
EVIDENTIAL REASONING

Chapter for the Handbook of Legal Reasoning

Marcello Di Bello & Bart Verheij – September 17, 2015

CONTENTS

1	Setting the stage	3
1.1	Eyewitness testimony	3
1.2	DNA profiling	4
1.3	Central questions	5
1.4	Three normative frameworks	6
1.4.1	Arguments	6
1.4.2	Probabilities	6
1.4.3	Scenarios	6
1.5	Paper plan	7
2	Conflicting evidence	7
2.1	Arguments	8
2.2	Scenarios	9
2.3	Probabilities	10
3	Evidential strength	13
3.1	Probability	13
3.2	Arguments	20
3.3	Scenarios	20
4	Coherently interpreting the evidence	20
4.1	Scenarios	20
4.2	Probability	22

5	Collection of the evidence and exclusionary rules	23
5.1	Compulsory process	23
5.2	Exclusionary rules	23
6	When can we convict?	25
7	Summary and conclusion	27

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 and structure the exposition that follows.

1 SETTING THE STAGE

Fact-finders, the jurors, judges, or both, aim to reconstruct what has happened in the crime on the basis of the evidence. We will use two 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.

Eyewitness testimony can provide detailed information about what has happened on the scene of the crime. Here is an example.

¹Source: CBS, the Dutch central bureau of statistics, publishing its data at www.cbs.nl.

²SOURCE TO BE ADDED

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 it really is. This is done by counting the number of occurrences of each allele at each core 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

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

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 is that there are no dependencies among the alleles at different loci. 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 births⁵ are not all that rare—, or a lab error.

1.3 Central questions

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

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

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 testi-

⁴Some 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.

⁵Source: <https://en.wikipedia.org/wiki/Twin#Statistics>.

monies 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

1.4.1 Arguments

Argument diagrams

Wigmore

Pollock

1.4.2 Probabilities

Bayesian approaches

LRs

1.4.3 Scenarios

Scenario comparison

Pennington and Hastie

Anchored narratives (Crombag, Van Koppen, Wagenaar)

1.5 *Paper plan*

A-1

B-2

C-3

(3x3 matrix with most on the diagonal)

D

E

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 DNA evidence that the genetic material found at the crime scene does not match the defendant's. When two pieces of evidence lead to contradictory conclusions, 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, not a monological, 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) Arguments are well suited to represent the dialectical aspect of reasoning in court. They are typically defined as a collection of statements in which one statement is the conclusion and the others are the premises functioning as evidence for the conclusion. But this definition is incomplete, for arguments can be better understood in the context of ongoing disagreements between two or more parties. Consider the collection of statements {I am getting wet; it is raining }. Which is the premise? Which is the conclusion? If the weather condition is at issue, getting wet functions as evidence that it is raining. If, instead, one's physical condition is at issue, the fact that it is raining functions as evidence 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 (REFERENCE) has distinguished two such ways.

Rebutting: Arg2 leads to a different conclusion from Arg1 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. When two arguments support contradictory conclusions, they conflict with one another in the manner of *rebutting*, to use Pollock's terminology. 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 could use the DNA match as a premise for the conclusion that Adam raped the victim. But, in the case we are considering, Adam had an alibi: his girlfriend claimed he was with her during the time of crime. So, the defense used Adam's alibi as a premise for the contradictory conclusion that Adam had nothing to do with the crime. This is an example of rebutting, for the two conclusions cannot be concurrently maintained.

Undercutting: Arg2 attacks the relation between Premises and Conclusion in Arg1 Argument can also conflict with one another in the manner of *undercutting*, to use again Pollock's terminology. Suppose A offers an argument for X consisting in premises P1, P2, ... leading to conclusion X, and B shows that the premises used by A in support of X do not actually support X, or at least, not as strongly as A thought. Here is a legal example. The

expert witness for the prosecutor asserts that the defendant's DNA matches with the crime scene DNA. This is a premise supporting the conclusion that the defendant visited the crime scene at some point. But, now suppose the defense shows that the defendant had a twin brother or that the genetic profile used to declare the DNA match is shared by several thousand people. The new information undercuts, or at least weakens, the support that the DNA match lends to the conclusion that the defendant was around the crime scene. Other individuals, after all, might have visited the crime scene instead of the defendant himself.

