EVIDENTIAL REASONING
Chapter for the Handbook of Legal Reasoning

Marcello Di Bello & Bart Verheij – February 10, 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

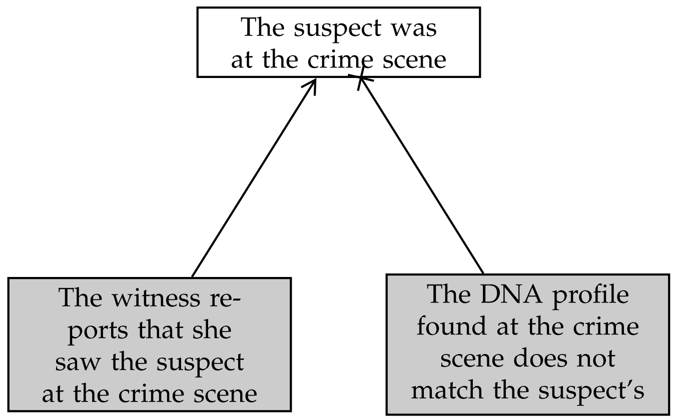


Figure 1: Arguments with supporting and attacking reasons

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:

S1: The victim's former partner killed the victim after a fight.

S2: 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 S1 and matching scenario S2.

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

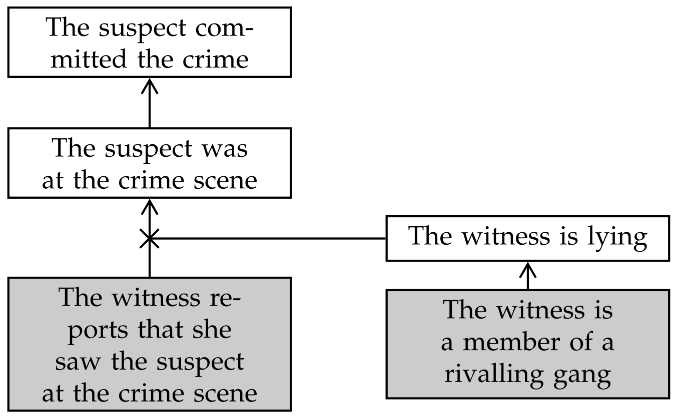


Figure 2: Arguments have structure

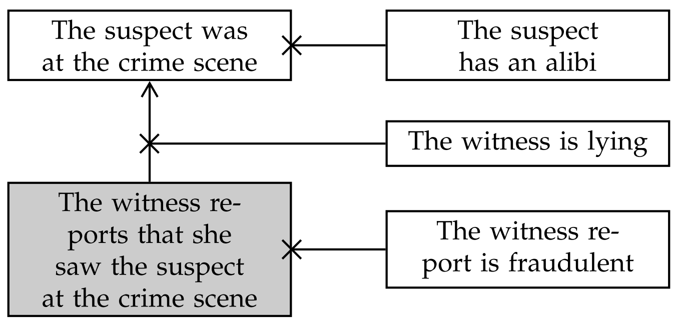


Figure 3: Three kinds of attack

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. 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

Also the scenario perspective sheds light on the treatment of conflicting evidence. We illustrate the use of conflicting evidence in a scenario perspective by going back to the example in Section 1.4.3. In the example, there were three main scenarios: the breakup murder scenario S_1, the innocent former partner scenario S_2, and the caught robber scenario S_3 (Figure 4).

Several scenarios and their relations In the scenario framework, it is natural to consider several mutually inconsistent scenarios simultaneously. Scenarios can have different relations with one another. They can be mutually inconsistent, such as the two murder scenarios S_1 and S_3 that each assume a different killer. They can be compatible, such as the innocent former partner scenario S_2 and the caught robber scenario S_3 that both can be true. They can have subscenario relations, such as the break up fight scenario S_4 (in the figure consisting of the events H_1, H_2 and H_3) that is a subscenario of the two scenarios S_1 and S_2 involving the victim's former partner.

