EVIDENTIAL REASONING
Chapter for the Handbook of Legal Reasoning

Marcello Di Bello & Bart Verheij – February 23, 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 it has reached as high a rate as 99%.[2] This does not mean that fact-finders deciding about the facts of a criminal 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? What to do with illegally obtained 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. Before diving into the literature, we set the stage by using two important and often encountered kinds of evidence as an illustration: eyewitness testimony and DNA profiling. Similarities and differences between these kinds of evidence are used to establish a list of central questions that structure the exposition that follows.

1 SETTING THE STAGE

Fact-finders—typically jurors and judges—aim to reconstruct what has happened in the crime on the basis of the evidence. We will use two central types of evidence to develop a list of central questions associated with evidential reasoning: eyewitness testimony and DNA analysis.

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 written reports of oral examinations in the pre-court stages of the criminal investigation, normally by prosecuting officers and judges.

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

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?
>
> 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 an FC Groningen t-shirt, which surprised me as we were in the Vondelpark.

On the basis of eyewitness testimony, we can form a hypothesis about what has happened. Sometimes this hypothesis contains specific detail—as in the example—, still it remains a hypothesis. There are many reasons why the hypothetical events reconstructed on the basis of the testimony may 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 profile is determined by analyzing a number of specific locations—the so-called loci—of a DNA molecule, and establish the type of structure found there. These types are called alleles, and typically consist of the number of repetitions of a small DNA structure at a location. For instance, one locus used in the profiles stored in forensic DNA databases in the USA is referred to as CSF1PO, and it can have alleles 5, 6, 7 and then up to 16, 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 USA CODIS system has 13 core loci. As said, each specific DNA profile is rare, and reference databases of profiles are used to numerically measure how rare a profile really is. This is done by counting the number of occurrences of each allele at each core

[3]See `http://www.cstl.nist.gov/strbase/str_CSF1PO.htm`.

locus in the reference database, which gives an estimate of the proportional frequency of that allele at that locus in the population. The measured proportional frequencies for the individual alleles at the core loci are then multiplied to compute what is called the Random Match Probability of the DNA profile.[4] 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, we 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[5] 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 crim-

[4]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.

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

inal has blond hair, but the suspect whose DNA matched that of the trace at the crime scene, has dark hair. What to do in case of such conflicts?

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

Question 3: How should we coherently interpret the available evidence? A DNA 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 collect, include, exclude evidence? During the collection of evidence all kinds of things can happen. A witness' answer to a question can be discarded when the prosecution's question is judged to have lead the witness to an unjustified position. The classic example is the question "When did you start hitting your wife?" before it has been established that the suspect has been hitting his wife in the first place. Also DNA material can have been collected illegally, for instance without the suspect's consent. Which rules exist that guide the collection, marshaling, inclusion and exclusion of the evidence?

Question 5: 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?

In the following sections, each of these questions is addressed. Before that, we discuss three normative tools that can help understand how to correctly handle the evidence.

1.4 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; Kaptein et al., 2009; Dawid et al., 2011). The first framework discussed uses arguments as primary tool, the second scenarios, and the third probabilities.

1.4.1 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

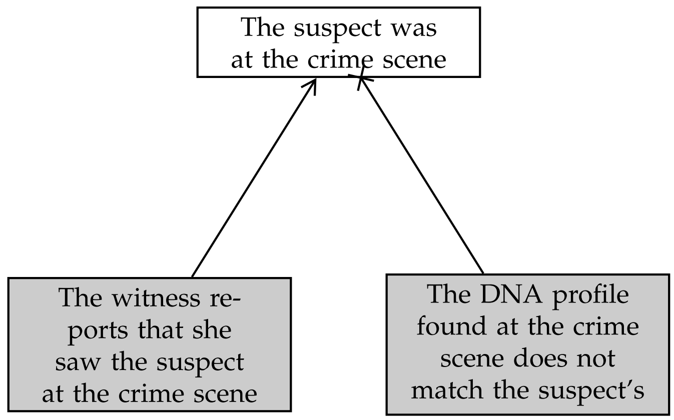

Figure 1: Arguments with supporting and attacking reasons

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). 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., 2014).

