EVIDENTIAL REASONING
Chapter for the Handbook of Legal Reasoning

Marcello Di Bello & Bart Verheij – March 8, 2017

Contents

When a suspect appears in front of a criminal court, there is a very high probability that he will be found guilty. In the United States, the conviction rate in federal courts has been roughly 90% and in Japan reaches as high a rate as 99%.[1] In the United Kingdom, 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 on 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. 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.

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 has 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, very 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 very white skin, and unusually blue eyes. He looked unhealthy, with bad teeth, like a drug addict. He was wearing a perfectly ironed Armani 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 profiling

DNA profiling has become an important tool in courts. The evidential relevance of a DNA profile stems from the fact that, although most of the structure of DNA is shared among all human beings (more than 99%), the variations that do exist are very specific for each individual.

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 United States uses 13 core loci, to be expanded to 20 core loci in 2017.[3] At each specific locus, a different *allele* might occur. For instance, one locus used in the profiles stored in forensic DNA databases in the United States is CSF1PO. This locus 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.

How is the statistical frequency of a DNA profile estimated? Many countries have created extensive reference databases that contain million of DNA profiles. Each specific DNA profile is rare, and reference databases of profiles are used to estimate how rare a profile is. This is a two step process. First, the number of occurrences of each allele at each core locus in the reference database are counted. This gives an estimate 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 gives an estimate of the frequency of the DNA profile, or to use a more common terminology, the *Random Match Probability* of the DNA profile. These Random Match Probabilities are the numbers reported by forensic experts in courts. The sets of core loci have been chosen such that Random Match Probabilities are typically very small, for instance, in the order of 1 in 50 billion, amply exceeding the number of people on our planet.

A key assumption underlying the model—used when multiplying the estimated probabilities 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 family relations. Scientists have also established certain dependencies among the profiles within ethnic groups. Moreover, testing the independence assumption can be hard, and require the assessment of more profiles than reasonably possible.

With this background in place, suppose now that a trace of blood is found on the crime scene, and that the DNA profile created from the trace matches the DNA profile of the suspect. Using the match, the fact finders can form the hypothesis that the suspect 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 can be formed on the basis of a DNA match 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 *guilty*, at least absent other information about how the trace got there. Further, the hypothesis itself that the suspect is the source need not be true. The suspect and the perpetrator, though different people, might share the same DNA profile, either because they are identical twins or because, though unrelated, they happen to share the same profile.[5] We should also be wary of laboratory errors and false positive matches.

1.3 Central questions

Using the two kinds of evidence as an illustration, we can now provide a list of central questions about evidential reasoning in the law.

Question 1: How should we handle conflicting evidence? It often occurs that the evidence provides conflicting perspectives on the crime. For instance, a witness claims that the criminal has blond hair, but the suspect whose DNA matched that of the trace at the crime scene, has dark hair. What to do in case of such conflicts?

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 come with different Random Match 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?

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 when a decision can be made and what that decision is. When is the burden of proof met? What is the meaning of "beyond a reasonable doubt"?

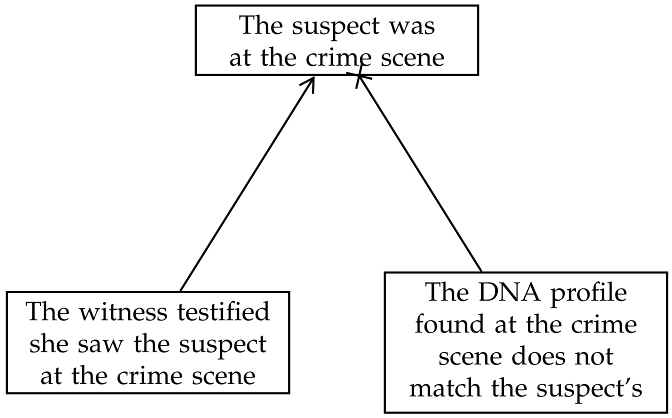

Figure 1: Arguments with supporting and attacking reasons

The plan for the paper is as follows. In the next section (Section 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, while emphasizing the role of the three normative frameworks (Sections 3, 4, 5 and 6).

2 THREE NORMATIVE FRAMEWORKS

In this section, we discuss three normative frameworks for the correct handling of the evidence: arguments; probabilities; and scenarios. We shall only emphasize their distinctive theoretical strengths. Arguments can naturally capture the dialogical dimension, by modeling relations of support and attack. Probabilities are better suited to quantify the value of the strength of the evidence. Finally, scenarios are best in offering a coherent and holistic interpretation of large bodies of evidence.

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

The analysis of the structure of arguments goes back to the early twentieth century when John Henry Wigmore developed his famous evidence charts (Wigmore, 1913). The work by the New Evidence Scholarship (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 computationally studied by the philosopher John Pollock (1987, 1995). The work by Pollock stimulated an extensive literature on the formal and computational study of arguments for and against conclusions (van Eemeren et al., 2014b).

2.2 Probabilities

The second normative framework uses probabilities as the 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 the famous 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[6]—shows how the conditional probability $\Pr(H|E)$ of hypothesis H given evidence E can be computed by the prior probability $\Pr(H)$ and the factor $\Pr(E|H)/\Pr(E)$.

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 (Tillers, 2011), and can in fact be traced back to early 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.

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, 1993a). 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.

Further readings The three normative frameworks (Anderson et al., 2005; Dawid et al., 2011; Kaptein et al., 2009). *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., 2014a). *Probabilities:* Evidence and probabilities (Mortera and Dawid, 2007; Schum, 1994) Statistics in the law (Fenton, 2011). 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), and more recently, the 2012 special issue of *Law, Probability and Risk*; Vol. 4, No. 2. *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, 2011; Bex et al., 2010). Bayesian networks for evidential reasoning (Fenton et al., 2013; Hepler et al., 2007). Combining arguments, scenarios and probabilities (Timmer et al., 2017; Verheij, 2014, 2017; Verheij et al., 2016; Vlek et al., 2016).

3 CONFLICTING EVIDENCE

In many situations, it is clear what the facts are. In a simple 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 support 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.

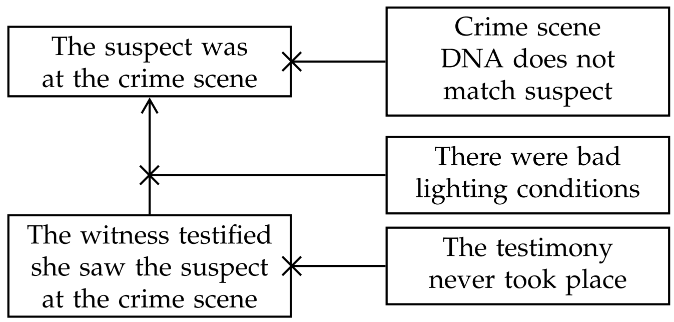

Figure 2: Three kinds of attack

3.1 Arguments

In the argument-based framework, the handling of conflicting evidence is analyzed in terms of reasons for and against a certain conclusion. Consider a crime case, where a witness testified 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 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. 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, although this is a bit more difficult to imagine. For instance, if the witness never actually testified that she saw the suspect at the crime scene, this attacks the very 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 Figure 2.