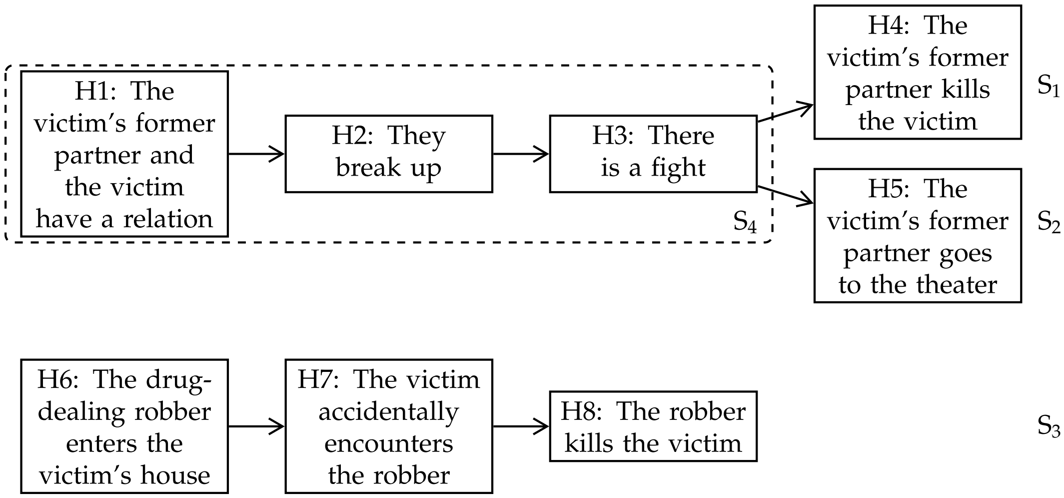


Figure 4: Scenarios

Scenarios as alternative explanations of the evidence Scenarios can be considered as alternative explanations of the evidence. Returning to the three pieces of evidence discussed in Section 1.4.3, we see that the breakup murder scenario $S1$ explains the skin trace $E1$, the innocent former partner scenario $S2$ explains the skin trace $E1$ and the bank card use $E2$, and the caught robber scenario explains the confession $E3$.

COPIED FROM THAT SECTION; REORGANIZE; my suggestion: that section even briefer. the breakup murder scenario $S1$ explains the skin trace $E1$, but is contradicted by the use of the bank card $E2$, and again by the confession $E3$. The innocent former partner scenario $S2$ explains the skin trace $E1$ and the bank card use $E2$, and is independent from the confession $E3$. The caught robber scenario explains the confession $E3$, and is independent from the skin trace $E1$ and the bank card use $E2$. Considering these scenarios and this evidence, breakup murder scenario $S1$ is hard to believe, the innocent former partner and caught robber scenarios $S2$ and $S3$ seem to be true.

Comparative adequacy of alternative scenarios/hypotheses in explaining the evidence

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]

[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 $1 - P(H|E) \geq$

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

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).

2.4 Further reading

Arguments: Wigmore, references Pollock 1995, Dung 1995; ?mention nonmonlog, Toulmin's anti-logicism

3 EVIDENTIAL VALUE

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

$1 - P(H)$. 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 $P(H|E) > P(H)$.

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.

Confusion sometimes arises with the RMP, generally known as prosecutor's fallacy There is an error people sometimes make while reasoning with the RMP. The mistake is to think that if the RMP is low—that is, the probability that a random person would match is low—the probability that the defendant is a random person (or that the defendant is innocent) is also low, and thus the probability of guilt is high. The inference from 'it is unlikely that a random (or innocent) person would match' to the 'it is unlikely that someone who matches is a random (or innocent) person' is wrong. This is a confusion of $P(M|\neg G)$, which is indeed low if the RMP is low (given our simplifying assumptions), with $P(\neg G|M)$, which need not be low at all. The application of Bayes' theorem can protect us from this confusion, as illustrated by the following table:

$P(\neg G)$	$P(M\vert\neg G)$ (or RMP)	$P(\neg G\vert M)$
0.9999999	1 in 100 million	0.1
0.99999999	1 in 100 million	0.5
0.999999999	1 in 100 million	0.9

If the probability that the defendant is *not* guilty is high prior to considering the DNA match, this probability can still remain high even after taking into account the DNA match, or at least, no identification can be made between the probability of $\neg G$ and the RMP. The confusion is an instance of a reasoning fallacy known in the probabilistic literature as the *inversion fallacy* and consists in confusing the probability of $P(E|H)$ and $P(H|E)$. In the legal literature, this confusion is known as the *prosecutor's fallacy* because this fallacy typically works to the advantage of prosecutors (**?**).