1.4.2 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.[6]

1.4.3 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, 1993). 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

[6] 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).

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).

1.5 Paper plan

The three normative frameworks for the handling of evidence, arguments, scenarios, and probabilities, are connected to the first three of the central questions that we have discussed:

Question 1: How should we handle conflicting evidence?

Question 2: How should we handle the strength of the evidence?

Question 3: How should we coherently interpret the available evidence? When are we done?

Although—as we shall see—each of the three normative frameworks provides relevant insights for answering each of these three questions, the first question about conflicting evidence is especially closely related to the arguments framework, the second question about strength of the evidence in particular to the probabilities framework, and the third question about coherently interpreting the evidence most strongly to the scenarios framework.

In the following sections, these three questions will be discussed, consecutively, while emphasising the role of the three normative frameworks (Sections 2, 3 and 4). The remaining two questions are less strongly connected to the normative frameworks, and are discussed in Sections 5 and 5:

Question 4: How should we collect, include, exclude evidence?

Question 5: How should we decide about the facts given the evidence? When are we done?

2 CONFLICTING EVIDENCE

Evidential reasoning in the law is a dialectical process involving reasons pro and con different reconstructions of the facts. In many situations, it is clear what the facts and their legal interpretation are. Consider for instance a routine traffic violation such as speeding. If you are caught driving at 100 km/h, the speed limit is 50 km/h, and a police officer issues you a ticket, there is little to dispute. Yet, cases that are litigated in court are typically more complicated either because the interpretation of the law is disputed or because there are conflicting reconstructions of the facts. (For disputes about matters of law, see OTHER CHAPTER IN HANDBOOK). Conflicting reconstructions of the facts emerge when the two parties in a trial—the defense and the prosecutor in a criminal trial or the plaintiff in a civil trial—introduce evidence that support conflicting conclusions. For example, a witness for the prosecutor may assert she saw the defendant around the crime scene at the time of the crime, while the defense may introduce evidence that the genetic material found at the crime scene does not match the defendant's. When two or more pieces of evidence support contradictory reconstructions of the facts, it can be hard to decide which piece of evidence to trust or which reconstruction to believe.

Legal trials therefore often take the form of adversarial confrontations. Each party is given the opportunity to make its case on the basis of the evidence she thinks important. But, trials are not confined to the mere presentation of the evidence by the interested parties. Since the parties will advance conflicting reconstructions of the facts, the dialectical testing of the evidence is also crucial. Although one party may make a strong case, backed up by good evidence, the other party may come up with a stronger case, backed up by even better evidence. In the law, more often than not, reasoning toward factual conclusions is a dialectical process. The examination and cross examination of the evidence is the legal machinery that is used to identify which party has the stronger case.

The three normative frameworks for handling the evidence provide different perspectives on how to handle conflicting evidence.

2.1 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 rivalling 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.

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 3, on the left). This argument can be attacked in three ways. First, its

