Burdens of Proof - Sample Chapter

Marcello Di Bello and Rafal Urbaniak

SAMPLE CHAPTER PLAN

In rethinking the sample chapter, we should perhaps stick to a simpler structure, trying to offer a more focused and compelling argument. Right now I think we have too many possible accounts under consideration, and the structure is not very tight or cohesive. It feels more like a literature review, especially the first few sections.

So here is how I proposed we do it:

1. Begin by stating the simplest probabilistic account based on a threshold for the posterior probability of guilt/liability. The threshold can be variable or not. Add brief description of decision-theoretic ways to fix the threshold. (Perhaps here we can also talk about intervals of posterior probabilities or imprecise probabilities.)
2. Formulate two common theoretical difficulties against ths posterior probability threshold view: (a) naked statistical evidence and (b) conjuction. (We should state these difficilties before we get into alternative probabilistic accounts, or else the reader might wonder why so many different variants are offerred of probabilistic accounts).

   R: Yes. That's what I thought.

   We might also want to add a third difficulty: (c) the problem of priors (if priors cannot be agreed upon then the posterior probability threshold is not functionally operative). Dahlman I think has quite a bit of stuff on the problem of priors.
3. As a first response to the difficulties, articulate the likelihood ratio account. This is the account I favor in my mind paper. Kaplow seems to do something similar. So does Sullivan. So it's a popular view, worth discusing in its own right. You say that Cheng account is one particular variant of this account, so we can talk about Cheng here, as well.
4. Examine how the likelihood ratio account fares against the two/three difficulties above. One could make an argument (not necessarily a correct one) that the likelihood ratio account can address all the two/three difficulties. So we should say why one might think so, even thought the argument will ultimately fail. I think this will help grab the reader's attention. This is what I have in mind:

   4a: the LR approach solves the naked stat problem because LR=1 (Cheng, Sullivan) or L1=unknown (Di Bello).

   4b: the LR approach solves the conjuction problem because – well this is Dawid's point that we will have to make sense of the best we can

   4c: the LR approach solves the priors problem b/c LR do not have priors.
5. Next, poke holes in the likelihood ratio account:

   against 4a: you do not believeLR=1 or LR=unknown , so we should talk about this

   against 4b: this is your cool argument against Dawid

   against 4c: do you believe the arguemt in 4c? we should talk about this

   In general, we will have to talk to see where we stand. As of now, I tentatively believe that the likelihood ratio account can solve (a) and (c), and you seem to disagree with that. Even if I am right, the account is still not good enough becaue it cannot solve (b).
6. Articulate (or just sketch?) a better probabilistic account overall. Use Bayesian networks, narratives, etc. I am not sure if this should be another paper. That will depend on how much we'll have to say here.

Contents

1 Introduction

After the evidence has been presented, examined and cross-examined at trial, trained judges or lay jurors must reach a decision. The decision criterion is defined by law and consists of a standard of proof, also called the burden of persuasion. So long as the evidence against the defendant is sufficiently strong to meet the requisite proof standard, the defendant should be found liable.

In criminal proceedings, the governing standard is 'proof beyond a reasonable doubt.' If the decision makers are persuaded beyond a reasonable doubt that the defendant is guilty, they should convict, or else they should acquit. In civil cases, the standard is typically 'preponderance of the evidence'. The latter is less demanding than the former, so the same body of evidence may be enough to meet the preponderance standard, but not enough to meet the beyond a reasonable doubt standard. A vivid example of this difference is the 1995 trial of O.J. Simpson who was charged with murdering his wife. He was acquitted of the criminal charges, but when the family of the victim brought a lawsuit against him, they prevailed. O.J. Simpson did not kill his wife according to the beyond a reasonable doubt standard, but he did according to the preponderance standard. An intermediate standard, called 'clear and convincing evidence', is sometimes used for civil proceedings in which the decision is particularly weighty, for example, a decision whether someone should be committed to a hospital facility.

This tripartite distinction of proof standards—beyond a reasonable doubt; preponderance; clear and convincing evidence—is common in Anglo-american jurisprudence. It is not universal, however. Different countries may use different standards. France, for example, uses the standard of 'intimate conviction' for both civil and criminal proceedings. Judges deciding cases 'must search their conscience in good faith and silently and thoughtfully ask themselves what impression the evidence given against the accused and the defence's arguments have made upon them' (French Code of Criminal Procedure, art. 353). German law is similar. Germany's Code of Civil Procedure, Sec. 286, states that 'it is for the court to decide, based on its personal conviction, whether a factual claim is indeed true or not.'

How to define standards of proof—and whether they should be even defined in the first place—remains

contentious (Diamond, 1990; Horowitz & Kirkpatrick, 1996; Laudan, 2006; Newman, 1993; Walen, 2015). Judicial opinions offer different paraphrases, sometimes conflicting, of what these standards mean. The meaning of 'proof beyond a reasonable doubt' is the most controversial. It has been equated to 'moral certainty' or 'abiding conviction' (Commonwealth v. Webster, 59 Mass. 295, 320, 1850) or to 'proof of such a convincing character that a reasonable person would not hesitate to rely and act upon it in the most important of his own affairs' (US Federal Jury Practice and Instructions, 12.10, at 354, 4th ed. 1987). But courts have also cautioned that there is no need to define the term because 'jurors know what is reasonable and are quite familiar with the meaning of doubt' and attempts to define it only 'muddy the water' (U.S. v. Glass, 846 F.2d 386, 1988).

We will not survey the extensive legal literature and case law about proof standards. We will examine, instead, whether or not probability theory can bring conceptual clarity to an otherwise heterogeneous legal doctrine. This chapter outlines different probabilistic approaches, formulates the most common challanges against them, and offers a number of responses from the perspective of legal probabilism. The legal and philosophical literature has focused on the theoretical and analytical challanges. We will do the same here. We will focus on two key theoretical challanges that have galvanized the philosophical and legal literature in the last thirty years: the problem of naked statistical evidence and the conjunction paradox. One reason to choose these two in particular is that it would be desirable to be able to handle basic conceptual difficulties before turning to more complex issues or attempting to implement probabilistic standards of proof in reality of trial proceedings.

2 Probability thresholds

Imagine you are a trier of fact, say a judge or a juror, who is expected to make a decision about the guilt of a defedantfacing criminal charges. The defendant denies the accusation. As a trier of fact, you are confronted with the question, does the totality of the evidence presented at trial, all things considered, prove guilt beyond a reasonable doubt?

Legal probabilists have proposed to interpret proof beyond a reasonable doubt as the requirement that the defendant's probability of guilt, given the evidence presented at trial, meet a threshold (see Bernoulli, 1713; Dekay, 1996; Kaplan, 1968; Kaye, 1979; Laplace, 1814; Laudan, 2006). In other words, so long as the guilt of the defendant is established with a sufficiently high probability, say 95%, guilt is proven beyond a reasonable doubt and the defendant should be convicted. If the probability of guilt does not reach the requisite threshold, the defendant should be acquitted. This intepretation can be spelled out more formally by means of conditional probabilities. That is, a body of evidence E establishes guilt G beyond a reasonable doubt if and only if $\mathsf{P}(G|E)$ is above a threshold. From this perspective, a conviction is justified whenever guilt is sufficiently probable given the evidence.

This interpretation is, in many respects, plausible. From a legal standpoint, the requirement that guilt be established with high probability, still short of 100%, accords with the principle that proof beyond a reasonable doubt is the most stringent standard but does not require—as the Supreme Court of Canada put it—'proof to an absolute certainty' and thus 'it is not proof beyond any doubt' (R v Lifchus, 1997, 3 SCR 320, 335). The plausibility of a probabilistic intepretation is further attested by the fact that such an intepretation of proof standards is taken for granted in psychological studies about people's understanding of proof beyond a reasonable doubt (Dhami, Lundrigan, & Mueller-Johnson, 2015). This research examines how high lay jurors set the bar—say a 75% or 95% probability threshold—when they decide about criminal charges, but does not question whether the standard functions as a probabilistic threshold of some kind.

Reliance on probability is even more explicit in the standard 'preponderance of the evidence'—also called 'balance of probabilities'—which governs decisions in civil disputes. This standard can be interpreted as the requirement that the plaintiff—the party making the complaint against the defendant in a civil case—establish its version of the facts with greater than 50% probability. The 50% threshold, as opposed to a more stringent threshold of 95% for criminal cases, reflects the fact that preponderance is less demanding than proof beyond a reasonable doubt. The intermediate standard 'clear and convincing evidence' is more stringent than the preponderance standard but not as stringent as the beyond a reasonable doubt standard. Since it lies in between the other two, it can be interpreted as the requirement that the plaintiff establish its versions of the facts with, say, 75-80% probability.

When appellate courts have examined the question whether standards of proof can be quantified using probabilities, they have often answered in the negative. One of the clearest opposition to quantification

was formulated by Germany's Supreme Court, the Federal Court of Justice, in the case of Anna Anderson who claimed to be a descendant of the Tsar family. In 1967, the Regional Court of Hamburg ruled that Anderson failed to present sufficient evidence to establish that she was Grand Duchess Anastasia Nikolayevna, the youngest daughter of Tsar Nicholas II, who allegedly escaped the murder of the Tsar family by the Bolsheviks in 1918. (Incidentally, DNA testing later demonstrated that Anna Anderson had no relationship with the Tsar family.) Anderson appealed to Germany's Federal Court, complaining that the Regional Court had set too demanding a proof standard. Siding with the lower court, the Federal Court made clear that '[t]he law does not presuppose a belief free of all doubts', thus recognizing the inevitable fallibility of trial decisions. The Court warned, however, that it would be 'wrong' to think that a trial decision could rest on 'a probability bordering on certainty' (Federal Court of Justice, February 17, 1970; III ZR 139/67).

This decision is all the more interesting as it applies to a civil case. It semes as though the German court did not think trial decisions could rest on a probability, not even in a civil case. Turnign from civil to criminal cases, Buchak (2014) has argued that an attribution of criminal culpability is an ascription of blame which requires a full belief in someone's guilt. One is left wondering, however. If a high probability of guilt short of 100% isn't enough but absolute certainty cannot be required either, how else could the standard of proof be met? The question becomes more pressing in civil cases. Anticipating this worry, Germany's Federal Court in the Anderson case endorsed a conception of proof standards that acknowledged the inveitable fallibility of trial decisions while at the same time pressed the need for certainty. The Federal Court wrote that a judge's decision must satisfy 'a degree of certainty which is useful for practical life and which makes the doubts silent without completely excluding them' (Federal Court of Justice, February 17, 1970; III ZR 139/67).

The words of Germany's Federal Court echo dillemmas that bedeviled early theorists of probability and evidence law. For instance, Jacob Bernoulli — one of the pionerres of probability theory – discusses the requirement for a criminal conviction in his *Ars Conjectandi*, and he writes that 'it might be determined whether 99/100 of probability suffices or whether 999/1000 is required' (part IV). This is one of the earliest suggestions that the criminal standard of proof be equated to a threshold probability of guilt. On the other hand, the Italian legal penologist Cesare Beccaria in his celebrated treatise *On Crimes and Punishments* remarks that the moral certainty needed to convict is 'nothing but a probability, though a probability of such a sort to be called certainty' (chapter 14). This admittedly quite elusive remark seems to indicate that the standard of decision in criminal trials should be understood as a blend of probability and certainty.

3 Practical worries and idealization

The resistance that some legal practioners display about quantifying standards of proof probabiistically can have many causes. Part of the story is certainly the conviction that a probabilistic intepretation of proof standards is unrealistic. After all, such probabilistic interpretation would face several practical challanges.[1] How are probabilities – say the probability of someone's guilt – going to be estimated? How will the triers of facts apply these probabilistic standards? Should the application of the threshold be mechanic – that is, if the evidence is above the requisite threshold, find against the defedant (say, convict in a criminal trial) and otherwise find for the defedant (say, acquit)? More generally, the challenge is how probabilistic thresholds can be operationalized as part of trial decisions. This is by no means clear. Judges or jurors do not necessarily assess evidence in a probabilistic way and do not use probability thresholds as the standard for their decisions.

But these practical worries should be exagerated. In interpreting proof standards as probability threhsolds, legal probabilists commit themselves to two claims, what we might call the 'quantification claim' and the 'threshold claim'. To illustrate, in the case of a criminal trial, these claims are as follows:

> QUANTIFICATION CLAIM: a probabilistic quantification of the defendant's guilt can be given through an appropriate weighing of all the evidence available (that is, of all the evidence against, and of all the evidencein defense of, the accused).
>
> THRESHOLD CLAIM: an appropriately high threshold guilt probability, say 95%, should be the decision criterion for criminal convictions.

[1] With a few exceptions, most of the arguments about legal probabilism in the early stage of the debate were concerned mostly with practicality. See [@ball1960moment; @kaplan1968decision; @cullison1969probability; @simon1970quantifying; @tribe1971trial; @tribe1970further; @lempert1977modeling @kaye1979paradox; @tillers1988probability].