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, ..., while B shows that one of the premises is false. Here is a legal example. The prosecutor argues for the conclusion that the defendant was at the crime scene on the basis of the premise that the defendant's DNA matches the crime scene DNA. The defense, on the other hand, points out that a laboratory error occurred, thereby suggesting that one of the premises in the prosecutor's argument, namely the DNA match, is false. This is a case of undermining because one of the premises in the proposed argument is shown to be false.

Wigmore charts John Wigmore has devised a systematic method to chart arguments by identifying the various pieces of evidence and the relations of support and attack. ILLUSTRATE REBUTTING, UNDERCUTTING AND UNDERMINING WITH WIGMORE CHARTS. But, despite their clarity and precision, Wigmore's charts have never been particularly popular among lawyers and practitioners. Part of the reason for this might be that charting arguments can often seem unnatural and needlessly complex. This becomes especially clear if we use argument charts to describe legal cases in their entirety. In a court of law, after all, the two parties in a trial aim to establish various conclusions by proposing various arguments, but the ultimate conclusion at issue is whether the defendant is guilty. The problem is that arguments aimed to establish the defendant's guilt can be extremely intricate, laborious and complex. Charting them might not be the best way to grasp a legal case as a whole. SHOW A VERY COMPLEX WIGMORE CHARTS TO MAKE THE POINT THAT THEY ARE OFTEN UNREADABLE AND NOT INTUITIVE.

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

2.2 Scenarios

Mutually inconsistent, different hypotheses/scenarios Instead of viewing evidential reasoning as an intricate process of argument construction, there is a more natural represen-

tation. Studies in psychology and cognitive science (REFERENCES) shows that we tend to comprehend a legal case and the evidence presented therein by constructing comprehensive scenarios (or stories, narratives). On this perspective, each party puts forward a comprehensive scenario that comprises a reasonably detailed and complete reconstruction of how the defendant got involved and carried out the crime. The scenario, of course, cannot be invented but should be backed up by good supporting evidence. In criminal trials, since the burden of proof is on the prosecutor, we should expect the prosecutor to put forward a comprehensive case that establishes beyond a reasonable doubt the defendant's guilt. (On proof beyond a reasonable doubt, see SECTION X). The defense, in turn, may respond with an alternative scenario, also backed up by good supporting evidence. When two alternative scenarios are proposed, this resembles a case of rebutting. The difference is that rebutting sometimes concerns small-scale arguments with small-scale conclusions, while the conflict between scenarios concerns the entire prosecutor's and defense's cases. It bears emphasizing that the defense is not legally required to offer a full-fledged alternative scenario because the burden of proof is on the prosecutor and never shifts to the defendant. This does not mean, however, that the defense can simply be inherit. If it did not offer any response to the prosecutor's case, this would most likely qualify as an ineffective defense, and constitutional guarantees in many countries grant defendants the right to an effective defense.⁶ The defense, then, must at least attempt to identify weaknesses in the prosecutor's case and raise reasonable doubts. ILLUSTRATE HOW THIS GOES.

Conflicting scenarios can offer alternative explanations for the evidence

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 by saying that E is evidence favoring hypothesis H whenever E raises the probability of H , or more precisely, whenever taking into consideration E makes it more likely that H is true than it would be without taking into consideration E . (REFERENCE TO FEDERAL RULE OF EVIDENCE 403 AND DEFINITION OF RELEVANCE) Note that the fact alone that the probability of H given E is high does not mean—according to the characterization of 'evidential favoring' just given—that E is evidence for H . There should be a probability shift upwards for E to be evidence favoring hypothesis H .

⁶US CONSTITUTION. RIGHT TO EFFECTIVE ASSISTANCE OF COUNSEL.

Different, incompatible outcomes In a criminal trial, the prosecutor and the defendant will put forward conflicting hypotheses. These can be as complex as the final proposition ‘guilty’ or ‘not-guilty’, but also more circumscribed such as ‘the defendant had contact with the victim’ or ‘the defendant was away from the crime when the crime occurred’. Recall the British case *R v Adams*. Adam’s DNA matched the traces of semen found on the victim’s body, but Adam also had an alibi provided by his girlfriend. How can we represent probabilistically the conflict between the DNA match and Adam’s alibi? For one, we should identify the two conflicting hypotheses. Let CV stand for the hypothesis that Adam had contact with the victim and AW stand for the hypothesis that Adam was away from the scene of the crime when it was committed. The hypotheses CV and AW are incompatible—after all, if the defendant was away from the crime scene, he could not have had contact with the victim, at least, not as the crime occurred. The next step is to identify the pieces of evidence of interest. Let M stand for the proposition that the crime scene DNA matches Adam’s DNA and A stand for Adam’s alibi. Note that M raises the probability of CV , while A raises the probability of AW . If Adam’s DNA matches the crime scene DNA, it is more probable that he visited the crime scene than if the DNA did not match. Similarly, if Adam has an alibi, it is more probable that he was away from the crime than if he did not have an alibi. So, M is evidence for CV because $P(CV|M) > P(CV)$ and A is evidence for AW because $P(AW|A) > P(AW)$. Pieces of M and A are in conflict because they favor incompatible hypotheses. More generally, two pieces of evidence $E1$ and $E2$ conflict with one another whenever they favor incompatible hypotheses $H1$ and $H2$, as follows:

CONFLICT (I): $P(H1|E1) > P(H1)$ and $P(H2|E2) > P(H2)$ with $H1$ and $H2$ incompatible.

This description is not much different from a case of rebuttal in which two arguments conflict with one another because their conclusions are incompatible. Instead of contradictory or incompatible conclusions, here we speak of contradictory or incompatible hypotheses.

Same hypothesis can receive different probability assignments depending on evidence (or interpretation thereof) We can represent—probabilistically—a conflict between two pieces evidence also in a different way. As discussed earlier, a piece of evidence favors a hypothesis whenever it raises the probability of the hypothesis. If, on the other hand, the probability of the hypothesis shifts downwards as we take into account the evidence, then the evidence disfavors the hypothesis. In short:

$$E \text{ favors } H \text{ iff } P(H|E) > P(H);$$

$$E \text{ disfavors } H \text{ iff } P(H|E) < P(H).$$

Let's now consider the hypothesis CV that Adam had contact with the victim (as the crime occurred). The probability of CV raises as we take into account the DNA evidence match M , but the probability of CV decreases as we take into account the Adam's alibi A . This, at least, is a plausible stipulation. In this sense, then, we can say that the two pieces of evidence M and A conflict with one another because one favors and other disfavors CV . More generally, two pieces of evidence $E1$ and $E2$ conflict with one another whenever one favors and the other disfavors the same hypothesis H , as follows:

$$\text{CONFLICT (II): } P(H|E1) > P(H) \text{ and } P(H|E2) < P(H).$$

What is the relation between CONFLICT (I) and CONFLICT (II)? It is easy to see that CONFLICT (II) entails CONFLICT (I) because if $P(H|E2) < P(H)$, then $P(\text{not-}H|E2) > P(\text{not-}H)$. WHAT ABOUT THE OTHER DIRECTION?

Relative to the same hypothesis, likelihood ratios for different pieces of evidence can be positive or negative. We can describe the relation of evidential favoring and disfavoring by means of likelihood ratios, as well. If E raises the probability of H , this is equivalent to a likelihood ratio $\frac{P(E|H)}{P(E|\neg H)}$ greater than 1. If E lowers the probability of H , this is equivalent to a likelihood ratio $\frac{P(E|H)}{P(E|\neg H)}$ smaller than 1. In short:

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

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

If LR for E1 relative to H is positive, while LR for E2 relative to H is negative, E1 and E2 are in conflict CONFLICT (I) can be suitably rewritten in terms of likelihood ratios, as follows:

$$\text{CONFLICT (I): } LR1 = \frac{P(E1|H1)}{P(E1|\neg H1)} > 1 \text{ and } LR2 = \frac{P(E2|H2)}{P(E2|\neg H2)} > 1,$$

where $H1$ and $H2$ are incompatible hypotheses.

The idea here is, once again, that two pieces of evidence are in conflict with one another when they favor incompatible hypotheses. Similarly, CONFLICT (II) can also be suitably rewritten in terms of likelihood ratios, as follows:

$$\text{CONFLICT (II): } LR1 = \frac{P(E1|H)}{P(E1|\neg H)} > 1 \text{ and } LR2 = \frac{P(E2|H)}{P(E2|\neg H)} < 1.$$

The idea here is that two pieces of evidence are in conflict with one another when one favors and the other disfavor the same hypothesis.

Comparison between the three frameworks

3 EVIDENTIAL STRENGTH

3.1 Probability

Likelihood ratios can model evidential strength vs “posterior probability” (difference between the two) So far we have only represented—probabilistically—the fact that a piece of evidence favors (or disfavors) a certain hypothesis. The probabilistic framework allows us to quantify the strength of this support relation between evidence and hypothesis. There are different approaches in the literature (Fitelson, 2006). Here are a couple. We can consider the difference $P(H|E) - P(H)$. The larger the positive difference, the stronger the support which E offers toward H . Alternatively, we can consider the likelihood ratio $\frac{P(E|H)}{P(E|\neg H)}$. The higher the positive likelihood ratio, the stronger the support which E offers toward H . Since it is common to speak of evidential strength in terms of likelihood ratios, we shall do the same here.

