EVIDENTIAL REASONING
Chapter for the Handbook of Legal Reasoning

Marcello Di Bello & Bart Verheij – October 6, 2016

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 Netherlands, for instance, the conviction rate of suspects that appear in criminal courts is reported to be around 95% year after year.[1] In the United States, the conviction rate in federal courts has been roughly 75% and in Japan has reached as high a rate as 99%.[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 'beyond a reasonable doubt'?

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

The fact-finders, lay jurors or professional judges, aim to reconstruct what has happened in the crime on the basis of the evidence. 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 detailed information about what has happened on the scene of the crime. Here is an example.

Q: Can you describe what happened that day?

[1] Source: CBS, the Dutch central bureau of statistics, publishing its data at www.cbs.nl.
[2] **SOURCE TO BE ADDED**

> 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 rich in specific details, and on its basis the fact finders can form a hypothesis—or, to use the terminology we adopt here, a scenario—about what happened. Still, this 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 a witness has wrongly interpreted what he saw, that time has distorted his memories, or that the witness is intentionally lying.

1.2 DNA profiling

DNA profiling has become an important tool in courts. DNA profiling has a strong scientific underpinning, and comes with precise statistical information. 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. Each locus is associated with a number of different *alleles* which may occur that location, where an allele differs from another because of the number of repetitions of a small DNA sequence. 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.[3] Different countries use different sets of—what are called—*core loci* for their forensic DNA profile databases. For instance, the CODIS system in the United States uses 13 core loci, to be expanded to 20 core loci beginning January 2017.[4]

How is the statistical frequency of a DNA profile estimated? As said, 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

[3]See `http://www.cstl.nist.gov/strbase/str_CSF1PO.htm`.
[4]See `https://www.fbi.gov/services/laboratory/biometric-analysis/codis`.

of that allele at that locus in the population. Second, the measured proportional frequencies for the individual alleles at the core loci are multiplied to compute what is called the Random Match Probability of the DNA profile.[5] These Random Match Probabilities—and numbers mathematically related to them—are the numbers reported by forensic experts in courts, and the smaller they are, the higher the evidential value of the profile is taken to be. 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. The use of more loci leads to smaller Random Match Probabilities. 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.

Suppose now that a trace of blood has been found on the scene of the crime, and that the found DNA profile matches that of the suspect's DNA. Using this evidence, 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. It is a common misunderstanding to equate this number with the probability that the suspect is not the source of the trace. This well-known misunderstanding is referred to as the prosecutor's fallacy. The probability that the suspect is not the source of the trace can only be determined from the Random Match Probability, after considering the prior odds that the suspect is the source.

The hypothesis that can be formed on the basis of a DNA match is very specific, and is limited to the suspect being the source of the trace. The hypothesis need not be true, in particular in the cases of an accidental match, the existence of an identical twin—that at a rate of a dozen or more twin births per 1000 live births[6] are not all that rare—, or a lab error.

1.3 Central questions

Using the two kinds of evidence as an illustration, we can now provide the list of central questions associated with evidential reasoning that we use to structure this chapter.

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?

[5] Some special care is needed to accommodate for the fact that an allele can be from either part of the double helix that comprises our DNA.

[6] Source: `https://en.wikipedia.org/wiki/Twin#Statistics`.

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 profile match can support 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"? When have we collected enough evidence to make a decision?

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, the four questions we set out above will be discussed, consecutively, 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, as distinguished in the scholarly literature (Anderson et al., 2005; Dawid et al., 2011; Kaptein et al., 2009). The first framework discussed uses arguments as primary tool, the second scenarios, and the third probabilities.

2.1 Arguments

The first normative framework for the handling of evidence that we discuss uses arguments as primary tool. Arguments contain reasons that support or attack the conclusions considered. For instance, when a witness reports that she saw the suspect at the crime scene, there is a reason for the suspect having been at the crime scene. There is a reason attacking that conclusion when the DNA profile found at the crime scene does not match the suspect's (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, 1931).

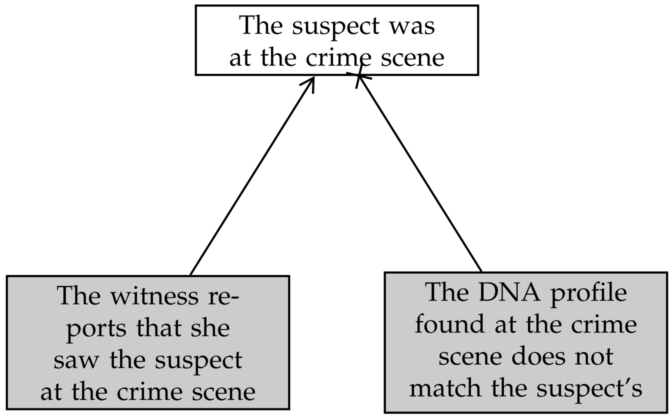

Figure 1: Arguments with supporting and attacking reasons

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 for the correct handling of the evidence uses probabilities as main tool. The probability calculus is used to connect the probabilities of evidence and events, conditioned on each other. Consider for instance a trace found at the crime scene with a rare DNA profile of estimated frequency 1 in a billion. Because the profile is rare, a match is not often found accidentally. This statement can be made precise in the probability calculus. When E expresses the evidence that the suspect's profile matches the trace's and H that the suspect is not the source, we write:

$$\Pr(E|H) = 1/10^9$$

Here $\Pr(E|H)$ denotes the conditional probability that the suspect's profile matches the trace's, given the condition that the suspect is not the source. Conditional probabilities obey the famous Bayes' theorem:

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

This formula shows how the posterior probability $\Pr(H|E)$ of the hypothesis given the evidence can be computed by multiplying the prior probability $\Pr(H)$ and the Bayes 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.[7]

2.3 Scenarios

The third normative framework for the correct handling of the evidence centers around scenario analysis. In a scenario, a coherent account of what may have happened in a case is made explicit. Different scenarios are contrasted, and evaluated, by considering their plausibility and by checking to what extent they match and contradict the available evidence.

For instance, consider a murder case with two suspects: the victim's former partner and a robber. For each suspect, a scenario is considered that explains the murder:

S_1: The victim's former partner killed the victim after a fight.

S_2: The robber killed the victim when caught during a robbery.

When the robber confesses having killed the victim during a robbery, there is evidence contradicting scenario S_1 and matching scenario S_2.

Scenario analysis proves helpful when considering a complex case and its evidence. The coherent explanation of the evidence provided by a scenario can be regarded as a sense-making tool for handling cases with a large dossier. In particular, legal psychology has contributed to our knowledge about the role of scenarios in handling the evidence (Bennett and Feldman, 1981; Pennington and Hastie, 1993a). Scenarios were shown to be misleading, as experiments showed that a false scenario told in a sensible chronological order was more easily believed than a true scenario that was told in a random order. Still, the legal psychologists Wagenaar et al. (1993) emphasised the usefulness of scenario analysis for the rational handling of the evidence, using the technique in their work on debunking dubious case decisions. Scenario analysis is connected with inference to the best explanation (Pardo and Allen, 2008).

Further readings The three normative frameworks (Anderson et al., 2005; Dawid et al., 2011; Kaptein et al., 2009). Wigmore charts (Wigmore, 1913, 1931). The New Evidence

[7] A recent instance of the debate concerns the R v T case, where the UK Court of Appeal restricted the use of Bayes' theorem in courts to cases with a solid statistical foundation such as DNA; see the 2012 special issue of Law, Probability and Risk; Vol. 4, No. 2. For a 1970s instance of the debate, see Finkelstein and Fairley (1970); Tribe (1971).

Scholarship (Anderson et al., 2005). Informal and formal argumentation theory (van Eemeren et al., 2014a). Evidence and probabilities Mortera and Dawid (2007); Schum (194) Statistics in the law (Fenton, 2011). Miscarriages of justice involving statistics (Dawid et al., 2011; Schneps and Colmez, 2013). 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). 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., 2015; Verheij, 2014; Verheij et al., 2016; Vlek et al., 2014).