Those who are worried about practcalities might reason as follows. If guilt cannot be quantified probabilistically—for example, in terms of the conditional probability of G given the total evidence E—no probabilistic threshold could ever be used as a decision criterion. Since the quantification claim is unfeasible, and the threshold claim rests on the quantification claim, the threshold claim rests on shaky grounds. But this conclusion is not as damning for legal probabilist as it might seem at first.

The quantification claim can be interpreted in at least two different ways. It can be the claim that a quantification of guilt—understood as an actual reasoning process—can be effectively carried out by the fact-finders. The quantification claim can also express an idealization or a regulative ideal. The latter interpretation is plausible. For instance, the authors of a
book on probabilistic inference in forensic science write that 'the ... [probabilistic] formalism should primarily be considered as an *aid to structure and guide one's inferences under uncertainty, rather than a way to reach precise numerical assessments*' (p. xv) (CITE TARONI). Even from a probabilist standpoint, then, the quantification of guilt can well be an idealization which has, primarily, a heuristic role.

Just as the quantification claim can be interpreted in two different ways, the same can be said of the threshold claim. For one thing, we can interpret it as describing an effective decision procedure, as though the fact-finders were required to mechanically convict whenever the defendant's probability of guilt happened to meet the desired probabilistic threshold. But there is a second, and less mechanistic, interpretation of the threshold claim. On the second interpretation, the threshold claim would only describe a way to understand, or theorize about, thestandard of proof or the rule of decision. The second interpretation of the threshold claim—which fits well with the "idealization interpretation" of the quantification claim—is the one which many legal probabilists endorse, and it is less likely to cause outrage among the traditionalists.

Lawrence Tribe, in his famous 1971 article 'Trial by Mathematics', expresses disdain for a trial process that were mechanically governed by numbers and probabilities. He claims that in this case the jurors would forget their humanizing function. He writes:

> Guided and perhaps *intimidated by the seeming inexorability of numbers*, induced by the persuasive force of formulas and the precision of decimal points to perceive themselves as performing a largely mechanical and automatic role, *few jurors ... could be relied upon to recall, let alone to perform, [their] humanizing function*. (CITE TRIBE)

But this worry does not apply if we interpret the threshold claim in a non-mechanistic way. This is also the interpretation we shall adopt for the purpose of this chapter. To avoid setting the bar for legal probabilism too high, we will not be concerned with practical issues that would arise if one wanted to deploy a probabilistic threshold directly. We will grant that, at least for now, successful deployments of such thresholds are not viable. For the time being, probabilistic thresholds are perhaps best understood as offerring an theoretical, analytical model of trial decisions. The fact that this theorethical model cannot be easily operationalized does not mean it is pointless. There are multiple ways in which such a model, even if unfit for direct deployment in trial proceedings, could be useful and offer insight into trial decision-making.

Here is an illustration of the analytic power of the probabilistic interpretation of proof standards. Intuitively, standards of proof can be ranked from the least demanding (preponderance of the evidece) to the most demanding (proof beyond a reasonable doubt). But why think this way? Expected utility theory explains why the choice of increasingly demanding thresholds – say 95% in criminal cases and 50% in civil cases – is well-justified (Kaye, 1986). To see how this works, let $cost(CI)$ and $cost(AG)$ be the costs associated with the two decisional errors that can be made in a criminal trial, convicting an innocent (CI) and acquitting a guilty defendant (AG). Let $Pr(G|E)$ and $Pr(I|E)$ be the guilt probability and the innocence probability estimated on the basis of the evidence presented at trial. Given a simple decision-theoretic model, a conviction should be preferred to an acquittal whenever the expected cost resulting from a mistaken conviction—namely, $Pr(I|E) \cdot cost(CI)$—is lower than the expected cost resulting from a mistaken acquittal—namely, $Pr(G|E) \cdot cost(AG)$. This yields the following rule of decision: %

$$\text{convict } \textit{only if } \frac{cost(CI)}{cost(AG)} < \frac{Pr(G|E)}{Pr(I|E)}. \quad \text{(R-post)}$$

[2] This follows from $Pr(I|E) \cdot cost(CI) < Pr(G|E) \cdot cost(AG)$; see [@Kaplan1968Decision]. This model assumes, simplistically, that correct decisions do not bring any positive utility. More complex models are also possible, but the basic idea is the same; see [@Dekay1996; @kaplow2012].

% For the inequality in (R-post) to be satisfied, the posterior ratio $\frac{Pr(G|E)}{Pr(I|E)}$ should exceed the cost ratio $\frac{cost(CI)}{cost(AG)}$. Once the cost ratio is fixed, this determines a guilt probability threshold as the rule of decision. For example, consider a cost ratio of nine according to which a mistaken conviction is nine times as costly as a mistaken acquittal. The corresponding guilt probability threshold will be as high as 90%.

Say the probability of guilt in a criminal case (or the probability of civil liability in a civil case) is p. The stringency of threshold shoul be a function of what is at stake in trial. What's at stake could be more significant in a criminal trial than in a civil trial. After all, a mistaken conviction following a criminal trial will injustly deprive the defedant of basic liberties or even life. Instaed, a mistaken decision in a civil trial would not encroach upon someone's basic liberties since decision in civil trials are mostly about imposing monetary compensation on one of the two litigants. To make this more precise, note that trial decisions can be factually mistaken in two ways. The decision could be a false positive – i.e. a decision to find against the defendant (convict, in a criminal case) even though the defedant committed no wrong (or committed no crime). The decision could also be a false negative – i.e. a decision to find in favor of the defendant (acquit, in a criminal case) even though the defendant did commit the wrong (or committed the crime). Say the cost of a false positive is c_{fp} and the costs of a false negative is c_{fn}. In general, expected utility theory reccomend that one take the action that maximizes expected utility or minimizes expected costs. So, on this view, the action 'find against the defedant' (the action 'convict' in a criminal case) is justified if and only if the expected costs of a false negative – i.e. p c_{fn} – is greater than the expected cost of a false positive – i.e. (1-p)c{fp}. This can be stated more formally with the following inequality:

$$pc_{fn} > (1-p)c_{fp}$$

Now, solving for p gives us:

$$\begin{aligned} pc_{fn} &> c_{fp} - pc_{fp} \\ pc_{fn} + pc_{fp} &> c_{fp} \\ p(c_{fn} + c_{fp}) &> c_{fp} \\ p &> \frac{c_{fp}}{c_{fn} + c_{fp}} \end{aligned}$$

So, as long as you can quantify these disutilities, the probability threshold can be determined. But since I want to focus on probabilistic considerations, I will not pursue this discussion.

Thresholds, however, can vary depending on the costs and benefits at stake in each case (see later discussion). So they need not be applied mechanically without considering the individual circumstances (CITE Hedden and Colyvan, 2019). Furthermore, if jurors are numerically literate, they should not lose sight of their humanizing function as they would no longer be intimated by numbers. So the force of the objection underscores the need to ensure that jurors are numerically literate, not to dispense with numerical thresholds altogether.

LP comes in various shapes. It is one thing to say that the standards of juridical proof are to be explicated in probabilistic terms, it is another to provide such an explication. The threshold-based legal probabilism has it that once the probability of guilt (or, to be more precise, the factual statement that according to law is equivalent to guilt) given the total evidence available is assessed, conviction is justified just in case this probability is above a certain threshold.[3]

Classical Legal Probabilism (CLP), stemming from (Bernoulli, 1713), keeps the threshold constant:[4]

(CLP) There is a certain probability of guilt threshold t, such that in any particular case, if the probability of guilt conditional on all the evidence is above t, convict; otherwise acquit.

A slightly weaker (and perhaps more common among evidence scholars) variant of this view, let us call it the *Sensitive Legal Probabilism* (SLP), also embraces the idea that what is to be evaluated is

[3] In the Anglo-Saxon tradition there is a distinction between decision standards in civil and in criminal cases. In the former, decision is to be made on the preponderance of probability, and in criminal cases, the guilt statement is supposed to be beyond reasonable doubt. Assuming these are to be modeled by probability thresholds different from 1, there is no essential difference here as far as the conceptual difficulties to be discussed in this paper are involved.

[4] Again, we are going to ignore the difference between civil and criminal litigation here. If one wants to keep this distinction in mind, CLP can be easily revised by positing one threshold for criminal cases, and one for civil ones.

the probability of guilt given the evidence, but abandons the requirement that there should be a single threshold for all cases; rather, SLP suggests that the context of each particular case will determine which threshold is appropriate for it.

(SLP) For any particular case, there is a contextually determined probability threshold t such that if the probability of guilt conditional on all the evidence is above t, convict; otherwise acquit.

3.1 Interval thresholds (Finkelstein)

The prior probability cannot be easily determined (Friedman, 2000). Even if it can be determined, arriving at a posterior probability might be impractical because of lack of adequate quantitative information. Perhaps, decision thresholds should not rely on a unique posterior probability but on an interval of admissible probabilities given the evidence (Finkelstein & Fairley, 1970). Perhaps, the assessment of the posterior probability of guilt can be viewed as an idealized process, a regulative ideal which can improve the precision of legal reasoning. (CITE BIEDERMAN TARONI).

4 Theoretical challenges

Another set of problems for a probabikistic intepretation of proof standards are analytical and theoretical. Even if the practical problems can be overcome, theoretical difficulties remain. We will focus on two in particular: naked statistical evidence and the difficulty with conjunction.

4.1 Naked statistical evidence

Here's another problem with TLP, the *paradox of the gatecrasher* (Cohen, 1977; Nesson, 1979). A variant of the paradox goes as follows:

> Suppose our guilt threshold is high, say at 0.99. Consider the situation in which 1000 fans enter a football stadium, and 991 of them avoid paying for their tickets. A random spectator is tried for not paying. The probability that the spectator under trial did not pay exceeds 0.99. Yet, intuitively, a spectator cannot be considered guilty on the sole basis of the number of people who did and did not pay.[5]

The thought experiment can be adapted to match any particular threshold that a proponent of TLP might suggest, as long as it is < 1. For any such a choice of a threshold, it seems, we can think of a situation where all available evidence increases the probability of guilt above it, and yet, conviction seems unjustified.

The problem is not only that TLP leads to a conviction that intuitively seems unjustified and might be wrong. Once we notice that our evidence about each spectator is exactly the same, TLP seems to commit us to the conclusion that all of them should be punished, including the nine that actually paid, as long as we can't tell them apart. And arguably, there is something disturbing in the idea of a system of justice which pretty much explicitly admits that some innocent people should be punished.

The gatecrasher paradox can be considered (or at least has been considered by some scholars) illustrative of a wider phenomenon. According to at least some approaches, there is an important distinction between *naked statistical evidence*, such as the evidence involved in the Gatecrasher Paradox, and *individualized evidence* (such as, say, eyewitness testimony) (Haack, 2014). Seemingly, judges and human subjects are less willing to convict based on naked statistical evidence than when individualized evidence is available, despite the subjective probability of guilt being the same (Wells, 1992).

Philosophers accepting this distinction have proposed many different explications of what this supposed difference consists in exactly, without much agreement being reached.[6] However, the underdevelopment of philosophical theories aside, as the gatecrasher paradox and some real cases based solely on DNA cold hits that got thrown away indicate, there are at least some cases in which the probability of guilt given the evidence might be high, and the conviction still is not justified. Arguably,

[5] The thought experiment that in the absence of any other evidence, the only source of probabilistic information is the statistics, and so that the probability of guilt corresponds to the frequency of unpaid admissions. If the reader does not agree, I ask her to play along, and to notice that in such a case a principled story of what the probability of guilt is and why is needed.

[6] See [@redmayne2008exploring] for a critical survey and [@enoch2015sense] and [@smith2017does] for more recent proposals.

a probabilistic explication of judiciary decision standard should at least allow for this possibility and specify the conditions under which this might happen.

4.2 Conjuction paradox

The *Difficulty About Conjunction* (DAC) proceeds as follows. Say we focus on a civil suit where a plaintiff is required to prove their case on the balance of probability, which for the sake of argument we construe as passing the 0.5 probability threshold.[7] Suppose the plaintiff's claim to be proven based on total evidence E is composed of two elements, A and B, independent conditionally on E.[8] The question is, what exactly is the plaintiff supposed to establish? It seems we have two possible readings:

Requirement 1	$\mathsf{P}(A \wedge B|E) > 0.5$
Requirement 2	$\mathsf{P}(A|E) > 0.5$ and $\mathsf{P}(B|E) > 0.5$

Requirement 1 says that the plaintiff should show that their *whole* claim is more likely than its negation. There are strong intuitions that this is what they should do. But the problem is, this requirement is not equivalent to **Requirement 2**. In fact, if we need $\mathsf{P}(A \wedge B|E) = \mathsf{P}(A|E) \times \mathsf{P}(B|E) > 0.5$ (the identity being justified by the independence assumption), satisfying **Requirement 2** is not sufficient for this purpose. For instance, if $\mathsf{P}(A|E) = \mathsf{P}(B|E) = 0.51$, $\mathsf{P}(A|E) \times \mathsf{P}(B|E) \approx 0.26$, and so the plaintiff's claim as a whole still fails to be established. This means that requiring the proof of $A \wedge B$ on the balance of probability puts an importantly higher requirement on the separate probabilities of the conjuncts.