MEASURING EVIDENTIAL STRENGTH: The evidential strength of E relative to H is proportional to the likelihood ratio $\frac{P(H|E)}{P(H|\neg E)}$.

The likelihood ratios is a good measure of evidential strength because it tells how much a piece of evidence E can impact the initial probability of a hypothesis H . Evidential strength of a piece of is not measured in terms of how high the probability of a certain hypothesis is given the evidence. If the probability of H given E is high, this does not mean that E strongly favors H .

This allows us to address a paradox that has been formulated against a probabilistic account of evidential strength (REFERENCE TO COHEN). Suppose the probability of H given E is low because, say, the evidence is meager. For example, the probability that it will rain tomorrow given that the stock market crashed seems rather low. If so, according to the probability calculus, the probability that it will *not* rain tomorrow given that the stock market crashed must be rather high. But how can that be? How can meager or irrelevant evidence make a certain hypothesis—in our case, that it will not rain tomorrow—highly probable? L.J. Cohen summarized the paradox here by saying that “the probability calculus creates evidence from ignorance” (REFERENCE). This paradox can be address by noting that evidential strength is not a function of the conditional probability of H given E , but a function of the likelihood ratio $\frac{P(E|H)}{P(E|\neg H)}$ or of the difference $P(H|E) - P(H)$. Now, learning about the stock market crash does not raise or lowers the probability that it will

rain or not rain tomorrow, so the difference $P(H|E) - P(H)$ must be roughly zero and the likelihood ratio must be roughly 1. This means that the information that the stock market crashed has no evidential strength relative to the hypothesis that it will rain (or not) tomorrow. The probability that it will rain or that it will not rain tomorrow, given that the stock market crashed, is whether probability the event of rain has regardless of considering the fact that the stock market crashed. The probability of rain, or not rain, tomorrow could be high or low. Whatever its value, it will be independent from the stock market crashing. The paradox, then, seems merely an artifact, but it is instructive to illustrate the difference between likelihood ratio and the conditional probability of an hypothesis given the evidence (sometimes also called the “posterior probability” because it is the probability resulting from taking into account a new item of evidence). It is certainly paradoxical to measure evidential strength in terms of mere conditional probability. The likelihood account must be adopted instead.

Illustration In the British case *R v Adams*, Adam’s DNA matched the traces of semen found on the victim’s body. Earlier we analyzed this probabilistically by saying that the DNA match favors the hypothesis that Adam had contact with the victim. The advantage of the probability-based framework is the analyst can be much more refined than that. As known, a DNA evidence is always accompanied by a probabilistic estimate, and in the Adam’s case, the DNA lab analyst testified that the probability that a random person would match the crime traces was estimated to be 1 in 200 million. This value can be incorporated in the likelihood ratio as a measure of the strength of the evidential strength of the DNA match. Let CV stand for the hypothesis that Adam had contact with the victim and M stand for the proposition that the crime scene DNA matches Adam’s DNA. The likelihood ratio for the DNA match, relative to hypothesis CV — can be described as follows:

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

The idea behind this assignment is as follows. If the Adam has in fact contact with the victim, we can assume that there the match would be found, so $P(M|CV) = 1$, and if instead Adam did not have contact with the victim, there would still be a 1 in 200 million chance that the DNA analysis would show match. This is because the probability that a random person—that is, someone who had nothing to do with the crime—would match is 1 in 200 million, so $P(M|\neg CV)$ is 1 in 200 million. The likelihood ratio $\frac{P(M|CV)}{P(M|\neg CV)}$ is therefore 200 million. This is a rather high value and therefore the strength of the evidence is high, at least relative to the hypothesis that Adam has contact with the victim.

