EVIDENTIAL REASONING
Chapter for the Handbook of Legal Reasoning

Marcello Di Bello & Bart Verheij – November 25, 2015

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. 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. Scientists have found that this is not entirely true, as some dependencies have been established, for instance among the profiles within ethnic groups. It is also accepted that the independence assumption is hard to test in full generality, as that would require assessing more profiles than 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 be determined from the Random Match Probability, after a correction for 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

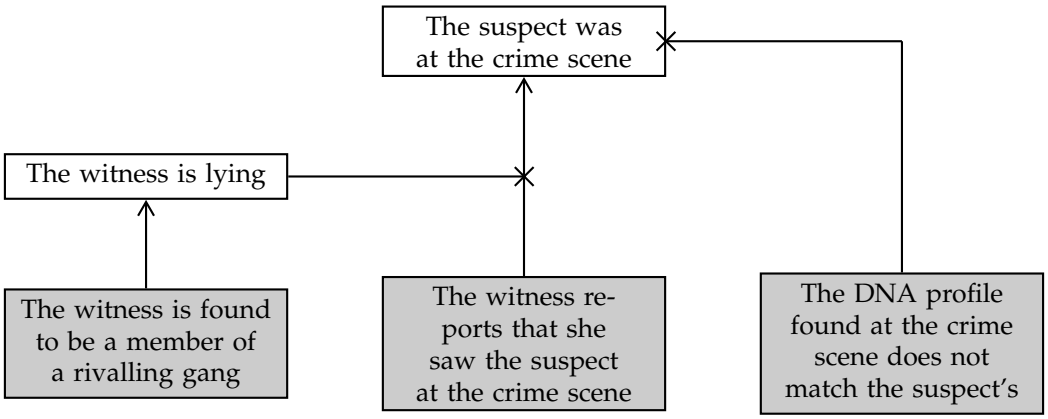


Figure 1: Arguments contain supporting and attacking reasons

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. When it is found that the witness is a member of a rivalling gang, it is concluded that the witness is lying. This resolves the conflict of reasons since the lying of the witness attacks the support for the suspect being at the crime scene on the basis of the witness testimony. Figure 1 summarizes these reasons for and against conclusions in a diagram.

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), who distinguished two kinds of attacking reasons. The first kind of attacking reasons—that he referred to as rebutting reasons—are reasons that support the opposite of the conclusion of the attacked reason. In the example, the non-matching DNA profile is a rebutting reason attacking the reason from the witness report since it supports that the suspect was not at the crime scene. The second kind of attacking reasons distinguished by Pollock, undercutting reasons, only attack the supporting connection of the attacked reason. In the example, the lying by the witness is an undercutting attack of the reason based on the witness report since it is not a reason that supports that the suspect was *not* at the crime scene. 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, and let E be the evidence that the suspect's profile matches the trace's. We are interested in the hypothetical event H that the suspect is the source of the trace. Because the profile is rare, a match is not often found accidentally, only when the suspect actually is the source of the trace. Intuitively, finding a match therefore has a high evidential value for establishing that the suspect is the source of the trace.

This intuition can be made precise in the probability calculus, as follows. If the suspect is not the source, written as $\neg H$, the probability of still finding the match by accident is 1 in a billion:

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

Ignoring sources of error, e.g., made in the lab, we take it that finding a match is certain when the suspect is the source of the trace:

$$\Pr(E|H) = 1$$

The ratio $\Pr(E|H)/\Pr(E|\neg H)$ of these two conditional probabilities is called the likelihood ratio, in the example 10^9. The posterior odds $\Pr(H|E)/\Pr(\neg H|E)$ of a hypothesis after the evidence is taken into account can be found by multiplying the likelihood ratio with the prior odds $\Pr(H)/\Pr(\neg H)$ before the evidence is taken into account:

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

This likelihood ratio formula explains why the likelihood ratio is a useful measure for the evidential value of new evidence: if the likelihood ratio is greater than 1, the posterior odds of the hypothesis are larger than the prior odds. In other words, the evidence makes the hypothesis more strongly supported, and the larger the likelihood ratio is, the larger the difference between posterior and prior. In our example, the likelihood ratio is large, 10^9, indicating a high evidential value of finding the match, as expected by the rarity of the profile. When the likelihood ratio is 1, the evidence does not change the odds; and when it is smaller than 1, the evidence decreases the support of the hypothesis. Then the posterior odds are smaller than the prior odds.

The likelihood ratio holds in the probability calculus, and is closely related to 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 goes in fact back to the early days of forensic science (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 in which some skin tissue is found under the victim's nails. The former partner of the victim is considered a suspect because of the recent breakup. A hypothetical scenario then can be that, before the killing, the suspect had a fight with the murdered victim, who scratched the suspect so heavily that some skin tissue was left under his nails. The prosecution tests this scenario finding out that the DNA profile of the suspect matches that of the trace. The suspect's defense can report an alternative scenario that indeed the suspect had had a fight with the victim on the occasion of their breaking up, but that she is not the murderer as she was in the theater that specific night. As evidence, she produces proof that here bank card was used at the theater's counter. The prosecution agrees that their guilty scenario cannot be considered true beyond a reasonable doubt, and cannot close the case, until a drug-using robber arrested in another case breaks, and confesses having killed the victim who accidentally entered during a robbery.

We now have three scenarios:

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

S2: The victim's former partner has had a fight with the victim, but has not killed him and was in the theater that night.

S3: The drug-dealing robber killed the victim when caught during a robbery.

[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).

We also have three pieces of evidence:

E1: There is a match of the DNA profile of the trace and that of the victim's former partner.

E2: The victim's former partner's bank card was used at the theater that night.

E3: The drug-dealing robber confesses having killed the victim during a robbery.

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.

Attention for scenario analysis rose when it was realized that scenarios are 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 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, 2008a).

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 6:

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

Legal cases arise because there are reasons pros and cons a position In many cases, the law and the facts are not in dispute. Consider a routine traffic violation such as speeding. If you are 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 usually 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 is not easy to decide which piece of evidence to trust or which reconstruction to believe. The need for a legal trial therefore arises.

evidential reasoning in the law is dialectical Legal trials 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.

2.1 Arguments

Args are for different, possibly conflicting positions (Van Eemeren et al 2014; not specific for evidence) ???