Moreover, what is required exactly for one of them depends on what has been achieved for the other. If I already established that $\mathsf{P}(A|E) = 0.8$, I need $\mathsf{P}(B|E) \geq 0.635$ to end up with $\mathsf{P}(A \wedge B|E) \geq 0.51$. If, however, $\mathsf{P}(A|E) = 0.6$, I need $\mathsf{P}(B|E) \geq 0.85$ to reach the same threshold. This would mean that standards of proof for a given claim could vary depending on how well a different claim has been argued for and on whether it is a part of a more complex claim that one is defending, and this does not seem very intuitive. At least, this goes strongly against the equal treatment requirement mentioned already in the introduction.

Should we then abandon **Requirement 1** and remain content with **Requirement 2**? [Cohen1977The-probable-an: 66] convincingly argues that we should not. Not evaluating a complex civil case as a whole is the opposite of what the courts themselves normally do. There are good reasons to think that every common law system subscribes to a sort of conjunction principle, which states that if A and B are established on the balance of probabilities, then so is $A \wedge B$.

So, on one hand, if we take our decision standard from **Requirement 2**, our acceptance standard will not involve closure under conjunction, and might lead to conviction in cases where $\mathsf{P}(G|E)$ is quite low, just because G is a conjunction of elements which separately satisfy the standard of proof – and this seems unintuitive. On the other hand, following Cohen, if we take our decision standard from **Requirement 1**, we will put seemingly unnecessarily high requirements sensitive to fairly contingent and irrelevant facts on the prosecution, and treat various elements to be proven unevenly. Neither seems desirable.

5 Likelihood thresholds

5.1 The likelihood strategy

At its simplest, decision theory based on the maximization of expected utility states that between a number of alternative courses of action, the one with the highest expected utility (or with the lowest expected cost) should be preferred. The decision-theoretic framework is very general and can be applied to a variety of situations, including criminal trials, where the two alternatives to choose from are an acquittal and a conviction.

To see how this works, let $cost(CI)$ and $cost(AG)$ be the costs associated with the two decisional errors that can be made at trial, convicting an innocent (CI) and acquitting a guilty defendant (AG). Let $Pr(G|E)$ and $Pr(I|E)$ be the guilt probability and the innocence probability estimated on the basis of the evidence presented at trial. Given a simple decision-theoretic model, a conviction should be

[7] This is a natural choice given that the plaintiff is supposed to show that their claim is more probable than the defendant's. The assumption is not essential. DAC can be deployed against any $\neq 1$ guilt probability threshold.

[8] These assumptions, again, are not too essential. In fact, the difficulties become more severe as the number of elements grows, and, extreme cases aside, do not tend to disappear if the elements are dependent.

preferred to an acquittal whenever the expected cost resulting from a mistaken conviction—namely, $Pr(I|E) \cdot cost(CI)$—is lower than the expected cost resulting from a mistaken acquittal—namely, $Pr(G|E) \cdot cost(AG)$. This yields the following rule of decision: %

$$\text{convict } \textit{only if } \frac{cost(CI)}{cost(AG)} < \frac{Pr(G|E)}{Pr(I|E)}.^{9} \tag{R-post}$$

% For the inequality in (R-post) to be satisfied, the posterior ratio $\frac{Pr(G|E)}{Pr(I|E)}$ should exceed the cost ratio $\frac{cost(CI)}{cost(AG)}$. Once the cost ratio is fixed, this determines a guilt probability threshold as the rule of decision. For example, consider a cost ratio of nine according to which a mistaken conviction is nine times as costly as a mistaken acquittal. The corresponding guilt probability threshold will be as high as 90%.

The first thing to note here is that the value of the posterior ratio $\frac{Pr(G|E)}{Pr(I|E)}$—except for rare cases such as the prisoner scenario—is hard to estimate upfront. Fortunately, this task can be broken down into two smaller tasks. By Bayes' theorem, we know that

$$\frac{Pr(G|E)}{Pr(I|E)} = \frac{Pr(E|G)}{Pr(E|I)} \cdot \frac{Pr(G)}{Pr(I)}.$$

% That is, the posterior ratio $\frac{Pr(G|E)}{Pr(I|E)}$ can be arrived at by estimating the likelihood ratio $\frac{Pr(E|G)}{Pr(E|I)}$ and the prior ratio $\frac{Pr(G)}{Pr(I)}$ separately, and then by multiplying the two together.

5.2 Cheng

Here is one way to think about the decision thresholds in terms of likelihoods, stemming from (Cheng, 2012). The idea is to conceptualize juridical decisions in analogy to statistical hypothesis testing. We have two hypotheses under consideration: defendant's H_Δ and plaintiff's H_Π, and we are to pick one: D_Δ stands for the decision for H_Δ and D_Π is the decision that H_Π. On this approach, rather than directly evaluating the probability of H_Π given the evidence and comparing it to a threshold, we compare the support that the evidence provides for these hypotheses, and decide for the one for which the evidence provides better support.

Cheng motivates this approach by the following considerations. Suppose that if the decision is correct, no costs result, but incorrect decisions have their price. Let us say that if the defendant is right and we find against them, the cost is c_1, and if the plaintiff is right and we find against them, the cost is c_2:

		Decision	
		D_Δ	D_Π
Truth	H_Δ	0	c_1
	H_Π	c_2	0

Intuitively, it seems that we want a decision rule which minimizes the expected cost. Say that given our total evidence E the relevant conditional probabilities are:

$$p_\Delta = \mathsf{P}(H_\Delta|E)$$
$$p_\Pi = \mathsf{P}(H_\Pi|E)$$

The expected costs for deciding that H_Δ and H_Π, respectively, are:

$$E(D_\Delta) = p_\Delta 0 + p_\Pi c_2 = c_2 p_\Pi$$
$$E(D_\Pi) = p_\Delta c_1 + p_\Pi 0 = c_1 p_\Delta$$

For this reason, on these assumptions, we would like to choose H_Π just in case $E(D_\Pi) < E(D_\Delta)$. This condition is equivalent to:

$$\begin{aligned} c_1 p_\Delta &< c_2 p_\Pi \\ c_1 &< \frac{c_2 p_\Pi}{p_\Delta} \\ \frac{c_1}{c_2} &< \frac{p_\Pi}{p_\Delta} \end{aligned} \tag{1}$$

[9] This follows from $Pr(I|E) \cdot cost(CI) < Pr(G|E) \cdot cost(AG)$; see **?**. This model assumes, simplistically, that correct decisions do not bring any positive utility. More complex models are also possible, but the basic idea is the same; see **??**.

Cheng (2012) (1261) insists:

> At the same time, in a civil trial, the legal system expresses no preference between finding erroneously for the plaintiff (false positives) and finding erroneously for the defendant (false negatives). The costs c_1 and c_2 are thus equal...

If we grant this assumption, $c_1 = c_2$, (1) reduces to:

$$1 < \frac{p_\Pi}{p_\Delta}$$
$$p_\Pi > p_\Delta \tag{2}$$

That is, in standard civil litigation we are to find for the plaintiff just in case H_Π is more probable given the evidence than H_Δ, which seems plausible.[10] Let's call this decision standard **Relative Legal Probabilism (RLP)**.[11]

Here is a slightly different perspective, due to Dawid (1987), that also suggests that juridical decisions should be likelihood-based. The focus is on witnesses for the sake of simplicity. Imagine the plaintiff produces two independent witnesses: W_A attesting to A, and W_B attesting to B. Say the witnesses are regarded as 70% reliable and A and B are probabilistically independent, so we infer $\mathsf{P}(A) = \mathsf{P}(B) = 0.7$ and $\mathsf{P}(A \wedge B) = 0.7^2 = 0.49$.

But, Dawid argues, this is misleading, because to reach this result we misrepresented the reliability of the witnesses: 70% reliability of a witness, he continues, does not mean that if the witness testifies that A, we should believe that $\mathsf{P}(A) = 0.7$. To see his point, consider two potential testimonies:

A_1	The sun rose today.
A_2	The sun moved backwards through the sky today.

Intuitively, after hearing them, we would still take $\mathsf{P}(A_1)$ to be close to 1 and $\mathsf{P}(A_2)$ to be close to 0, because we already have fairly strong convictions about the issues at hand. In general, how we should revise our beliefs in light of a testimony depends not only on the reliability of the witness, but also on our prior convictions.[12] And this is as it should be: as indicated by Bayes' Theorem, one and the same testimony with different priors might lead to different posterior probabilities.

So far so good. But how should we represent evidence (or testimony) strength then? Well, one pretty standard way to go is to focus on how much it contributes to the change in our beliefs in a way independent of any particular choice of prior beliefs. Let a be the event that the witness testified that A. It is useful to think about the problem in terms of *odds, conditional odds (O)* and *likelihood ratios (LR)*:

$$O(A) = \frac{\mathsf{P}(A)}{\mathsf{P}(\neg A)}$$
$$O(A|a) = \frac{\mathsf{P}(A|a)}{\mathsf{P}(\neg A|a)}$$
$$LR(a|A) = \frac{\mathsf{P}(a|A)}{\mathsf{P}(a|\neg A)}.$$

Suppose our prior beliefs and background knowledge, before hearing a testimony, are captured by the prior probability measure $\mathsf{P}_{prior}(\cdot)$, and the only thing that we learn is a. We're interested in what our *posterior* probability measure, $\mathsf{P}_{posterior}(\cdot)$, and posterior odds should then be. If we're to proceed with Bayesian updating, we should have:

$$\frac{\mathsf{P}_{posterior}(A)}{\mathsf{P}_{posterior}(\neg A)} = \frac{\mathsf{P}_{prior}(A|a)}{\mathsf{P}_{prior}(\neg A|a)} = \frac{\mathsf{P}_{prior}(a|A)}{\mathsf{P}_{prior}(a|\neg A)} \times \frac{\mathsf{P}_{prior}(A)}{\mathsf{P}_{prior}(\neg A)}$$

that is,

$$O_{posterior}(A) = O_{prior}(A|a) = \underbrace{LR_{prior}(a|A)}_{\text{conditional likelihood ratio}} \times O_{prior}(A) \tag{3}$$

[10] Notice that this instruction is somewhat more general than the usual suggestion of the preponderance standard in civil litigation, according to which the court should find for the plaintiff just in case $\mathsf{P}(H_\Pi|E) > 0.5$. This threshold, however, results from (2) if it so happens that H_Δ is $\neg H_\Pi$, that is, if the defendant's claim is simply the negation of the plaintiff's thesis. By no means, Cheng argues, this is always the case.

[11] We were not aware of any particular name for Cheng's model so we came up with this one. We're not particularly attached to it, and it is not standard terminology.

[12] An issue that Dawid does not bring up is the interplay between our priors and our assessment of the reliability of the witnesses. Clearly, our posterior assessment of the credibility of the witness who testified A_2 will be lower than that of the other witness.

The conditional likelihood ratio seems to be a much more direct measure of the value of a, independent of our priors regarding A itself. In general, the posterior probability of an event will equal to the witness's reliability in the sense introduced above only if the prior is $1/2$.[13]

Quite independently, a similar approach to juridical decisions has been proposed by Kaplow (2014) – we'll call it **decision-theoretic legal probabilism (DTLP)**. It turns out that Cheng's suggestion is a particular case of this more general approach. Let $LR(E) = \mathsf{P}(E|H_\Pi)/\mathsf{P}(E|H_\Delta)$. In whole generality, DTLP invites us to convict just in case $LR(E) > LR^\star$, where $LR^\star$ is some critical value of the likelihood ratio.

Say we want to formulate the usual preponderance rule: convict iff $\mathsf{P}(H_\Pi|E) > 0.5$, that is, iff $\frac{\mathsf{P}(H_\Pi|E)}{\mathsf{P}(H_\Delta|E)} > 1$. By Bayes' Theorem we have:

$$\frac{\mathsf{P}(H_\Pi|E)}{\mathsf{P}(H_\Delta|E)} = \frac{\mathsf{P}(H_\Pi)}{\mathsf{P}(H_\Delta)} \times \frac{\mathsf{P}(E|H_\Pi)}{\mathsf{P}(E|H_\Delta)} > 1 \Leftrightarrow$$
$$\Leftrightarrow \frac{\mathsf{P}(E|H_\Pi)}{\mathsf{P}(E|H_\Delta)} > \frac{\mathsf{P}(H_\Delta)}{\mathsf{P}(H_\Pi)}$$

So, as expected, $LR^\star$ is not unique and depends on priors. Analogous reformulations are available for thresholds other than 0.5.