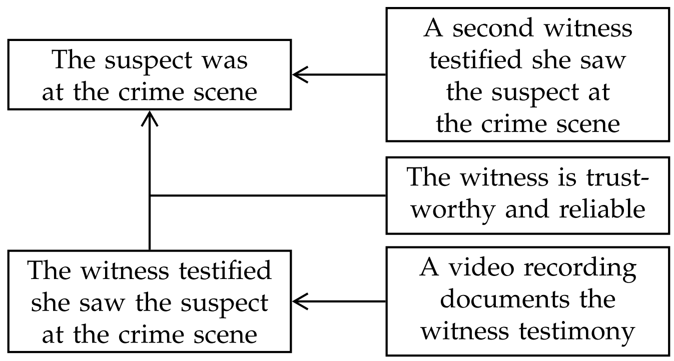

Figure 3: Three kinds of (further) support

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 unnamed kind of support of the connection between reason and conclusion.

Arguments can involve complex structures of supporting and attacking reasons. So far we have looked at a simple argument, consisting of a reason and a conclusion, along with three types of attacking reasons and three types of symmetric supporting reasons. But an

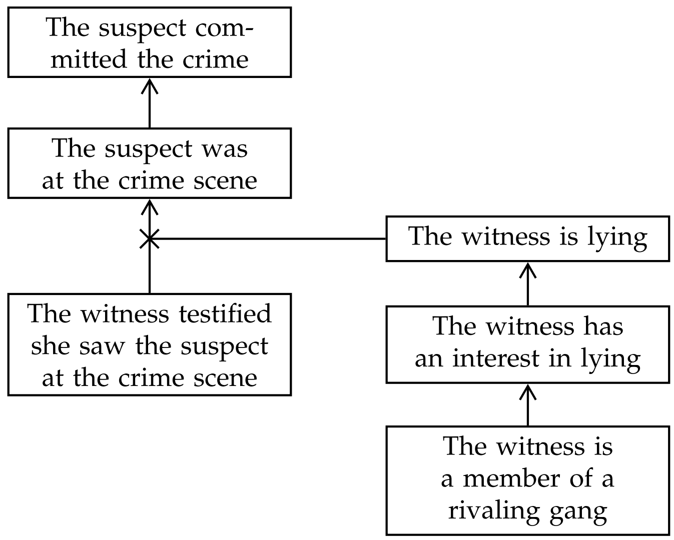

Figure 4: Supporting and attacking reasons can be chained.

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 Figure 4, on the left. 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 (Figure 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.

Further readings Argument structure and diagrams (Freeman, 1991; Toulmin, 1958; Wigmore, 1913). Defeasible reasoning and nonmonotonic logic (Gabbay et al., 1994; Pollock, 1987). Rebutting and undercutting attack (Pollock, 1987, 1995). Undermining attack (Bondarenko et al., 1997). Formal evaluation of defeasible arguments (Dung, 1995; Pollock, 1987, 1995; Prakken, 2010). Argumentative dialogue (Hage, 2000; Prakken, 1997; Toulmin, 1958; Walton and Krabbe, 1995). Accrual of reasons and weighing (Hage, 1997; Pollock, 1995; Prakken, 2005; Verheij, 1996). Argument diagramming and evaluation software (Gordon et al., 2007; Kirschner et al., 2003; Pollock, 1995; Reed and Rowe, 2004; van Gelder, 2003; Verheij, 2005).

3.2 Scenarios

In the scenario-based framework, the handling of conflicts is 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 has 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 a perpetrator other than the defendant. 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.

Consider, for example, the conflicting scenarios S_1 and S_2. 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. 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 exhibit 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. Scenario S_1 can explain this evidence, but not S_2. All in all, scenarios S_1 and S_2 are not only inconsistent on their face, 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 other 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.

Further readings 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).

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, $\Pr(H|E) > P(H)$. 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 guilty. This can be described probabilistically, as follows:

$$P(\textit{guilt}|\textit{testimony}) > P(\textit{guilt}).$$

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. Intuitively, this means that the presence of the evidence is more probable if the hypothesis is true than if the hypothesis is false. For the example considered earlier, we have:

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

These two characterizations of evidential support—in terms of probability increase and positive likelihood ratio—are in fact equivalent. For the following statements hold:

$$P(H|E) > P(H) \text{ iff } \frac{\Pr(E|H)}{\Pr(E|\neg H)} > 1.$$[7]

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, $\Pr(H|E) < P(H)$. 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 guilty. Probabilistically,

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

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 likely if the hypothesis is true than if the hypothesis is false. For the 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:

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

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).

Further readings On confirmation theory and accounts of evidential support (Bovens and Hartmann, 2003a; Carnap, 1950; Crupi, 2015; Fitelson, 1999; Hacking, 2001; Skyrms, 1999). Probabilistic accounts of evidential support in the law (Lempert, 1977).

4 EVIDENTIAL VALUE

The evidence found in a criminal investigation has different levels of evidential value: some evidence is very 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, $\Pr(H|E) - P(H)$. The larger the positive difference, the higher the value of the evidence for the hypothesis. An alternative approach is given by the likelihood ratio $\Pr(E|H)/P(E|\neg H)$. 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 $\Pr(H|E) - P(H)$ 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 $E_1, \ldots, E_n$, the overall evidential value of the evidence is measured as $\Pr(H|E_1, \ldots, E_n)$.

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 considred, 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 and there is a misalignment between the two values. 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. Then $\Pr(\neg H)$ is high and also $\Pr(\neg H|E_2)$ is high. 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 an estimated Random Match Probability (RMP). One way to interpret this probability is as the probability that a random person, who had nothing to do with the crime, would match. Now, with some simplifications (on these later), the evidential value of the DNA match M in favor of the suspect being the source of the sample S, in terms of a likelihood ratio, is as follows

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

The numerator $\Pr(M|S)$ equals 1 because we assume that if the defendant is the source of the sample, the lab test will report a match. As for the denominator, putting $\Pr(M|\neg S) = RMP$ is plausible because the probability that a match would be reported assuming that the defendant was *not* the source is roughly the same as the chance that a random person—someone who had no contact with the victim—would match anyway. For example, if the RMP 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 in favor of the suspect being the source has a high evidential value. More generally, a low RMP corresponds to a match with a rare profile, hence has a high evidential value.

Still, even with a low RMP 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 *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 undermine the inference along the way.

First, the inference from 'reported match' to 'true match' depends on the question whether the laboratory has made a mistake. A key source of lab mistakes originates in human error, much less rare than a good DNA profile. Second, the inference from 'true match' to 'source' can go wrong in several ways. Of course, even a rare match can be accidental, as measured by the RMP. But another cause of a match without the suspect being the source occurs in cases of close family relations. For instance, a suspect's genetically identical twin has identical DNA. 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 'guilt' can go wrong in many ways, because having visited a crime scene is not nearly the same as having committed the crime investigated.

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

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' 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 very 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 Section 3.1 (see in particular Figure 2, page 9). 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 questions will weaken the support for the conclusion that the suspect was at the crime scene, perhaps up to the point of making it non-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.

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 Figure 5, these situations are illustrated by two opposite witness testimonies.