Dialectical aspect of arguments and Argument can support a certain conclusion (i.e. premises support a conclusion) An argument is a collection of statements in which one statement is the conclusion and the others are the premises functioning as evidence for the conclusion. Premises and conclusions, however, are contextual notions. Consider the collection of statements {I am getting wet; it is raining}. Which is the premise? Which is the conclusion? This depends on what is at issue. If the weather condition is at issue, getting wet is the premise functioning as evidence for the conclusion that it is raining. If, instead, one's physical condition is at issue, the fact that it is raining is the premise functioning as evidence for the conclusion that one is getting wet. In this sense, the conclusion of an argument is what the interlocutors disagree about, while the premises represent what the interlocutors take as evidence that can prove or disprove the conclusion.

Arguments can attack other arguments There are different ways in which two interlocutors can disagree as they put forward conflicting arguments. The argumentation theorist Pollock has distinguished two such ways, typically referred to in the literature as rebutting and undercutting (REFERENCE). These forms of conflict between arguments occur in everyday discussions, but occur also in a court of law while the prosecution and defense argue a case.

Rebutting: Arg2 leads to a different conclusion from Arg1 Let us begin with *rebutting*. Suppose A offers an argument for conclusion X, and B responds with an argument for conclusion Y, while X and Y cannot be both true, or in other words, X and Y are contradictory. When two arguments support contradictory conclusions, they rebut one another. Here is a legal illustration. In the British case, R v Adams [1996] 2 Cr App R 467, the victim was raped and the defendant's DNA matched with the traces of semen found on the victim's body. The prosecutor used the DNA match as a premise to support the conclusion that Adam raped the victim. But, Adam had an alibi—his girlfriend claimed he was with her when the crime occurred—and so the defense used the alibi as a premise to support the contradictory conclusion that Adam had nothing to do with the crime. This is an example of rebutting, because the two conclusions, each supported by a different piece of evidence, cannot be concurrently maintained.

Undercutting: Arg2 attacks the relation between Premises and Conclusion in Arg1 Turning now to *undercutting*, suppose A offers an argument for X consisting in premises P1,

P2, etc., while B shows that the premises do not support X, or at least, not as strongly as A thought. Here is an example. A prosecutor expert testifies that the defendant's DNA matches the crime scene DNA, and the prosecutor uses the expert testimony as evidence for the conclusion that the defendant visited the crime scene. But, suppose the expert for the defense testifies that the genetic profile used for the match is shared by millions of individuals. This information undercuts, or at least significantly weakens, the support that the DNA match lends to the conclusion that the defendant visited the crime scene.

Undermining: Arg2 attacks the premises on which Arg1 is based Besides rebutting and undercutting, some argumentation theorists (REFERENCES) identify a third way in which two arguments can conflict. This form of conflict is sometimes called *undermining*. Suppose A offers an argument for conclusion X consisting in premises P1, P2, etc., while B shows that one of the premises is false. For example, given the premise that the defendant matches the crime scene DNA, the prosecutor argues that the defendant visited the crime scene. The defense, on the other hand, points out that a laboratory error occurred and thus alleges that the premise in the prosecutor's argument, namely the DNA match, is false. If the laboratory made a mistake, the DNA match declared by the laboratory need not be a true match. This is a case of undermining because one of the premises in the proposed argument is false, or at least this is what one party in the dispute claims.

Wigmore's charts John Wigmore, as early as the beginning of the 20th century, devised a systematic method to chart arguments by identifying the various pieces of evidence and the relations of support and attack (REFERENCE). ILLUSTRATE REBUTTING, UNDERCUTTING AND UNDERMINING WITH WIGMORE CHARTS. In a court of law, the two parties in a trial aim to establish various conclusions, but the ultimate conclusion at issue is whether the defendant is guilty. The difficulty is that arguments aimed to establish the defendant's guilt can be intricate, and charting them with the Wigmore's method can be extremely laborious and complicated. SHOW A VERY COMPLEX WIGMORE CHARTS TO MAKE THE POINT THAT THEY ARE OFTEN UNREADABLE AND NOT INTUITIVE. Charting the arguments in a legal cases with Wigmore's charts might not be the best way to grasp a case as a whole. This might explain why, despite their clarity and precision, Wigmore's charts have never been popular among lawyers and practitioners.

references (Pollock 1995, Dung 1995; ?mention nonmonlog, Toulmin's anti-logicism)

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

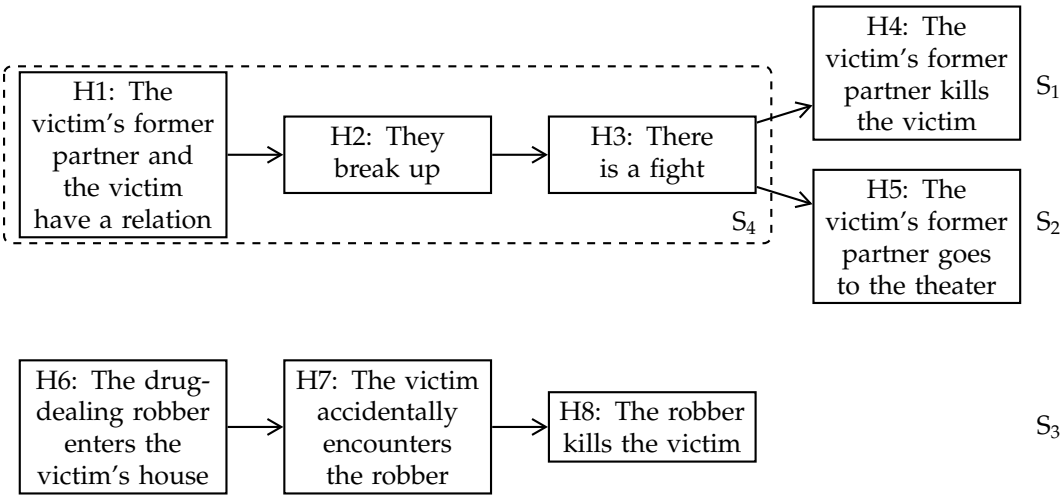


Figure 2: Scenarios

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 2).

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.