Kaplow's point is not that we can reformulate threshold decision rules in terms of priors-sensitive likelihood ratio thresholds. Rather, he insists, when we make a decision, we should factor in its consequences. Let G represent potential gain from correct conviction, and L stand for the potential loss resulting from mistaken conviction. Taking them into account, Kaplow suggests, we should convict if and only if:

$$\mathsf{P}(H_\Pi|E) \times G > \mathsf{P}(H_\Delta|E) \times L \tag{4}$$

Now, (4) is equivalent to:

$$\frac{\mathsf{P}(H_\Pi|E)}{\mathsf{P}(H_\Delta|E)} > \frac{L}{G}$$
$$\frac{\mathsf{P}(H_\Pi)}{\mathsf{P}(H_\Delta)} \times \frac{\mathsf{P}(E|H_\Pi)}{\mathsf{P}(E|H_\Delta)} > \frac{L}{G}$$
$$\frac{\mathsf{P}(E|H_\Pi)}{\mathsf{P}(E|H_\Delta)} > \frac{\mathsf{P}(H_\Delta)}{\mathsf{P}(H_\Pi)} \times \frac{L}{G}$$
$$LR(E) > \frac{\mathsf{P}(H_\Delta)}{\mathsf{P}(H_\Pi)} \times \frac{L}{G} \tag{5}$$

This is the general format of Kaplow's decision standard.

5.3 Likelihood and DAC

But how does our preference for the likelihood ratio as a measure of evidence strength relate to DAC? Let's go through Dawid's reasoning.

[13] Dawid gives no general argument, but it is not too hard to give one. Let $rel(a) = \mathsf{P}(a|A) = \mathsf{P}(\neg a|\neg A)$. We have in the background $\mathsf{P}(a|\neg A) = 1 - \mathsf{P}(\neg a|\neg A) = 1 - rel(a)$. We want to find the condition under which $\mathsf{P}(A|a) = \mathsf{P}(a|A)$. Set $\mathsf{P}(A) = p$ and start with Bayes' Theorem and the law of total probability, and go from there:

$$\mathsf{P}(A|a) = \mathsf{P}(a|A)$$
$$\frac{\mathsf{P}(a|A)p}{\mathsf{P}(a|A)p + \mathsf{P}(a|\neg A)(1-p)} = \mathsf{P}(a|A)$$
$$\mathsf{P}(a|A)p = \mathsf{P}(a|A)[\mathsf{P}(a|A)p + \mathsf{P}(a|\neg A)(1-p)]$$
$$p = \mathsf{P}(a|A)p + \mathsf{P}(a|\neg A) - \mathsf{P}(a|\neg A)p$$
$$p = rel(a)p + 1 - rel(a) - (1 - rel(a))p$$
$$p = rel(a)p + 1 - rel(a) - p + rel(a)p$$
$$2p = 2rel(a)p + 1 - rel(a)$$
$$2p - 2rel(a)p = 1 - rel(a)$$
$$2p(1 - rel(a)) = 1 - rel(a)$$
$$2p = 1$$

First we multiplied both sides by the denominator. Then we divided both sides by $\mathsf{P}(a|A)$ and multiplied on the right side. Then we used our background notation and information. Next, we manipulated the right-hand side algebraically and moved $-p$ to the left-hand side. Move $2rel(a)p$ to the left and manipulate the result algebraically to get to the last line.

A sensible way to probabilistically interpret the 70% reliability of a witness who testifies that A is to take it to consist in the fact that the probability of a positive testimony if A is the case, just as the probability of a negative testimony (that is, testimony that A is false) if A isn't the case, is 0.7:[14]

$$\mathsf{P}_{prior}(a|A) = \mathsf{P}_{prior}(\neg a|\neg A) = 0.7.$$

$\mathsf{P}_{prior}(a|\neg A) = 1 - \mathsf{P}_{prior}(\neg a|\neg A) = 0.3$, and so the same information is encoded in the appropriate likelihood ratio:

$$LR_{prior}(a|A) = \frac{\mathsf{P}_{prior}(a|A)}{\mathsf{P}_{prior}(a|\neg A)} = \frac{0.7}{0.3}$$

Let's say that *a provides (positive) support* for A in case

$$O_{posterior}(A) = O_{prior}(A|a) > O_{prior}(A)$$

that is, a testimony a supports A just in case the posterior odds of A given a are greater than the prior odds of A (this happens just in case $\mathsf{P}_{posterior}(A) > \mathsf{P}_{prior}(A)$). By (3), this will be the case if and only if $LR_{prior}(a|A) > 1$.

One question that Dawid addresses is this: assuming reliability of witnesses 0.7, and assuming that a and b, taken separately, provide positive support for their respective claims, does it follow that $a \wedge b$ provides positive support for $A \wedge B$?

Assuming the independence of the witnesses, this will hold in non-degenerate cases that do not involve extreme probabilities, on the assumption of independence of a and b conditional on all combinations: $A \wedge B$, $A \wedge \neg B$, $\neg A \wedge B$ and $\neg A \wedge \neg B$.[15,~16]

Let us see why the above claim holds. The calculations are my reconstruction and are not due to Dawid. The reader might be annoyed with me working out the mundane details of Dawid's claims, but it turns out that in the case of Dawid's strategy, the devil is in the details. The independence of witnesses gives us:

$$\begin{aligned}
\mathsf{P}(a \wedge b|A \wedge B) &= 0.7^2 = 0.49 \\
\mathsf{P}(a \wedge b|A \wedge \neg B) &= 0.7 \times 0.3 = 0.21 \\
\mathsf{P}(a \wedge b|\neg A \wedge B) &= 0.3 \times 0.7 = 0.21 \\
\mathsf{P}(a \wedge b|\neg A \wedge \neg B) &= 0.3 \times 0.3 = 0.09
\end{aligned}$$

Without assuming A and B to be independent, let the probabilities of $A \wedge B$, $\neg A \wedge B$, $A \wedge \neg B$, $\neg A \wedge \neg B$ be $p_{11}, p_{01}, p_{10}, p_{00}$. First, let's see what $\mathsf{P}(a \wedge b)$ boils down to.

By the law of total probability we have:

$$\begin{aligned}
\mathsf{P}(a \wedge b) = {} & \mathsf{P}(a \wedge b|A \wedge B)\mathsf{P}(A \wedge B) + \\
& + \mathsf{P}(a \wedge b|A \wedge \neg B)\mathsf{P}(A \wedge \neg B) \\
& + \mathsf{P}(a \wedge b|\neg A \wedge B)\mathsf{P}(\neg A \wedge B) + \\
& + \mathsf{P}(a \wedge b|\neg A \wedge \neg B)\mathsf{P}(\neg A \wedge \neg B)
\end{aligned} \tag{6}$$

which, when we substitute our values and constants, results in:

$$= 0.49p_{11} + 0.21(p_{10} + p_{01}) + 0.09p_{00}$$

Now, note that because p_{ii}s add up to one, we have $p_{10} + p_{01} = 1 - p_{00} - p_{11}$. Let us continue.

$$\begin{aligned}
&= 0.49p_{11} + 0.21(1 - p_{00} - p_{11}) + 0.09p_{00} \\
&= 0.21 + 0.28p_{11} - 0.12p_{00}
\end{aligned}$$

[14] In general setting, these are called the *sensitivity* and *specificity* of a test (respectively), and they don't have to be equal. For instance, a degenerate test for an illness which always responds positively, diagnoses everyone as ill, and so has sensitivity 1, but specificity 0.

[15] Dawid only talks about the independence of witnesses without reference to conditional independence. Conditional independence does not follow from independence, and it is the former that is needed here (also, four non-equivalent different versions of it).

[16] In terms of notation and derivation in the optional content that will follow, the claim holds if and only if $28 > 28p_{11} - 12p_{00}$. This inequality is not true for all admissible values of p_{11} and p_{00}. If $p_{11} = 1$ and $p_{00} = 0$, the sides are equal. However, this is a rather degenerate example. Normally, we are interested in cases where $p_{11} < 1$. And indeed, on this assumption, the inequality holds.

Next, we ask what the posterior of $A \wedge B$ given $a \wedge b$ is (in the last line, we also multiply the numerator and the denominator by 100).

$$\begin{aligned} \mathsf{P}(A \wedge B | a \wedge b) &= \frac{\mathsf{P}(a \wedge b | A \wedge B)\mathsf{P}(A \wedge B)}{\mathsf{P}(a \wedge b)} \\ &= \frac{49 p_{11}}{21 + 28 p_{11} - 12 p_{00}} \end{aligned}$$

In this particular case, then, our question whether $\mathsf{P}(A \wedge B | a \wedge b) > \mathsf{P}(A \wedge B)$ boils down to asking whether

$$\frac{49 p_{11}}{21 + 28 p_{11} - 12 p_{00}} > p_{11}$$

that is, whether $28 > 28 p_{11} - 12 p_{00}$ (just divide both sides by p_{11}, multiply by the denominator, and manipulate algebraically).

Dawid continues working with particular choices of values and provides neither a general statement of the fact that the above considerations instantiate nor a proof of it. In the middle of the paper he says:

> Even under prior dependence, the combined support is always positive, in the sense that the posterior probability of the case always exceeds its prior probability... When the problem is analysed carefully, the 'paradox' evaporates [pp. 95-7]

where he still means the case with the particular values that he has given, but he seems to suggest that the claim generalizes to a large array of cases.

The paper does not contain a precise statement making the conditions required explicit and, *a fortriori*, does not contain a proof of it. Given the example above and Dawid's informal reading, let us develop a more precise statement of the claim and a proof thereof.

Fact 1. *Suppose that* $rel(a), rel(b) > 0.5$ *and witnesses are independent conditional on all Boolean combinations of A and B (in a sense to be specified), and that none of the Boolean combinations of A and B has an extreme probability (of 0 or 1). It follows that* $\mathsf{P}(A \wedge B | a \wedge b) > \mathsf{P}(A \wedge B)$. *(Independence of A and B is not required.)*

Roughly, the theorem says that if independent and reliable witnesses provide positive support of their separate claims, their joint testimony provides positive support of the conjunction of their claims.

Let us see why the claim holds. First, we introduce an abbreviation for witness reliability:

$$\begin{aligned} \mathbf{a} &= rel(a) = \mathsf{P}(a|A) = \mathsf{P}(\neg a | \neg A) > 0.5 \\ \mathbf{b} &= rel(b) = \mathsf{P}(b|B) = \mathsf{P}(\neg b | \neg A) > 0.5 \end{aligned}$$

Our independence assumption means:

$$\begin{aligned} \mathsf{P}(a \wedge b | A \wedge B) &= \mathbf{ab} \\ \mathsf{P}(a \wedge b | A \wedge \neg B) &= \mathbf{a}(\mathbf{1} - \mathbf{b}) \\ \mathsf{P}(a \wedge b | \neg A \wedge B) &= (\mathbf{1} - \mathbf{a})\mathbf{b} \\ \mathsf{P}(a \wedge b | \neg A \wedge \neg B) &= (\mathbf{1} - \mathbf{a})(\mathbf{1} - \mathbf{b}) \end{aligned}$$

Abbreviate the probabilities the way we already did:

$$\begin{aligned} \mathsf{P}(A \wedge B) &= p_{11} \qquad \mathsf{P}(A \wedge \neg B) = p_{10} \\ \mathsf{P}(\neg A \wedge B) &= p_{01} \qquad \mathsf{P}(\neg A \wedge \neg B) = p_{00} \end{aligned}$$

Our assumptions entail $0 \neq p_{ij} \neq 1$ for $i, j \in \{0, 1\}$ and:

$$p_{11} + p_{10} + p_{01} + p_{00} = 1 \tag{7}$$

So, we can use this with (6) to get:

$$\begin{aligned} \mathsf{P}(a \wedge b) &= \mathbf{ab} p_{11} + \mathbf{a}(\mathbf{1} - \mathbf{b}) p_{10} + (\mathbf{1} - \mathbf{a})\mathbf{b} p_{01} + (\mathbf{1} - \mathbf{a})(\mathbf{1} - \mathbf{b}) p_{00} \\ &= p_{11}\mathbf{ab} + p_{10}(\mathbf{a} - \mathbf{ab}) + p_{01}(\mathbf{b} - \mathbf{ab}) + p_{00}(1 - \mathbf{b} - \mathbf{a} + \mathbf{ab}) \end{aligned} \tag{8}$$

Let's now work out what the posterior of $A \wedge B$ will be, starting with an application of the Bayes' Theorem:

$$\begin{aligned} \mathsf{P}(A \wedge B | a \wedge b) &= \frac{\mathsf{P}(a \wedge b | A \wedge B)\mathsf{P}(A \wedge B)}{\mathsf{P}(a \wedge b)} \\ &= \frac{\mathbf{ab} p_{11}}{p_{11}\mathbf{ab} + p_{10}(\mathbf{a} - \mathbf{ab}) + p_{01}(\mathbf{b} - \mathbf{ab}) + p_{00}(1 - \mathbf{b} - \mathbf{a} + \mathbf{ab})} \end{aligned} \tag{9}$$