3.2 Arguments

Difference between deductive and presumptive arguments [inductive, abductive, ampliative, defeasible; what have you] Deductive arguments, when valid, ensure that the truth of the conclusion necessarily follows from the truth of the premises. Indeed, the premises might be altogether false, but in so far as they are true, the conclusion must unquestionably be true. Arguments in a court of law, however, can hardly be of this sort. If a witness asserts that she saw the defendant on 5th Avenue, the assertion does not establish, unquestionably, that the defendant was in fact on 5th Avenue. The argument 'the witness asserts p, *therefore* p' is not deductively valid. At the same time, the argument is not a worthless either, and there are two ways to see this. First, if a witness asserts that p is the case, the assertion establishes—at least tentatively, *prima facie*, absent contrary evidence—the truth of p. Second, if a witness asserts that p is the case, the assertion establishes—with a high probability—the truth of p (at least, provided the witness is a reliable one). In the former case, we speak of defeasibly valid arguments, and in the latter case we speak of inductively valid arguments. REFERENCE. Both types must be distinguished from deductively valid arguments in which the truth of the premises guarantees—with no exception whatsoever—the truth of the conclusion. In defeasibly or inductively valid arguments, instead, it is still possible for the conclusion to be false and the premises true.

Defeasible or inductive arguments can always be transformed in deductive ones by adding the appropriate premises. For example, the defeasibly valid argument

> The witness asserts p
>
> ---
>
> Therefore, p

can be rewritten as the deductively valid argument

> The witness asserts p
>
> (+) There is no contrary evidence
>
> (++) Absent contrary evidence, what a witness asserts is true
>
> ---
>
> Therefore, p

In the deductive version, the weaknesses of the inductive argument are explicitly laid out. It is therefore instructive to rewrite an inductive argument in the deductively valid form, so that the ways in which the inductive argument could go wrong are made explicit. REFERENCES.

Strength In assessing the strength of arguments, two factors must be considered. The first is how strongly the premises support the conclusion. In this respect, deductively valid arguments have maximal strength because the premises unquestionably support the conclusion. But, as seen earlier, any argument can be turned into a deductively valid argument by adding the appropriate premises. This cannot mean, however, that any argument can be turned into an argument with maximal strength; this would be absurd. The strength of an argument—and this is the second factor to consider—must therefore also depend on whether the premises are well supported. We consider a number of approaches to make these ideas more precise.

Args are good when surviving scrutiny under critical questions (arg schemes; Walton et al ...) Critical questions can be used to assess the strength of an argument. They can attack the premises, the inferential link between premises and conclusion or the conclusion itself. Of course, in a deductively valid argument, only the premises and the conclusion can be attacked, while in a defeasibly valid argument the inferential link between premises and conclusion can also be attacked.

Consider first a defeasibly or inductively valid argument based on eyewitness testimony:

> The witness asserts she saw the defendant run away from the crime scene.
>
> ---
>
> The defendant run away from the crime scene.

The critical questions that attack the inferential link between premise and conclusion can include: Were the visibility conditions good at the time? Does the witness have good memory? The critical questions that directly attack the conclusion itself can include: Was the defendant physically able to run away? Was defendant with his girlfriend when, allegedly, he run away from the crime scene? Critical questions are pointers toward potential counterarguments. If the questions target the inferential link, they are pointers to undercutting arguments, and if they target the conclusion itself, they are pointers to rebutting arguments.

In the above argument, it would be pointless to formulate a critical question that attacks the premise itself. If the witness is speaking in court, the assertion cannot be questioned. Consider now an argument based on expert option:

> The expert asserts that Benedict intake during pregnancy causes birth defects among animals.
>
> Benedict's effects on animals are sufficiently similar to its effects on humans.
>
> ---
>
> Benedict intake during pregnancy causes birth defects in humans. REFERENCE SUPREME COURT CASE ON THAT, DAUBERT

The critical questions targeting the inferential link here include: Is the expert competent in the domain? Is the expert credible? Is the expert reliable? Does the expert agree with other experts? REFERENCE Other questions can be used to attack the premises themselves. To be sure, it is hard to attack the first premise because the statement was presumably made in court. Still, the second premise can be doubted. Does Benedict affect animals and humans in the same way? What if animals and humans react very differently? Don't animal and humans have very different digestive apaarrati?

