridge: Cambridge University Press.
Tribe, L. 1971. Trial by mathematics: Precision and ritual in the legal process. *Harvard Law Review* 84: 1329–1393.
van Eemeren, F.H., B. Garssen, E.C.W. Krabbe, A.F. Snoeck Henkemans, B. Verheij, and J.H.M. Wagemans. 2014a. *Chapter 11: Argumentation in artificial intelligence*. In Handbook of argumentation theory. Berlin: Springer.
van Eemeren, F.H., B. Garssen, E.C.W. Krabbe, A.F. Snoeck Henkemans, B. Verheij, and J.H.M. Wagemans. 2014b. *Handbook of argumentation theory*. Berlin: Springer.
van Gelder, T. 2003. Enhancing deliberation through computer supported argument visualization. In *Visualizing argumentation: Software tools for collaborative and educational sense-making*, ed. P.A. Kirschner, S.J.B. Shum, and C.S. Carr, 97–115. New York, N.Y.: Springer.
Velleman, D. 2003. Narrative explanation. *The Philosophical Review* 112 (1): 1–25.
Verheij, B. 1996. *Rules, reasons, arguments. Formal studies of argumentation and defeat.*. Maastricht: Dissertation Universiteit Maastricht.
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*. The Hague: T.M.C. Asser Press.
Verheij, B. 2014. To catch a thief with and without numbers: Arguments, scenarios and probabilities in evidential reasoning. *Law, Probability and Risk* 13: 307–325.
Verheij, B. 2017. Proof with and without probabilities. correct evidential reasoning with presumptive arguments, coherent hypotheses and degrees of uncertainty. *Artifical Intelligence and Law* 25(1):127–154.
Verheij, B., F.J. Bex, S.T. Timmer, C.S. Vlek, J.J. Meyer, S. Renooij, and H. Prakken. 2016. Arguments, scenarios and probabilities: Connections between three normative frameworks for evidential reasoning. *Law, Probability and Risk* 15 (1): 35–70.
Vlek, C.S., H. Prakken, S. Renooij, and B. Verheij. 2014. Building Bayesian Networks for legal evidence with narratives: a case study evaluation. *Artifical Intelligence and Law* 22 (4): 375–421.
Vlek, C.S., H. Prakken, S. Renooij, and B. Verheij. 2016. A method for explaining Bayesian Networks for legal evidence with scenarios. *Artifical Intelligence and Law* 24 (3): 285–324.
Vreeswijk, G.A.W. 1997. Abstract argumentation systems. *Artificial Intelligence* 90: 225–279.
Vrij, A. 2008. *Detecting lies and deceit: The psychology of lying and the implications for professional practice*. Chichester: Wiley.
Wagenaar, W.A., P.J. van Koppen, and H.F.M. Crombag. 1993. *Anchored narratives: The psychology of criminal evidence*. London: Harvester Wheatsheaf.
Walton, D.N., and E. Krabbe. 1995. *Commitment in dialogue. Basic concepts of interpersonal reasoning*. Albany (New York): State University of New York Press.
Walton, D.N., C. Reed, and F. Macagno. 2008. *Argumentation schemes*. Cambridge: Cambridge University Press.
Wasserman, D. 2008. Forensic DNA typing. In *A companion to genethics*, ed. J. Burley, and J. Harris. Malden, MA: Blackwell.
Weir, B.S. 2007. The rarity of DNA profiles. *The Annals of Applied Statistics* 1: 358–370.
Wells, G.L. 1992. Naked statistical evidence of liability: Is subjective probability enough? *Journal of Personality and Social Psychology* 62: 793–752.
Wells, G.L., A. Memon, and S.D. Penrod. 2006. Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest* 7 (2): 45–75.
Whitman, J.Q. 2008. *The origins of reasonable doubt: Theological roots of the criminal trial*. New Haven, CT: Yale University Press.

Wigmore, J.H. 1913. *The principles of judicial proof as given by logic, psychology, and general experience, and illustrated in judicial trials. Second edition 1931, third edition "The science of judicial proof" 1937*. Boston, MA: Little, Brown and Company.
Woodward, J. 2014. Scientific explanation. In *The Stanford encyclopedia of philosophy*, ed. E.N. Zalta. Stanford University.
Zabell, S.L. 2005. Fingerprint evidence. *Journal of Law and Policy* 13: 143–179.