Confusions may arise between likelihood (ratios) and posterior probability (i.e. prosecutor's fallacy) Given that the likelihood ratio is such a high number, the DNA match relative to hypothesis S has a great deal of evidential strength. But, we should not be fooled by the high likelihood ratio. 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. The inference is a confusion of $P(M|\neg CV)$, which is indeed low, with $P(\neg CV|M)$, which need not be low at all. This confusion is known in the literature as the *inversion fallacy* or the *prosecutor's fallacy* (Thompson and Shumann, 1987). Bayes' theorem can protect us from this fallacy. By applying the theorem, with different prior odds and a $LR=100,000,0000$, the posterior odds are as follows:

Prior Odds	Likelihood Ratios	Posterior Odds	$P(S M)$
1: 10 million	100 million	10	0.9
1: 100 million	100 million	1	0.5
1: 1 billion	100 million	0.1	0.09

Likelihood ratios can model sources of weakness/error/uncertainty (illustrate this through DNA evidence) The analyzes of the strength of a DNA match can be more refined than what we have done so far. Consider a generic criminal case in which there is a DNA evidence against the defendant. Laboratory analyses show that the defendant matches with the genetic traces found at the crime scene. Call this piece of evidence M . Let us suppose the prosecutor's hypothesis is that the defendant is the source of the genetic material found at the crime; call this hypothesis S . (Note that hypothesis S is not quite the same as the hypothesis that the defendant is guilty. The fact that the defendant is the source of the crime scene genetic material need not entail that he is guilty. He could have left the genetic traces accidentally or in a way consistent with his innocence.) What should we conclude about S given M ? How are we to assess the evidential strength of M relative to S ? As said at the beginning, we can identify three major sources of uncertainty while drawing inferences from a DNA evidence match:

1. an accidental match
2. a laboratory error
3. a twin brother or a close relative with the same DNA as the defendant

We shall now see how likelihood ratios can represent these three uncertainties and assess the evidential strength of a DNA evidence match accordingly.

Let's start with the possibility of a *coincidental match*. Suppose the defendant matches with the crime scene DNA; there was no laboratory error; and no close relative had the same DNA. Even under these assumptions, the match could still be a pure coincidence.

This is why DNA evidence matches typically come with a Random Match Probability. The RMP expresses the probability that a random person would show a match, so a RMP conveys information about the possibility of a purely coincidental match. The RMP is often an astronomically low value such as 1 in 50 billion. When we have DNA evidence matches, it is therefore very unlikely that a random person would show a match. The probability of a coincidental match, in other words, is typically very low.

If we set aside the other two sources of uncertainty, the evidential strength of a DNA evidence match, whose Random Match Probability equals a probability p , is as follows:

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

We put $P(M|S) = 1$, that is, given that the defendant is in fact the source, the probability of finding a DNA match for him is one. Intuitively, that is true. If the defendant actually left the traces, we would expect him to match with the traces. Second, we put $P(M|\neg S) = p$. That is intuitively true, as well. The probability p is the value of the RMP, which represents the probability that someones picked at random, namely someone who is *not* the source of the crime scene DNA, would match. Suppose, in our example, the RMP is estimated to be 1 in 100,000,000. The likelihood ratio would be as follows:

$$\frac{P(M|S)}{P(M|\neg S)} = \frac{1}{\frac{1}{100,000,000}} = 100,000,000.$$

Beside a coincidental match, another source of uncertainty with a DNA evidence match is the possibility of a *laboratory error*. If the RMP equals 1 in 100,000,000, assigning a likelihood ratio of 100,000,000 rests on two simplifications. For one, $P(\text{match}|\neg S) = \frac{1}{100,000,000}$, or more generally $P(M|\neg S) = p$. Intuitively, given a RMP=100,000,000 the probability that someone who is not the source would show as a match is 1 in 100,000,000. This is a simplification because the defendant might *not* possess the DNA profile in question, yet laboratory analyses could show that he does (what is called a *false positive*). To take false positives 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:

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

Now, $P(M_t|\neg S)$ equals p , and thus $P(\neg M_t|\neg S)$ equals $1-p$. 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

⁷This simple analysis can be found, among others, in (Dawid, 2002) and (Balding, 2005).

⁸For an explanation, see (Thompson et al., 2003).

FN . So, we have:

$$P(M|\neg S) = [(1 - FN) \times p] + [FP \times (1 - p)].^9$$

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).

The second simplification is $P(M|S) = 1$, that is, given that the defendant is in fact the source, the probability of finding a DNA match for him is one. This is a simplification because the laboratory analyses might fail to show a match even though the defendant did leave the traces (what is called a *false negative*). In our new notation, this second simplification can be stated as $P(M_r|S) = 1$. To take the possibility of a false negative into account, we write:

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

In the equation, $P(M_t|S) = 1$ and $P(\neg M_t|S) = 0$. Thus, $P(M_r|S) = 1 - FN$.