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 **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

We have seen how the premises of an argument can function as evidence favoring the conclusion. We can represent this relation probabilistically. We say that E is evidence favoring or supporting hypothesis H whenever E raises the probability of H, or more precisely, whenever taking into consideration E makes it more probable that H is true than it would be without taking into consideration E. The relation of evidential favoring or evidential support can also be described by means of likelihood ratios, that is, E is evidence favoring or supporting hypothesis H whenever the likelihood ratio $\frac{P(E|H)}{P(E|\neg H)}$ is greater than one. In fact, the two characterizations are equivalent, since the following holds:

$$P(H|E) > P(H) \textit{ iff } \frac{P(E|H)}{P(E|\neg H)} > 1.^{7}$$

Different, incompatible outcomes In a criminal trial, the prosecutor and the defense often disagree at trial as they put forward pieces of evidence that support contradictory hypotheses about what happened during the crime or who committed the crime. These can be as complex as the final proposition 'the defendant is guilty' or 'the defendant is not guilty', but also more circumscribed such as 'the defendant had contact with the victim' or 'the defendant was away when the crime occurred'. Let H_p and H_d be contradictory hypotheses and let E_p and E_d be pieces of evidence, respectively put forward by the prosecutor and the defense. The disagreement between the two parties can be represented probabilistically, as follows:

$$P(H_p|E_p) > P(H_p) \text{ and } P(H_d|E_d) > P(H_d),$$

or in terms of likelihood ratios,

$$\frac{P(E_p|H_p)}{P(E_p|\neg H_p)} > 1 \text{ and } \frac{P(E_d|H_d)}{P(E_d|\neg H_d)} > 1.$$

To illustrate, recall the earlier case in which Adams' DNA matched the traces of semen found on the victim's body (abbreviated M), but Adams also had an alibi provided by his girlfriend (abbreviated A). The match M favors the hypothesis CV that Adams has contact with the victim because M makes CV more probable. Similarly, the alibi A favors the hypothesis AW that Adams was away from the crime scene because A makes AW more

[7] This holds because of Bayes' theorem:

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

probable. So, we have

$$P(CV|M) > P(CV) \text{ and } P(AW|A) > P(AW),$$

or in terms of likelihood ratios,

$$\frac{P(M|CV)}{P(M|\neg CV)} > 1 \text{ and } \frac{P(A|AW)}{P(A|\neg AW)} > 1.$$

Note that CV and AW are contradictory hypotheses. If the defendant was away from the crime scene, he could not have had any contact with the victim (at least, not when the crime occurred). This closely mirrors a case of *rebutting* between arguments. The only difference is that instead of having contradictory conclusions, here we have contradictory hypotheses, but this is only a difference in terminology.

Same hypothesis can receive different probability assignments depending on evidence (or interpretation thereof) We have seen how two parties in a trial can disagree when they offer evidence that support contradictory hypotheses. But disagreements may also arise with respect to the same hypothesis. As the prosecutor introduces evidence in support of an incriminating hypothesis, the defense can argue that the prosecutor evidence does not—in fact—support the hypothesis. This disagreement can be represented probabilistically, as follows:

$$P(H_p|E_p) > P(H_p) \text{ but it is not the case that } P(H_p|E_p, E_d) > P(H_p),$$

or in terms of likelihood ratios,

$$\frac{P(E_p|H_p)}{P(E_p|\neg H_p)} > 1 \text{ but it is not the case that } \frac{P(E_p|H_p, E_d)}{P(E_p|\neg H_p, E_d)} > 1.$$

To illustrate, consider the hypothesis V that the defendant visited the crime scene at the time of the crime. Suppose the prosecutor witness testifies that she saw the defendant around the crime scene while the crime occurred (call this W_p), while a defense witness testifies that the prosecutor bribed the witness (call this W_d). We have:

$$P(V|W_p) > P(V) \text{ but it is not the case that } P(V|W_p, W_d) > P(V),$$

or in terms of likelihood ratios,

$$\frac{P(W_p|V)}{P(W_p|\neg V)} > 1 \text{ but it is not the case that } \frac{P(W_p|V, W_d)}{P(W_p|\neg V, W_d)} > 1.$$

This resembles a case of *undercutting* between arguments. Although the testimony by the prosecutor witness increases the probability that the defendant visited the crime scene, the testimony by the defense witness undercuts the reliability of this testimony.

Relative to the same hypothesis, likelihood ratios for different pieces of evidence can be positive or negative. ALREADY ADDRESSED EARLIER

If LR for E1 relative to H is positive, while LR for E2 relative to H is negative, E1 and E2 are in conflict ALREADY ADDRESS EARLIER

Comparison between the three frameworks IS THIS NEEDED?

3 EVIDENTIAL STRENGTH

3.1 Probability

Likelihood ratios can model evidential strength vs "posterior probability" (difference between the two) The probabilistic framework allows us to quantify the strength of the evidential favoring or support relation between evidence and hypothesis. The are different approaches in the literature (FITELSON REFERENCE HERE). One approach is to look at the difference between the probability of the hypothesis with and without considering the evidence, that is, $P(H|E) - P(H)$. The larger the difference, the stronger the evidential support. 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 stronger the evidential support. Both the difference $P(H|E) - P(H)$ and the likelihood ratio tell how much a piece of evidence E can impact, upwards or downwards, the initial probability of a hypothesis H. There are subtle differences between the two approaches, but these should not concern us here.

Evidential strength is not measured in terms of the conditional probability of an hypothesis given the evidence, often referred to as *posterior probability*. It is not true that a strong evidential strength amounts to a high posterior probability. This depends on the probability of the hypothesis regardless of the evidence, often referred to as *prior probability*. The following table shows how the posterior probability varies depending on the prior probability, with a fixed likelihood ratio of 1,000:

$P(H)$	Likelihood Ratio	$P(H\|E)$
0.0001	1,000	0.09
0.001	1,000	0.5
0.01	1,000	0.9
0.1	1,000	0.99

The posterior probability of the hypothesis is not high if its prior probability is particularly low, and in particular, if the likelihood ratio is n and the prior probability is $1/n$ or lower, the posterior probability will not be greater than 0.5.

ASIDE: LJ Cohen?s challenge (negation, conjunction paradox) and the Bayesian response Identifying evidential strength with conditional probability leads to counterintuitive results, one of them being the *negation paradox* (REFERENCE TO LJ COHEN). To illustrate, if the probability that it will rain given that the stock market crashed is low, the probability that it will *not* rain given that the stock market crashed must be high. This is a fact about probability. If the probability of H given E is low, the probability of $\neg H$ given E must be high. This fact, however, led LJ Cohen to conclude that the probability calculus "creates evidence from ignorance" and thus cannot offer an adequate account of evidential strength. But this conclusion is too quick. Indeed, if we identify evidential strength with conditional probability, the market crashing would be strong evidence that it will not rain, and this is patently absurd. Yet, according to the probabilistic account of evidential strength presented here, the information that the stock market crashed is not strong evidence that it will not rain. Learning about the market crash neither raises nor lowers the probability that it will (not) rain. In other words, the difference $P(H|E) - P(H)$ is roughly zero and the likelihood ratio roughly one. Cohen's negation paradox, then, only suggests that we should not characterize evidential strength in terms of conditional probability.