Further readings Nonmonotonic reasoning (Gabbay et al., 1994). Prima facie reasons, undercutting and rebutting defeaters (Pollock, 1987, 1995). Warrants and backings (Toul-

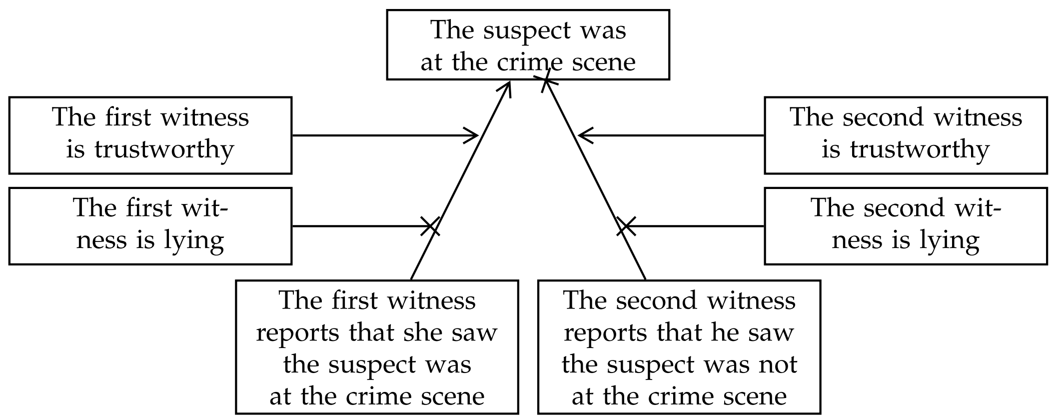

Figure 5: Arguments about whether a reason is supporting or attacking

min, 1958). Argument schemes and critical questions (Walton et al., 2008). Formal and computational argumentation (van Eemeren et al., 2014b).

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.

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 the very least, a scenario should not violate the laws of nature or common sense. 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 can *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 simple 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 defendant's. 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 towards 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 version.

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 6PM, 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.

Further readings 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 (Bex, 2011; Pennington and Hastie, 1993a; Wagenaar et al., 1993).

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.

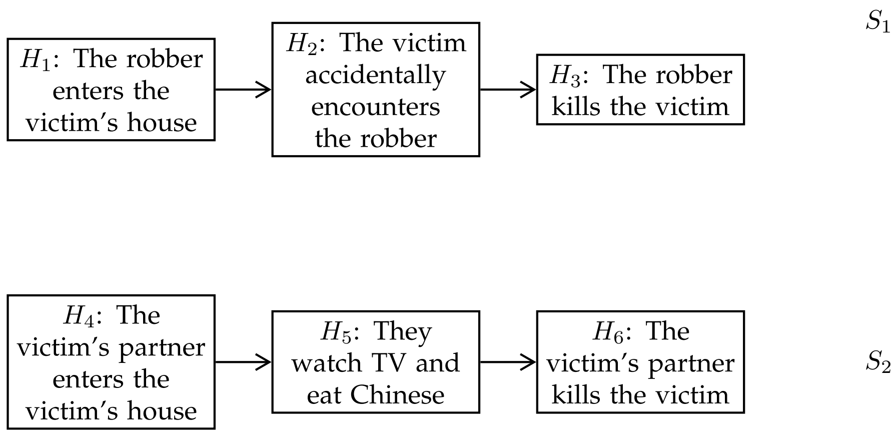

Figure 6: Scenarios and their structure. The second scenario lacks in causal structure.

Scenarios are coherent clusters of events, ordered in time and with causal relations. Earlier we encountered a simple scenario in a murder case, 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.

Contrast 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 take-out, 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 take-out (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.

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 structure or scripts. For example, in most violent crimes, we can identify an initial moment of conflict, which triggers a specific 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 vistkm. 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 take-out in front of the TV. But now suppose that the pieces of evidence are as follows: the investigators find Chinese take-out 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 victim's house (E_3). While the robbery scenario was more coherent, the Chinese take-out 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 take-out scenario, but inferior on another dimension, for example, the robbery scenario has less explanatory power than the Chinese take-out scenario.

Further readings 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, 1993b). 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., 2016). Worries about scenarios in law (Velleman, 2003). Scenarios shifting the legal perspective (Bex and

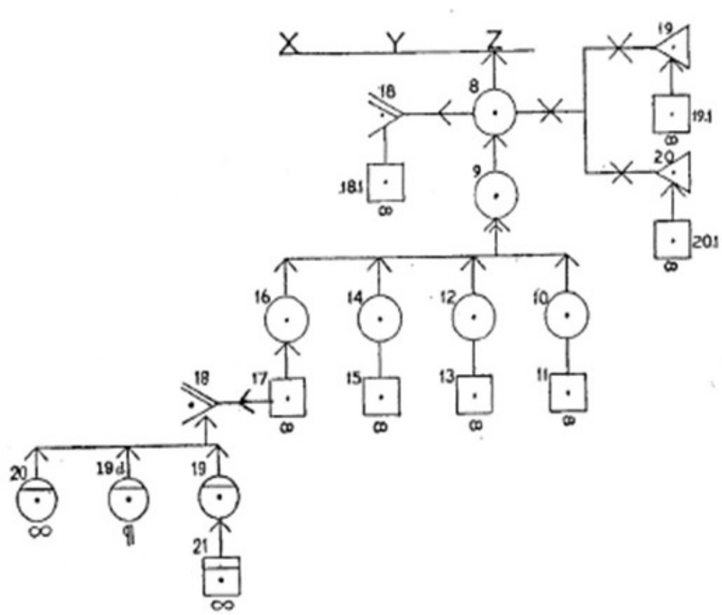

Figure 7: A Wigmore chart

Verheij, 2013).

5.2 Arguments

An analysis of a case in terms of arguments can become very complex. This was already noted by Wigmore, when he developed his charting method for analyzing the evidence in a criminal case (Wigmore, 1913). Figure 7 provides an example of a Wigmore diagram. Here Wigmore has analyzed the murder case *Commonwealth v. Umilian* (1901). The diagram contains some two dozen nodes. Diagrams for more complex cases can contain many more nodes.

The evaluation of an argument can depend on its subarguments. Given an argumentative analysis of the case, one would like to know which conclusions follow, and which don't. The structure of a complex of arguments influences the evaluation of the arguments, and in particular, the subarguments of a larger argument determine the evaluation of the whole. For example, consider the argument in Figure 4 (page 11). There the witness testimony supported the intermediate conclusion that the suspect was at the crime scene, which in turn supported the conclusion that the suspect committed the crime. So there is an argument consisting of two steps. In the example, the first of these steps is attacked by a counterargument involving the lying of the witness. As a result of this attack, the argument to the intermediate conclusion that the suspect was at the crime scene breaks down and its conclusion does not follow. As a result, also the larger two-step argument no longer supports its conclusion, which hence does not follow. Since the one-step subargument does not successfully support the intermediate conclusion, also the whole two-step argument for the final conclusion does not successfully support its conclusion.