To answer our question we therefore have to compare the content of (9) to p_{11} and our claim holds just in case:

$$\frac{\mathbf{ab}p_{11}}{p_{11}\mathbf{ab} + p_{10}(\mathbf{a} - \mathbf{ab}) + p_{01}(\mathbf{b} - \mathbf{ab}) + p_{00}(1 - \mathbf{b} - \mathbf{a} + \mathbf{ab})} > p_{11}$$

$$\frac{\mathbf{ab}}{p_{11}\mathbf{ab} + p_{10}(\mathbf{a} - \mathbf{ab}) + p_{01}(\mathbf{b} - \mathbf{ab}) + p_{00}(1 - \mathbf{b} - \mathbf{a} + \mathbf{ab})} > 1$$

$$p_{11}\mathbf{ab} + p_{10}(\mathbf{a} - \mathbf{ab}) + p_{01}(\mathbf{b} - \mathbf{ab}) + p_{00}(1 - \mathbf{b} - \mathbf{a} + \mathbf{ab}) < \mathbf{ab} \tag{10}$$

Proving (10) is therefore our goal for now. This is achieved by the following reasoning:[17]

1.	$\mathbf{b} > 0.5,\ \mathbf{a} > 0.5$	assumption
2.	$2\mathbf{b} > 1,\ 2\mathbf{a} > 1$	from 1.
3.	$2\mathbf{ab} > \mathbf{a},\ 2\mathbf{ab} > \mathbf{b}$	multiplying by **a** and **b** respectively
4.	$p_{10}2\mathbf{ab} > p_{10}\mathbf{a},\ \ p_{01}2\mathbf{ab} > p_{01}\mathbf{b}$	multiplying by p_{10} and p_{01} respectively
5.	$p_{10}2\mathbf{ab} + p_{01}2\mathbf{ab} > p_{10}\mathbf{a} + p_{01}\mathbf{b}$	adding by sides, 3., 4.
6.	$1 - \mathbf{b} - \mathbf{a} < 0$	from 1.
7.	$p_{00}(1 - \mathbf{b} - \mathbf{a}) < 0$	From 6., because $p_{00} > 0$
8.	$p_{10}2\mathbf{ab} + p_{01}2\mathbf{ab} > p_{10}\mathbf{a} + p_{01}\mathbf{b} + p_{00}(1 - \mathbf{b} - \mathbf{a})$	from 5. and 7.
9.	$p_{10}\mathbf{ab} + p_{10}\mathbf{ab} + p_{01}\mathbf{ab} + p_{01}\mathbf{ab} + p_{00}\mathbf{ab} - p_{00}\mathbf{ab} > p_{10}\mathbf{a} + p_{01}\mathbf{b} + p_{00}(1 - \mathbf{b} - \mathbf{a})$	8., rewriting left-hand side
10.	$p_{10}\mathbf{ab} + p_{01}\mathbf{ab} + p_{00}\mathbf{ab} > -p_{10}\mathbf{ab} - p_{01}\mathbf{ab} + p_{00}\mathbf{ab} + p_{10}\mathbf{a} + p_{01}\mathbf{b} + p_{00}(1 - \mathbf{b} - \mathbf{a})$	9., moving from left to right
11.	$\mathbf{ab}(p_{10} + p_{01} + p_{00}) > p_{10}(\mathbf{a} - \mathbf{ab}) + p_{01}(\mathbf{b} - \mathbf{ab}) + p_{00}(1 - \mathbf{b} - \mathbf{a} + \mathbf{ab})$	10., algebraic manipulation
12.	$\mathbf{ab}(1 - p_{11}) > p_{10}(\mathbf{a} - \mathbf{ab}) + p_{01}(\mathbf{b} - \mathbf{ab}) + p_{00}(1 - \mathbf{b} - \mathbf{a} + \mathbf{ab})$	11. and equation (7)
13.	$\mathbf{ab} - \mathbf{ab}p_{11} > p_{10}(\mathbf{a} - \mathbf{ab}) + p_{01}(\mathbf{b} - \mathbf{ab}) + p_{00}(1 - \mathbf{b} - \mathbf{a} + \mathbf{ab})$	12., algebraic manipulation
14.	$\mathbf{ab} > \mathbf{ab}p_{11} + p_{10}(\mathbf{a} - \mathbf{ab}) + p_{01}(\mathbf{b} - \mathbf{ab}) + p_{00}(1 - \mathbf{b} - \mathbf{a} + \mathbf{ab})$	13., moving from left to right

%\end{adjustbox}

The last line is what we have been after.

Optional Content Ends

Now that we have as a theorem an explication of what Dawid informally suggested, let's see whether it helps the probabilist handling of DAC.

5.4 Kaplow

On RLP, at least in certain cases, the decision rule leads us to (14), which tells us to decide the case based on whether the likelihood ratio is greater than 1.

[18] While Kaplow did not discuss DAC or the gatecrasher paradox, it is only fair to evaluate Kaplow's proposal from the perspective of these difficulties.

Add here stuff from Marcello's Mind paper about the prisoner hypothetical. Then, discuss Rafal's critique of the likelihood ratio threshold and see where we end up.

5.5 p-value (Cheng?)

6 Challenges (again)

6.1 Dawid's likelihood strategy doesn't help

Recall that DAC was a problem posed for the decision standard proposed by TLP, and the real question is how the information resulting from Fact 1 can help to avoid that problem. Dawid does not mention any decision standard, and so addresses quite a different question, and so it is not clear that '`the`paradox' evaporates", as Dawid suggests.

What Dawid correctly suggests (and we establish in general as Fact 1) is that the support of the conjunction by two witnesses will be positive as soon as their separate support for the conjuncts is positive. That is, that the posterior of the conjunction will be higher that its prior. But the critic of probabilism never denied that the conjunction of testimonies might raise the probability of the conjunction if the testimonies taken separately support the conjuncts taken separately. Such a critic can still insist that Fact 1 does nothing to alleviate her concern. After all, at least *prima facie* it still might be the case that:

[17] Thanks to Pawel Pawlowski for working on this proof with me.

[18] Again, the name of the view is by no means standard, it is just a term I coined to refer to various types of legal probabilism in a fairly uniform manner.

- the posterior probabilities of the conjuncts are above a given threshold,
- the posterior probability of the conjunction is higher than the prior probability of the conjunction,
- the posterior probability of the conjunction is still below the threshold.

That is, Fact 1 does not entail that once the conjuncts satisfy a decision standard, so does the conjunction.

At some point, Dawid makes a general claim that is somewhat stronger than the one already cited:

> When the problem is analysed carefully, the 'paradox' evaporates: suitably measured, the support supplied by the conjunction of several independent testimonies exceeds that supplied by any of its constituents.
> [p. 97]

This is quite a different claim from the content of Fact 1, because previously the joint probability was claimed only to increase as compared to the prior, and here it is claimed to increase above the level of the separate increases provided by separate testimonies. Regarding this issue Dawid elaborates (we still use the p_{ij}-notation that we've already introduced):

> "More generally, let $\mathsf{P}(a|A)/\mathsf{P}(a|\neg A) = \lambda$, $\mathsf{P}(b|B)/\mathsf{P}(b|\neg B) = \mu$, with $\lambda, \mu > 0.7$, as might arise, for example, when there are several available testimonies. If the witnesses are independent, then
>
> $$\mathsf{P}(A \wedge B|a \wedge b) = \lambda\mu p_{11}/(\lambda\mu p_{11} + \lambda p_{10} + \mu p_{01} + p_{00})$$
>
> which increases with each of λ and μ, and is never less than the larger of $\lambda p_{11}/(1 - p_{11} + \lambda p_{11}), \mu p_{11}/(1 - p_{11}1 + \mu p_{11})$, the posterior probabilities appropriate to the individual testimonies." [p. 95]

This claim, however, is false.

Optional Content Starts

Let us see why. The quoted passage is a bit dense. It contains four claims for which no arguments are given in the paper. The first three are listed below as (11), the fourth is that if the conditions in (11) hold, $\mathsf{P}(A \wedge B|a \wedge b) > max(\mathsf{P}(A|a), \mathsf{P}(B|b))$. Notice that $\lambda = LR(a|A)$ and $\mu = LR(b|B)$. Suppose the first three claims hold, that is:

$$\begin{aligned} \mathsf{P}(A \wedge B|a \wedge b) &= \lambda\mu p_{11}/(\lambda\mu p_{11} + \lambda p_{10} + \mu p_{01} + p_{00}) \\ \mathsf{P}(A|a) &= \frac{\lambda p_{11}}{1 - p_{11} + \lambda p_{11}} \\ \mathsf{P}(B|b) &= \frac{\mu p_{11}}{1 - p_{11} + \mu p_{11}} \end{aligned} \tag{11}$$

Is it really the case that $\mathsf{P}(A \wedge B|a \wedge b) > \mathsf{P}(A|a), \mathsf{P}(B|b)$? It does not seem so. Let $\mathbf{a} = \mathbf{b} = 0.6$, $pr = \langle p_{11}, p_{10}, p_{01}, p_{00}\rangle = \langle 0.1, 0.7, 0.1, 0.1\rangle$. Then, $\lambda = \mu = 1.5 > 0.7$ so the assumption is satisfied. Then we have $\mathsf{P}(A) = p_{11} + p_{10} = 0.8$, $\mathsf{P}(B) = p_{11} + p_{01} = 0.2$. We can also easily compute $\mathsf{P}(a) = \mathbf{a}\mathsf{P}(A) + (1 - \mathbf{a})\mathsf{P}(\neg A) = 0.56$ and $\mathsf{P}(b) = \mathbf{b}\mathsf{P}(B) + (1 - \mathbf{b})\mathsf{P}(\neg B) = 0.44$. Yet:

$$\begin{aligned} \mathsf{P}(A|a) &= \frac{\mathsf{P}(a|A)\mathsf{P}(A)}{\mathsf{P}(a)} = \frac{0.6 \times 0.8}{0.6 \times 0.8 + 0.4 \times 0.2} \approx 0.8571 \\ \mathsf{P}(B|b) &= \frac{\mathsf{P}(b|B)\mathsf{P}(B)}{\mathsf{P}(b)} = \frac{0.6 \times 0.2}{0.6 \times 0.2 + 0.4 \times 0.8} \approx 0.272 \\ \mathsf{P}(A \wedge B|a \wedge b) &= \frac{\mathsf{P}(a \wedge b|A \wedge B)\mathsf{P}(A \wedge B)}{\begin{array}{l}\mathsf{P}(a \wedge b|A \wedge B)\mathsf{P}(A \wedge B) + \mathsf{P}(a \wedge b|A \wedge \neg B)\mathsf{P}(A \wedge \neg B) + \\ \quad + \mathsf{P}(a \wedge b|\neg A \wedge B)\mathsf{P}(\neg A \wedge B) + \mathsf{P}(a \wedge b|\neg A \wedge \neg B)\mathsf{P}(\neg A \wedge \neg B)\end{array}} \\ &= \frac{\mathbf{ab}p_{11}}{\mathbf{ab}p_{11} + \mathbf{a}(1 - \mathbf{b})p_{10} + (1 - \mathbf{a})\mathbf{b}p_{01} + (1 - \mathbf{a})(1 - \mathbf{b})p_{00}} \approx 0.147 \end{aligned}$$

The posterior probability of $A \wedge B$ is not only lower than the larger of the individual posteriors, but also lower than any of them!

So what went wrong in Dawid's calculations in (11)? Well, the first formula is correct. However, let

us take a look at what the second one says (the problem with the third one is pretty much the same):

$$\mathsf{P}(A|a) = \frac{\frac{\mathsf{P}(a|A)}{\mathsf{P}(\neg a|A)} \times \mathsf{P}(A \wedge B)}{\mathsf{P}(\neg(A \wedge B)) + \frac{\mathsf{P}(a|A)}{\mathsf{P}(\neg a|A)} \times \mathsf{P}(A \wedge B)}$$

Quite surprisingly, in Dawid's formula for $\mathsf{P}(A|a)$, the probability of $A \wedge B$ plays a role. To see that it should not take any B that excludes A and the formula will lead to the conclusion that *always* $\mathsf{P}(A|a)$ is undefined. The problem with Dawid's formula is that instead of $p_{11} = \mathsf{P}(A \wedge B)$ he should have used $\mathsf{P}(A) = p_{11} + p_{10}$, in which case the formula would rather say this:

$$\begin{aligned}
\mathsf{P}(A|a) &= \frac{\frac{\mathsf{P}(a|A)}{\mathsf{P}(\neg a|A)} \times \mathsf{P}(A)}{\mathsf{P}(\neg A) + \frac{\mathsf{P}(a|A)}{\mathsf{P}(\neg a|A)} \times \mathsf{P}(A)} \\
&= \frac{\frac{\mathsf{P}(a|A)\mathsf{P}(A)}{\mathsf{P}(\neg a|A)}}{\frac{\mathsf{P}(\neg a|A)\mathsf{P}(\neg A)}{\mathsf{P}(\neg a|A)} + \frac{\mathsf{P}(a|A)\mathsf{P}(A)}{\mathsf{P}(\neg a|A)}} \\
&= \frac{\mathsf{P}(a|A)\mathsf{P}(A)}{\mathsf{P}(\neg a|A)\mathsf{P}(\neg A) + \mathsf{P}(a|A)\mathsf{P}(A)}
\end{aligned}$$

Now, on the assumption that witness' sensitivity is equal to their specificity, we have $\mathsf{P}(a|\neg A) = \mathsf{P}(\neg a|A)$ and can substitute this in the denominator:

$$= \frac{\mathsf{P}(a|A)\mathsf{P}(A)}{\mathsf{P}(a|\neg A)\mathsf{P}(\neg A) + \mathsf{P}(a|A)\mathsf{P}(A)}$$

and this would be a formulation of Bayes' theorem. And indeed with $\mathsf{P}(A) = p_{11} + p_{10}$ the formula works (albeit its adequacy rests on the identity of $\mathsf{P}(a|\neg A)$ and $\mathsf{P}(\neg a|A)$), and yields the result that we already obtained:

$$\begin{aligned}
\mathsf{P}(A|a) &= \frac{\lambda(p_{11} + p_{10})}{1 - (p_{11} + p_{10}) + \lambda(p_{11} + p_{10})} \\
&= \frac{1.5 \times 0.8}{1 - 0.8 + 1.5 \times 0.8} \approx 0.8571
\end{aligned}$$

The situation cannot be much improved by taking **a** and **b** to be high. For instance, if they're both 0.9 and $pr = \langle 0.1, 0.7, 0.1, 0.1\rangle$, the posterior of A is ≈ 0.972, the posterior of B is ≈ 0.692, and yet the joint posterior of $A \wedge B$ is 0.525.