DNA evidence illustration As an illustration, we now describe the evidential strength of a DNA match in probabilistic terms. The first step is to decide on the hypothesis of interest. If the hypothesis is that the defendant is guilty, the evidential strength will be expressed by the likelihood ratio $\frac{P(M|G)}{P(M|\neg G)}$, where M is the DNA match and G is the hypothesis of guilt. But, it is not obvious that a DNA match bears directly on guilt itself. This depends on what kind of genetic traces we are considering. If the presence of the traces is incompatible with the innocent conduct of whoever left them, the DNA match would bear directly on guilt. But that is not always the case. The psychologist Jonathan Koheler has distinguished different hypotheses that are progressively more circumscribed: the defendant is guilty; the defendant partook in the crime; the defendant visited the crime scene; the defendant left the crime traces; the defendant is the source of the traces. The truth of a more circumscribed hypothesis does not necessarily entail the truth of a less circumscribe one. The defendant could have visited the crime scene without partaking in the crime or could have being the source of the traces without leaving them. (REFERENCE)

In what follows, we shall focus on the most circumscribed hypothesis that the defendant is the source of the crime traces. The extent to which the DNA match support this hypothesis is expressed by the likelihood ratio $\frac{P(M|S)}{P(M|\neg S)}$, where M is the DNA match and

S is the hypothesis that the defendant is the source. The higher the value of the ratio, the stronger the evidential support. Let us consider the numerator first. It is customary to put $P(M|S) = 1$ because if the defendant is the source of the crime traces, we expect that the genetic analyst would declare a match. This, to be sure, is a simplification, and we shall say more later in the chapter. Turning now to the denominator, it is customary to put $P(M|\neg S) = RMP$, where RMP is the probability that a random person, unrelated to the crime and to the defendant, would be found to have a matching DNA. This identification is plausible in the sense that 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 nothing to do with the crime or had no contact with the victim—would match anyway. The lower the RMP, the lower the denominator and thus the higher the likelihood ratio.

If we set aside certain complications on which we shall expand later, the evidential strength of a DNA match relative to the hypothesis S is as follows (Dawid, 2002; Balding, 2005):

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

In the Adams case we considered earlier, the RMP was 1 in 200 million, so the likelihood ratio would be:

$$\frac{P(M|S)}{P(M|\neg S)} = 200,000,000.$$

In the case, given that the likelihood ratio is such a high number, the DNA match relative to hypothesis S has a great deal of evidential strength, although this does not immediately translate into a high probability of guilty or in a high probability of the hypothesis 'source'.

The likelihood ratio analysis presented here can be extended to other forms of trace evidence. For example, it can be extended to fingerprints: instead of a genetic match, we will have a match between fingerprint profiles. The difference with DNA evidence is that we do not have a Random Match Probability for fingerprint profiles (Zabell, 2005). Besides fingerprints, the same analysis can be used for blood types, glass evidence, and any other type of trace evidence. But, again, we would need an estimate of the Random Match Probability which we often lack.

Confusions may arise between likelihood (ratios) and posterior probability (i.e. prosecutor?s fallacy) There is an error people routinely make while reasoning with the RMP. The mistake is to think that since the RMP is low—that is, the probability that a random person would match is low—it follows that there is a low probability that the defendant is a random person (or that the defendant has nothing to do with the crime or is innocent). The inference from the premise 'it is unlikely that a random person would match' to the conclusion 'it is unlikely that someone who matches is a random person' is wrong. This is a

confusion of $P(M|\neg S)$, which is indeed low if the RMP is low, with $P(\neg S|M)$, which need not be low at all. The application of Bayes' theorem can protect us from this confusion, as illustrated by the following table:

$P(\neg S)$	$P(M\|\neg S)$ (or RMP)	$P(\neg S\|M)$
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* the source 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 $\neg S$ and the RMP. (EXPLAIN MORE?) 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 (Thompson and Shumann, 1987). We have just seen an example of it in DNA evidence cases, namely the confusion was between $P(M|\neg S)$ and $P(\neg S|M)$. MENTION COLLINS AS ANOTHER EXAMPLE?

Likelihood ratios can model sources of weakness/error/uncertainty (illustrate this through DNA evidence) Let us now address some of the complications we left out as we provided a probabilistic analysis of the evidential strength of a DNA match. The likelihood ratio $\frac{P(S|M)}{P(\neg S|M)}$ was equated to $\frac{1}{RMP}$ but this overlooks certain important details. What M stands for is the lab report of a match, but the lab technician could make a mistake. She could not declare a match while in fact the defendant does match (*false negative*) or declare a match while in fact the defendant does not match (*false positive*). If we take false negatives into account, the numerator $P(M|S)$ must be lower than one. This has a negative impact on the strength of the evidence insofar as a lower value in the numerator will result in a lower likelihood ratio. Turning to false positives, since the laboratory or the technician performing the test could falsely report a match, the probability $P(M|\neg S)$ should be higher than the RMP. For example, a likelihood ratio as high as 1 billion reduces to about 100 if the laboratory error rate is just 1%.[8] More generally, false positives can have a negative

[8]To take lab errors into account, instead of a match M, we should properly speak of a reported match M_r. A reported match M_r can occur in two situations: when there is a true, factual match M_t or when there is a true, factual non-match $\neg M_t$. So, we have (Thompson et al., 2003):

$$P(M_r|\neg S) = P(M_r|M_t)P(M_t|\neg S) + P(M_r|\neg M_t)P(\neg M_t|\neg S).$$

Now, $P(M_t|\neg S)$ equals RMP, and thus $P(\neg M_t|\neg S)$ equals $1 - RMP$. Also, $P(M_r|\neg M_t)$ describes the laboratory's false positive rate, abbreviated by FP. Finally, $P(M_r|M_t)$ describes the laboratory's true positive rate, or the inverse of the laboratory's false negative rate, abbreviated FN. So, we have (Thompson et al., 2003) and (Buckleton, 2005a):

$$P(M|\neg S) = [(1 - FN) \times RMP] + [FP \times (1 - RMP)].$$

impact on the strength of the evidence insofar as a higher value in the denominator will result in a lower likelihood ratio

There are other reasons why $P(M|\neg S)$ should be higher than RMP. If the DNA traces were synthesized and implanted, the defendant will not be the source, yet an identical DNA could still be found at the crime scene and a match would be reported (absent any laboratory error or random match). This is not a far-fetched possibility and is likely to become more common in the future (REFERENCE). Another reason why $P(M|\neg S)$ should be higher than the RMP is that the defendant could have an identical twin. The RMP does not take into account this possibility because it typically describes the possibility that a random person who is *unrelated to the defendant* would match. Interestingly enough, the probability that any defendant has an identical twin is not negligible because the frequency of identical twins is estimated to be 3 every 1000 births.