The evaluation of an argument can depend on chains of attacks. When an argument is successfully attacked, it no longer successfully supports its conclusion. But attack can be chained, since the attack itself can be countered by a further attack. When an attack is successfully attacked, the original argument can be reinstated, in the sense that it again successfully supports its conclusion. Figure 9 shows an example. A first witness, Witness A, reports that the suspect was at the crime scene. Given only this information, there is good reason to assume that the suspect was at the crime scene. However, there is a second witness, Witness B, who reports that Witness A is lying. Given these two reasons, based on the witness reports by A and B, it is no longer successfully supported that the suspect was at the crime scene. If now there is a third witness, Witness C, who reports that Witness B is lying, the attack is countered. Witness B is no longer believed, so there is no reason to conclude that Witness A is lying. As a result, A's report can again support its conclusion that the suspect was at the crime scene. The original argument based on A's report is reinstated.

Conflicts between reasons can be addressed by exceptions, preferences and weighing. The counterarguments to a reason that result from asking critical questions give rise to conflicts between reasons. Sometimes conflicts of reasons can be resolved in the sense that it can be determined which conclusions follow from the conflicting reasons. We distinguish three kinds of addressing conflicts.

In the first kind of addressing conflicts between reasons, there is one reason for a conclusion and another against the conclusion, but there is an exception that excludes one of them. For instance, there are two witnesses with opposite reports about whether the suspect was at the crime scene or not, and one of them is lying (see the top of Figure 8). The exceptional situation is shown as an undercutting reason, i.e. an attack that goes against the connection between the reason and its conclusion (cf. Section 3.1). In the example, the conflict of reasons is resolved by an exception that has the effect that the conflicting reason does not support its conclusion.

In the second kind of addressing conflicts between reasons, there is also one reason for a conclusion and another against the conclusion, but one is preferred over the other. For instance, there are two witnesses with opposite reports about whether the suspect was at the crime scene or not, but one of them is more reliable (see the middle of Figure 8). This is an example of a rebutting attack (Section 3.1). The conflict of reasons is unresolved given

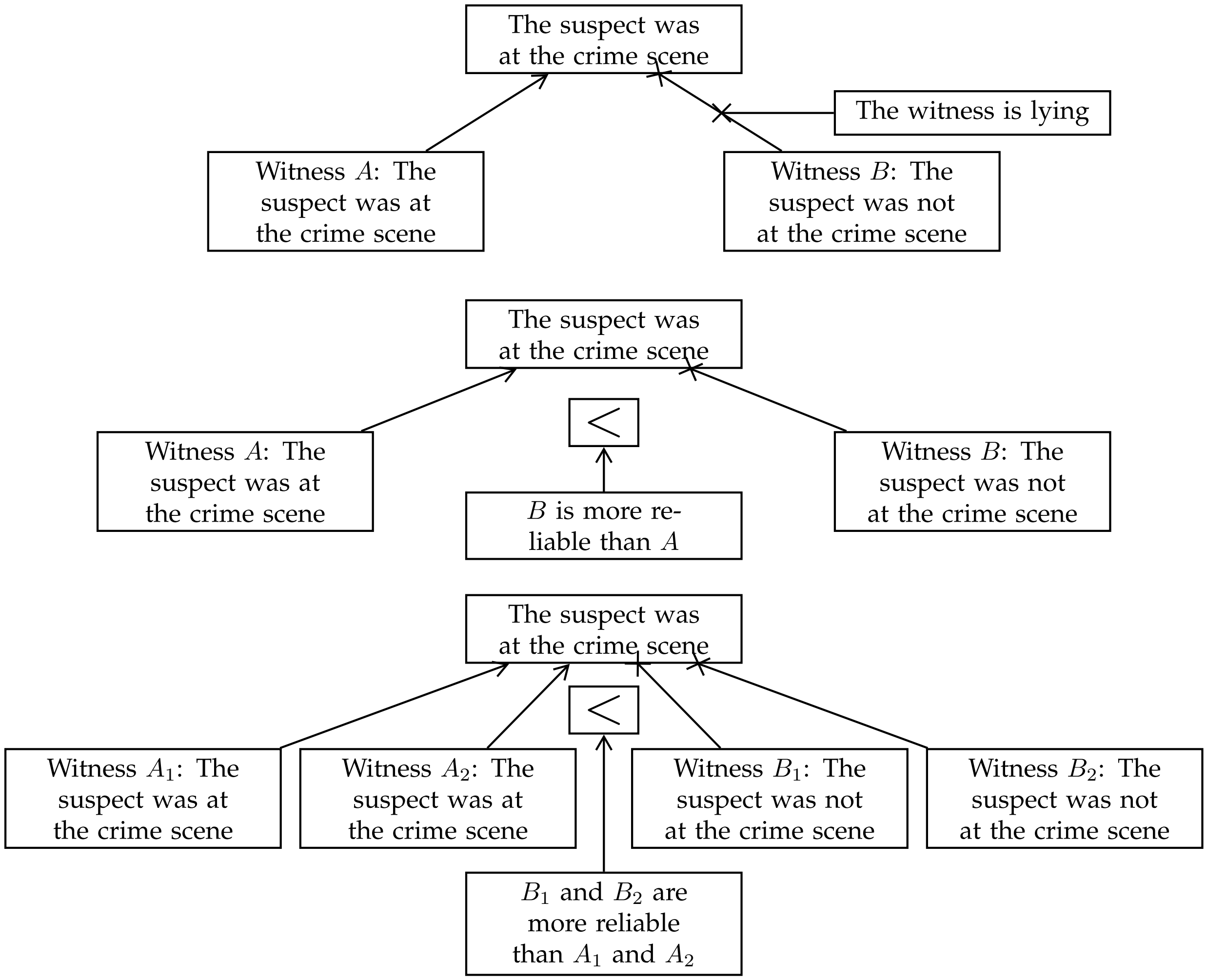

Figure 8: Three kinds of addressing conflicts of reasons

only the two reasons involved. Further information about the preference of one over the other resolves the conflict. A reason can be preferred over another, for instance when it is stronger. In the example, a preference (in the figure indicated by the >-sign) 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 of the more reliable witness will be drawn, thereby resolving the conflict.

The third kind of addressing conflicts between reasons discussed here involves more than two reasons. For instance, there can be more than two witnesses, with conflicting reports (Figure 8, bottom). Resolving such conflicts can be thought of as a weighing of reasons, generalizing the preference between two conflicting reasons.

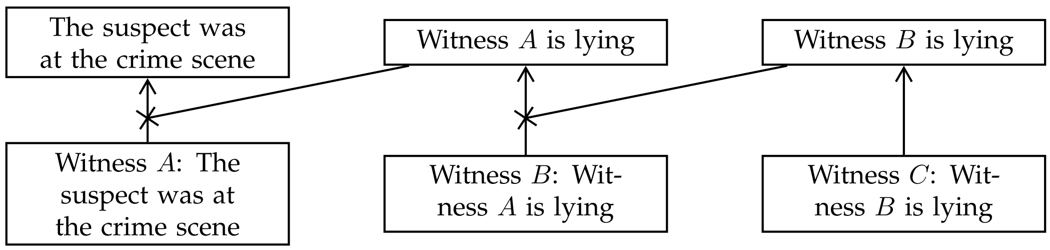

Figure 9: Reinstatement

Further readings Argument structure and their evaluation (Pollock, 1995). Formalizing argumentation (Prakken and Vreeswijk, 2002). Evaluating argument attack (Dung, 1995). Formal argumentation models (Gordon et al., 2007; Prakken, 2010; Simari and Loui, 1992; Verheij, 2003; Vreeswijk, 1997). Informal and formal argumentation theory (van Eemeren et al., 2014a).