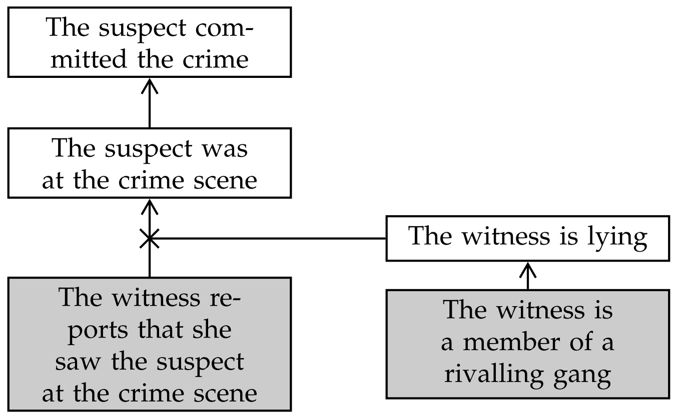

Figure 2: Arguments have structure

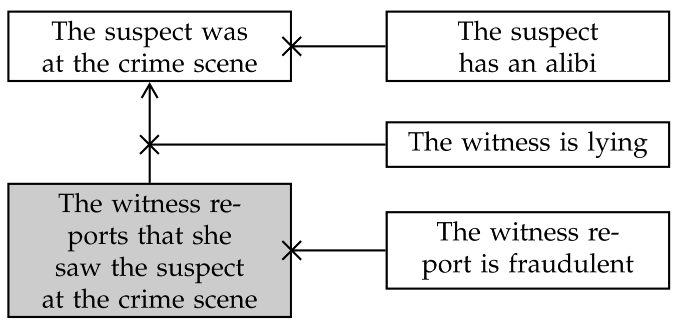

Figure 3: Three kinds of attack

conclusion can be attacked. The suspect can for instance have an alibi, showing that he 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 3, on the right.

In an argumentative dialogue, parties take positions supported by reasons that can be challenged by attacking reasons. Arguments have a dialogical counterpart, in which parties exchange reasons for and against the positions they endorse. The arguments shown in Figure 2 for instance form the backbone of the following argumentative dialogue—here presented as a fictitious, stylized discussion between judge, prosecution and defense:

Prosecution: The suspect committed the crime.

Judge: Why do you believe that?

Prosecution: The suspect was at the crime scene.

Judge: Do you have evidence supporting that position?

Prosecution: There is a witness reporting that she saw the suspect at the crime scene.

Defense: Objection, your honor! The witness is lying.

Judge: Why do you believe that?

Defense: The witness is a member of a rivalling gang.

Models of argumentative dialogue involve specifications of the kinds of moves parties can make, the commitments these moves imply for parties, and the rules that determine the allowed sequences of dialogue moves.

2.2 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 1.4.3). Scenario S_1 can be made explicit as a sequence of four hypothetical consecutive events (Figure 4). 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.

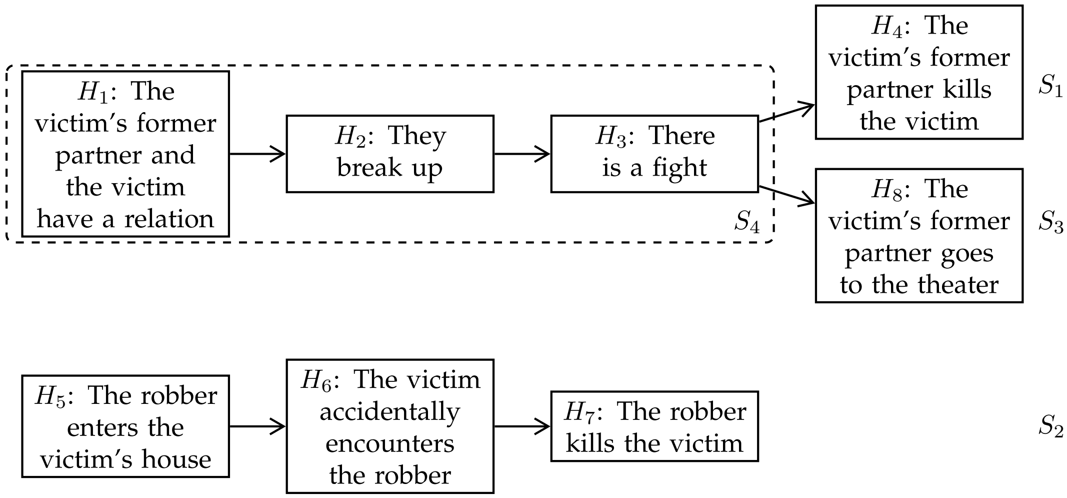

Figure 4: Scenarios and their structure

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 4). 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 4).

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.