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.

3.1 Arguments

In the argumentative normative framework, the handling of conflicting evidence is analyzed in terms of the arguments for and against the different positions considered.

The arguments for and against different positions have structure, involving complexes of reasons supporting and attacking positions. Consider a crime case, where a witness reports that she saw the suspect at the crime scene. Then there is a reason supporting that the suspect indeed was at the crime scene, which in turn provides some support to the conclusion that the suspect committed the crime. This chain of supporting reasons is graphically depicted in Figure 2, on the left. When it is now discovered that the witness is a member of a rivaling gang, there is reason to believe that she is lying (Figure 2, on the right). The reason that the witness is lying attacks the argument on the left, in the sense that the witness testimony no longer provides support for the suspect being at the crime scene.

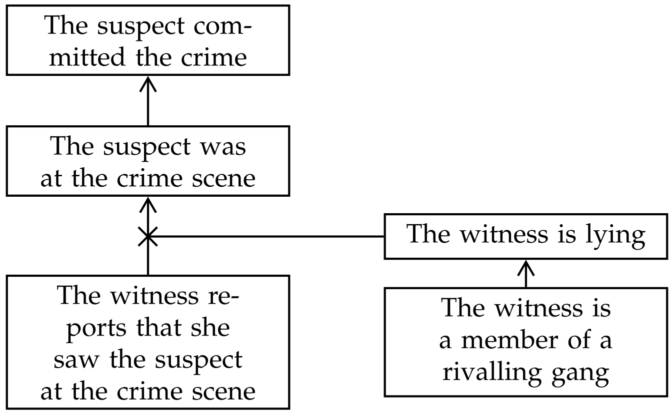

Figure 2: Arguments have structure

Three kinds of support can be distinguished: multiple, subordinated and coordinated support. If we consider the argument thet the suspect was at the crime scene because the witness reports that she saw the suspect at the crime scene, each of the three parts of the argument can be supported: the conclusion, the reason and the connection between the reason and the conclusion. For instance, the conclusion can be supported by a second witness testimony. The reason can be supported by the police report that documents the witness testimony. The connection between the reason and the conclusion can be supported by the trustworthiness of the witness report. Additional support of the conclusion is referred to as multiple support. Additional support of the reason is called subordinating support. Additional support of 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 more than one element that in their conjunctive combination provide support for the conclusion. Coordinated support is distinguished from multiple support since in the latter each element of the support for the conclusion provides support by itself. Figure 3 shows the three named kinds of support. Note that multiple and subordinated support are graphically visualized with an arrow, whereas coordinated support is shown with a line. An arrow would indicate the unnamed kind of support of the connection between reason and conclusion.

Three kinds of attack can be distinguished: rebutting, undercutting, and undermining attack. Consider again the argument for the position that the suspect was at the crime scene, as supported by the reason of a witness reporting that she saw the suspect at the crime scene (Figure 4, on the left). This argument can be attacked in three ways. First, its conclusion can be attacked. The suspect can for instance have an alibi, showing that he

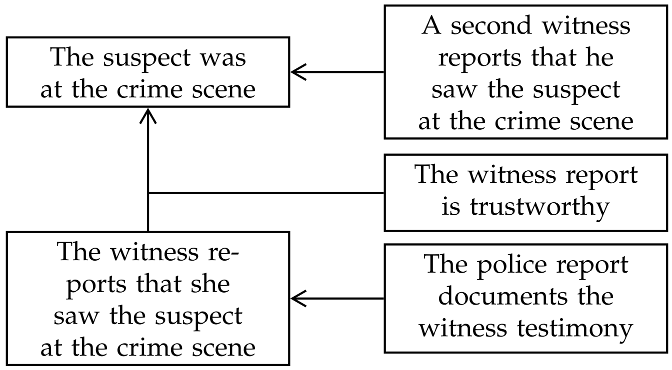

Figure 3: Three kinds of support

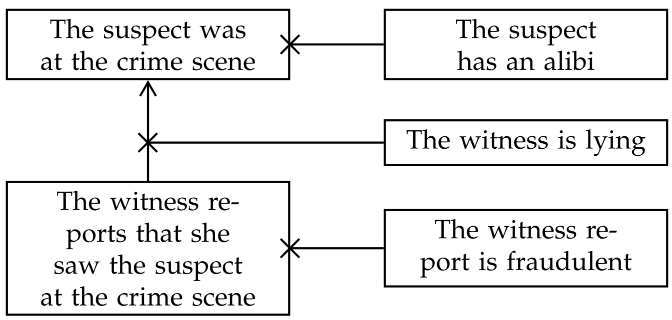

Figure 4: Three kinds of attack

was not at the crime scene. Such an attacking reason that supports an opposite conclusion is called a rebutting attack. Second, the connection between the reason and the conclusion can be attacked. The lying of the witness 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, when the witness is lying, there is no reason supporting that the suspect was not at the crime scene. The attack statement, here the lying of the witness, is also referred to as an exclusionary reason. Third, the reason itself can be attacked. For instance, when the witness report is fraudulent, it can be supported that the witness did not report that she saw the suspect at the crime scene. This kind of attack is referred to as undermining attack. The three examples of the different kinds of attack are shown in Figure 4, on the right.

Further readings Argument structure and diagrams (Freeman, 1991; Toulmin, 1958; Wigmore, 1913, 1931). 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;

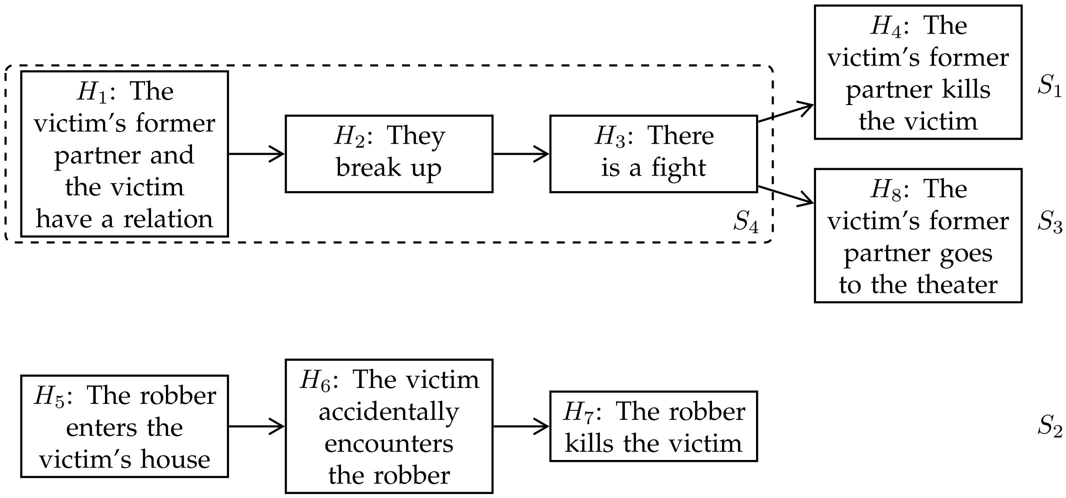

Figure 5: Alternative scenarios and their structure

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 scenarios normative framework, the handling of conflicting evidence is analyzed by considering different scenarios about what may have happened.

Scenarios are clusters of events, ordered in time and connected by causal relations. Consider again the murder example with two suspects: the victim's former partner, who killed the victim after a fight (S_1), and a robber who killed the victim when caught during a robbery (S_2) (used in Section 2.3). Scenario S_1 can be made explicit as a sequence of four hypothetical consecutive events (Figure 5). First, the victim and his former partner have a relation (H_1); then they break up (H_2); subsequently, there is a fight (H_3); and finally, the victim is killed by his former partner (H_4). Scenario S_2 can be analyzed as consisting of three events. First, the robber enters the victim's house (H_5); then the victim accidentally encounters the robber (H_6); and finally, the robber kills the victim (H_7). Some of the events in these chronologically ordered scenarios are causally connected. For instance, in the first scenario, the killing is caused by the break up and the fight; and in the second scenario by the accidental encounter.