5.3 Probability

The probability calculus provides formal rules for the coherent interpretation of the evidence.

The likelihood ratio formula shows how to find the posterior odds given the evidence. The odds of a hypothesis H are defined as the ratio $\Pr(H)/P(\neg H)$ of the probability of the hypothesis and the probability of its negation. Given evidence E, the odds $\Pr(H)/P(\neg H)$ of the hypothesis unconditioned on the evidence are called the *prior odds* of the hypothesis, and the odds $\Pr(H|E)/P(\neg H|E)$ of the hypothesis conditioned on the evidence the *posterior odds*.It follows from Bayes' theorem (Section 2.2) that the posterior odds can be found by multiplying the prior odds with the likelihood ratio:

$$\frac{\Pr(H|E)}{\Pr(\neg H|E)} = \frac{\Pr(E|H)}{\Pr(E|\neg H)} \cdot \frac{\Pr(H)}{\Pr(\neg H)}.$$[8]

This formula shows that an estimate of 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 an estimate of the posterior odds of a hypothesis. One needs an estimate of the prior odds.

To arrive at the *posterior probability* $\Pr(H|E)$ of a hypothesis conditional on the evidence, from the posterior odds $\frac{\Pr(E|H)}{\Pr(E|\neg H)}$ of the hypothesis given the evidence, the following formula applies:

$$\Pr(H|E) = \frac{\frac{\Pr(H|E)}{\Pr(\neg H|E)}}{1 + \frac{\Pr(H|E)}{\Pr(\neg H|E)}}.$$[9]

Consider an example. The incremental evidential value of a DNA match, relative to the hypothesis that the defendant is guilty, is given by the 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:

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

So the poster probability of the hypothesis is as follows:

$$\Pr(G|M) = \frac{\frac{\Pr(G|M)}{\Pr(\neg G|M)}}{1 + \frac{\Pr(G|M)}{\Pr(\neg G|M)}} \approx \frac{20}{1+20} \approx 0.95.$$

A generalization of the formula shows how to handle more pieces of evidence. So far we have considered only one piece of evidence. For two pieces of evidence E_1 and E_2, the formula can be generalized by multiplying the likelihood ratios of the individual pieces of evidence, as follows:

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

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

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 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, $\Pr(W|G) = P(W|G \wedge M)$.

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:

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

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

$$\Pr(G|M \wedge W) \approx \frac{20,000}{1 + 20,000} \approx 0.99.$$

Compare this probability with $\Pr(G|M)$, which was 0.95, 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:

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

It is easy to see the first, simple 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.

For the combination of multiple pieces of evidence and multiple hypotheses, more complex analytic tools can be used, in particular Bayesian networks. Probabilistic analyses as above become more and more complex when more elements are involved. Since many realistic crime cases involve many pieces of evidence and several hypotheses, probabilistic calculations quickly become unmanageable without appropriate modeling tools. As a prominent example of a modeling tool that allows for complex probabilistic representations and calculations, we discuss Bayesian networks. Bayesian networks have been proposed in computer science and artificial intelligence, and exploit probabilistic independencies to simplify representations and calculations.

A Bayesian network has a graphical and a numeric part. The graphical part is a directed, acyclic graph of the relevant probabilistic variables. The numeric part consists of conditional probability tables. In these tables, the conditional probabilities of a variable are specified, conditioned on the different instantiations of its parent variables. Consider for instance the Bayesian network shown in Figure 10. At the top, two hypotheses are shown, one expressing the suspect's guilt, the other that the suspect has an alibi. By the arrow

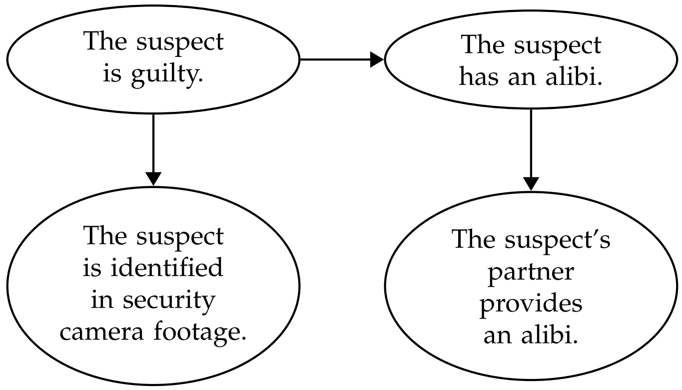

Figure 10: An example Bayesian network: directed acyclic graph

Guilt	Pr(Guilt)
False	0.999
True	0.001

Guilt	Alibi	Pr(Alibi|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

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

Table 1: An example Bayesian network: conditional probability tables

between them, the alibi variable has the guilt variable as a parent. The arrow expresses the dependence between the hypotheses: when one is true, the other is false, and vice versa. At the bottom, two pieces of evidence are shown: security camera footage that shows the suspect, and the testimony of the suspect's partner who provides an alibi. The evidence depends on the hypotheses, hence the evidence variables have the hypotheses as a parent.

Table 1 contains the numeric part of the Bayesian network, and contains four tables of conditional probabilities, one per variable. The prior probability of guilt has been set to 1 in a 1000 (top left table). The two hypotheses logically exclude one another (top right table). The two tables at the bottom specify how the occurrence of the evidence depends on the truth of the hypotheses. At the bottom left, it is shown how the camera identification depends on the guilt of the suspect. Given that the suspect is not guilty, there is a 1% chance that the suspect is identified in security camera footage. Given that the suspect is guilty, there is a 30% chance that the suspect is not identified, e.g., because he is not recognizable. At the bottom left, it is shown how the partner's testimony depends on the truth of the suspect's alibi. Given that the alibi is false, there is a 10% chance that the suspect's partner still testifies for the alibi. Given that the alibi is true, the partner certainly testifies for the alibi.

We illustrate how to calculate the probability $\Pr(\text{Guilt}|\text{Partner}\wedge\text{Camera})$ of guilt given the partner's alibi testimony and the camera identification from the Bayesian network. We use this equality:

$$\Pr(\text{Guilt} \mid \text{Partner} \wedge \text{Camera}) = \Pr(\text{Guilt} \wedge \neg\text{Alibi} \mid \text{Partner} \wedge \text{Camera})$$
$$= \frac{\Pr(\text{Partner} \wedge \text{Camera} \wedge \text{Alibi} \wedge \neg\text{Guilt})}{\Pr(\text{Partner} \wedge \text{Camera} \wedge \text{Alibi} \wedge \neg\text{Guilt}) \ + \ \Pr(\text{Partner} \wedge \text{Camera} \wedge \neg\text{Alibi} \wedge \text{Guilt})}$$

A Bayesian network provides a representation of a full probability distribution over the variables, by chaining conditional probabilities and using the dependencies and independencies represented in the graph. For instance, we have the following:

$\Pr(\text{Partner}\wedge\text{Camera}\wedge\text{Alibi}\wedge\neg\text{Guilt})$
$= \Pr(\text{Partner}|\text{Camera}\wedge\text{Alibi}\wedge\neg\text{Guilt}) \Pr(\text{Camera}|\text{Alibi}\wedge\neg\text{Guilt}) \Pr(\text{Alibi}|\neg\text{Guilt}) \Pr(\neg\text{Guilt})$
$= \Pr(\text{Partner}|\text{Alibi}) \Pr(\text{Camera}|\neg\text{Guilt}) \Pr(\text{Alibi}|\neg\text{Guilt}) \Pr(\neg\text{Guilt})$
$= 1*0.01*1*0.999 = 0.00999 \sim 1\%$

Here the first equality holds by the probability calculus, while the second equality uses independency relations as modeled in the Bayesian network. In particular, we have for this network:

$\Pr(\text{Partner}|\text{Camera}\wedge\text{Alibi}\wedge\neg\text{Guilt}) = \Pr(\text{Partner}|\text{Alibi})$

$\Pr(\text{Camera}|\text{Alibi}\wedge\neg\text{Guilt}) = \Pr(\text{Camera}|\neg\text{Guilt})$

These independencies follow from the graph of the Bayesian network as the node representing the partner's testimony has only the alibi variable node as a parent, and the camera node only the guilt node.

The following example calculation shows the probability that the suspect is guilty, while the partner testifies for the alibi and he is not identified.

$\Pr(\text{Partner}\wedge\neg\text{Camera}\wedge\neg\text{Alibi}\wedge\text{Guilt})$
$= \Pr(\text{Partner}|\neg\text{Camera}\wedge\neg\text{Alibi}\wedge\text{Guilt}) \Pr(\neg\text{Camera}|\neg\text{Alibi}\wedge\text{Guilt}) \Pr(\neg\text{Alibi}|\text{Guilt}) \Pr(\neg\text{Guilt})$
$= \Pr(\text{Partner}|\neg\text{Alibi}) \Pr(\neg\text{Camera}|\text{Guilt}) \Pr(\neg\text{Alibi}|\text{Guilt}) \Pr(\text{Guilt})$
$= 0.1*0.3*1*0.001 = 0.00003$

As the two hypotheses exclude one another, some probabilities are equal to 0:

$\Pr(\text{Partner}\wedge\text{Camera}\wedge\text{Alibi}\wedge\text{Guilt})$
$= \Pr(\text{Partner}|\text{Camera}\wedge\text{Alibi}\wedge\text{Guilt}) \Pr(\text{Camera}|\text{Alibi}\wedge\text{Guilt}) \Pr(\text{Alibi}|\text{Guilt}) \Pr(\text{Guilt})$

= Pr(Partner|Alibi) Pr(Camera|Guilt) Pr(Alibi|Guilt) Pr(Guilt)
= 1*0.7*0*0.001 = 0

For probability functions of many variables that have many independencies, a Bayesian network representation can be significantly more compact than a general full probability distribution in 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).

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

6 REASONING AND DECISION MAKING

So far we have focused on how the evidence can be evaluated and combined, and how inferences can be drawn. 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 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 (1954), 348 U.S. 121, 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 narratives—can be used to characterize more precisely the standard of proof, although they are not immune from shortcomings, as we shall soon see.

Further readings Evidence law manuals (Fisher, 2008; Méndez, 2008). Criminal Procedure manuals (Allen et al., 2005). Character evidence and its exclusion (Redmayne, 2015).

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 estimated by weighing the evidence with the probability calculus. On the probabilistic framework, the goal is to estimate the probability of the defendant's guilt based on all the available evidence. The estimation begins with the lowest possible value for the guilt probability, prior to considering any evidence. As more evidence is presented, the guilt probability moves upwards or downwards depending on whether the evidence is incriminating or exculpatory. When all the evidence is considered, a final guilt probability value is reached, all things considered. 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 estimated frequencies, but often they are not. For example, $\Pr(G)$ is required to calculate $\Pr(G|E)$ using Bayes' theorem. $\Pr(G)$ is the probability of the defendant's guilt regardless of the evidence presented at trial. What should $\Pr(G)$ be? It is subject to debate how to estimate that number, and even whether it makes sense to estimate it.

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,

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

The inequality holds just in case

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

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 hold 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 estimate all the required probabilities. The characterization is simple, crisp and elegant, but a too literal interpretation of it is 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 this value could be known, other probability values might remain unknown. 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 estimation 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 estimate 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.

Further readings Probabilistic accounts of the burden of proof (Cheng, 2013; Hamer, 2004; Kaplan, 1968; Kaye, 1986, 1999). Critique of probabilistic accounts (Cohen, 1977; Haack, 2014; Ho, 2008; Nesson, 1979; Pardo and Allen, 2008; Stein, 2005; Thomson, 1986). 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 the proof beyond a reasonable doubt as understood in the law (Larry Laudan, 2006). History of beyond a reasonable doubt standard (Shapiro, 1991; Whitman, 2008). Other measures, weight, resiliency and completeness of the evidence (Kaye, 1999; Nance, 2016; Stein, 2005).

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 reasons, 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, allow 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 the principle of expected utility theory be applied to the argument framework as well?

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

6.3 Scenarios

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

Competing scenarios are collected and compared. On the narrative 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, etc. 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. A decision rule could stipulate that the best explanatory scenario is 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 etc. 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.

Further readings 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).

7 SUMMARY AND CONCLUSION

We have discussed evidential reasoning in the law. For this, we have distinguished three normative frameworks: 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 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.

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 has 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.

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.

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. For the combination of multiple pieces of evidence and multiple hypotheses, more complex analytic tools can be used, in particular Bayesian networks.

Reasoning and decision making

Probabilities The guilt probability is estimated by weighing the evidence with the probability calculus. The decision criterion is a guilt probability threshold. It is not obvious how to estimate 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.

ACKNOWLEDGMENT

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 programme 'Probability and Statistics in Forensic Science' which was supported by EPSRC Grant Number EP/K032208/1.

NOTES

[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, Chapter 3, Section 1, available at `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`.

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

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

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

[6]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. Hence, $\Pr(H \wedge E) = \Pr(E|H) \cdot \Pr(H)$. 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 suggested by the terminology of prior and posterior probability. The terminology is standard usage in approaches of Bayesian updating.

[7]To see why, recall 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

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

For one direction, if $\Pr(H|E) > P(H)$, then $1-P(H|E) < 1-P(H)$. This means that $\frac{\Pr(H|E)}{1-P(H|E)} > \frac{\Pr(H)}{1-P(H)}$, 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 $\frac{\Pr(H|E)}{1-P(H|E)} > \frac{\Pr(H)}{1-P(H)}$. To establish $\Pr(H|E) > P(H)$, suppose for contradiction that $\Pr(H|E) \leq P(H)$, which implies $1 - P(H|E) \geq 1 - P(H)$. This means that $\frac{\Pr(H|E)}{1-P(H|E)} \leq \frac{\Pr(H)}{1-P(H)}$. This contradicts $\frac{\Pr(H|E)}{1-P(H|E)} > \frac{\Pr(H)}{1-P(H)}$, and thus $\Pr(H|E) > P(H)$.

[8]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, 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.