2.3 Probabilities

Evidential favoring and disfavoring can be characterized as "probability difference" A piece of evidence E favors 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 favors the hypothesis that the defendant is guilty. This can be described probabilistically, as follows:

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

By contrast, a piece of evidence E disfavors an 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 disfavors the hypothesis that the defendant is guilty. Probabilistically,

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

Evidential favoring and disfavoring can be characterized as "likelihood ratio" There is another characterization of evidential favoring and disfavoring. 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 $\frac{P(E|H)}{P(E|\neg H)}$ can also be used. On this account, E favors 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 likely if the hypothesis is true than if the hypothesis is false. By contrast, a piece of evidence E disfavors an 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. In terms of the two examples considered earlier, we have:

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

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

Interestingly enough, the two characterizations of evidential favoring/disfavoring—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:

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

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

The conflict between two pieces of evidence can be described probabilistically Two pieces of evidence come into conflict with one another insofar as one favors an hypothesis and the other disfavors 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).

[7] 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 $P(H|E) > P(H)$.

2.4 Further reading

Arguments: Wigmore, references Pollock 1995, Dung 1995; ?mention nonmonlog, Toulmin's anti-logicism, FvE/Walton Defeat by seq weakening, par strengthening Accrual Preferences, weighing

3 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.

3.1 Probability

Evidential value can be quantified as probability difference, likelihood ratio or overall probability The value of the evidence for, or against, an 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 $\frac{P(E|H)}{P(E|\neg H)}$. 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* an 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.

Intuitively, the value of a piece of evidence for or against an hypothesis can also be quantified by the probability of the hypothesis given the evidence. The higher, or lower, the probability $P(H|E)$, the higher the value of the evidence for, or against, the hypothesis. The probability $P(H|E)$, however, should not be confused with the probability difference or the likelihood ratio. To illustrate, suppose a witness testifies that the defendant beat the victim to death, but as it turns out, the witness was drunk at the time, so the evidential value of the testimony in favor of guilt is low. If the testimony by the drunk witness is the only incriminating evidence, the probability $P(\textit{guilt}|\textit{drunk witness})$ should also be low. Now, it follows that $P(\textit{innocence}|\textit{drunk witness})$ should be high insofar as *innocence* is the negation of *guilt*. But this seems problematic. A drunk witness, in fact, is of little help in establishing innocence (just as it is of little help in establishing guilt).

It is instructive to quantify the value of the testimony in favor of guilt by means of the probability difference and the likelihood ratio. Since the witness was drunk, the value of the evidence in favor of guilt is low, that is:

$$\text{the positive difference } P(\textit{guilt}|\textit{drunk witness}) - P(\textit{guilt}) \text{ is small, and}$$

$$\text{the likelihood ratio } \frac{P(\textit{guilt}|\textit{drunk witness})}{P(\textit{guilt}|\neg\textit{drunk witness})} \text{ is only slightly above one.}$$

Similarly, the value of the evidence in favor of innocence is also low, that is:

$$\text{the negative difference } P(\textit{innocence}|\textit{drunk witness}) - P(\textit{guilt}) \text{ is small, and}$$

$$\text{the likelihood ratio } \frac{P(\textit{innocence}|\textit{drunk witness})}{P(\textit{innocence}|\neg\textit{drunk witness})} \text{ is only slightly below one.}$$

So, the value of the testimony in favor of guilt, and innocence, is low in both cases. This means that $P(\textit{guilt}|\textit{drunk witness})$ and $P(\textit{guilt})$ are roughly the same value, and so are $P(\textit{innocence}|\textit{drunk witness})$ and $P(\textit{innocence})$. Consequently, if $P(\textit{innocence})$ is high, and thus $P(\textit{guilt})$ low, $P(\textit{innocence}|\textit{drunk witness})$ will be high, and thus $P(\textit{guilt}|\textit{drunk witness})$ low. We should thus not be surprised that $P(\textit{innocence}|\textit{drunk witness})$ is high. This is because, by Bayes' theorem, $P(H|E)$ depends on $P(H)$. Even if E does not change significantly the probability of H, or the likelihood ratio (positive or negative) $\frac{P(E|H)}{P(E|\neg H)}$ is small, $P(H|E)$ could still be high or low, insofar as $P(H)$ itself is high or low. All in all, we should be careful in not confusing a high probability $P(H|E)$ with a high evidential value in terms of a large probability difference or a high likelihood ratio.