Scenarios can explain a piece of evidence or be contradicted by it. The starting point of the murder investigation is the victim's body found at the crime scene (evidence E_0). There are two alternative scenarios that explain this evidence: Both the former partner murder scenario S_1 and the robbery murder scenario S_2 explain evidence E_0, in the sense that finding the body is expected assuming the scenario to be true. Assume now that a match is found between the DNA profile of a tissue trace found under the victim's finger nails and that of his former partner (evidence E_1). This evidence is explained by scenario S_1, but not by S_2, of which evidence E_1 is independent.

When the former partner is confronted with the DNA evidence, she provides a different scenario that explains evidence E_1: Indeed there was a fight about the break up, but she did not kill the victim, and instead subsequently went to the theater (scenario S_3; Figure 5). So we have two scenarios explaining the DNA profile match E_1: the former partner murder scenario S_1 and the former partner alibi scenario S_3. In fact, the two scenarios share a subscenario about the break up fight, that explains the finding of the match (scenario S_4 in Figure 5).

When checking the alibi scenario, it is found that the victim's former partner's bank card was used at the theater that night (evidence E_2), contradicting the former partner murder scenario S_1. Scenario S_3 explains both the DNA evidence E_1 and the use of the bank card E_2.

Scenarios considered may or may not solve a case, and show which evidence is legally relevant. A criminal case is only solved when the legally relevant circumstances can be proven. For instance, in murder cases, it should be proven who killed the victim and why. Of the scenarios considered, the former partner murder scenario S_1 and the robbery murder scenario S_2 explain how and why the murder happened, but the alibi scenario S_3 does not. In a crime investigation, it can be hard to find evidence that proves a sufficiently detailed scenario about what has happened. For instance, in the example, it is initially clear that a murder had happened, but not by who. Because of the break up, a scenario that answers the why-question is considered: the former partner murder scenario S_1. That scenario is initially corroborated by the DNA match E_1, but then breaks down by the use of the bank card E_2 that proves the alibi scenario S_3. Only when the robber is caught and confesses having killed the victim (evidence E_3)—possibly much later, it becomes clear what has happened. The body found E_0 and the confession E_3 are relevant for answering the legally relevant questions who killed the victim and why. The other evidence considered, the DNA match E_1 and the use of the bank card E_2, have played a relevant role in the investigation, but do not support or contradict robbery murder scenario S_2.

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 probabilistic normative framework, the handling of conflicting evidence is analyzed in terms of the conditional probabilities that connect the evidence and the hypotheses.

Evidential support and attack can be characterized as "probability change". A piece of evidence E supports an hypothesis H whenever E raises the probability of H, or in symbols, $P(H|E) > P(H)$. 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:

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

By contrast, a piece of evidence E attacks a hypothesis H whenever E lowers the probability of H, or in symbols, $P(H|E) < P(H)$. 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,

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

Evidential support and attack can be characterized as "likelihood ratio". There is another characterization of evidential support and attack. Instead of comparing the initial probability $P(H)$ and the probability $P(H|E)$ of the hypothesis given the evidence, a so-called likelihood ratio of the form $P(E|H)/P(E|\neg H)$ can also be used. On this account, E supports H whenever the likelihood ratio $\frac{P(E|H)}{P(E|\neg H)}$ 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. By contrast, 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 two examples considered earlier, we have:

$$\frac{P(\mathit{testimony}|\mathit{guilt})}{P(\mathit{testimony}|\neg\mathit{guilt})} > 1, \text{ and}$$

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

These two characterizations of evidential support/attack—in terms of probability increase/decrease, and in terms of a likelihood ratio greater/lower than one—are in fact equivalent. For the following statements hold:

$$P(H|E) > P(H) \text{ iff } \frac{P(E|H)}{P(E|\neg H)} > 1.$$[8]

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

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 (Carnap, 1950; Crupi, 2015; Fitelson, 1999).

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.

[8] To see why, recall that

$$\frac{P(H|E)}{P(\neg H|E)} = \frac{P(E|H)}{P(E|\neg H)} \times \frac{P(H)}{P(\neg H)},$$

which implies

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

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

4.1 Probability

In the probabilistic framework, evidential value is measured using the numeric calculus of probability theory.

The evidential value of a piece of evidence is measured by probabilistic difference or likelihood ratio ('incremental evidential value'). 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, $P(H|E) - P(H)$. The larger the positive difference, the higher the value of the evidence for the hypothesis. An alternative approach is given by the likelihood ratio $P(E|H)/P(E|\neg H)$. 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 $P(H|E) - P(H)$ 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.

The evidential value of all the evidence, in its totality, is measured by the overall conditional probability ('overall evidential value'). 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 $P(H|E)$, the higher the overall value of the evidence for, or against, the hypothesis. If there is a series of pieces of evidence $E_1, \ldots, E_n$, the overall evidential value of the evidence is measured as $P(H|E_1, \ldots, E_n)$.

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 $P(H|E_1)$ of the evidence is high. If this is the only evidence, then also the incremental evidential value is high: before the evidence the hypothesis is not strongly supported, i.e., $P(H)$ is low, whereas after the evidence, the hypothesis is strongly supported, i.e., $P(H|E_1)$ is high. In this case, the overall and incremental evidential value of E_1 align. Now suppose that 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 $P(H|E_1, E_2)$ of the two pieces of evidence together. This will not have changed much when compared to $P(H|E_1)$. As a result, the incremental evidential value of E_2 is low, while still the overall evidential value $P(H|E_1, E_2)$ 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 $P(\neg H)$ is high and also $P(\neg H|E_2)$ is high. Uncritically interpreted, the high value of $P(\neg H|E_2)$ 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.

Likelihood ratio can quantify the value of DNA evidence. As an illustration, we discuss the likelihood ratio of a DNA match in favor of guilt. When introduced in court, a DNA match comes with an estimated Random Match Probability (RMP). As explained earlier, this is the probability that a random person, who had nothing to do with the crime, would match. The RMP can also be viewed as the estimated frequency of the DNA profile across a sample population. Now, with some simplifications (on these later), the evidential value of the DNA match in favor of guilt, in terms of a likelihood ratio, is as follows

$$\frac{P(M|G)}{P(M|\neg G)} = \frac{1}{RMP}.$$

The numerator $P(M|G)$ equal 1 because if the defendant is guilty, the lab test will report a match. As for the denominator, putting $P(M|\neg G) = RMP$ 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

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

Since the likelihood ratio in question is a high number, the DNA match in favor of guilt has a high evidential value. More generally, the lower the RMP, the higher the evidential value of the match. This is because the lower the RMP, the less likely a random person would match.

The likelihood ratio might not capture every source of uncertainty. It is, however, important to keep track of the simplifying assumptions that were made. To see what they were, note that the hypothesis of guilt is not equivalent to other hypotheses, such as the following:

- the *lab reports* that the defendant's genetic profile matches with the crime traces;

- the defendant's genetic profile *truly matches* with the crime traces;
- the defendant is the *source* of the traces; and
- the defendant is *guilty*.

The inference from 'reported match' to 'guilt', passing through the intermediate steps 'true match' and 'source', is a complex one, and many sources of error may undermine the inference along the way. Above we have made three simplifying assumptions.

First, we assumed that the inference from 'reported match' to 'true match' was unobjectionable, or in other words, we assumed the DNA test reporting a match (or a non-match) was infallible. This, of course, need be the case because laboratories make mistakes. A more sophisticated probabilistic analysis will show that a RMP as low as 1/1 billion—-which in our simplistic analysis would give rise to a likelihood ratio as high as 1 billion—reduces to about 100 likelihood ratio if the laboratory error rate is 1%.