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

REFERENCES

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

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

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

R. J. Allen, J. L. Hoffmann, D. Livingston, and W. Stuntz. *Comprehensive Criminal Procedure (2nd ed.)*. Aspen, 2005.

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

D. J. Balding. *Weight-of-evidence for forensic DNA profiles*. John Wiley and Sons, 2005.

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

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

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

F. J. Bex and B. Verheij. Legal stories and the process of proof. *Artificial Intelligence and Law*, 21(3):253–278, 2013.

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

A. Bondarenko, P. M. Dung, R. A. Kowalski, and F. Toni. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence*, 93:63–101, 1997.

L. BonJour. *The Structure of Empirical Knowledge*. Harvard University Press, 1985.

L. Bovens and S. Hartmann. *Bayesian Epistemology*. Oxford University Press, 2003a.

L. Bovens and S. Hartmann. Solving the riddle of coherence. *Mind*, 112:601–633, 2003b.

J. Buckleton. A framework for interpreting evidence. In J. Buckleton, C. M. Triggs, and S. J. Walsh, editors, *Forensic DNA Evidence Interpretation*. CRC Press, 2005.

R. Carnap. *Logical Foundations of Probability*. University of Chicago Press, 1950.

E. Cheng. Reconceptualizing the burden of proof. *Yale Law Journal*, 122(5):1104–1371, 2013.

L. J. Cohen. *The Probable and the Provable*. Clarendon Press, Oxford, 1977.

R. Cook, I. W. Evett, G. Jackson, P. J. Jones, and J. A. Lambert. A hierarchy of propositions: deciding which level to address in casework. *Science and Justice*, 38(4):231–239, 1998.

V. Crupi. Confirmation. In E. N. Zalta, editor, *Stanford Encyclopedia of Philosophy*. Stanford University, 2015.

A. Darwiche. *Modeling and Reasoning with Bayesian networks*. Cambridge University Press, Cambridge, 2009.

A. P. Dawid. The difficulty about conjunction. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 36(2/3):91–92, 1987.

A. P. Dawid. Bayes's theorem and weighing evidence by juries. In *Bayes's Theorem*, volume 113, pages 71–90. Oxford University Press, 2002.

A. P. Dawid. Beware of the DAG! In I. Guyon, D. Janzing, and B. Schölkopf, editors, *JMLR Workshop and Conference Proceedings: Volume 6. Causality: Objectives and Assessment (NIPS 2008 Workshop)*, pages 59–86. jmlr.org, 2010.

A. P. Dawid, W. Twining, and M. Vasiliki, editors. *Evidence, Inference and Enquiry*. Oxford University Press, Oxford, 2011.

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

I. Evett, G. Jackson, J. A. Lambert, and S. McCrossan. The impact of the principles of evidence interpretation on the structure and content of statements. *Science and Justice*, 40 (4):233–239, 2000.

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

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

N. E. Fenton, M. D. Neil, and D. A. Lagnado. A general structure for legal arguments about evidence using Bayesian Networks. *Cognitive Science*, 37:61–102, 2013.

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

G. Fisher. *Evidence (2nd Edition)*. Foundation Press, 2008.

B. Fitelson. The plurality of Bayesian measures of confirmation and the problem of measure sensitivity. *Philosophy of Science*, 66:362–378, 1999.

J. B. Freeman. *Dialectics and the Macrostructure of Arguments. A Theory of Argument Structure*. Foris, Berlin-New York, 1991.

M. Friedman. Explanation and scientific understanding. *Journal of Philosophy*, 71:5–19, 1974.

R. D. Friedman. Route analysis of credibility and hearsay. *The Yale Law Journal*, 97(4): 667–742, 1987.

R. D. Friedman. A presumption of innocence, not of even odds. *Stanford Law Review*, 52: 873–887, 2000.

D. Frumkin, Waserstrom A., A. Davidson, and A. Grafit. Authentication of forensic dna samples. *Forensic Science International: Genetics*, 4(2):95–103, 2009.

D. M. Gabbay, C. J. Hogger, and J. A. Robinson, editors. *Handbook of Logic in Artificial Intelligence and Logic Programming. Volume 3. Nonmonotonic Reasoning and Uncertain Reasoning*. Clarendon Press, Oxford, 1994.

T. F. Gordon and D. N. Walton. Proof burdens and standards. In I. Rahwan and G. R. Simari, editors, *Argumentation in Artificial Intelligence*, pages 239–258. Springer, 2009.

T. F. Gordon, H. Prakken, and D. N. Walton. The Carneades model of argument and burden of proof. *Artificial Intelligence*, 171(10–15):875–896, 2007.

L. K. Griffin. Narrative, truth, trial. *Georgetown Law Journal*, 101:281–335, 2013.

S. Haack. Warrant, causation, and the atomist of evidence law. *Journal of Social Epistemology*, 5:253–265, 2008.

S. Haack. Legal probabilism: An epistemological dissent. In *Evidence Matters. Science, Proof, and Truth in the Law*, pages 47–77. Cambridge University Press, Cambridge, 2014.

I. Hacking. *An Introduction to Probability and Inductive Logic*. Cambridge University Press, Cambridge, 2001.

J. C. Hage. *Reasoning with Rules. An Essay on Legal Reasoning and its Underlying Logic*. Kluwer, Dordrecht, 1997.

J. C. Hage. Dialectical models in artificial intelligence and law. *Artificial Intelligence and Law*, 8:137–172, 2000.

D. Hamer. 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, 2004.

C. Hempel and P. Oppenheim. Studies in the logic of explanation. *Philosophy of Science*, 15: 135–175, 1948.

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

H. L. Ho. *Philosophy of Evidence Law*. Oxford University Press, 2008.

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

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

J. Kaplan. Decision theory and the fact-finding process. *Stanford Law Review*, 20(1065-1092), 1968.

L. Kaplow. Burden of proof. *Yale Law Journal*, 121(4):738–1013, 2012.

H. Kaptein, H. Prakken, and B. Verheij, editors. *Legal Evidence and Proof: Statistics, Stories, Logic (Applied Legal Philosophy Series)*. Ashgate, Farnham, 2009.

D. H. Kaye. Do we need a calculus of weight to understand proof beyond a reasonable doubt? *Boston University Law Review*, 66:657–672, 1986.

D. H. Kaye. DNA evidence: Probability, population genetics and the courts. *Harvard Journal of Law and Technology*, 7:101–172, 1993.

D. H. Kaye. Clarifying the burden of persuasion: What Bayesian rules do and not do. *International Commentary on Evidence*, 3:1–28, 1999.

D. H. Kaye. Beyond uniqueness: the birthday paradox, source attribution and individualization in forensic science. *Law, Probability and Risk*, 12(1):3–11, 2013.

D. H. Kaye and G. F. Sensabaugh. Reference guide on DNA evidence. In *Reference Manual on Scientific Evidence (2dn ed.)*, pages 576–585. Federal Judicial Center, 2000.

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

J. Keppens and B. Schafer. Knowledge based crime scenario modelling. *Expert Systems with Applications*, 30(2):203–222, 2006.

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

J. J. Koehler. Error and exaggeration in the presentation of DNA evidence in trial. *Jurimetrics Journal*, 34:21–39, 1993.