Depending on the domain, more specific questions can be used, attacking the premises, the inferential link or the conclusion. If an expert testifies about a DNA match, the appropriate critical questions will also include: Could the match be coincidental? Did the expert comply with the laboratory protocols? Could the DNA material have been implanted? Does the DNA match have anything to do with the guilt of the defendant? And so on. In section REFERENCE TO EARLIER SECTION, we identified different sources of uncertainty for DNA evidence. These sources correspond to specific critical questions that can be posed for expert opinions about a DNA match.

The more critical questions an argument survives, the stronger the argument. It is not easy to say what it takes for an argument to *survive* critical questions. First recall that, as noted earlier, critical questions are pointers toward a variety of counterarguments. A critical question that challenges the inferential link between premises and conclusion is a pointer to a potentially undercutting argument. A critical question that challenges one of the premises is a pointer to a potentially undermining argument. A critical questions that is directed against the conclusion itself is a pointer to a potentially rebutting argument. Given the parallelism between types of critical questions and types of attacks, to survive a critical question just means to survive an attack by a counterargument. To clarify this, in what follows, we shall look more closely at systems of arguments in which arguments attack and counterattack one another. REFERENCE The notion of attack and counterattack will be taken to be primitive. We shall next examine the notion of attack and counterattack more closely and offer a theory of when attack or counterattack is successful. REFERENCE

Args win when they can defend themselves against attacks (Dung 1995) Dung REFERENCE provided an abstract framework to analyze systems of arguments. The notion of an argument attacking another argument is taken as primitive. We can think of systems of argument as an alliance of arguments, and in particular, following Dung's framework, the system S of arguments is a *successful alliance* provided:

> (*conflict-free*) each argument in S is not attacked by any argument in S; and
>
> (*defense*) for any argument (outside S) that attacks any argument in S, there is an argument inside S that attacks the attacker.

So, a system of arguments forms a successful alliance provided there is no internal conflict and any external attack can be counterattacked by arguments within the alliance. The notions of a winning and losing argument can now be defined:

> (*winning*) A *wins* provided A is part of a successful alliance; and
>
> (*losing*) A *loses* provided another argument A' attacks A and A' wins.

These definitions are intuitive. If an argument can defend itself from any attack by appealing to its allies, it wins. If, instead, there is a winning argument that attacks an argument, the latter loses. A problem here, however, is that an argument can be both winning and losing. Consider four arguments A, B and A', B' such that A' attacks A and vice versa, and B attacks B' and vice versa. The system $\{A, B\}$ is a successful alliance because there are no internal conflicts and any attack against A or B is counterattacked within the alliance $\{A, B\}$. It follows that arguments A and B both win. At the same time, the system $\{A', B'\}$ is also a successful alliance, so arguments A' and B' win. This means that A and B both lose because each is attacked by a winning argument, and the same holds for A' and B'.

(As far as evidential reasoning in the law is concerned, something is unsettling about this state of affairs. There is something unnatural about the systems $\{A, B\}$ and $\{A', B'\}$ and the fact that A attack A' and vice versa, and the same holds for B and B'. Suppose an eyewitness asserts she saw the defendant stab the victim and another witness asserts the eyewitness is biased because she has reasons to hate the defendant. The second witness attacks the testimony of the eyewitness. This exemplifies an attack against A by A'. There is nothing problematic about that. But now suppose the eyewitness defended herself against the second witness by asserting that the second witness is biased against her because the witness has reasons to hate her. This exemplifies a situation in which A attacks A'. There is nothing problematic about that either. But, although the fact that A attacks A' and vice versa are fine as isolated attacks, when they are considered together, they become problematic. Suppose we engage in a conversation in which I assert conclusion C on the basis of evidence e and you challenge me by challenging e on the basis of other evidence e'. Next, I defend myself by challenging your evidence e' and I do so by using my evidence e. You challenge targets e on the basis of e' and I target e' on the basis of e. This is a stale mate and a circle of attack and counterattack. Ideally, we would need further evidence, besides e and e', to resolve the controversy. In the court example, a third witness is needed who can tell us whether the first eyewitness or the second witness is biased. So, an easy fix to the above problem, especially when it comes to evidential reasoning in the law, is to require that stalemate situations or circles of attacks and counterattacks be avoided.)