Second, consider the inference from 'true match' to 'source'. There are many ways this inference can go wrong. Even if one is a true match, one need not be the source. A genetically identical twin or someone else who just so happens to have an identical DNA could be the source. It could also be that synthetic genetic material, perfectly matching one's DNA, was planted. In this background, the second simplifying assumption we made was that the inference from 'true match' to 'source' can be undermined by one source of error only, namely the possibility that another random individual could coincidentally match. This possibility of error is captured by the Random Match Probability. Other sources of error, for example, a twin brother or an artificially synthesized matching DNA, were disregard.

The third assumption we made was that whoever is the source of the crime traces must be the perpetrator, so 'guilt' and 'source' were treated as equivalent. This, of course, need not be the case. One could be the source of the traces without participating in the crime because, for example, one visited the crime scene before or after the crime took place. Relaxing the identification of the two hypotheses further weakens the value of a DNA match in favor of guilt.

From a probabilistic point of view, laboratory errors can be quantified if the lab error rates are available, but other sources of error are more difficult to quantify. How often does it happen that synthesized matching DNA is implanted? Or what is the probability of error in the inference from 'source' to 'guilt'? The moral is that, in weighting the value of DNA match with a high likelihood ratio, one should always be wary of the sources of error that were, and were not, taken into account in the calculations. Even if the RMP is very low, this does not mean that the evidential value of the DNA match in favor of guilt will be very high.

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 Shumann, 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, 194; Schum and Starace, 2001).

4.2 Arguments

The evidential value of arguments can be analyzed in terms of the reasons they are built from.

The reasons used in arguments have different evidential value. Some are conclusive, others defeasible. A reason is conclusive when, given the reason, its conclusion is guaranteed. The main type of conclusive reasons correspond to deductive, logically valid reasoning. An example of a conclusive reason occurs in the logically valid argument from the reason 'The witness saw the suspect commit the crime and the suspect denies having been at the crime scene' to the conclusion 'The suspect denies having been at the crime scene'. Its logical validity is connected to the underlying logical structure of the argument: From A AND 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, as—for instance—the witness can have been lying or have made a mistake. A defeasible reason can provide prima facie justification for a conclusion, that is withdrawn in light of further information.

Some other terms are used in connection with the difference between conclusive and defeasible arguments. For instance, there is the triplet of deductive, abductive, and inductive arguments. Consider the rule 'If someone is shot, he dies'. A deductive argument involving this rule applies the rule to an instance of its antecedent: From 'John is shot' and 'If someone is shot, he dies', conclude 'John dies'. An abductive argument using this rule uses an instance of the rule's consequent as a starting point, to infer the antecedent: From 'John dies' and 'If someone is shot, he dies', conclude 'John is shot'. Abductive arguments can be thought of as providing an explanation. Abductive arguments are typically defeasible, since there often are alternative explanations. Inductive arguments generalize from an instance of the rule's antecedent and consequent (or several such instances) to the rule: From 'John is shot' and 'John dies', conclude: 'If someone is shot, he dies'. Inductive arguments are also typically defeasible, as the inferred rule often does not hold, at least not in full generality. Deductive arguments are also contrasted with ampliative arguments. In that distinction, deductive arguments only lead to conclusions that were already implicit in their premises, whereas ampliative arguments go beyond their premises. Arguing from A AND B to B is deductive in this sense, and from A to B ampliative.

An argument can be tested by asking critical questions and the evaluation of the argument depends on the answers given. 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 that can be asked about the argument include 'Are there reasons to doubt the suspect was not at the crime scene, such as an alibi?', 'Are there reasons to doubt that the witness testimony supports the conclusion that the suspect was at the crime scene, for instance is the witness lying?' and 'Are there reasons to doubt 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 4, page 11). 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 it no being believable.

It can be subject to debate whether a reason supports or attacks a conclusion. The supporting or attacking relation between a reason and its conclusion can be 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 6, these situations are illustrated by two opposite witness testimonies.

Whether a reason supports a conclusion is associated with 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) is connected to 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.

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

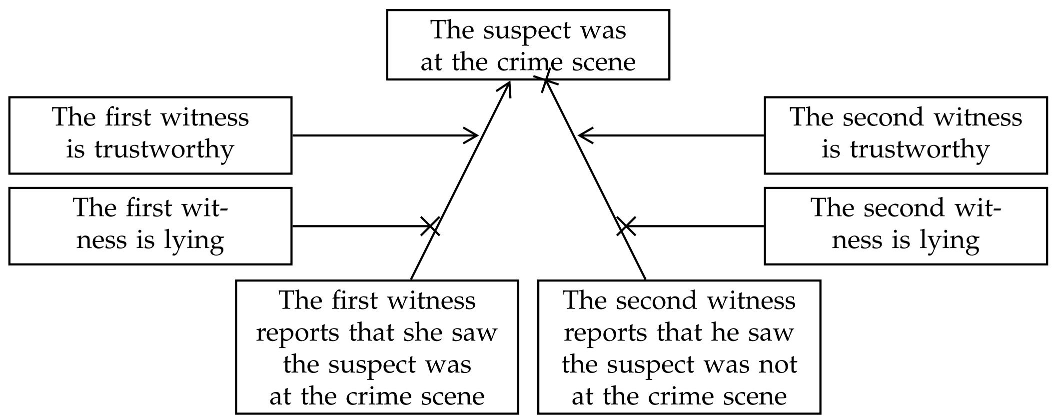

Figure 6: Arguments about whether a reason is supporting or attacking

1958). Argument schemes and critical questions (Walton et al., 2008).

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: a scenario's consistency with the evidence; its power to explain the evidence; and its evidential support. Let us consider each in turn.

Scenarios must be consistent with the evidence, taken at face value. We can evaluate a scenario by checking whether it is consistent with the evidence presented in a case. 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. More precisely, 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. A *prima facie* inconsistency between the evidence and a proposed scenario need not be damning for the scenario's credibility, but at least, it calls for further investigating. It might, of course, turn out that the witness was untruthful or simply confused about the timing. If so, the evidence will be discarded, not the scenario. But, there are cases in which the evidence cannot be easily discarded. Suppose a video recording shows the defendant went into a bank and committed a robbery, and instead, the scenario proposed by the defense denies precisely that. The defense will be hard pressed to justify the proposed scenario given the inconsistency with the evidence.

Scenarios must explain the evidence. The fact that a scenario is consistent with the evidence, is not by itself a good enough reason to believe the scenario. Scenarios should also *explain* the evidence. Two senses of 'explanation' are relevant here. First, a scenario explains the evidence in the sense that it *predicts* the evidence. This can expressed by the statement *if* S, *then* e, where S is a scenario and e the evidence, in the sense that e follows deductively from S. The relation of prediction can also be probabilistic, in the sense that given the truth of the scenario, the truth of the evidence is highly probable. For example, let V be a scenario that comprises the event that the defendant visited the crime scene, and let M be evidence that certain genetic material matching the defendant was found at the crime scene. Clear V predicts M, in the sense that from the truth of V the match M follows almost deductively or at least with a high probability.

There is another sense of explanation that deserves mentioning here. A scenario explains the evidence in the sense it exhibits the *causal process* by which the evidence was brought about. This applies most naturally to physical and testimonial evidence. In a case in which traces were found at the crime scene, it is more plausible that an individual visited the crime scene and left the traces by physical contact. Here the scenario exhibits the causal process (physical contact) by which the existence of the evidence (traces) was brought about.

Scenarios can be evaluated against two directions of fit. Scenarios can be evaluated according to two directions of it: whether they adequately explain the evidence, so that the fit is from the scenario to the evidence; and whether the evidence supports the scenario, so that the fit is from the evidence to the scenario. We have already discussed the first direction of fit. The other direction—from the evidence to the scenario—is equally important in evaluating the credibility of a scenario. In the language of probability, besides the conditional probability $P(e|S)$ of the evidence given the scenario (which is closer to explanatory power), we should also consider the conditional probability of the scenario given the evidence, $P(S|e)$ (which is closer to evidential support).