More generally, DNA profiles are not distributed across the population of the earth with a fixed stable probability. It is not the case that each individual has the same probability of been assigned a genetic profile. Rather, DNA profiles are part of our genetic make-up, which we receive through evolution and natural selection. We did not receive our DNA through a random process. Naturally enough, people who are related are more likely to share the same DNA profile than people who are unrelated (Buckleton, 2005b; Weir, 2007). This is why forensic experts must qualify their statements by saying that the RMP applies to a population of individuals *unrelated to the defendant*. Another complication is that the RMP probability varies depending on the population of interest, Caucasians, African-Americas, Asians etc (REFERENCE).

To summarize, the different sources of uncertainty that may weaken the strength of a DNA match relative to the hypothesis that the defendant is the source of the crime traces are as follows:

(a) the match could be coincidental, that is, a random person could match;

(b) the laboratory test result could be mistaken (cf. false positives and false negatives);

(c) the genetic material could be synthesized and implanted;

The formula above shows that in order to properly estimate the value of $P(M_r|\neg S)$, we need a statistical estimate of the DNA profile's RMP *as well as* the laboratory's error rates (false positive and negative rates). Now suppose a DNA profile has a frequency as low as of 1 in 1 billion. Suppose laboratory false positive and false negative rates equal 0.01. If the likelihood ratio is simply equivalent to $1/RMP$, then it would be 1 billion. With the new formulation, we have:

$$\frac{P(M_r|S)}{P(M_r|\neg S)} = \frac{1}{[(1-0.01)\times 0.000000001] + [0.01\times(1-0.000000001)]} \approx 100.$$

The calculation is as follows:

$= \frac{1}{[(1-0.01)\times 0.000000001]+[0.01\times(1-0.000000001)]}$

$= \frac{1}{0.99\times 0.000000001+[0.01\times(0.999999999)]}$

$= \frac{1}{0.00000000099+0.00999999999}$

$= \frac{1}{0.01000000098} \approx 100$ See (Thompson et al., 2003) for more details.

(d) the defendant could have an identical twin; and

(e) the DNA profile could be more or less frequent depending on the racial group.

If we considered only the first factor, the likelihood ratio $\frac{P(S|M)}{P(\neg S|M)}$ for a DNA match would simply be $1/RMP$. As we take into account the other factors, the likelihood ratio becomes more complex. Part of the problem in considering the other factors is that the required numbers for assigning the right probabilities might not be available. The rate of identical twins is well-known and available and the false positive and negative rates of a laboratory might also be available, but we certainly do not know how often genetic material is synthesized and implanted on the crime scene. It is also not clear how to incorporate into a single likelihood ratio the fact that a DNA profile has varying frequencies depending in the racial group.[9]

Mention the fiction of chains of independent evidence; perhaps postpone discussion to section on coherence

References (Carnap, Skyrms, Fitelson, Schum, Kaye, Thompson)

[The problem of the old evidence? The issue of evidential strength of a DNA match after a database search?]

3.2 Arguments

Difference between deductive and presumptive arguments [inductive, abductive, ampliative, defeasible; what have you]

Args are good when surviving scrutiny under critical questions (arg schemes; Walton et al ...)

Args win when they can defend themselves against attacks (Dung 1995)

[9] Furthermore, likelihood ratios allow for an even more precise statement of the strength of DNA evidence toward the proposition that the defendant was the source of the crime traces. So far we spoke as though there is such a thing as a genetic match between the defendant's DNA and a the crime trace DNA. As a matter of fact, laboratory results do not yield any reported match. Rather, they yield a congruence result between two DNA profiles, one from the defendant and one from the crime traces. The two profiles are never perfectly identical; they are more or less congruent (Kaye, 1993). Instead of a reported match, what we really have are two pieces of evidence: one is the laboratory test showing that the defendant has the genetic profile d and the other is a laboratory test showing that the traces have the genetic profile t; call them G_d and G_t. Instead of M, we can use the conjunction $G_d \wedge G_t$. The likelihood ratio can therefore written more explicitly as follows:

$$\frac{P(G_d \wedge G_t|S)}{P(G_d \wedge G_t|\neg S)}.$$

This new formulation might not change the actual calculations, but it does make the reasoning clearer (Robertson and Vignaux, 1995).

Args win when they are better/stronger than/preferred over conflicting args

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.

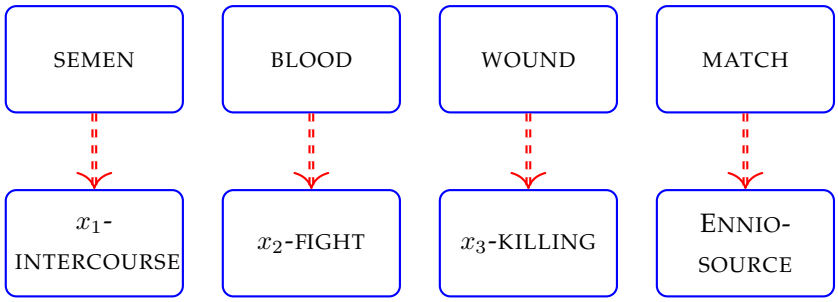
Imagine you are the prosecutor and you are constructing an incriminating case against Ennio. To a first rough approximation, four basic pieces of evidence are available:

(*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

Multiplying likelihood ratios and question of independence among difference pieces of evidence We first need a general account of how to model the *combination* of two pieces of evidence E1 and E2, each supporting, with its own degree of strength, the same hypothesis H. If the strength of a single pieces of evidence E depends on how much it impacts $P(H)$, the same applies to the combination of two pieces of evidence. Bayes' theorem shows us how the combination of E1 and E2 impacts the probability $P(H)$, as follows:

$$\frac{P(H|E1,E2)}{P(\neg H|E1,E2)} = \frac{P(E1 \wedge E2|H)}{P(E1 \wedge E2|\neg H)} \times \frac{P(H)}{P(\neg H)}.$$

If E1 and E2 are independent and do not influence one another, we have:

$$\frac{P(H|E1,E2)}{P(\neg H|E1,E2)} = \frac{P(E1|H)}{P(E1|\neg H)} \times \frac{P(E2|H)}{P(E2|\neg H)} \times \frac{P(H)}{P(\neg H)}.$$

$$\frac{P(H|E1,E2)}{P(\neg H|E1,E2)} = LR1 \times LR2 \times \frac{P(H)}{P(\neg H)}.$$

Suppose, in a criminal case, DNA evidence shows that the crime traces match with defendant's DNA. The DNA match has a likelihood ratio, relative to hypothesis S, of 1,000. Another piece of evidence, independent of the DNA evidence, shows that the crime scene blood type is the same as the defendant's blood type. Suppose the likelihood ratio is now just 10. So, the combined evidential strength of the two pieces of evidence is $1,000 \times 10 = 10,000$. This is an example of how two pieces of evidence converge toward the same hypothesis.