The situation cannot also be improved by saying that at least if the threshold is 0.5, then as soon as **a** and **b** are above 0.7 (and, *a fortriori*, so are λ and μ), the individual posteriors being above 0.5 entails the joint posterior being above 0.5 as well. For instance, for $\mathbf{a} = 0.7$ and $\mathbf{b} = 0.9$ with $pr = \langle 0.1, 0.3, 0.5, 0.1\rangle$, the individual posteriors of A and B are ≈ 0.608 and ≈ 0.931 respectively, while the joint posterior of $A \wedge B$ is ≈ 0.283.

Optional Content Ends

The situation cannot be improved by saying that what was meant was rather that the joint likelihood is going to be at least as high as the maximum of the individual likelihoods, because quite the opposite is the case: the joint likelihood is going to be lower than any of the individual ones.

Optional Content Starts

Let us make sure this is the case. We have:

$$\begin{aligned}
LR(a|A) &= \frac{\mathsf{P}(a|A)}{\mathsf{P}(a|\neg A)} \\
&= \frac{\mathsf{P}(a|A)}{\mathsf{P}(\neg a|A)} \\
&= \frac{\mathbf{a}}{\mathbf{1} - \mathbf{a}}.
\end{aligned}$$

where the substitution in the denominator is legitimate only because witness' sensitivity is identical to their specificity.

With the joint likelihood, the reasoning is just a bit more tricky. We will need to know what $\mathsf{P}(a \wedge b|\neg(A \wedge B))$ is. There are three disjoint possible conditions in which the condition holds: $A \wedge \neg B, \neg A \wedge B$, and $\neg A \wedge \neg B$. The probabilities of $a \wedge b$ in these three scenarios are respectively $\mathbf{a}(\mathbf{1}-\mathbf{b}), (\mathbf{1}-\mathbf{a})\mathbf{b}, (\mathbf{1}-\mathbf{a})(\mathbf{1}-\mathbf{b})$ (again, the assumption of independence is important), and so on the assumption $\neg(A \wedge B)$ the probability of $a \wedge b$ is:

$$\begin{aligned}\mathsf{P}(a \wedge b|\neg(A \wedge B)) &= \mathbf{a}(1-\mathbf{b}) + (1-\mathbf{a})\mathbf{b} + (1-\mathbf{a})(1-\mathbf{b}) \\ &= \mathbf{a}(1-\mathbf{b}) + (1-\mathbf{a})(\mathbf{b}+1-\mathbf{b}) \\ &= \mathbf{a}(1-\mathbf{b}) + (1-\mathbf{a}) \\ &= \mathbf{a} - \mathbf{ab} + 1 - \mathbf{a} = 1 - \mathbf{ab}\end{aligned}$$

So, on the assumption of witness independence, we have:

$$\begin{aligned}LR(a \wedge b|A \wedge B) &= \frac{\mathsf{P}(a \wedge b|A \wedge B)}{\mathsf{P}(a \wedge b|\neg(A \wedge B))} \\ &= \frac{\mathbf{ab}}{\mathbf{1}-\mathbf{ab}}\end{aligned}$$

With $0 < \mathbf{a}, \mathbf{b} < 1$ we have $\mathbf{ab} < \mathbf{a}$, $1 - \mathbf{ab} > 1 - \mathbf{a}$, and consequently:

$$\frac{\mathbf{ab}}{\mathbf{1}-\mathbf{ab}} < \frac{\mathbf{a}}{\mathbf{1}-\mathbf{a}}$$

which means that the joint likelihood is going to be lower than any of the individual ones.

Optional Content Ends

Fact 1 is so far the most optimistic reading of the claim that if witnesses are independent and fairly reliable, their testimonies are going to provide positive support for the conjunction,\footnote{And this is the reading that Dawid in passing suggests: "the combined support is always positive, in the sense that the posterior probability of the case always exceeds its prior probability." (Dawid, 1987: 95) and any stronger reading of Dawid's suggestions fails. But Fact 1 is not too exciting when it comes to answering the original DAC. The original question focused on the adjudication model according to which the deciding agents are to evaluate the posterior probability of the whole case conditional on all evidence, and to convict if it is above a certain threshold. The problem, generally, is that it might be the case that the pieces of evidence for particular elements of the claim can have high likelihood and posterior probabilities of particular elements can be above the threshold while the posterior joint probability will still fail to meet the threshold. The fact that the joint posterior will be higher than the joint prior does not help much. For instance, if $\mathbf{a} = \mathbf{b} = 0.7$, $pr = \langle 0.1, 0.5, 0.3, 0.1\rangle$, the posterior of A is ≈ 0.777, the posterior of B is ≈ 0.608 and the joint posterior is ≈ 0.216 (yes, it is higher than the joint prior = 0.1, but this does not help the conjunction to satisfy the decision standard).

To see the extent to which Dawid's strategy is helpful here, perhaps the following analogy might be useful.

Imagine it is winter, the heating does not work in my office and I am quite cold. I pick up the phone and call maintenance. A rather cheerful fellow picks up the phone. I tell him what my problem is, and he reacts:

— Oh, don't worry.
— What do you mean? It's cold in here!
— No no, everything is fine, don't worry.
— It's not fine! I'm cold here!
— Look, sir, my notion of it being warm in your office is that the building provides some improvement to what the situation would be if it wasn't there. And you agree that you're definitely warmer than you'd be if your desk was standing outside, don't you? Your, so to speak, posterior warmth is higher than your prior warmth, right?

Dawid's discussion is in the vein of the above conversation. In response to a problem with the adjudication model under consideration Dawid simply invites us to abandon thinking in terms of it and to abandon requirements crucial for the model. Instead, he puts forward a fairly weak notion of support (analogous to a fairly weak sense of the building providing improvement), according to which,

assuming witnesses are fairly reliable, if separate fairly reliable witnesses provide positive support to the conjuncts, then their joint testimony provides positive support for the conjunction.

As far as our assessment of the original adjudication model and dealing with DAC, this leaves us hanging. Yes, if we abandon the model, DAC does not worry us anymore. But should we? And if we do, what should we change it to, if we do not want to be banished from the paradise of probabilistic methods?

Having said this, let me emphasize that Dawid's paper is important in the development of the debate, since it shifts focus on the likelihood ratios, which for various reasons are much better measures of evidential support provided by particular pieces of evidence than mere posterior probabilities.

Before we move to another attempt at a probabilistic formulation of the decision standard, let us introduce the other hero of our story: the gatecrasher paradox. It is against DAC and this paradox that the next model will be judged.

Optional Content Starts

In fact, Cohen replied to Dawid's paper (Cohen, 1988). His reply, however, does not have much to do with the workings of Dawid's strategy, and is rather unusual. Cohen's first point is that the calculations of posteriors require odds about unique events, whose meaning is usually given in terms of potential wagers – and the key criticism here is that in practice such wagers cannot be decided. This is not a convincing criticism, because the betting-odds interpretations of subjective probability do not require that on each occasion the bet should really be practically decidable. It rather invites one to imagine a possible situation in which the truth could be found out and asks: how much would we bet on a certain claim in such a situation? In some cases, this assumption is false, but there is nothing in principle wrong with thinking about the consequences of false assumptions.

Second, Cohen says that Dawid's argument works only for testimonial evidence, not for other types thereof. But this claim is simply false – just because Dawid used testimonial evidence as an example that he worked through it by no means follows that the approach cannot be extended. After all, as long as we can talk about sensitivity and specificity of a given piece of evidence, everything that Dawid said about testimonies can be repeated *mutatis mutandis*.

Third, Cohen complaints that Dawid in his example worked with rather high priors, which according to Cohen would be too high to correspond to the presumption of innocence. This also is not a very successful rejoinder. Cohen picked his priors in the example for the ease of calculations, and the reasoning can be run with lower priors. Moreover, instead of discussing the conjunction problem, Cohen brings in quite a different problem: how to probabilistically model the presumption of innocence, and what priors of guilt should be appropriate? This, indeed, is an important problem; but it does not have much to do with DAC, and should be discussed separately.

6.2 Problems with Cheng's relative likelihood

How is RLP supposed to handle DAC? Consider an imaginary case, used by Cheng to discuss this issue. In it, the plaintiff claims that the defendant was speeding (S) and that the crash caused her neck injury (C). Thus, H_Π is $S \wedge C$. Suppose that given total evidence E, the conjuncts, taken separately, meet the decision standard of RLP:

$$\frac{\mathsf{P}(S|E)}{\mathsf{P}(\neg S|E)} > 1 \qquad\qquad \frac{\mathsf{P}(C|E)}{\mathsf{P}(\neg C|E)} > 1$$

The question, clearly, is whether $\frac{\mathsf{P}(S \wedge C|E)}{H_\Delta|E} > 1$. But to answer it, we have to decide what H_Δ is. This is the point where Cheng's remark that H_Δ isn't normally simply $\neg H_\Pi$. Instead, he insists, there are three alternative defense scenarios: $H_{\Delta_1} = S \wedge \neg C$, $H_{\Delta_2} = \neg S \wedge C$, and $H_{\Delta_3} = \neg S \wedge \neg C$. How does H_Π compare to each of them? Cheng (assuming independence) argues:

$$\begin{aligned}
\frac{\mathsf{P}(S \wedge C|E)}{\mathsf{P}(S \wedge \neg C|E)} &= \frac{\mathsf{P}(S|E)\mathsf{P}(C|E)}{\mathsf{P}(S|E)\mathsf{P}(\neg C|E)} = \frac{\mathsf{P}(C|E)}{\mathsf{P}(\neg C|E)} > 1 \\
\frac{\mathsf{P}(S \wedge C|E)}{\mathsf{P}(\neg S \wedge C|E)} &= \frac{\mathsf{P}(S|E)\mathsf{P}(C|E)}{\mathsf{P}(\neg S|E)\mathsf{P}(C|E)} = \frac{\mathsf{P}(S|E)}{\mathsf{P}(\neg S|E)} > 1 \\
\frac{\mathsf{P}(S \wedge C|E)}{\mathsf{P}(\neg S \wedge \neg C|E)} &= \frac{\mathsf{P}(S|E)\mathsf{P}(C|E)}{\mathsf{P}(\neg S|E)\mathsf{P}(\neg C|E)} > 1
\end{aligned} \tag{12}$$

It seems that whatever the defense story is, it is less plausible than the plaintiff's claim. So, at least in this case, whenever elements of a plaintiff's claim satisfy the decision standard proposed by RLP, then so does their conjunction.

Similarly, RLP is claimed to handle the gatecrasher paradox. It is useful to think about the problem in terms of odds and likelikoods, where the *prior odds* (before evidence E) of H_Π as compared to H_Δ, are $\frac{\mathsf{P}(H_\Pi)}{\mathsf{P}(H_\Delta)}$, the posterior odds of H_Δ given E are $\frac{\mathsf{P}(H_\Pi|E)}{\mathsf{P}(H_\Delta|E)}$, and the corresponding likelihood ratio is $\frac{\mathsf{P}(E|H_\Pi)}{\mathsf{P}(E|H_\Delta)}$.

Now, with this notation the *odds form of Bayes' Theorem* tells us that the posterior odds equal the likelihood ratio multiplied by prior odds:

$$\frac{\mathsf{P}(H_\Pi|E)}{\mathsf{P}(H_\Delta|E)} = \frac{\mathsf{P}(E|H_\Pi)}{\mathsf{P}(E|H_\Delta)} \times \frac{\mathsf{P}(H_\Pi)}{\mathsf{P}(H_\Delta)}$$

[@cheng2012reconceptualizing: 1267] insists that in civil trials the prior probabilities should be equal. Granted this assumption, prior odds are 1, and we have:

$$\frac{\mathsf{P}(H_\Pi|E)}{\mathsf{P}(H_\Delta|E)} = \frac{\mathsf{P}(E|H_\Pi)}{\mathsf{P}(E|H_\Delta)} \tag{13}$$

This means that our original task of establishing that the left-hand side is greater than 1 now reduces to establishing that so is the right-hand side, which means that RLP tells us to convict just in case:

$$\mathsf{P}(E|H_\Pi) > \mathsf{P}(E|H_\Delta) \tag{14}$$

Thus, (14) tells us to convict just in case $LR(E) > 1$.