The two directions of fit are not unrelated. Setting aside wildly implausible scenarios such as 'God did it', if a scenario explains or predicts the evidence, it will also be well supported by the evidence. Conversely, if a scenario is supported by the evidence, it will also explain it. This is not always true but often so. Suppose a witness asserts that she saw the defendant around the crime scene during the time of the crime. This evidence supports a scenario that contains the event 'the defendant was around the crime scene at the time of the crime'. Conversely, this scenario would also explain why the witness testified the way she did, that is, it was because the witness saw what the defendant was doing. However, suppose that the defendant was alone with the victim on a island during a time when it was impossible for anyone else to reach the island. During this time the victim dies of causes that cannot be natural. The fact that the defendant was alone on the island with the victim is evidence that the defendant killed the victim, but this scenario does not explain the evidence. It does not explain, in terms of prediction or causality, why or how the defendant found himself alone with the victim on the island.

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).

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

The more evidence a scenario can explain, the better. When a case comprises several items of evidence, a scenarios's credibility depends on how it can account for the whole mass of evidence. The more items of evidence a scenario can account for, preferably from both the prosecutor and the defense, the better the scenario. This depends on the scenario's consistency with the evidence, explanatory power and evidential support. These criteria were already discussed earlier. The difference now is that consistency, explanatory power and evidential support should be measured relative to the totality of the evidence.

Suppose two items of evidence must be explained: first, the presence of fingerprints at the crime scene; and second the fact that those fingerprints match with the defendant. Suppose scenario S_1 comprises the event that the defendant visited the crime scene, while S_2 compromises the event that some other individual visited the crime scene. Both scenarios explain the presence of fingerprint traces, but only S_1 can explain the fact that the traces match with the defendant. Now, a third scenario S_3 in which another individual visited the crime and implanted fingerprints that match the defendant can explain both items of evidence. As far as explanatory power goes, scenarios S_1 and S_3 are equal, although further evidence may distinguish them. For example, if a witness testified that she saw the defendant walk towards the location of the crime when the crime was committed, scenario S_3 could not easily explain the testimony, while S_1 could. So, absent other evidence, scenario S_1 explains more evidence than the other competing scenarios.

Other more openly holistic criteria such as a scenario's plausibility and completeness are also important. We discuss them now.

Scenarios can be more or less plausible, logically consistent, cohesive, normal While consistency with the evidence, explanatory power and evidential support measure how well a scenario matches up with the evidence, *plausibility* measures how well a scenario matches up with our background assumptions and knowledge of the world. The events that constitute a scenario must be linked by relations of temporal order and causality. These relations, at the very least, 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. 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

The plausibility of a scenario is weakened when the components of the scenario, its sub-scenarios, cannot be easily combined together. This could also be understood as a lack of *cohesiveness* among the components of a scenario. Suppose the defendant is said to have a lovely relation with the victim the day before, while the next day the defendant is depicted as killing the victim out of anger. The scenario lacks cohesiveness because it cannot explain why anger followed a peaceful interaction the day before. So, not only the scenario should explain the evidence, but certain components of a scenario should explain others. Typically we expect that events prior in time explain events later in time.

Plausibility should not be reduced to *normality*. Plausibility has certainly something to do with what happens most of the time, but criminal cases are often about odd coincidences, unexpected and improbable occurrences. Indeed, a scenario according to which the defendant covered 500 mile of distance by car is more normal than a scenario in which the defendant covered the same distance by foot. But the former need not be more plausible than the latter. Consistency with the laws of nature and common sense, whether the parts of the scenario "hang together" nicely (cohesiveness and internal consistency) are also indicators of plausibility.

Scenarios can be more or less complete. Another criterion to evaluate scenarios is their *completeness*. Since scenarios are discursive arrangement 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.

In a murder case, the identity of the perpetrator, the motive, the modus operandi of the crime and the weapon are relevant items which a scenario should specify. The lack of any of these from a proposed scenario would make it incomplete. The law itself sometimes requires that specific items be proven, for example, in criminal cases both the *mens rea* and *actus reus* must be proven. Still, the law does not say whether the defendant's whereabouts the day before or a month before the murder are relevant. They could be relevant to establish motive or they might not. This depends on the circumstances of the case. 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. The notion of completeness, then, overlaps with that of plausibility and cohesiveness of a scenario.

All in all, evaluating a scenario required different levels of analysis: consistency with the evidence; explanatory power (predictive power and causal fit); evidential support; plausibility (and also, cohesiveness, internal consistency and normality); completeness.

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). The story model (Pennington and Hastie, 1993b). Scenarios as scripts (Wagenaar et al., 1993). Scenarios in legal cases (Griffin, 2013). Worries about scenarios in law (Velleman, 2003).

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, 1931). Figure 7 provides an example of a Wigmore diagram (1931). Here Wigmore has analyzed a murder case (Commonwealth v. Umilian, 1901). Jedrusik, the victim, was the author of a letter in which he falsely advised a priest that Umilian had a wife and children (back in the country from which they emigrated), conflicting with Umilians intention to marry. In the chart, Z stands for the charge that Umilian killed Jedrusik. The node at 8 represents a revengeful murderous emotion toward Jedrusik. At 18, it is represented that the marriage is in the end performed, reducing his feelings of revenge. This claim is supported by the (unspecified) evidence at 18.1. 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

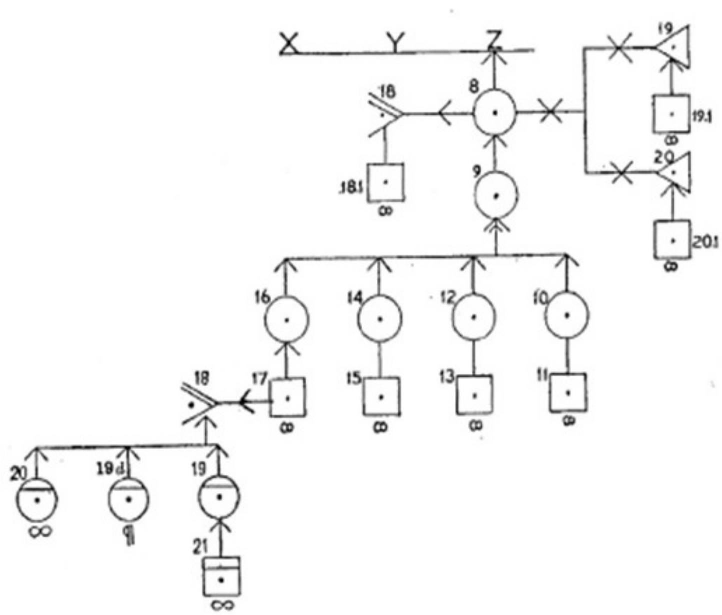

Figure 7: A Wigmore chart

don't. We already discussed how arguments consisting of conflicts of reasons can be evaluated in this sense, when we discussed the role of exceptions, preferences and weighing (Section 4.2). More generally, the structure of a complex of arguments influences the evaluation of the arguments. In particular, the subarguments of a larger argument determine the evaluation of the whole. For an example, we go back to the argument in Figure 2 (page 10). There the witness report 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 become 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 conflicts.

In the first kind of conflicts between reasons, there is a reason attacked by an exclusionary reason, i.e., an attack of the undercutting kind that goes against the connection between the reason and its conclusion (cf. Section 3.1). This situation is shown at the top of Figure 8: There is a witness reporting that the suspect was at the crime scene, but there is evidence that the witness is lying. In that case, the conclusion that the suspect was at the crime scene does not follow, since there is no supporting reason for it. The reason (the witness report itself) and the exception (the lying of the witness) both hold as they are assumptions. In this situation of a reason with an undercutting attack by an exclusionary reason, the conflict of reasons is resolved by the exception expressed in the exclusionary reason.