Args win when they are better/stronger than/preferred over conflicting args Dung's framework considers system of arguments and the relations of attack and counterattack among arguments, but does not examine the internal structure of arguments. The notions of attack and counterattack are left unspecified, and the notion of a successful attack is also left unspecified. To be sure, all we can say from Dung's framework is that an attack is successful when it consists of a winning argument. We are told nothing more about the internal structure of arguments.

Prakken REFERENCE has provided a unified theory of argumentation which builds on Dung's insights but also examines the internal structure of arguments. As seen earlier, argument can be attacked in three ways. An argument can be attacked by offering another argument with the opposite conclusion (rebutting). An argument can be attacked by offering an argument that weakens the inferential link between premises and conclusion (undercutting). An argument can be attacked by challenging its premises (undermining). The remaining question is, when is an attack successful?

Let us begin with an example of one argument rebutting another. A witness asserts that the defendant stabbed the victim, while DNA evidence shows no match between the defendant and the crime traces. The argument based on an witness statements is rebutted by—but also rebuts—the argument based on the DNA match. The attack works in both direction in the sense that bother arguments attack one another regardless of which argument is made first. Is the attack successful? In Dung's framework, the answer is that both arguments win and both arguments lose (provided both arguments are part of a successful alliance). This is hardly satisfying.

A step forward is a ranking among arguments. A rebutting argument A' is considered a successful attack on an argument A whenver A is ranked higher than A'. How can arguments be ranked? The ranking will depend on the relative strength we assign to the premises and to the inferential link between premises and conclusion. Consider an abstract example. The first argument consists of premises A and the defeasible inference $A \Rightarrow B$, while the second argument consists of premises A' and the defeasible inference $A' \Rightarrow \neg B$. The conclusion of the first argument is B and the conclusion of the second argument is the negation of B. Which argument ranks higher? This will depend on whether we consider A a better premise—for example, more probable—than A'. It will also depend on whether we consider $A \Rightarrow B$ a better inference—for example, more likely to preserve the truth—than $A' \Rightarrow \neg B$. Clearly, other things being equal, if A is better than A', the first argument is ranked higher. However, if A is a better premise but $A' \Rightarrow \neg B$ is a better inference, it is not easy to adjudicate the ranking of the two arguments.

Consider now a case in which an argument undercuts another argument. There is nothing peculiar about this case. We can apply the same ideas as in the rebutting case. In order for the attack to be successful, the undercutting argument must be ranked higher than the attacked argument. If not, the attack launched by the undercutting argument must fail.

Finally, the third type of attack: undermining. An attack against a premise can consists in a one line statement that the premise is false or can be a more elaborate argument resulting in the denial of the premise. How can we tell which attacks on the premises are successful? At first blush, one mighty say that the mere statement that the premise is false would not do, but this would be too simplistic. Premises come in different guises and sometimes it might be enough to merely disagree with a premise and some other times a more sustained argument might be necessary. To illustrate, consider three types of premises. REFERENCES TO PRAKKEN AND GORDON

First, premises can be axioms or self evident statements. If the premises are axioms or self evident, they cannot be attacked. An attack against an axiom or a self evident premise will therefore always fail.

Second, premises can also be assumptions or statements that are taken for granted without explicitly supporting reasons. Assumptions are peculiar in that that they hold until contrary evidence shows they are false. Assumptions are close to what in the law are known as presumptions. Suppose the prosecutor argues that a mail package containing drugs was received by the defendant. The prosecutor has proof that the package was mailed and that it reached the defendant's address. The prosecutor has no explicitly proof that the packed was delivered to the defendant. The prosecutor *assumes* that if the package was sent and received at the defendant's address, it was the defendant who received it. What if the defendant challenges the assumption alleging that he was not at home when the package was delivered? This raises the question of who has the burden of proof. Should the prosecutor give evidence that the assumption is correct or should the defendant give evidence that the assumption is incorrect, and in absence of contrary evidence, should the assumption be taken to be correct? If we are in fact dealing with an assumption, it is up to the party who challenges the assumption to show that it is correct. Assumptions are true until contrary evidence is provided.

Finally, we have ordinary premises. These must be defended with adequate evidence and arguments and cannot be assumed the be true. If the a party attacks an argument by challenging an ordinary premise, the burden of proof is on the party proposing the argument to back up the premise. Ordinary premises, in this sense, are very different from assumptions. In short, axioms or self evident premises cannot be challenged. Assumptions are assumed to hold until contrary evidence is presented. The burden of proof is on the attacking party. Ordinary premises can only be believed on the basis of supporting evidence or arguments. The burden of prof here is on the attacked party.