Now, in the case of the gatecrasher paradox, our evidence is statistical. In our variant E=`991 out of 1000 spectators gatecrashed''. Now pick a random spectator, call him Tom, and let $H_\Pi$`=Tom gatecrashed." (Cheng, 2012: 1270) insists:

> But whether the audience member is a lawful patron or a gatecrasher does not change the probability of observing the evidence presented.

So, on his view, in such a case, $\mathsf{P}(E|H_\Pi) = \mathsf{P}(E|H_\Delta)$, the posterior odds are, by (13), equal to 1, and conviction is unjustified.

There are various issues with how RLP has been deployed to resolve the difficulties that CLP and TLP run into.

First of all, to move from (1) to (2), Cheng assumes that the costs of wrongful decision is the same, be it conviction or acquittal. This is by no means obvious. If a poor elderly lady sues a large company for serious health damage that it supposedly caused, leaving her penniless if the company is liable is definitely not on a par with mistakenly making the company lose a small percent of their funds. Even in cases where such costs are equal, careful consideration and separate argument is needed. If, for instance, $c_1 = 5c_2$, we are to convict just in case $5 < \frac{p_\Pi}{p_\Delta}$. This limits the applicability of Cheng's reasoning about DAC, because his reasoning, if correct (and I will argue that it is not correct later on), yields only the result that the relevant posterior odds are greater than 1, not that they are greater than 5. The difficulty, however, will not have much impact on Cheng's solution of the gatecrasher paradox, as long as $c_1 \leq c_2$. This is because his reasoning, if correct (and I will argue that it is not correct later on), establishes that the relevant posterior odds are below 1, and so below any higher threshold as well.

Secondly, Cheng's resolution of DAC uses another suspicious assumption. For (12) to be acceptable we need to assume that the following pairs of events are independent conditionally on E: $\langle S, C\rangle$, $\langle S, \neg C\rangle$, $\langle \neg S, C\rangle$, $\langle \neg S, \neg C\rangle$. Otherwise, Cheng would not be able to replace conditional probabilities of corresponding conjunctions with the result of multiplication of conditional probabilities of the conjuncts. But it is far from obvious that speeding and neck injury are independent. If, for instance, the evidence makes it certain that if the car was not speeding, the neck injury was not caused by the accident, $\mathsf{P}(\neg S \wedge C|E) = 0$, despite the fact that $\mathsf{P}(\neg S|E)\mathsf{P}(C|E)$ does not have to be 0!

Without independence, the best that we can get, say for the first line of (12), is:

$$\mathsf{P}(S \wedge C|E) = \mathsf{P}(C|E)\mathsf{P}(S|C \wedge E)$$
$$\mathsf{P}(S \wedge \neg C|E) = \mathsf{P}(\neg C|E)\mathsf{P}(S|\neg C \wedge E)$$

and even if we know that $\mathsf{P}(C|E) > \mathsf{P}(\neg C|E)$, this tells us nothing about the comparison of $\mathsf{P}(S \wedge C|E)$ and $\mathsf{P}(S \wedge \neg C|E)$, because the remaining factors can make up for the former inequality.

Perhaps even more importantly, much of the heavy lifting here is done by the strategic splitting of the defense line into multiple scenarios. The result is rather paradoxical. For suppose $\mathsf{P}(H_\Pi|E) = 0.37$ and the probability of each of the defense lines given E is 0.21. This means that H_Π wins with each of the scenarios, so, according to RLP, we should find for the plaintiff. On the other hand, how eager are we to convict once we notice that given the evidence, the accusation is rather false, because $\mathsf{P}(\neg H_\Pi|E) = 0.63$?

The problem generalizes. If, as here, we individualize scenarios by boolean combinations of elements of a case, the more elements there are, into more scenarios $\neg H_\Pi$ needs to be divided. This normally would lead to the probability of each of them being even lower (because now $\mathsf{P}(\neg H_\Pi)$ needs to be "split" between more different scenarios). So, if we take this approach seriously, the more elements a case has, the more at disadvantage the defense is. This is clearly undesirable.

In the process of solving the gatecrasher paradox, to reach (13), Cheng makes another controversial assumption: that the prior odds should be one, that is, that before any evidence specific to the case is obtained, $\mathsf{P}(H_\Pi) = \mathsf{P}(H_\Delta)$. One problem with this assumption is that it is not clear how to square this with how Cheng handles DAC. For there, he insisted we need to consider *three different* defense scenarios, which we marked as $H_{\Delta_1}, H_{\Delta_2}$ and H_{Δ_3}. Now, do we take Cheng's suggestion to be that we should have

$$\mathsf{P}(H_\Pi) = \mathsf{P}(H_{\Delta_1}) = \mathsf{P}(H_{\Delta_2}) = \mathsf{P}(H_{\Delta_3})?$$

Given that the scenarios are jointly exhaustive and pairwise exclusive this would mean that each of them should have prior probability 0.25 and, in principle that the prior probability of guilt can be made lower simply by the addition of elements under consideration. This conclusion seems suboptimal.

If, on the other hand, we read Cheng as saying that we should have $\mathsf{P}(H_\Pi) = \mathsf{P}(\neg H_\Pi)$, the side-effect is that even a slightest evidence in support of H_Π will make the posterior probability of H_Π larger than that of $\neg H_\Pi$, and so the plaintiff can win their case way too easily. Worse still, if $\mathsf{P}(\neg H_\Pi)$ is to be divided between multiple defense scenarios against which H_Π is to be compared, then as soon as this division proceeds in a non-extreme fashion, the prior of each defense scenario will be lower than the prior of H_Π, and so from the perspective of RLP, the plaintiff does not have to do anything to win (as long as the defense does not provide absolving evidence), because his case is won without any evidence already!

Finally, let us play along and assume that in the gatecrasher scenario the conviction is justified just in case (14) holds. Cheng insists that it does not, because $\mathsf{P}(E|H_\Pi) = \mathsf{P}(E|H_\Delta)$. This supposedly captures the intuition that whether Tom paid has no impact on the statistics that we have.

But this is not obvious. Here is one way to think about this. Tom either paid the entrance fee or did not. Consider these two options, assuming nothing else about the case changes. If he did pay, then he is among the 9 innocent spectators. But this means that if he had not paid, there would have been 992 gatecrashers, and so E would be false (because it says there was 991 of them). If, on the other hand, Tom in reality did not pay (and so is among the 991 gatecrashers), then had he paid, there would have been only 990 gatecrashers and E would have been false, again!

So whether conviction is justified and what the relevant ratios are depends on whether Tom really paid. Cheng's criterion (14) results in the conclusion that Tom should be penalized if and only if he did not pay. But this does not help us much when it comes to handling the paradox, because the reason why we needed to rely on E was exactly that we did not know whether Tom paid.

If you are not buying into the above argument, here is another way to state the problem. Say your priors are $\mathsf{P}(E) = e$, $\mathsf{P}(H_\Pi) = \pi$. By Bayes' Theorem we have:

$$\mathsf{P}(E|H_\Pi) = \frac{\mathsf{P}(H_\Pi|E)e}{\pi}$$
$$\mathsf{P}(E|H_\Delta) = \frac{\mathsf{P}(H_\Delta|E)e}{1-\pi}$$

Assuming our posteriors are taken from the statistical evidence, we have $\mathsf{P}(H_\Pi|E) = 0.991$ and

$\mathsf{P}(H_\Delta|E) = 0.009$. So we have:

$$\begin{aligned} LR(E) &= \frac{\mathsf{P}(H_\Pi|E)e}{\pi} \times \frac{1-\pi}{\mathsf{P}(H_\Delta|E)e} \\ &= \frac{\mathsf{P}(H_\Pi|E) - \mathsf{P}(H_\Pi|E)\pi}{\mathsf{P}(H_\Delta|E)\pi} \\ &= \frac{0.991 - 0.991\pi}{0.009\pi} \end{aligned} \tag{15}$$

and $LR(E)$ will be > 1 as soon as $\pi < 0.991$. This means that contrary to what Cheng suggested, in any situation in which the prior probability of guilt is less than the posterior probability of guilt, RLP tells us to convict. This, however, does not seem desirable.

6.3 Problem's with Kaplow's stuff

Kaplow does not discuss the conceptual difficulties that we are concerned with, but this will not stop us from asking whether DTLP can handle them (and answering to the negative). Let us start with DAC.

Say we consider two claims, A and B. Is it generally the case that if they separately satisfy the decision rule, then so does $A \wedge B$? That is, do the assumptions:

$$\frac{\mathsf{P}(E|A)}{\mathsf{P}(E|\neg A)} > \frac{\mathsf{P}(\neg A)}{\mathsf{P}(A)} \times \frac{L}{G}$$
$$\frac{\mathsf{P}(E|B)}{\mathsf{P}(E|\neg B)} > \frac{\mathsf{P}(\neg B)}{\mathsf{P}(B)} \times \frac{L}{G}$$

entail

$$\frac{\mathsf{P}(E|A \wedge B)}{\mathsf{P}(E|\neg(A \wedge B))} > \frac{\mathsf{P}(\neg(A \wedge B))}{\mathsf{P}(A \wedge B)} \times \frac{L}{G}?$$

Alas, the answer is negative.

Optional Content Starts

This can be seen from the following example. Suppose a random digit from 0-9 is drawn; we do not know the result; we are told that the result is < 7 ($E =$'the result is < 7'), and we are to decide whether to accept the following claims:

A	the result is < 5.
B	the result is an even number.
$A \wedge B$	the result is an even number < 5.

Suppose that $L = G$ (this is for simplicity only — nothing hinges on this, counterexamples for when this condition fails are analogous). First, notice that A and B taken separately satisfy (5). $\mathsf{P}(A) = \mathsf{P}(\neg A) = 0.5$, $\mathsf{P}(\neg A)/\mathsf{P}(A) = 1$ $\mathsf{P}(E|A) = 1$, $\mathsf{P}(E|\neg A) = 0.4$. (5) tells us to check:

$$\frac{\mathsf{P}(E|A)}{\mathsf{P}(E|\neg A)} > \frac{L}{G} \times \frac{\mathsf{P}(\neg A)}{\mathsf{P}(A)}$$
$$\frac{1}{0.4} > 1$$

so, following DTLP, we should accept A.
For analogous reasons, we should also accept B. $\mathsf{P}(B) = \mathsf{P}(\neg B) = 0.5$, $\mathsf{P}(\neg B)/\mathsf{P}(B) = 1$ $\mathsf{P}(E|B) = 0.8$, $\mathsf{P}(E|\neg B) = 0.6$, so we need to check that indeed:

$$\frac{\mathsf{P}(E|B)}{\mathsf{P}(E|\neg B)} > \frac{L}{G} \times \frac{\mathsf{P}(\neg B)}{\mathsf{P}(B)}$$
$$\frac{0.8}{0.6} > 1$$

But now, $\mathsf{P}(A \wedge B) = 0.3$, $\mathsf{P}(\neg(A \wedge B)) = 0.7$, $\mathsf{P}(\neg(A \wedge B))/\mathsf{P}(A \wedge B) = 2\frac{1}{3}$, $\mathsf{P}(E|A \wedge B) = 1$, $\mathsf{P}(E|\neg(A \wedge B)) = 4/7$ and it is false that:

$$\frac{\mathsf{P}(E|A \wedge B)}{\mathsf{P}(E|\neg(A \wedge B))} > \frac{L}{G} \times \frac{\mathsf{P}(\neg(A \wedge B))}{\mathsf{P}(A \wedge B)}$$
$$\frac{7}{4} > \frac{7}{3}$$

The example was easy, but the conjuncts are probabilistically dependent. One might ask: are there counterexamples that involve claims which are probabilistically independent?[19]

Consider an experiment in which someone tosses a six-sided die twice. Let the result of the first toss be X and the result of the second one Y. Your evidence is that the results of both tosses are greater than one ($E =: X > 1 \wedge Y > 1$). Now, let A say that $X < 5$ and B say that $Y < 5$.

The prior probability of A is $2/3$ and the prior probability of $\neg A$ is $1/3$ and so $\frac{\mathsf{P}(\neg A)}{\mathsf{P}(A)} = 0.5$. Further, $\mathsf{P}(E|A) = 0.625$, $\mathsf{P}(E|\neg A) = 5/6$ and so $\frac{\mathsf{P}(E|A)}{\mathsf{P}(E|\neg A)} = 0.75$ Clearly, $0.75 > 0.5$, so A satisfies the decision standard. Since the situation with B is symmetric, so does B.