In the second kind of conflicts between reasons, there is one reason for a conclusion and another against the conclusion. For instance, there are two witnesses with opposite reports about whether the suspect was at the crime scene or not (see the middle of Figure 8). This is an example of a rebutting attack (Section 3.1). The conflict of reasons is unresolved given only the reasons. Further information about the reasons is required to resolve the conflict. The resolution of such a conflict of reasons can be thought of as the preference of one over the other. A reason can be preferred over another, for instance when it is stronger. In the example, the preference (in the figure indicated by the $<$-sign) is argued for by the claim that one of the witnesses is more reliable than the other. In this case, the conclusion follows that the suspect was not at the crime scene, given this conflict of reasons and its resolution.

The third kind of conflicts between reasons discussed here involves more than two reasons. For instance, there are more than two witnesses, with conflicting reports (Figure 8, bottom). In the figure, both sides are supported by two witness reports. Resolving such a conflict can be thought of as a weighing of the reasons, generalizing the preference between

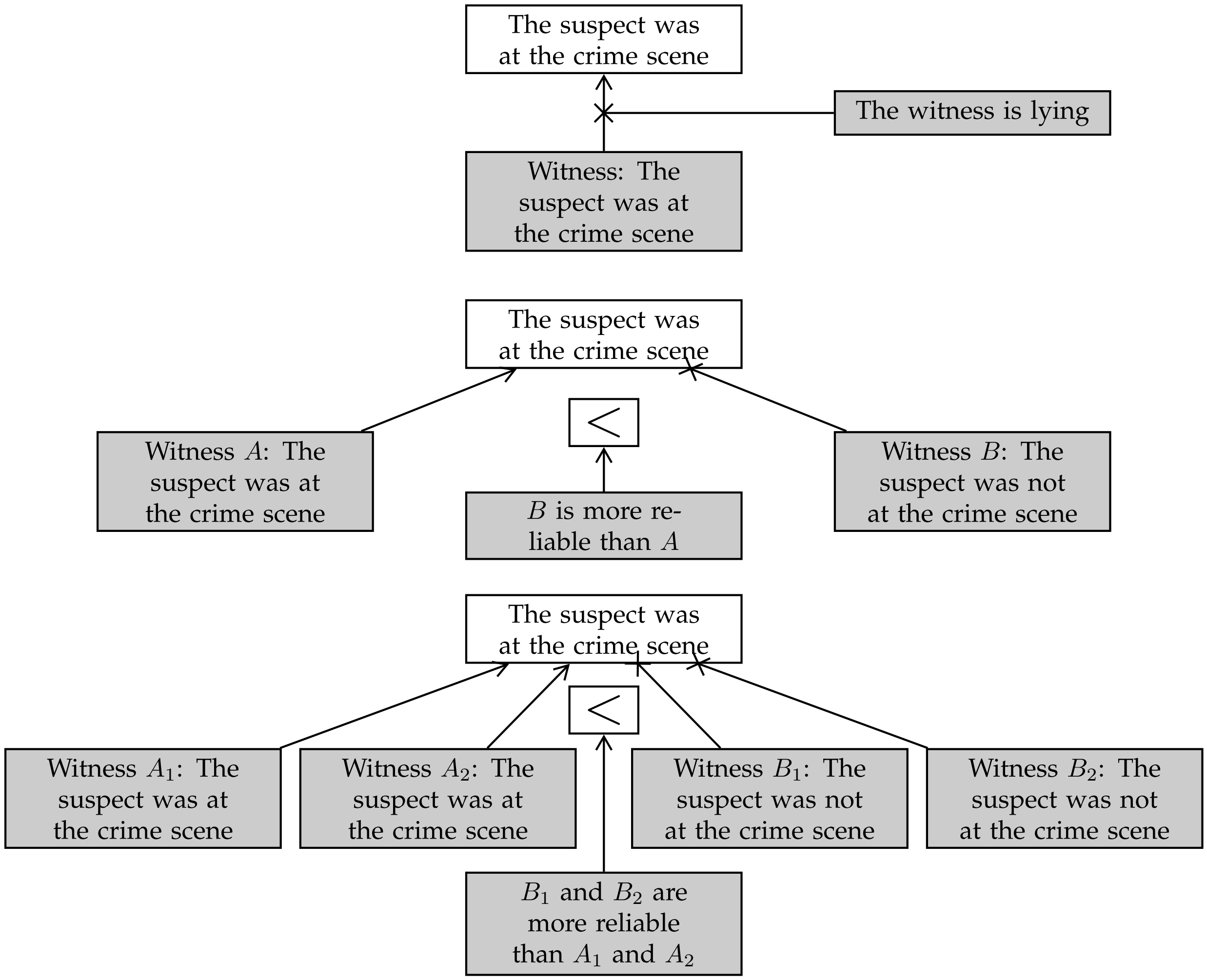

Figure 8: Three kinds of conflicts and their resolution

two conflicting reasons. Again, it is concluded that the suspect was not at the crime scene, given this resolution of this conflict of reasons.

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).

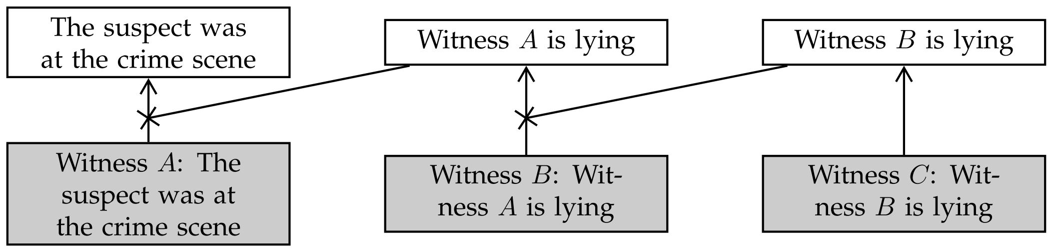

Figure 9: Reinstatement

5.3 Probability

Independent items of evidence can be combined by multiplying the likelihood ratios. In a criminal case, there may more than one piece of evidence, for example, there may be an eyewitness testimony and a DNA match. Suppose $P(guilt|testimony) = 0.7$ and $P(guilt|match) = 0.7$. What is the probability $P(guilt|testimony \wedge match)$ resulting from combing the two pieces of evidence? It would be a mistake to say that, assuming that the two items of evidence are independent, $P(guilt|match \wedge testimony)$ equals $0.7 \times 0.7 = 0.49$. The combination of two pieces of evidence, each favoring guilt to some extent, should strengthen the case for guilt, not weaken it. Observe that

$$\frac{P(guilt|match \wedge testimony)}{P(\neg guilt|match \wedge testimony)} = \frac{P(match \wedge testimony|guilt)}{P(match \wedge testimony|\neg guilt)} \times \frac{P(guilt)}{P(\neg guilt)},$$

and if *match* and *testimony* are independent, conditional on *guilt*, then

$$\frac{P(match \wedge testimony|guilt)}{P(match \wedge testimony|\neg guilt)} = \frac{P(match|guilt)}{P(match|\neg guilt)} \times \frac{P(testimony|guilt)}{P(testimony|\neg guilt)}.$$

The point is that the likelihood ratios should be multiplied, not the probabilities of each hypothesis given the evidence. Suppose DNA evidence shows that the crime traces match with defendant, and the match has a likelihood ratio $\frac{P(match|guilt)}{P(match|\neg guilt)}$ of 36. Suppose the witness testimony favors the hypothesis of guilt and again has a likelihood ratio $\frac{P(testimony|guilt)}{P(testimony|\neg guilt)}$ of 36. These numbers are purely illustrative. The combined evidential value of the two pieces of evidence is $36 \times 36 = 1296$. This is a higher value than the two pieces of evidence considered separately, so $P(guilt|match \wedge testimony)$ will be greater than $P(guilt|match)$ or $P(guilt|testimony)$ considered separately, as expected.