An attack against the premises of an argument will succeed or fail depending on the targeted premises. Attacks against self evident premises always fail. Attacks against assumptions are successful only if evidence is introduced that the assumption is indeed false. Attacks against ordinary premises succeeds even though no evidence that the premise is false is introduced and provided that the other party has introduced no evidence for the truth of the premise. If the other party has introduced evidence for the truth of the premise, the premise is the conclusion of an argument, and thus the attack must take form of a rebuttal.

Pros and cons can be weighed (accrual)

(Mention ?Pollock's anti-probabilism)

3.3 Scenarios

4 COHERENTLY INTERPRETING THE EVIDENCE

4.1 Scenarios

Often criminal cases form a mosaic. A DNA match can support the hypothesis that the suspect is the source, and a witness can add information about how the crime was committed. 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? Consider a fairly realistic, though still quite simplified, scenario:

> *Ennio*. A woman is found dead in the woods. The investigators recover remnants of semen on her body, which is severely wounded; they also recover blood stains in a parking lot near the woods. The blood type matches with victim's blood type and DNA. From the semen on the woman's body, a DNA profile is created. Forensic experts estimate that the DNA profile in question has a statistical frequency of 1 in 100 million. Through a database search, it turns out that an individual in the neighborhood, Ennio, has a matching DNA profile. Ennio is arrested and charged with murder.

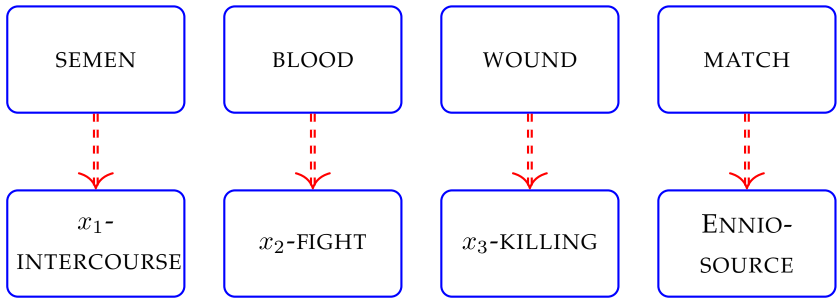
Imagine you are the prosecutor and you are constructing an incriminating case against Ennio. To a first rough approximation, four basic pieces of evidence are available:

(*semen*) Semen traces on the victim's body;
(*blood*) Blood stains in the parking lot;
(*wound*) Wounds from a knife on victim's body;
(*genetic match*) DNA match between Ennio and semen.

From the prosecutor's point of view, the four pieces of evidence above tend to establish, or at least they make a *prima facie* case in support of four hypotheses, as follows:

(x_1-*intercourse*) individual x_1 had intercourse with victim;
(x_2-*fight*) individual x_2 had violent fight with victim in parking lot;
(x_3-*killing*) individual x_3 killed victim with a knife;
(Ennio-*source*) Ennio is the source of the crime scene DNA.

The prosecutor, in constructing his incriminating case, can advance the four inferences:



The inference from the genetic match to the hypothesis that Ennio is the source of the crime scene DNA was discussed earlier while examining the evidential strength of DNA evidence. The acceptability of the other hypotheses (i.e. intercourse; fight; killing) depends on a number of case-by-case details: whether the semen on the victim's body was in a quantity, location, and arrangement that indicate intercourse; whether the wound on the victim's body was caused by a human artifact, e.g. a knife, and whether this caused the victim's death; etc. Without dwelling on unnecessary details, let's assume that forensic experts, on the basis of their experience, are willing to claim that the hypotheses in question can be established, at least *prima facie* and until objections are raised by the defense.

In order to make a unified case, the prosecutor can advance advance a *unification hypothesis*, as follows:

$$x_1 = x_2 = x_3 = \textit{Ennio}$$

The four initial hypotheses existed in isolation. They still left open the possibility that whoever had intercourse with the victim might be a different person from whoever fought with or killed the victim. They still left open the possibility that different people participated in the crime. The unification hypothesis, instead, identifies Ennio as the (only or main) perpetrator. The unification hypothesis, together with the others, asserts that Ennio had sexual intercourse with the victim, fought with the victim and finally killed her with a knife.