Likelihood ratio can quantify the value of DNA evidence The most widely used measure of evidential value is the likelihood ratio. As an illustration, let us determine the likelihood ratio of a DNA match in favor of guilt. First, it is important to realize that the hypothesis of guilt is not equivalent to other propositions, 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 though the intermediate steps 'true match' and 'source', is a complex one, and many sources of error may undermine the inference along the way. To ease exposition, we make three simplifying assumptions. The first is that the inference from 'reported match' to 'true match' is unobjectionable, or in other words, we assume that the DNA test reporting a match (or a non-match) is infallible. This, of course, need be the case because laboratories make mistakes. The second assumption concerns the inference from 'true match' to 'source'. 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. Now, the second simplifying assumption is that the inference from 'true match' to 'source' can be undermined by one source of error only, namely the possibility that another individual could coincidentally match. This possibility is expresses through the so-called Random Match Probability, that is, the probability that a random person would match. Other sources of error—for example, a twin brother or an artificially synthesized DNA could be the source—will be disregard. Finally, the third simplifying assumption is that whoever is the source of the crime traces must be the perpetrator, so 'guilt' and 'source' are 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.

Suppose now that the DNA test reports a match between the crime traces and the defendant. Given our simplifying assumptions, the evidential value of the DNA match in favor of guilt, in terms of a likelihood ratio, is as follows

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

If the RMP is 1 in 200 million, the likelihood ratio would be

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

The numerator $P(M|G)$ equal 1 because of our simplifying assumption that 'source' and guilt' are interchangeable and the laboratory test is infallible. If the defendant is guilty and thus the source of the crime traces, the infallible lab test must report a match. As for the denominator, putting $P(M|\neg G) = RMP$ is plausible because (1) 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, and because (2) 'source' and 'guilt' are, by assumption, equivalent.

Since the likelihood ratio in question is a high number, the DNA match in favor of guilt has a high evidential value. We should not forget, however, our simplifying assumptions. For example, a likelihood ratio as high as 1 billion reduces to about 100 if the laboratory error rate is 1% (**?**). Also, the identification of 'source' and 'guilt' must also be taken into account, and relaxing this identification further weakens the value of a DNA match in favor of guilt.

Further readings Introductions to using probability for weighing evidence (Finkelstein and Fairley, 1970; Dawid, 2002; Mortera and Dawid, 2007). Critique of the probabilistic approach (Tribe, 1971; Cohen, 1977; Allen and Pardo, 2007). Prosecutor's fallacy (Thompson and Shumann, 1987). Introduction to DNA evidence (Wasserman, 2008; Kaye and Sensabaugh, 2000). Different hypotheses for evaluating DNA evidence (Koehler, 1993; Cook et al., 1998; Evett et al., 2000). Probabilistic analyses of DNA evidence (Robertson and Vignaux, 1995; Buckleton, 2005; Balding, 2005). 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 (NRC, 1996; Balding and Donnely, 1996). Comparison between DNA evidence and fingerprints (Zabell, 2005). Probabilistic analyses of eyewitness testimony (Friedman, 1987; Schum and Starace, 2001).

3.2 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 strengths. 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 arguments. 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.

Conclusive and defeasible reasons correspond to one-step conclusive and defeasible arguments. 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 beyong their premises. Arguing from A AND B to B is deductive in this sense, and from A to B ampliative.

The reasons in arguments can be tested using critical questions. Defeasible reasons are characterized by the possibility of circumstances that have the effect that the conclusion of the reason does not follow, given the reason. The occurrence of such defeating circumstances can be guided by asking critical questions. The answers to those critical questions provide insight into the evidential value of the reasons. Consider for instance 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 include 'Are there reasons showing that the suspect was not at the crime scene, such as an alibi?', 'Was the witness lying?' and 'Are there reasons to think that the witness report is fraudulent?'. The first of these questions is directed at the argument's conclusion, the second at the argument step, and the third at the argument's reason. These different kinds of critical questions are connected to the kinds of argument attack discussed in Section 2.1 (see in particular Figure 3).