Multiplying the likelihood ratios increases the evidential value if the likelihood ratios are greater than one, and decreases the evidential value if the likelihood ratio are lower than one. If one likelihood ratio is greater and the other lower than one, that is, one item of evidence favors the hypothesis and the other disfavors it, their combined evidential value will vary. To illustrate, consider a case in which there are two conflicting testimonies. One witness asserts that the defendant was around the scene of the crime when the crime was committed. This incriminating testimony favors the hypothesis of guilt, for example,

$$\frac{P(\textit{incriminating witness}|\textit{guilt})}{P(\textit{incriminating witness}|\neg\textit{guilt})} = \frac{0.9}{0.1} = 9,$$

where the numbers are purely illustrative. The other witness offers an alibi for the defendant and asserts that she was with the defendant during the time of the crime. The alibi disfavors the hypothesis of guilt, so that

$$\frac{P(\textit{alibi}|\textit{guilt})}{P(\textit{alibi}|\neg\textit{guilt})} = \frac{0.1}{0.9} = 1/9,$$

where the numbers are, once again, purely illustrative. Absent further information about the trustworthiness of the testimonies, they should cancel one another. The combined evidential value of the two testimonies is null, as expected, for

$$\frac{P(\textit{alibi}|\textit{guilt})}{P(\textit{alibi}\neg\textit{guilt})} \times \frac{P(\textit{incriminating witness}|\textit{guilt})}{P(\textit{incriminating witness}|\neg\textit{guilt})} = 9 \times \frac{1}{9} = 1.$$

The multiplication of the likelihood ratios as a procedure to combine items of evidence can be used for more than two items of evidence. Assuming independence, the general formula is as follows:

$$LR_1 \times \cdots \times LR_k,$$

where $LR_i = \frac{P(E_i|H)}{P(E_i|\neg H)}$, for $i \in \{1, \dots, k\}$, and H is the hypothesis of interest.

Items of evidence can be combined even if they are not independent. So far we have assumed that the items of evidence to be combined are independent, conditional on guilt or on whatever hypothesis of interest. If this assumption is dispensed with, the combined likelihood ratio is no longer the simple multiplication of the individual likelihood ratios, but rather, it becomes:

$$\frac{P(E_1 \wedge E_2|H)}{P(E_1 \wedge E_2|\neg H)} = \frac{P(E_1|H)}{P(E_1|\neg H)} \times \frac{P(E_2|E_1 \wedge H)}{P(E_2|E_1 \wedge \neg H)},$$

and for more than two items of non independent evidence, the formula becomes:

$$\frac{P(E_1 \wedge \cdots \wedge E_k|H)}{P(E_1 \wedge \cdots \wedge E_k|\neg H)} = \frac{P(E_1|H)}{P(E_1|\neg H)} \times \cdots \times \frac{P(E_k|E_1 \wedge \cdots \wedge E_{k-1} \wedge H)}{P(E_k|E_1 \wedge \cdots \wedge E_{k-1} \wedge \neg H)}.$$

Different items of evidence for different hypotheses can also be combined. Another assumption that can be dispensed with is that the items of evidence favor or disfavor the *same* hypothesis. Suppose a DNA match links the defendant to the crime scene, while an email written by the defendant indicates that he had a plan to kill the victim. Each of these item of evidence support a different hypothesis, abbreviated *contact* and (*plan*. Suppose now $P(contact|match) = 0.8$ and $P(plan|\neg email) = 0.8$. It is tempting to say that if the two hypothesis are independent, by the product rule, the probability of both hypotheses taken as a conjunction will be $0.8 \times 0.8 = 0.64$, which is lower than the probability of each hypothesis considered separately. In general, by increasing the number of conjuncts, the probability of their conjunction can be made arbitrarily low. Paradoxically, it would seem that even if each conjunct has a high probability on the available evidence, their conjunction may have a very low probability given the evidence combined. This is known as the conjunction paradox.

To deflect the paradox, observe that:

$$\frac{P(contact \wedge plan|match \wedge email)}{P(\neg(contact \wedge plan)|match \wedge email)} = \frac{P(match \wedge email|contact \wedge plan)}{P(match \wedge email|\neg(contact \wedge plan))} \times \frac{P(contact \wedge plan)}{P(\neg(contact \wedge plan))}.$$

Assuming that the two items of evidence are independent, conditional on the hypothesis of interest, and assuming that the two hypotheses are also independent, we have:

$$\begin{aligned}\frac{P(match \wedge email|contact \wedge plan)}{P(match \wedge email|\neg(contact \wedge plan))} &= \frac{P(match|contact \wedge plan)}{P(match|\neg(contact \wedge plan))} \times \frac{P(email|contact \wedge plan)}{P(email|\neg(contact \wedge plan))} \\ &= \frac{P(match|contact)}{P(match|\neg contact)} \times \frac{P(email|plan)}{P(email|\neg plan)}\end{aligned}$$

Once again, the likelihood ratios, not the probabilities of each hypothesis, should be multiplied. Suppose, for the sake of argument, the evidential value of the DNA match in favor of the hypothesis *contact*, and the evidential value of the defendant's email in favor of the hypothesis *plan*, are as follows:

$$\frac{P(match|contact)}{P(match|\neg contact)} = 36 \text{ and } \frac{P(email|plan)}{P(email|\neg plan)} = 36$$

where the numbers are, as usual, illustrative. The combined evidential value of *match* and *email* in favor of the overall hypothesis, including *contact* and *plan*, is $36 \times 36 = 1296$, so

$$\frac{P(contact \wedge plan|match \wedge email)}{P(\neg(contact \wedge plan)|match \wedge email)} = 1296 \times \frac{P(contact \wedge plan)}{P(\neg(contact \wedge plan))}$$

For the value of the ratio $\frac{P(contact \wedge plan)}{P(\neg(contact \wedge plan))}$, assume that

$$\frac{P(contact)}{P(\neg contact)} = 0.1/0.9 \text{ and } \frac{P(plan)}{P(\neg plan)} = 0.1/0.9.$$ [9]

Now, putting everything together, we get

$$\begin{aligned}
\frac{P(contact \wedge plan|match \wedge email)}{P(\neg(contact \wedge plan)|match \wedge email)} &= \frac{P(match \wedge email|contact \wedge plan)}{P(match \wedge email|\neg(contact \wedge plan))} \times \frac{P(contact \wedge plan)}{P(\neg(contact \wedge plan))} \\
0.93/0.07 &= 1296 \times \frac{0.1 \times 0.1}{1-(0.1 \times 0.1)}
\end{aligned}$$

So, the $P(contact \wedge plan|match \wedge email)$ turns out to be higher, albeit somewhat only slightly, than $P(contact|match)$ or $P(plan|email)$. By combing different items of evidence and different hypotheses, the probability of the conjunction need not be lower than the probability of each conjunct, *contra* the conjunction of paradox.

Further readings The conjunction paradox (Cohen, 1977) and a response (Dawid, 1987). Coherence and probability (Bovens and Hartmann, 2003). Bayesian networks (Taroni et al., 2006). Probabilistic analysis of an entire legal case (Kadane and Schum, 1996). On the use of probability in law (Fenton, 2011).