We now have a coherent and unifying case against the defendant Ennio. What is the ground to assert the unification claim? A number of considerations could weigh in here: when examined by forensic experts, the crime traces suggest the presence of one perpetrator only; the crime could, in principle, be committed by one person only; Ennio would have had the physical force to commit the crime alone; etc. Depending on the information available, the unification hypothesis will be more or less strongly supported. The evidential support for the unification hypothesis cannot be traced back to any single piece of evidence. The support for the unification claim heavily rests on holistic considerations based on cohesiveness. Putting everything together, the prosecutor's unified case against Ennio can be described as follows:

> The perpetrator had or attempted to have sexual intercourse with the victim in the parking lot (which explains the perpetrator's semen on the victim's body); a

> fight ensued during which the victim was wounded (which explains the blood stains in the parking lot); finally, the perpetrator killed the woman with a knife (which explains the wounds on the victim's body) and hid her body in the woods. Ennio is the perpetrator: he has a matching DNA profile whose frequency is as low as 1 in 100 millon.

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

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

Hierarchy of propositions (Evett et al)

Combining difference pieces of evidence using 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 $\frac{P(\textit{match}|\textit{guilt})}{P(\textit{match}|\neg\textit{guilt})}$ of 36. Suppose the witness testimony favors the hypothesis of guilt and again has a likelihood ratio $\frac{P(\textit{testimony}|\textit{guilt})}{P(\textit{testimony}|\neg\textit{guilt})}$ 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 $P(\textit{guilt}|\textit{match} \wedge \textit{testimony})$ will be greater than $P(\textit{guilt}|\textit{match})$ or $P(\textit{guilt}|\textit{testimony})$ 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 $i \in \{1, \ldots, k\}$, 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 $P(\mathit{contact}|\mathit{match}) = 0.8$ and $P(\mathit{plan}|\neg \mathit{email}) = 0.8$. 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:

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

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:

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

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.

The BN connection

[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, at least in theory, is to reach an estimate of the probability $P(G|E)$ of guilt based on all the available evidence. This process begins with the lowest possible value for $P(G)$, prior to considering any evidence. As more evidence is presented, the probability of guilt moves upwards or downward depending on whether the evidence is incriminating or exculpatory. When all the available evidence is considered—which means all the admissible evidence that was presented at trial—a guilt probability value, all things considered, is reached. This forms the basis of the decision to convict or acquit.

It is the view that establishing guilt beyond a reasonable doubt means to establish that the defendant's *probability of guilt*, given the total evidence presented at trial, meets a threshold, say, 0.99 or 0.999 (**???**). Consequently, a doubt would be reasonable or unreasonable depending on a measurable probability. Expected utility theory can be used to identify a precise numerical probability threshold. David Kaplan used a relative assessment of the disutilities associated with convicting an innocent, D_i, and with acquitting a guilty defendant, D_g. Suppose the probability of guilt and innocence equals P_g and P_i such that $P_g = 1 - P_i$. To convict—Kaplan suggests—the jury must be believe that

[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.

$P_g D_g > (1 - P_g) D_i$.

The inequality represents a situation in which the expected disutility resulting from acquitting a guilty defendant is larger than the disutility resulting from convicting an innocent defendant. So, given the inevitable possibility of error, such a situation would be one in which convicting is less harmful than acquitting, so that conviction is justified. As one can see, this is an application of the more general statistical decision theory based on maximization of expected utility or the minimization of expected loss. REFERENCES HERE.

The inequality holds only if P_g reaches a certain value. From the inequality, by algebra, we have

$\frac{P_g}{1-P_g} > \frac{D_i}{D_g}$.

This formula gives a precise indication of how high the probability of guilt must be to justify a guilty verdict, relative to the ratio between D_i and D_g. If $\frac{D_i}{D_g} = \frac{9}{1}$—as it might be more appropriate in a criminal case—, the lower bound for P_g must be at least 0.9. More complicated models are also possible, but the basic idea is that the probability required for a conviction will be a function of weighing certain social and economic costs that would result from erroneous decisions.

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 **?** 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.

6 SUMMARY AND CONCLUSION

REFERENCES

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

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

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

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.

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

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.

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.

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

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

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.

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.