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 2.1). This situation is shown at the top of Figure 5: There is a witness reporting that the suspect was at teh crime scene, but 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 5). The

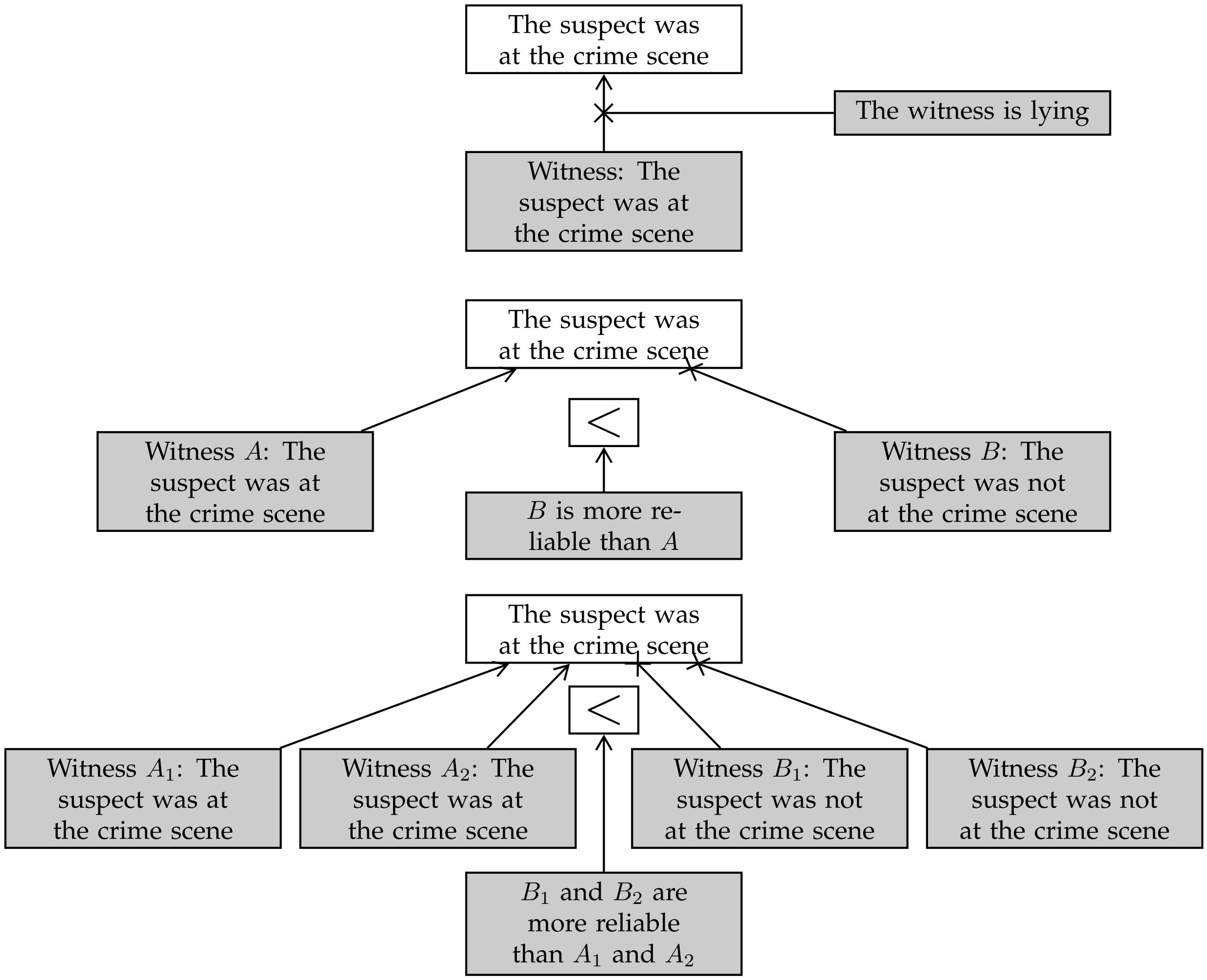

Figure 5: Three kinds of conflicts and their resolution

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.