This seems a largely unsolved problem for probabilistic accounts (and for the others accounts as well)

The BN connection

5 INTRODUCTION, EXCLUSION AND EXAMINATION OF THE EVIDENCE

So far we have examined how inferences can be drawn from the evidence. We now turn to the process of acquiring and presenting evidence in the context of trial proceedings. A system in which the parties have complete freedom in the acquisition and presentation of the evidence at trial would be very simple and efficient, and yet, not without problems. When they are left unchecked, the divergent interests of the parties can hinder the discovery of the truth and put indigent defendants at an unfair disadvantage. Suppose a prosecutor or plaintiff found evidence favoring the defendant but decided not to disclose it nor present it at trial because of reasons of self interest. The withholding of the evidence will hinder the discovery of the truth and be unfair toward the defendant. The latter, to be sure, can discover the evidence himself but that would require considerable monetary resources which an indigent defendant typically does not have. Because of this, legal rules are be in place to correct the imbalances between the parties.

This section illustrates certain legal rules for the acquisition, introduction and exclusion of the evidence. Since an exhaustive examination of the topic cannot be undertaken here, a few illustrative examples, mostly from the federal law of the United States, will be discussed. In particular, we will discuss examples of rules of discovery and examples of rules of exclusion.

Rules of discovery The *rules of discovery* regulate the acquisition of the evidence by the parties and the disclosure of the evidence between the parties, but do not regulate the introduction of the evidence at trial. The process of discovery and disclosure happens prior to the trial.

The rules of discovery can be broad and symmetric. Rule 26(b)(1) of the Federal Rules of Civil Procedure, for example, mandates that any relevant evidence, both in favor and

against the defendant, must be disclosed by the party who possesses it upon request of the other party. This rule is broad insofar as any relevant relevant evidence is subject to disclosure notwithstanding certain limitations.[10] The rule is symmetric insofar as both prosecution and defense are under the obligation to disclose relevant information upon request. This arrangement, at least in theory, encourages the widest possible scrutiny of the facts prior to the confrontation at trial. A party, however, may abuse Rule 26(b)(1) and use it to artificially delay the proceedings or force a settlement. Consider a plaintiff who is suing a hospital for malpractice and requests under rule Rule 26(b)(1) that the defendant disclose all email conversations of the past ten years. Retrieving and disclosing this information has staggering costs, in the order of millions of dollars, given the high number of email messages that are exchanged every day. Since the party complying with the request of discovery must also bear the costs, the plaintiff might well make the discovery request not because relevant information could be found, but simply to force the defendant to settle as a way to avoid the discovery costs (Beisner, 2010)

Discovery rules in criminal cases are less broad and asymmetric. In Brady v. Maryland (1963), 373 U.S. 83, the US Supreme Court ruled that, in criminal cases and upon request, the prosecution must disclose exculpatory evidence that is crucial for the defendant's case, while the defense is under no such obligation. The asymmetry between defense and prosecution is justified by the different degrees of harm that may result if an error is committed at the detriment of the defendant or the prosecutor. A wrongful conviction, after all, will result in the defendant's unjust loss of liberty or even life, whereas a wrongful acquittal is thought to be less harmful. In civil trials, by contrast, the stakes are not so high for defendants, so the asymmetry between the parties is not so pronounced. EXAMPLE OF ABUSE OF DISCOVERY RULES IN CRIMINAL CASES.

As the above discussion suggests, anyone devising legal rules for the acquisition of the evidence faces at least two complications. The first is that acquiring evidence is costly. Acquiring as much evidence as possible, both for and against the defendant, is a worthy objective, and yet, discovery costs can be significant and should be allocated fairly. Should the costs be allocated to the parties according to the resources available to each? Should the party who initiated the litigation bear the costs? Should the losing party pay? The answers here are not straightforward. The second complication is that, in criminal cases, the protection of innocent defendants against wrongful convictions might take priority against other trial goals such as truth and accuracy. According to the Brady rule, it is the prosecution, not the defense, who should disclose evidence of innocence. If the discovery of the truth and the scrutiny of the evidence were paramount, why not force the defense to disclose

[10] Rule 26(b)(1) states that discovery requests must be "proportional to the needs of the case, considering the importance of the issues at stake in the action, the amount in controversy, the parties' relative access to relevant information, the parties' resources, the importance of the discovery in resolving the issues, and whether the burden or expense of the proposed discovery outweighs its likely benefit".

evidence of guilt? As far as discovery rules are concerned, the objective of collecting evidence and discovering the truth is often weighed against the pragmatic limitations of the system, such as the limited economic resources of the litigants, but also against the moral imperative of protecting innocent defendants from wrongful convictions.

Relevance and admissibility Rules of discovery regulate the acquisition and disclosure of the evidence prior to trial. This is often enough for the parties to reach a settlement or plea bargain. In a few cases, however, the parties elect to go to trial, and if this happens, a new question arises. What evidence can be admitted to trial and what evidence is instead inadmissible? The common answer here is that only relevant evidence is admissible. It is customary to distinguish two components in the notion of relevance: probative value and materiality (Fisher, 2008; Méndez, 2008). Probative value concerns whether the evidence has any tendency to establish the proposition that it is intended to prove. This can be explained probabilistically. A piece of evidence is positively (or negatively) probative of a proposition insofar as the evidence makes the proposition more (or less) probable than it would be in absence of the evidence. This characterization mirrors the notion of evidential favoring (or disfavoring) that we discussed earlier. (SEE EARLIER SECTION OF THIS CHAPTER).

Besides probative value, materiality concerns the fit between the evidence and the case. If the evidence is offered to establish a proposition that has nothing to do with the case, the proposition is immaterial and the evidence irrelevant. The evidence might still be probative, but if the proposition it proves is immaterial, the evidence counts as irrelevant. In general, materiality depends on the substantive law of the case which identifies the propositions to be proven. In a murder case, for example, the prosecutor must prove that the perpetrator caused the victim's death and did so with premeditated intent. Premeditated intent and cause of death are both material to the case. However, a proposition can be material to a case even though it does not explicitly fall under the scope of the substantive law. For example, it is material whether the witness for the prosecutor was in fact present at the time of the crime. This has little to do with what the perpetrator did or with the crime itself, but it is an evidentiary issue that is important to ascertain. REFERENCE MISSING.