Now, $\mathsf{P}(A \wedge B) = (2/3)^2 = 4/9$ and $\mathsf{P}(\neg(A \wedge B)) = 5/9$. So $\frac{\mathsf{P}(\neg(A \wedge B))}{\mathsf{P}(A \wedge B)} = 5/4$. Out of 16 outcomes for which $A \wedge B$ holds, E holds in 9, so $\mathsf{P}(E|A \wedge B) = 9/16$. Out of 20 remaining outcomes for which $A \wedge B$ fails, E holds in 16, so $\mathsf{P}(E|\neg(A \wedge B)) = 4/5$. Thus, $\frac{\mathsf{P}(E|A \wedge B)}{\mathsf{P}(E|\neg(A \wedge B))} = 45/64 < 5/4$, so the conjunction does not satisfy the decision standard.

Optional Content Ends

Let us turn to the gatecrasher paradox.

Suppose $L = G$ and recall our abbreviations: $\mathsf{P}(E) = e$, $\mathsf{P}(H_\Pi) = \pi$. DTLP tells us to convict just in case:

$$LR(E) > \frac{1 - \pi}{\pi}$$

From (15) we already now that

$$LR(E) = \frac{0.991 - 0.991\pi}{0.009\pi}$$

so we need to see whether there are any $0 < \pi < 1$ for which

$$\frac{0.991 - 0.991\pi}{0.009\pi} > \frac{1 - \pi}{\pi}$$

Multiply both sides first by 009π and then by π:

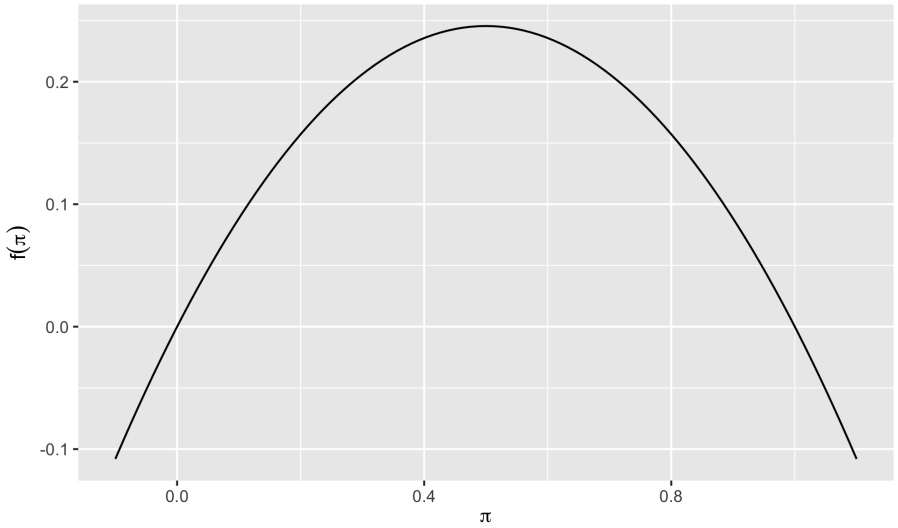
$$0.991\pi - 0.991\pi^2 > 0.09\pi - 0.009\pi^2$$

Simplify and call the resulting function f:

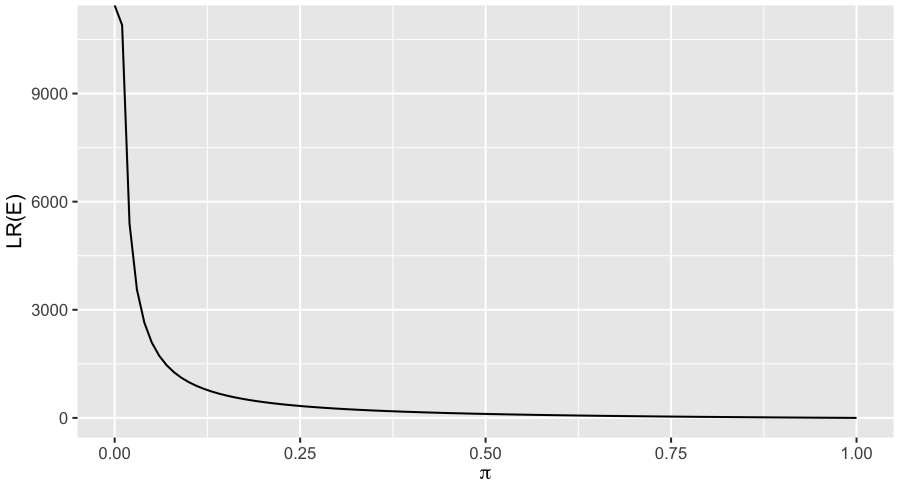
$$f(\pi) = -0.982\pi^2 + 0.982\pi > 0$$

The above condition is satisfied for any $0 < \pi < 1$ (f has two zeros: $\pi = 0$ and $\pi = 1$). Here is a plot of f:

[19] Thanks to Alicja Kowalewska for pressing me on this.



Similarly, $LR(E) > 1$ for any $0 < \pi < 1$. Here is a plot of $LR(E)$ against π:



Notice that $LR(E)$ does not go below 1. This means that for $L = G$ in the gatecrasher scenario DTLP wold tell us to convict for any prior probability of guilt $\pi \neq 0, 1$.

One might ask: is the conclusion very sensitive to the choice of L and G? The answer is, not too much.

Optional Content Starts

How sensitive is our analysis to the choice of L/G? Well, $LR(E)$ does not change at all, only the threshold moves. For instance, if $L/G = 4$, instead of f we end up with

$$f'(\pi) = -0.955\pi^2 + 0.955\pi > 0$$

and the function still takes positive values on the interval $(0, 1)$. In fact, the decision won't change until L/G increases to ≈ 111. Denote L/G as ρ, and let us start with the general decision standard, plugging

in our calculations for $LR(E)$:

$$\begin{aligned}
LR(E) &> \frac{\mathsf{P}(H_\Delta)}{\mathsf{P}(H_\Pi)}\rho \\
LR(E) &> \frac{1-\pi}{\pi}\rho \\
\frac{0.991-0.991\pi}{0.009\pi} &> \frac{1-\pi}{\pi}\rho \\
\frac{0.991-0.991\pi}{0.009\pi}\frac{\pi}{1-\pi} &> \rho \\
\frac{0.991\pi-0.991\pi^2}{0.009\pi-0.009\pi^2} &> \rho \\
\frac{\pi(0.991-0.991\pi)}{\pi(0.009-0.009\pi)} &> \rho \\
\frac{0.991-0.991\pi}{0.009-0.009\pi} &> \rho \\
\frac{0.991(1-\pi)}{0.009(1-\pi)} &> \rho \\
\frac{0.991}{0.009} &> \rho \\
110.1111 &> \rho
\end{aligned}$$

Optional Content Ends

So, we conclude, in usual circumstances, DTLP does not handle the gatecrasher paradox.

7 Probabilistic Thresholds Revised

7.1 Likelihood ratios and naked statistical evidence

7.2 Conjcution paradox and Bayesian networks

8 Conclusions

Where are we, how did we get here, and where can we go from here? We were looking for a probabilistically explicated condition Ψ such that the trier of fact, at least ideally, should accept any relevant claim (including G) just in case $\Psi(A, E)$.

From the discussion that transpired it should be clear that we were looking for a Ψ satisfying the following desiderata:

conjunction closure If $\Psi(A, E)$ and $\Psi(B, E)$, then $\Psi(A \wedge B, E)$.

naked statistics The account should at least make it possible for convictions based on strong, but naked statistical evidence to be unjustified.

equal treatment the condition should apply to any relevant claim whatsoever (and not just a selected claim, such as G).

Throughout the paper we focused on the first two conditions (formulated in terms of the difficulty about conjunction (DAC), and the gatecrasher paradox), going over various proposals of what Ψ should be like and evaluating how they fare. The results can be summed up in the following table:

View	Convict iff	DAC	Gatecrasher
Threshold-based LP (TLP)	Probability of guilt given the evidence is above a certain threshold	fails	fails
Dawid's likelihood strategy	No condition given, focus on $\frac{P(H\|E)}{P(H\|\neg E)}$	- If evidence is fairly reliable, the posterior of $A \wedge B$ will be greater than the prior. - The posterior of $A \wedge B$ can still be lower than the posterior of any of A and B. - Joint likelihood, contrary do Dawid's claim, can also be lower than any of the individual likelihoods.	fails
Cheng's relative LP (RLP)	Posterior of guilt higher than the posterior of any of the defending narrations	The solution assumes equal costs of errors and independence of A and B conditional on E. It also relies on there being multiple defending scenarios individualized in terms of combinations of literals involving A and B.	Assumes that the prior odds of guilt are 1, and that the statistics is not sensitive to guilt (which is dubious). If the latter fails, tells to convict as long as the prior of guilt < 0.991.
Kaplow's decision-theoretic LP (DTLP)	The likelihood of the evidence is higher than the odds of innocence multiplied by the cost of error ratio	fails	convict if cost ratio < 110.1111

Thus, each account either simply fails to satisfy the desiderata, or succeeds on rather unrealistic assumptions. Does this mean that a probabilistic approach to legal evidence evaluation should be abandoned? No. This only means that if we are to develop a general probabilistic model of legal decision standards, we have to do better. One promising direction is to go back to Cohen's pressure against **Requirement 1** and push against it. A brief paper suggesting this direction is (Di Bello, 2019), where the idea is that the probabilistic standard (be it a threshold or a comparative wrt. defending narrations) should be applied to the whole claim put forward by the plaintiff, and not to its elements. In such a context, DAC does not arise, but **equal treatment** is violated. Perhaps, there are independent reasons to abandon it, but the issue deserves further discussion. Another strategy might be to go in the direction of employing probabilistic methods to explicate the narration theory of legal decision standards (Urbaniak, 2018), but a discussion of how this approach relates to DAC and the gatecrasher paradox lies beyond the scope of this paper.

9 References

Bernoulli, J. (1713). *Ars conjectandi*.

Buchak, L. (2014). Belief, credence, and norms. *Philosophical Studies*, *169*(2), 285–311.

Cheng, E. (2012). Reconceptualizing the burden of proof. *Yale LJ*, *122*, 1254. HeinOnline.

Cohen, J. (1977). *The probable and the provable*. Oxford University Press.

Cohen, L. J. (1988). The difficulty about conjunction in forensic proof. *The Statistician*, *37*(4/5), 415. JSTOR. Retrieved from https://doi.org/10.2307/2348767

Dawid, A. P. (1987). The difficulty about conjunction. *The Statistician*, 91–97. JSTOR.

Dekay, M. L. (1996). The difference between Blackstone-like error ratios and probabilistic standards of proof. *Law and Social Inquiry*, *21*, 95–132.

Dhami, M. K., Lundrigan, S., & Mueller-Johnson, K. (2015). Instructions on reasonable doubt: Defining the standard of proof and the jurors task. *Psychology, Public Policy, and Law, 21(2), 169178*, *21*(2), 169–178.

Diamond, H. A. (1990). Reasonable doubt: To define, or not to define. *Columbia Law Review*, *90*(6), 1716–1736.

Di Bello, M. (2019). Probability and plausibility in juridical proof. *International Journal of Evidence and Proof*.

Finkelstein, M. O., & Fairley, W. B. (1970). A bayesian approach to identification evidence. *Harvard Law Review*, 489–517. JSTOR.

Friedman, R. D. (2000). A presumption of innocence, not of even odds. *Stanford Law Review*, *52*(4), 873–887.

Haack, S. (2014). Legal probabilism: An epistemological dissent. In *Haack2014-HAAEMS* (pp. 47–77).

Horowitz, I. A., & Kirkpatrick, L. C. (1996). A concept in search of a definition: The effect of reasonable doubt instrcutions on certainty of guilt standards and jury verdicts. *Law and Human Behaviour*, *20*(6), 655–670.

Kaplan, J. (1968). Decision theory and the fact-finding process. *Stanford Law Review*, *20*(6), 1065–1092.

Kaplow, L. (2014). Likelihood ratio tests and legal decision rules. *American Law and Economics Review*, *16*(1), 1–39. Oxford University Press.

Kaye, D. H. (1979). The laws of probability and the law of the land. *The University of Chicago Law Review*, *47*(1), 34–56.

Kaye, D. H. (1986). Do we need a calculus of weight to understand proof beyond a reasonable doubt? *Boston University Law Review*, *66*(3-4).

Laplace, P. (1814). *Essai philosophique sur les probabilités*.

Laudan, L. (2006). *Truth, error, and criminal law: An essay in legal epistemology*. Cambridge University Press.

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

Newman, J. O. (1993). Beyon "reasonable doub". *New York University Law Review*, *68*(5), 979–1002.

Urbaniak, R. (2018). Narration in judiciary fact-finding: A probabilistic explication. *Artificial Intelligence and Law*, 1–32.

Walen, A. (2015). Proof beyond a reasonable doubt: A balanced retributive account. *Louisiana Law Review*, *76*(2), 355–446.

Wells, G. (1992). Naked statistical evidence of liability: Is subjective probability enough? *Journal of Personality and Social Psychology*, *62*(5), 739–752. American Psychological Association.