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 5, 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 two conflicting reasons.

Further readings deductive, inductive, abductive deductive, ampliative

Mention ?Pollock's anti-probabilism

Argument scheme collections

Accrual

Reason-Based Logic

3.3 Scenarios

4 COHERENTLY INTERPRETING THE EVIDENCE

4.1 Scenarios

Good properties of scenarios: completeness, plausibility (here??), coverage (see Pennington and Hastie) causal rules

Scenario schemes (Schank, Bex)

4.2 Arguments

Inference to the best explanation (Allen/Pardo?)

WORRY: Confused about this one; why is it under coherence?) [Answer BV: Otherwise nothing remains here.]

4.3 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 $\frac{P(\mathit{match}|\mathit{guilt})}{P(\mathit{match}|\neg \mathit{guilt})}$ of 36. Suppose the witness testimony favors the hypothesis of guilt and again has a likelihood ratio $\frac{P(\mathit{testimony}|\mathit{guilt})}{P(\mathit{testimony}|\neg \mathit{guilt})}$ 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 $P(\mathit{guilt}|\mathit{match} \wedge \mathit{testimony})$ will be greater than $P(\mathit{guilt}|\mathit{match})$ or $P(\mathit{guilt}|\mathit{testimony})$ 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,

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

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

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

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

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

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 $i \in \{1, \ldots, k\}$, 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 $P(\textit{contact}|\textit{match}) = 0.8$ and $P(\textit{plan}|\neg \textit{email}) = 0.8$. 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 conjuction 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:

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

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:

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

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

$$\frac{P(contact)}{P(\neg contact)} = 0.1/0.9 \text{ and } \frac{P(plan)}{P(\neg plan)} = 0.1/0.9.$$[8]

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 Bayesian networks XXX, The condition paradox LJ Cohen and Dawid response

[8]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

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

5 REASONING AND DECISION MAKING

So far we have focused on how the evidence can be evaluated and combined, and how inferences can be drawn. This does not take place in vacuum. The legal system contains rules for the discovery, admission and exclusion of the evidence. The legal system also contains procedures and guarantees available to defendants. In most countries, for example, criminal defendants enjoy the right to cross examine their accusers and scrutinize the evidence presented against them, and all defendants are presumed innocent until proven guilty. 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. Currently, there is no strict regulation of how the fact finders should reason or reach conclusions on the basis of the evidence. By contrast, the decision criterion is defined by law and consists of a standard of proof, sometimes also called burden of persuasion.[9] The criterion in common law countries is *guilt beyond a reasonable doubt*, and a similar criterion exists in other countries outside the common law. If the fact finders are persuaded of the defendant's guilt beyond a reasonable doubt, they should convict, or else they should acquit. The three frameworks we considered—probability; arguments; and narratives—can be used to theorize how evidential reasoning should be conducted and how the decision criterion should be characterized.

Probability BLA BLA 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 $P(G|E)$ is reached, all things considered. This forms the basis for the decision to convict or acquit.

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,

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

[9] This is not to be confused with the burden of proof, which includes the burden of persuasion as well as the burden of production. REFERENCE.

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.

Arguments BLA BLA 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.

Giovanni Sartor and Henry Prakken develop an argumentation based framework to theorize about evidence and its use at trial. From their framework, we shall extrapolate a few ideas about how to characterize standards of proof, although they themselves never offer such characterizations. 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, as discussed in PREVIOUS SECTION while distinguishing between undermining, undercutting, and rebutting. The process is complex also because it can iterated. An argument can be attacked by a counteragent, and the latter in turn can be attached by a 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. For Sartor and Prakken, an argument is justified if it survives all the attacks by its counterarguments, or else the argument is either overruled or defensible. Meeting a standard of proof, then, simply means offering a claim that is supported by a justified argument. Interestingly enough, arguments are justified to different degrees. Suppose that the argument is slightly stronger than its counterargument. The argument survives the attack as far as the preponderance standard goes. In the case of the standard of proof beyond a reasonable doubt, an argument must be significantly stronger than its counterargument, given some suitably defined threshold. So, meeting a certain standard of proof amounts to offering a justified argument for a claim, that is, an argument that survives *all* the attacks by counterarguments in accordance with the applicable standard. A problem with this account 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 that a reasonable objector could in principle put forward, not just the counterarguments that in fact are put forward.