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. This is not to be confused with the burden of proof, which includes the burden of persuasion as well as the burden of production. The criterion for criminal cases in common law countries is *proof beyond*

[9] These assumptions are coherent with our earlier ones, so that

$$\begin{aligned}
\frac{P(contact|match)}{P(\neg contact|match)} &= \frac{P(match|contact)}{P(match|\neg contact)} \times \frac{P(contact)}{P(\neg contact)} \\
0.8/0.2 &= 36 \times 0.1/0.9
\end{aligned}$$

and also

$$\begin{aligned}
\frac{P(plan|email)}{P(\neg plan|email)} &= \frac{P(email|plan)}{P(email|\neg plan)} \times \frac{P(plan)}{P(\neg plan)} \\
0.8/0.2 &= 36 \times 0.1/0.9
\end{aligned}$$

a reasonable doubt, and a similar criterion exists outside the common law. If the decision makers are persuaded of the defendant's guilt beyond a reasonable doubt, they should convict, or else they should acquit.

In Commonwealth v. Massachusetts Webster (1850), proof beyond a reasonable doubt is equated to 'reasonable and moral certainty'. In R. v. Lifchus (1997), the Supreme Court of Canada writes that the 'the standard of proof beyond a reasonable doubt is inextricably intertwined with . . . the presumption of innocence', that it is connected with 'the evidence or absence of evidence', and also that 'it does not involve proof to an absolute certainty' and so 'it is not proof beyond any doubt' (335). Explanations abound in the case law. And yet, it is unclear whether they improve our understudying. 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 either, 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

The guilt probability is estimated by weighing the evidence within 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 downward 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 are known because they are based on estimated frequencies, but sometimes they are not. For example, $P(G)$ is required to calculate $P(G|E)$. This is the probability of the defendant's guilt regardless of the evidence presented at trial. What should $P(G)$ be? For technical reasons, it cannot be zero, but it also cannot be 50% because of the presumption of innocent. Arguably, $P(G)$ should be relatively low, but how low? This remains hotly debated.

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,

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

The inequality holds just in case

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

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 $P(G)$ 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 clear 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 $P(G)$ 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. In this spirit, setting a probabilistic criterion for criminal convictions would only be a way to theorize about the meaning and ideal of the criminal standard of proof.

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 evidential basis 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.

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, 2011; 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; Stein, 2005).

6.2 Arguments

Arguments and counterarguments are collected and weighed. In a court of law, the prosecutor or plaintiff puts forward a claim and offers supporting arguments. The opposing parts responds by offering counterarguments. The dialectical process can be complex. There are different counterarguments: undermining, undercutting and rebutting. The process is complex also because it can iterated. An argument can be attacked by a counteargument, and the latter in turn can be attached by a further counterargument. 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 argument can be assessed.

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

Defeating all counterarguments is the criterion for meeting the standard of proof. In order to establish the defendant's guilt beyond a reasonable doubt, all counterarguments to the contrary must be responded, or in other words, all the attacks against the argument for guilt must be defeated. Now, whether a counteargument or an attack is defeated is not always an all or nothing affair. It is often a matter of degrees. If the argument for guilt is slightly stronger than all its counterarguments, this would not be enough yet. To meet the demands of the standard of proof beyond a reasonable doubt, an argument must be significantly stronger than all its counterarguments. On the other hand, defeating all its counterarguments with absolute certainty would be too much to expect. So, more realistically, all counterattacks must be defeated in an almost definitive way. Here, it seems, we need to reintroduce some threshold, even though not in an explicitly probabilistic or numerical way.

It is not clear when to stop collecting arguments and counterarguments. The argumentation framework is rather realistic. The dialectic between arguments and counterarguments effectively takes place in court, and the idea that meeting the standard of proof requires to answer all counterargument against guilt is natural enough. A problem is that if the opposing party puts forward no counterarguments, meeting the standard of proof would be effortless. A possible response here is that the counterarguments must be all the counterarguments which a reasonable objector could in principle put forward, not just the counterarguments that in fact are put forward.

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 Carneades model (Gordon and Walton, 2009; Gordon et al., 2007). Justified argument model (Prakken and Sartor, 2007, 2009).

6.3 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, cohesiveness, 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 scenario 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 one 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), evidential support, plausibility, completeness, etc. Other criteria might also play a role, such the simplicity of the scenario. So, the best explanation is the scenario that fares better on some combination of these criteria. This is a matter of degrees. The scenarios get higher or lower scores relative to the applicable evaluation criteria. The scenario that gets the best score, and that meets a suitable threshold level, should be selected.

It is not clear how to find the scenarios. The process of scenario selection resembles how jurors reason in trial proceedings, whereas despite its clear mathematical underpinnings, it is hard to relate probability to judicial proceedings: jurors do not naturally quantify guilt, and it is difficult to quantify it even if we wanted to. Still, a problem with the scenario approach is that the method by which a scenario is selected is not entirely transparent. 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? One might wonder whether the criterion for the best scenario should simply be this: the most probable scenario which meets a sufficiently high probability threshold. If so, this would not be very different from the probability based account of the standard of proof.

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 summarize our discussions for each theme, using the highlighted phrases in the preceding sections.

Conflicting evidence

Arguments The arguments for and against different positions have structure, involving complexes of reasons supporting and attacking positions. Three kinds of support can be distinguished: multiple, subordinated and coordinated support. Three kinds of attack can be distinguished: rebutting, undercutting, and undermining attack.

Scenarios Scenarios are clusters of events, ordered in time and connected by causal relations. Scenarios can explain a piece of evidence or be contradicted by it. Scenarios considered may or may not solve a case, and show which evidence is legally relevant.

Probabilities Evidential support and attack can be characterized as "probability changes". Evidential support and attack can be characterized as "likelihood ratio". The conflict between two pieces of evidence can be described probabilistically.

Evidential value

Probabilities Evidential value can be quantified as probability difference, likelihood ratio or overall probability. Likelihood ratio can quantify the value of DNA evidence. Likelihood ratio might not capture every source of uncertainty.

Arguments The reasons used in arguments have different evidential value. Some are conclusive, others defeasible. An argument can be tested by asking critical questions and the evaluation of the argument depends on the answers given. It can be subject to debate whether a reason supports or attacks a conclusion.

Scenarios Scenarios must be consistent with the evidence, taken at face value. Scenarios must explain the evidence. Scenarios can be evaluated against two directions of fit.

Coherently interpreting the evidence

Scenarios The more evidence a scenario can explain, the better. Scenarios can be more or less plausible, logically consistent, cohesive, normal. Scenarios can be more or less complete.

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 Independent items of evidence can be combined by multiplying the likelihood ratios. Items of evidence can be combined even if they are not independent. Different items of evidence for different hypotheses can also be combined.

Reasoning and decision making

Probabilities The guilt probability is estimated by weighing the evidence within the probability calculus. The decision criterion is a guilt probability threshold. It is not clear how to estimate all the required probabilities.

Arguments Arguments and counterarguments are collected and weighed. Defeating all counterarguments is the criterion for meeting the standard of proof. It is not clear when to stop collecting arguments and counterarguments.

Scenarios Competing scenarios are collected and compared. The best explanatory scenario is the rule of decision. It is not clear how to find the scenarios.

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 this way, we hope to contribute to the further development of the three frameworks. 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 each other, and how they can be integrated into an overall normative perspective on evidential reasoning.

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, 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. Solving the riddle of coherence. *Mind*, 112:601–633, 2003.

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

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

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

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, 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, 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 (manuscript). 2011.

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

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.

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, 2006.

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.

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.

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, 1997.

H. Prakken. Proceedings of the tenth international conference on artificial intelligence and law. In *A Study of Accrual of Arguments, with Applications to Evidential Reasoning*, pages 85–94. ACM Press, 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.

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.

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. Shumann. Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and the defense attorney's fallacy. *Law and Human Behaviour*, 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. Explaining Bayesian Networks using argumentation. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty - 13th European Conference, ECSQARU 2015, Compigne, France, July 15-17, 2015. Proceedings*, pages 83–92. Springer, Berlin, 2015.

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, 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, 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*, 15:35–70, 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.

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 or the Process of Proof as Given by Logic, Psychology, and General Experience, and Illustrated in Judicial Trials. (Second edition 1931.)*. Little, Brown and Company, Boston (Massachusetts), 1913.

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

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.