Putting everything together, a formulation of the likelihood ratio that takes into account laboratory errors is as follows:

$$\frac{P(M_r|S)}{P(M_r|\neg S)} = \frac{1 - FN}{[(1 - FN) \times p] + [FP \times (1 - p)]}.$$

Laboratory error rates do impact the likelihood ratio quite dramatically. 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 $1/p$, then it would be 1 billion. With the new formulation, we have:

$$\frac{P(M_r|S)}{P(M_r|\neg S)} = \frac{1 - 0.01}{[(1 - 0.01) \times 0.000000001] + [0.01 \times (1 - 0.000000001)]} \approx 100.^{10}$$

Relatively small error rates can therefore significantly diminish the evidential strength of a DNA evidence match by dramatically reducing the likelihood ratio. In the example, we went from one billion to only one hundred.¹¹ The impact of laboratory error rates is an indication that no matter how low the RMP turns out to be, it will be of relatively little significance when it is unaccompanied by an estimate of the laboratory error rates. But

⁹See (Thompson et al., 2003) and (Buckleton, 2005a).

¹⁰ The calculation is as follows:

$$\begin{aligned} &= \frac{1-0.01}{[(1-0.01) \times 0.000000001] + [0.01 \times (1-0.000000001)]} \\ &= \frac{0.99}{0.99 \times 0.000000001 + [0.01 \times (0.999999999)]} \\ &= \frac{0.99}{0.00000000099 + 0.00999999999} \\ &= \frac{0.99}{0.01000000098} \approx 98 \approx 100 \end{aligned}$$

¹¹For a amore systematic discussion of this point, see Thompson et al. (2003) who show how even a small margin of laboratory error can substantially diminish the probative value of DNA evidence.

the important point here is that the likelihood ratio, as a measure of the strength of the evidence, allows us to take this complication into account.

There is a third source of uncertainty in assessing the evidential strength of a DNA evidence match. There could be a *twin brother* or a *close relative* who share the same DNA as the defendant. After all, 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 being 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*. The RMP probability varies depending on the population of interest, Caucasians, African-Americans, Asians etc. In a court of law, different RMPs are often presented, each relative to a certain reference population. There might be variations across different reference populations such as Caucasians, African-Americans, etc. Importantly, a low RMP will be significantly higher if the reference population is that of the defendant's close relatives. How can we represent this variety of reference populations and their impact on the RMP?

HOW CAN WE REPRESENT THIS WITH LR? SUBPOPULATION EFFECTS?

All in all, likelihood ratios are a promising way for assessing the evidential strength of a DNA match relative to a hypothesis of interest, typically the hypothesis that the defendant is the source of the traces. Complications aside, if the underlying Random Match Probability is 1 in 10,000,000, the likelihood ratio, as a measure of evidential strength, is 10,000,000. If the RMP is 1 in 1,000,000, the evidential strength of the DNA match would be much lower. The evidential strengths of the two DNA matches can be easily compared. As seen earlier, likelihood ratios can also incorporate more subtle information such as laboratory error rates.¹²

The likelihood ratio analysis 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

¹²Furthermore, 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 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 be 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).

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 do need an estimate of the Random Match Probability which we often lack.

This can be visualized by means of Bayes' theorem:

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

The formulation in terms of prior and posterior odds might appear cumbersome at first, but it is the easiest formulation as far as calculations are concerned. When the hypotheses in the likelihood ratio are one the negation of the other—such as H and $\neg H$ —there is a formula that allows us to derive the posterior probability of a hypothesis H given the new evidence E from the posterior odds PO , as follows:

$$\text{posterior probability of } H \text{ given } E = \frac{PO}{1 + PO}.$$

The table below shows how a likelihood ratio of 1,000 can impact the prior odds and lead to different values of the posterior odds and the posterior probability of the hypothesis of interest.

Prior Odds	LR	Posterior Odds	$P(H E)$
1:10,000	1,000	1:10	0.09
1: 1,000	1,000	1:1	0.5
1: 100	1,000	10:1	0.909
1: 1	1,000	1000:1	0.999

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

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

ASIDE: LJ Cohen's challenge (negation, conjunction paradox) and the Bayesian response

[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)

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.