Thomas Gordon and Douglas Walton develop an argumentation based framework to define different standards of proof. The lowest standard is scintilla of evidence. For them, this standard is met whenever there is at least one argument in favor of the claim. The preponderance standard, instead, requires a comparison of the arguments for and those against the claim. If the arguments for the claim are stronger, even slightly stronger, than the arguments against, the preponderance standard is met. Finally, the criminal standard of proof beyond a reasonable doubt is met whenever the preponderance standard is met, and in addition, there is at least one argument for the claim which has a weight that is greater than a suitable defined threshold t. This account leaves open how weights are assigned to arguments and what value the threshold t should take. While the probability base account could identify a specific probability threshold, at least in theory, by applying the principle of expected utility theory, the argumentation based framework cannot. After all, the weight of an argument cannot be understood as a probability, or at least, it not clear how weights of argument can be translated into probabilities.

Scenarios BLA BLA Some scholars suggested that we should elaborate a theory of legal reasoning which departs from the probabilistc approach and which does not ignore trial procedures broadly construed The key notion here seems to be plausibility rather than probability. Ronald Allen recently suggested that that '[no] plausible alternative to a plausible story of guilt [should be] the rule of decision in criminal cases' and that '[i]n criminal cases, fact finders find guilt if there is a plausible story of guilt and no plausible story of innocence; otherwise, they find innocence.'

The plausibility-based approach is appealing, but the obvious problem is that the notion of a 'plausible story' or of a 'plausible alternative to a plausible story' is wholly underdefined. SAY MORE HERE Still, if the notion of plausibility is hard to define precisely, it seems closer to how jurors actually reason in trial proceedings, whereas the notion of probability, despite its mathematical underpinnings, is hard to relate to actual trial proceedings: jurors do not naturally quantify guilt, and it is difficult to quantify it even if we wanted to. We face a trade-off: probability is a formally developed notion, but it is removed from trial proceedings and common-sense reasoning; plausibility lacks a well-established theory, but it is closer to trial practice.

Further readings Evidence law manuals (Fisher, 2008; Méndez, 2008). Criminal Procedure manuals (Allen et al., 2005). Character evidence and its exclusion (Redmayne, 2015). Free proof, legal arithmetic and rules of weight

Attempts to definite of beyond a reasonable doubt by courts in the US and elsewhere. Commonwealth v. Massachusetts Webster (1850) (proof beyond a reasonable doubt is equated to 'reasonable and moral certainty'). Holland v. United States (1954), 348 U.S. 121, 140 ('attempts to explain the term "reasonable doubt" do not result in making it any clearer') Supreme Court of Canada in R. v. Lifchus (1997), at 335 ('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')

Probabilistic accounts of the burden of proof (Kaplan, 1968; Kaye, 1986, 1999; Hamer, 2004; Cheng, 2013). Critique of probabilistic accounts (Cohen, 1977; Nesson, 1979; Thomson, 1986; Stein, 2005; Ho, 2008; Pardo and Allen, 2008; Haack, 2011). 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).

Narrative based account of proof beyond a reasonable doubt (Allen, 2010; Allen and Stein, 2013).

Argumentation based accounts of proof beyond a reasonable doubt ().

6 SUMMARY AND CONCLUSION

REFERENCES

R. J. Allen. No plausible alternative to a plausible story of guilt as the rule of decision in criminal cases. In Juan Cruz and Larry 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.

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.

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.

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

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

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.

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.

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.

S. Haack. Legal probabilism: An epistemological dissent (manuscript). 2011.

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.

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

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

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

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

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

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

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

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

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

J. 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.

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, 1993.

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.

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

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

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 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.

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.

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.

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. Chapter 11: Argumentation in Artificial Intelligence. In *Handbook of Argumentation Theory*. Springer, Berlin, 2014.

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

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.

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.

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