Exclusionary rules Once the question of relevance is settled, a new questions arises. Should all relevant evidence be admissible? Unlike countries in continental Europe, countries in the common law tradition, especially the United States, have a very elaborate system of exclusionary rules. There exists a specific branch of law, called Evidence Law, which consists of rules for the exclusion of otherwise relevant evidence. There are historical reasons why exclusionary rules developed in common law countries, but not in others. One theory is that this development has mostly to do with the institution of the jury. REFER-

ENCE. As we shall see, exclusionary rules are in place to contain the prejudicial effects that the introduction of certain items of evidence may cause, and such prejudicial effects are more likely to occur when lay jurors rather than experienced judges consider the evidence.

We shall now look at two well-established exclusionary rules in the common law: character evidence and hearsay.

Character evidence Character evidence concerns the previous conduct of the defendant, not the specific conduct under examination at trial. The Federal Rules of Evidence, for example, mandate that "[e]vidence of other crimes, wrongs, or acts is not admissible to prove the character of a person in order to show action in conformity therewith." FRE 404(b). This means that the prosecutor or the plaintiff cannot use evidence of other wrongdoings to establish the defendant's bad character and on this basis his involvement in the alleged wrongdoing. At first blush, one might think that character evidence is inadmissible because it is irrelevant. Insofar as a trial is not about prior acts, character evidence seems irrelevant. But this need not be the reason why character evidence is excluded. In criminal cases, for example, we know that those who committed an offense in the past are more likely to reoffend than those without a criminal record REFERENCE NEEDED HERE. If this is correct, character evidence must count as relevant, at least insofar as the characterization of relevant evidence we considered earlier. Despite its relevance, however, character evidence can have significant prejudicial effects. This is true especially in a jury trial in which the the jurors might take the defendant's record as unduly strong evidence that the defendant must have engaged in the wrongdoing again. The problem with character evidence, then, is that the lay jurors will exaggerate its probative value. On the other hand, it is far from uncontroversial that character evidence should be excluded from trials. The exclusion of character is mostly a feature of the common law tradition, and even within the common law there are exceptions to the rule. EXAMPLES. REFERENCES. More generally, some scholars have argued that character evidence is relevant and thus should be admissible (Redmayne, 2002).

Hearsay The exclusion of hearsay evidence is another well-entrenched rule in the common law. Hearsay evidence is a statement made out of court which is introduced at trial to prove the truth of the matter asserted in the statement. If a witness says in a bar "Mr. Walton was running down the street with a bloody knife in his hand" and this statement is introduced as evidence in court by a third party, the statement would count as hearsay evidence. There are at least a couple of reasons why hearsay evidence is inadmissible. One is that it is unreliable. A witness who saw the crime and can recount what she saw in court is a more reliable source of information than a second hand-witness who can only recount what another witness told her about the crime. It is because of the more indirect chain

of transmission of information that hearsay evidence is less reliable. The more intermediaries in the chain, the more corrupt the information at the end of the chain. REFERENCES NEEDED. This is not say that hearsay evidence is irrelevant, but that its probative value is tenuous. Another reason why hearsay evidence is excluded is that it cannot be scrutinized or cross examined. If A testifies in court that she heard B say that the defendant did so-and-so during the crime, the person who should be cross examined is B because B has first hand knowledge of what happened. Instead, the person who is testifying in court is A who has no first hand knowledge. Hearsay evidence, then, would undermine the right of defendants to directly confront their accusers and cross examine them. REFERENCES. MORE EXPLANATION NEEDED HERE? TALK ABOUT EXCEPTION TO HEARSAY RULE

Cross examination MAYBE A PARAGRAPH ABOUT CROSS EXAMINATION HERE IS NEDDED?

6 WHEN CAN WE CONVICT?

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 make a decision. They must decide whether to convict or acquit the defendant in a criminal case, and whether to find against or in favor of the defendant in a civil case. There is no denying that the decision must be based on the evidence presented and the law governing the case, and not human feelings. Still, this leaves a crucial question unanswered. What criterion should govern the decision?

In continental Europe until the 18th century, strict rules of decision applied. For example, the testimony of one incriminating witness was deemed insufficient to prove that the defendant committed a crimea as serious as murder. This is often called the 'two witness rule', inherited from Roman Law, because two witnesses were required for a conviction, or alternatively, the defendant's own confession. More generally, an elaborate system of legal arithmetic was in place, dividing legal proofs in full and half proofs and detailing how proofs could be added to one another. REFERENCES. With the age of the Enlightenment, however, the idea of free proof gained momentum and the system of legal arithmetic fell in disrepute. The rationale for free proof was that since each juror was endowed with reason and common sense, she should be able to weigh the evidence and come to a reasoned conclusion without following strict, obscure and ultimately arbitrary rules for evidential reasoning. The legal history in England went a different way, mostly because of the institution of the jury. But, as it turns out, common law countries and countries in continental Europe both embraced the principles of free proof. REFERENCES. The academic debate is still open today. Given the high number of wrongful convictions, some scholars have

championed the adoption of rules of weight, some of which reminiscent of the two witness rule. One proposal, for example, is that no conviction can solely be based on one piece of incriminating evidence. REFERENCES

Academic proposals and history aside, as things stand currently, the decision criterion simply consists of a standard of proof, sometimes also called burden of persuasion.[11] The standard of proof identifies, in a somewhat verbally imprecise manner, how strong the evidence should be for warranting a finding of criminal or civil liability. Failure to meet the standard of proof must result in an acquittal or a finding in favor of the defendant. The decision criterion, for criminal trials 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. In civil trials, instead, the standard is less demanding. A finding of civil liability against the defendant requires that the claim be established with *preponderance of the evidence*, or else the fact finders must find for the defendant.

Informal clarifications What does it mean to establish guilt beyond a reasonable doubt? When is a doubt reasonable or unreasonable? Can the standard be given a more precise meaning and characterization? We can only begin to scratch at the surface of these questions and point the reader toward existing literature. Explications of when a doubt is reasonable or unreasonable abound. In Commonwealth v. Massachusetts Webster (1850), proof beyond a reasonable doubt is equated to 'reasonable and moral certainty' (295, 59 Mass., 320). Another paraphrase is that proof beyond a reasonable doubt is such that 'a reasonable person would not hesitate to act upon it in the most important of his own affairs,' or again, proof beyond a reasonable doubt must cause 'an abiding conviction in the minds of the jurors.' One might worry whether these paraphrases do enhance our understand or instead muddle the waters further. Even the U.S. Supreme Court at some pint remarked that 'attempts to explain the term "reasonable doubt" do not result in making it any clearer' (Holland v. United States (1954), 348 U.S. 121, 140).