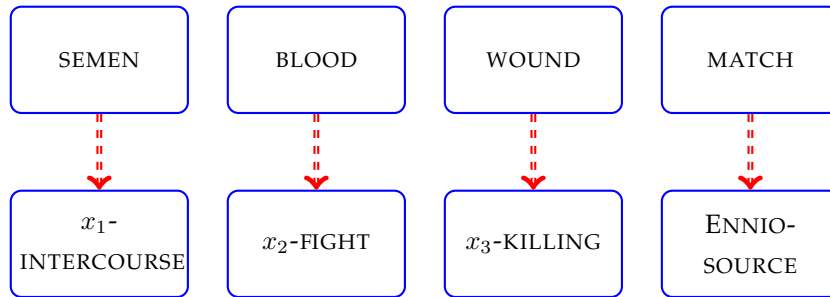
Imagine you are the prosecutor and you are constructing an incriminating case against Ennio. To a first rough approximation, four basic pieces of evidence are available:

- (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:

- (*x₁-intercourse*) individual *x₁* had intercourse with victim;
- (*x₂-fight*) individual *x₂* had violent fight with victim in parking lot;
- (*x₃-killing*) individual *x₃* 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 a *unification hypothesis*, as follows:

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

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 million.

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

Scenario schemes (Schank, Bex)

Arguments

Inference to the best explanation (Allen/Pardo?)

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

4.2 Probability

Hierarchy of propositions (Evetts et al)

Combining different pieces of evidence using probability

Multiplying likelihood ratios and question of independence among different pieces of evidence We first need a general account of how to model the *combination* of two pieces of evidence E_1 and E_2 , each supporting, with its own degree of strength, the same hypothesis H . If the strength of a single piece 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 E_1 and E_2 impacts the probability $P(H)$, as follows:

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

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 COLLECTION OF THE EVIDENCE AND EXCLUSIONARY RULES

5.1 *Compulsory process*

5.2 *Exclusionary rules*

In general, any relevant evidence is admissible unless it is deemed inadmissible on the basis of specific rules of exclusion of which we shall see some examples below.

All relevant evidence is admissible, except as otherwise provided by the Constitution of the United States, by Act of Congress, by these rules, or by other rules prescribed by the Supreme Court pursuant to statutory authority. Evidence which is not relevant is not admissible. (F.R.E. 402)

To determine admissibility, it is sufficient to determine whether a piece of evidence is relevant, and whether it is deemed inadmissible by some exclusionary rule found in the F.R.E., the U.S. Constitution, an Act of congress, or rules prescribed by the U.S. Supreme Court.

Rule 401 of the Federal Rules of Evidence (F.R.E., hereafter) defines relevant evidence as follows:

Relevant evidence means evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence. (F.R.E. 401)

It is customary to distinguish two dimensions operating in this definition: materiality and probative value (Fisher, 2008; Méndez, 2008). A piece of evidence E is probative of a fact F , provided E makes F more (or less) probable than it would be in absence of E . On the other hand, a fact F is material for the action or crime C under examination if it is of consequence to the determination of C . Whether or not a fact is of consequence to the determination of C depends on the substantive law which determines what facts should be proven in order to establish the crime.¹³ By combining materiality and probative value, it follows that a piece of evidence E is relevant to crime C if and only if it is probative of F , where F is material for C .

Once the question of relevance is settled, admissibility becomes a strictly legal question. It depends on the law of the particular jurisdiction. For instance, in federal cases, F.R.E. 403 establishes that:

Although relevant, evidence may be excluded if its probative value is substantially outweighed by the danger of unfair prejudice, confusion of the issues, or misleading the jury, or by considerations of undue delay, waste of time, or needless presentation of cumulative evidence. (F.R.E. 403)

Rule 403 requires the judge to perform a balancing test between the probative value of the evidence and certain negative or detrimental effects (prejudice, confusions, waste of time, etc.) that are more or less likely to occur if the contested piece of evidence is introduced. Rule 403 is a curious rule, and its standard of application is far from being well defined. For instance, what are the items that are intended to be balanced? The degree of probative value with the likelihood of prejudice, confusions, waste of time, etc.? Or, the degree of probative value with the degree of “detrimental seriousness” associated with prejudice, confusion, waste of time? Or does the balancing include all three variables? One way to interpret the rule is to say that it requires the judge to compare and balance a purely epistemic value—i.e. the probative value of a piece of evidence—with other non-epistemic values or dis-values, such as prejudice, confusion, and waste of time.¹⁴ The rule states that probative value is not always enough to guarantee admissibility, and hence a balancing test between epistemic and non-epistemic dimensions must be performed by the judge. The balancing-test of the two dimensions assumes that they can be compared using some common metric. Some have doubted that any balancing test between the two dimensions is possible because it would be like comparing apples and oranges, as it were; see Taruffo (2009). On the other hand, balancing tests are ubiquitous in the law, especially to reconcile conflicting values and goals. For instance, in 4th Amendment law about (un)reasonable searches and seizures two competing goals must be reconciled. One goal is the state’s

¹³This is a simplification, however. More precisely, a fact can be material if (a) it is an ultimate issue (one that the governing law says should be proven to establish the crime charged), or (b) it is an intermediary fact, or (c) it is an evidentiary fact. The governing law establishes materiality only for ultimate issues, not for the intermediary and evidentiary facts.