L. Larry Laudan. *Truth, Error, and Criminal Law: An Essay in Legal Epistemology*. Cambridge University Press, Cambridge, 2006.

R. O. Lempert. Modeling relevance. *Michigan Law Review*, 75(5/6):1021–1057, 1977.

P. Lipton. *Inference to the Best Explanation*. Routledge, 1991.

M. A. Méndez. *Evidence: The California Code and the Federal Rules (4th edition)*. Thomson West, 2008.

J. Mortera and P. Dawid. Probability and evidence. In T. Rudas, editor, *Handbook of Probability Theory*. Sage Handbook, 2007.

D. A. Nance. *The Burdens of Proof: Discriminatory Power, Weight of Evidence, and Tenacity of Belief*. Cambridge University Press, Cambridge, 2016.

C. R. Nesson. Reasonable doubt and permissive inferences: The value of complexity. *Harvard Law Review*, 92(6):1187–1225, 1979.

NRC. *The Evaluation of Forensic DNA Evidence*. National Academy Press, 1996.

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

J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo (California), 1988.

J. Pearl. *Causality: Models, Reasoning and Inference. Second Edition*. Cambridge University Press, Cambridge, 2000/2009.

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

N. Pennington and R. Hastie. *Inside the Juror*, chapter The Story Model for Juror Decision Making, pages 192–221. Cambridge University Press, Cambridge, 1993b.

F. Picinali. Two meanings of “reasonableness”: Dispelling the “floating” reasonable doubt. *Modern Law Review*, 76(5):845–875, 2013.

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

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

H. Prakken. *Logical Tools for Modelling Legal Argument. A Study of Defeasible Reasoning in Law*. Kluwer, Dordrecht, 1997.

H. Prakken. A study of accrual of arguments, with applications to evidential reasoning. In *Proceedings of the Tenth International Conference on Artificial Intelligence and Law*, pages 85–94. ACM Press, New York (New York), 2005.

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

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

H. Prakken and G. Sartor. A logical analysis of burdens of proof. In H. Kaptein, H. Prakken, and B. Verheij, editors, *Legal Evidence and Proof: Statistics, Stories, Logic*, chapter 9, pages 223–253. Ashgate, 2009.

H. Prakken and G. A. W. Vreeswijk. Logics for defeasible argumentation. In D. M. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic, Second Edition*, volume 4, pages 218–319. Kluwer Academic Publishers, Dordrecht, 2002.

M. Redmayne. *Character Evidence in the Criminal Trial*. Oxford University Press, 2015.

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

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

W. Salmon. *Scientific Explanation and the Causal Structure of the World*. Princeton University Press, 1984.

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

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

D. A. Schum. *The Evidential Foundations of Probabilistic Reasoning*. John Wiley and Sons, New York (New York), 1994.

D. A. Schum and S. Starace. *The Evidential Foundations of Probabilistic Reasoning*. Northwestern University Press, 2001.

B. Shapiro. *Beyond Reasonable Doubt and Probable Cause: Historical Perspectives on the Anglo-American Law of Evidence*. University of California Press, 1991.

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

B. Skyrms. *Choice and Chance: An Introduction to Inductive Logic (4th edition)*. Wadsworth Publishing, 1999.

A. Stein. *Foundations of Evidence Law*. Oxford University Press, 2005.

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

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

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

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

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

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

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

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

S. T. Timmer, J. J. Meyer, H. Prakken, S. Renooij, and B. Verheij. A two-phase method for extracting explanatory arguments from Bayesian Networks. *International Journal of Approximate Reasoning*, 80:475–494, 2017.

S. E. Toulmin. *The Uses of Argument*. Cambridge University Press, Cambridge, 1958.

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

F. H. van Eemeren, B. Garssen, E. C. W. Krabbe, A. F. Snoeck Henkemans, B. Verheij, and J. H. M. Wagemans. *Handbook of Argumentation Theory*. Springer, Berlin, 2014a.

F. H. van Eemeren, B. Garssen, E. C. W. Krabbe, A. F. Snoeck Henkemans, B. Verheij, and J. H. M. Wagemans. Chapter 11: Argumentation in Artificial Intelligence. In *Handbook of Argumentation Theory*. Springer, Berlin, 2014b.

T. van Gelder. Enhancing deliberation through computer supported argument visualization. In P. A. Kirschner, S. J. B. Shum, and C. S. Carr, editors, *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*, pages 97–115. Springer, New York (New York), 2003.

D. Velleman. Narrative explanation. *The Philosophical Review*, 112(1):1–25, 2003.

B. Verheij. *Rules, Reasons, Arguments. Formal studies of argumentation and defeat.* Dissertation Universiteit Maastricht, Maastricht, 1996.

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

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

B. Verheij. To catch a thief with and without numbers: Arguments, scenarios and probabilities in evidential reasoning. *Law, Probability and Risk*, 13:307–325, 2014.

B. Verheij. Proof with and without probabilities. correct evidential reasoning with presumptive arguments, coherent hypotheses and degrees of uncertainty. *Artifical Intelligence and Law*, 2017.

B. Verheij, F. J. Bex, S. T. Timmer, C. S. Vlek, J. J. Meyer, S. Renooij, and H. Prakken. Arguments, scenarios and probabilities: Connections between three normative frameworks for evidential reasoning. *Law, Probability and Risk*, 15(1):35–70, 2016.

C. S. Vlek, H. Prakken, S. Renooij, and B. Verheij. Building Bayesian Networks for legal evidence with narratives: a case study evaluation. *Artifical Intelligence and Law*, 22(4): 375–421, 2014.

C. S. Vlek, H. Prakken, S. Renooij, and B. Verheij. A method for explaining Bayesian Networks for legal evidence with scenarios. *Artifical Intelligence and Law*, 24(3):285–324, 2016.

G. A. W. Vreeswijk. Abstract argumentation systems. *Artificial Intelligence*, 90:225–279, 1997.

W. A. Wagenaar, P. J. van Koppen, and H. F. M. Crombag. *Anchored Narratives. The Psychology of Criminal Evidence*. Harvester Wheatsheaf, London, 1993.

D. N. Walton and E. Krabbe. *Commitment in Dialogue. Basic Concepts of Interpersonal Reasoning*. State University of New York Press, Albany (New York), 1995.

D. N. Walton, C. Reed, and F. Macagno. *Argumentation Schemes*. Cambridge University Press, Cambridge, 2008.

D. Wasserman. Forensic DNA typing. In J. Burley and J. Harris, editors, *Companion to Genetics*. Blackwell, 2008.

B. S. Weir. The rarity of DNA profiles. *The Annals of Applied Statistics*, 1:358–370, 2007.

G. L. Wells. Naked statistical evidence of liability: Is subjective probability enough? *Journal of Personality and Social Psychology*, 62:793–752, 1992.

J. Q. Whitman. *The Origins of Reasonable Doubt: Theological Roots of the Criminal Trial*. Yale University Press, 2008.

J. H. Wigmore. *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.* Little, Brown and Company, Boston (Massachusetts), 1913.

J. Woodward. Scientific explanation. In E. N. Zalta., editor, *The Stanford Encyclopedia of Philosophy*. Stanford University, 2014.

S. L. Zabell. Fingerprint evidence. *Journal of Law and Policy*, 13:143–179, 2005.