Another strategy for clarifying proof beyond a reasonable doubt is to identify certain functional constraints that the standard should satisfy, avoiding the trouble of proving a precise definition. Following this strategy, the Supreme Court of Canada in R. v. Lifchus (1997) writes:

> (-) the standard of proof beyond a reasonable doubt is inextricably intertwined with that principle fundamental to all criminal trials, the presumption of innocence;

[11] 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 burden of proof rests on the prosecution throughout the trial and never shifts to the accused;
>
> (-) a reasonable doubt is not a doubt based upon sympathy or prejudice; rather, it is based upon reason and common sense;
>
> (-) it is logically connected to the evidence or absence of evidence;
>
> (-) it does not involve proof to an absolute certainty; it is not proof beyond any doubt nor is it an imaginary or frivolous doubt; and
>
> R. v. Lifchus (1997) at 335.

Most would find the Court's remarks unobjectionable. The Court explicitly linked the criminal standard to other related notions, such as the presumption of innocence, the burden of proof, the presence and absence of evidence, reason and common sense. These connections are important and help us see the role that the standard proof beyond a reasonable doubt plays in the criminal trial.

Legal historians have investigated how the standard proof beyond a reasonable doubt came about. Some might think of it as an epistemological standard because it defines a bar on how strong the evidence should be to support a finding of guilt. REFERENCES. One legal historian, however, has suggested that the standard emerged as an instrument of moral conform. It emerged as a tool to make convictions easier rather than harder. When religious beliefs were more entrenched than today, jurors were afraid of making mistakes. After all, the punishment for convicting an innocent could be as terrifying as the flames of hell. Frightened by this prospect, jurors could hardly convict, thereby making convicting nearly impossible. The standard of proof beyond a reasonable doubt was then introduced to shield jurors from their anguish. If they found no reasonable doubt in their minds about the defendant's guilt, they would be allowed to convict without risking the flames of hell should the decision turn out to be incorrect. REFERENCES. This contrasts markedly without our current view. We typically think of the standard as a tool that protects innocent defendants from wrongful convictions, as a tool that makes convicting harder and not easier. So, as we theorized about the standard of proof, we benefit from being aware that a conceptual, cultural and moral shift has occurred, and this shifts has affected the meaning and function we assign to the standard.

Each of these analyses—informal paraphrases, functional and historical—provide us with some insights. For those who seek a more analytic characterization, however, these informal analyses might be unsatisfactory. In what follows, we present three analytical characterizations of the standard of proof, each inspired by one of our three normative frameworks: probability, arguments and scenarios.

Probability-based account In the existing literature, one view is legal probabilism. 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 (Kaplan, 1968; Kaye, 1999; Tillers and Gottfried, 2007). Consequently, a doubt would be reasonable or unreasonable depending on a measurable probability.

This definition is simple, crisp and elegant, but a too literal interpretation of it is obviously problematic. If a probabilistic threshold is understood as a criterion which the fact-finders should mechanically apply whenever they confront the decision to convict or acquit, two difficulties arise. The first difficulty is that assigning a probability value to guilt itself might not be feasible. This difficulty, however, can be sidestepped if we understand legal probabilism less mechanistically. The legal probabilists can defend their proposal by conceding that they are not offering a recipe that should be directly implementable in court. Assigning probabilities to propositions, they could say, is an idealized process, a regulative ideal which can improve trial proceedings. In this spirit, setting a probabilistic criterion for criminal convictions would only be a way to theorize about the meaning and function of the criminal standard of proof. REFERENCE HERE.

The second difficulty is that it is not clear where, exactly, the threshold should be placed: is it 0.99, 0.89, 0.899, 0.999, or what? The difficult here is not so much with identifying a precise numerical value. After all, a range of acceptable values instead of a precise threshold could still be used. The problem here is rather, where to place the bar, whether the bar corresponds to a precise value or an interval of acceptable values? IN response to this question, 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

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

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 we consider that the

disutility of convicting an innocent is as harmful as the disutility of acquitting an innocent, i.e., $D_g = D_i$—as it might be the case in a civil case—, the lower bound for P_g must be at least $\frac{1}{2}$. If, instead,we think that $\frac{D_i}{D_g} = \frac{9}{1}$—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.

MENTION PROOF PARADOXES HERE.

Scenario-based account 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 (Cohen, 1977; Nesson, 1979; Thomson, 1986; Walton, 2002; Stein, 2005; Pardo and Allen, 2008b; Ho, 2008; Haack, 2011). The key notion here seems to be plausibility rather than probability. Ronald Allen (2010) 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.

Argumentation-based account 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 argu-

ment 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.

7 SUMMARY AND CONCLUSION

REFERENCES

Ronald J. Allen. No plausible alternative to a plausible story of guilt as the rule of decision in criminal cases. 2010.

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

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

John H. Beisner. Discovering a better way: The need for effective civil litigation reform. *Duke Law Journal*, 60:547–596, 2010.

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

John Buckleton. A framework for interpreting evidence. In John Buckleton, Christopher M. Triggs, and Simon J. Walsh, editors, *Forensic DNA Evidence Interpretation*. CRC Press, 2005a.

John Buckleton. Population genetic models. In John Buckleton, Christopher M. Triggs, and Simon J. Walsh, editors, *Forensic DNA Evidence Interpretation*. CRC Press, 2005b.

Jonathan L. Cohen. *The Probable and the Provable*. Oxford University Press, 1977.

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

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

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.

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

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

Hock Lai Ho. *Philosophy of Evidence Law*. Oxford University Press, 2008.

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

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

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

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

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

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

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

Michael S. Pardo and Ronald J. Allen. Judicial proof and the best explanation. *Law and Philosophy*, 27(223-268), 2008b.

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.

Mike Redmayne. The relevance of bad character. *Cambridge Law Journal*, 61(3):684–714, 2002.

Bernard 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.

Alex 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.

William C. Thompson and Edward 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.

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

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

Peter Tillers and Jonathan Gottfried. A collateral attack on the legal maxim that proof beyond a reasonable doubt is unquantifiable. *Law, Probability and Risk*, 5:135–157, 2007.

Petr Tillers. Trial by mathematics – reconsidered. 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.

Douglas N. Walton. *Legal Argumentation and Evidence*. Penn State University Press, 2002.

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

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.

John Henry Wigmore. *The Principles of Judicial Proof (as Given by Logic, Psychology, and General Experience and Illustrated in Judicial Trials)*. Little, Brown, and Company, 1913.

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