¹⁴To be sure, prejudice and confusion might be epistemic (dis)values.

need to collect evidence and the other is the citizen's right not to have their privacy rights violated. The Supreme Court solution, in some cases, is to adopt a balancing-test. It is an interesting question whether the "balancing strategy" is the only, or the most appropriate strategy to undertake in case of a value conflict.

Another exclusionary rule contained in the F.R.E. is rule 404(b), which establishes:

Evidence of other crimes, wrongs, or acts is not admissible to prove the character of a person in order to show action in conformity therewith. (F.R.E., 404(b))

Character evidence concerns the previous conduct of the defendant, not the specific conduct under examination in the trial. Interestingly, some have recently argued for the relevance and admissibility of character evidence (Redmayne, 2002).

HEARSAY EXCLUSIONS

6 WHEN CAN WE CONVICT?

What is the meaning of the criminal standard of proof? What does it take to establish guilt beyond a reasonable doubt? When is a doubt reasonable or unreasonable? 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.' These paraphrases might not be that unhelpful, and if so, the U.S. Supreme Court is quite right when it says 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). Nevertheless, we still have a grasp of what the criminal standard is supposed to *do*, though we cannot articulate what it is supposed to *mean*. The Supreme Court of Canada in *R. v. Lifchus* (1997) listed some platitudes about proof beyond a reasonable doubt. They are worth stating:

- (-) the standard of proof beyond a reasonable doubt is inextricably intertwined with that principle fundamental to all criminal trials, the presumption of innocence;
- (-) 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 did not only explicate when a doubt is 'reasonable,' but it also 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.

In the existing literature, one prominent 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). This definition is simple, crisp and elegant, but a too literal interpretation of it is obviously problematic. Consequently, a doubt would be reasonable or unreasonable depending on a measurable probability. 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. First, it is not clear where, exactly, the threshold should be placed: is it 0.99, 0.89, 0.899, 0.999, or what? A second difficulty is that assigning a probability value to guilt itself might not be feasible. The second difficulty 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.

As for the first difficulty—how to identify the threshold—David Kaplan used a relative assessment of the disutilities associated with convicting an innocent, D_i , and with acquitting a guilty defendant, D_g . Now, suppose that 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. But the inequality holds only if P_g reaches a certain value. From the above 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

Some scholars suggested that we should elaborate a theory of legal reasoning which departs from the probabilistic approach and which does not ignore trial procedures broadly construed (Cohen, 1977; Nesson, 1979; Thomson, 1986; Walton, 2002; Stein, 2005; Pardo and Allen, 2008; 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 under-defined. Instead, the theory of probability is well-established and fully worked out mathematically. We should make sure, then, that in abandoning the probabilistic framework, we do not embark into a ill-conceived enterprise. On the other hand, while the notion of plausibility is hard to define precisely, it seems closer to how jurors actually reason in trial proceedings. In contrast, 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.

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.
- David J. Balding. *Weight-of-evidence for forensic DNA profiles*. John Wiley and Sons, 2005.
- 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.
- Philip Dawid. Bayes's theorem and weighing evidence by juries. In *Bayes's Theorem*, volume 113, pages 71–90. Oxford University Press, 2002.
- George Fisher. *Evidence (2nd Edition)*. Foundation Press, 2008.
- Branden Fitelson. Likelihoodism, Bayesianism, and relational confirmation. *Synthese*, 16(47):1–22, 2006.
- 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.
- David H. Kaye. DNA evidence: Probability, population genetics and the courts. *Harvard Journal of Law and Technology*, 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.
- Michael S. Pardo and Ronald J. Allen. Judicial proof and the best explanation. *Law and Philosophy*, 27(223–268), 2008.
- 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.
- Alex Stein. *Foundations of Evidence Law*. Oxford University Press, 2005.
- Michele Taruffo. *La semplice verità*. Laterza, 2009.
- 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.

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.

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