Likelihood ratio and decision thresholds

Marcello Di Bello and Rafal Urbaniak

SAMLE CHAPTER PLAN UPDATE

I am now realizing that perhpas the structure of the chapter could be further broken down into three chapters:

1. A chapter that shows how probabilistc thresholds are good as analytical tools, despite implementation or practical difficulties with them. This chapter would include discussions of expected utility, minimizing errors, signal detection theory, etc. A lot of this stuff is already in the extended version of the SEP entry. So main claim of this chayer is: yes, probabilistic threshold are not good practically, but they can still be good as analytical tools. Title: "Probability Thresholds as Analytical Models of Trial Decision Making"
2. Two chapters that looks at the two theoretical difficulties (naked stats and conjunction paradox, ad also problem of priors). One chapter on naked statistical evidence and our informal solutions to it, based either on LR or on specific narratives (this should be followed by another chapter with the formal details).
3. Another chapter on conjunction paradox and our informal solution to it, maybe in terms of LR, BF or narratives (followed by another chapter in which the formal details are spelled out).
4. A chapter that formally addresses the two theoretical difficulties, perhaps using Bayesian Networks. This need not be included in the sample chapters we sent out. Title: "Adressing the Proof Paradoxes with Bayesian Networks".

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 SAMPLE CHAPETR TITLE - "Probability Thresholds as Analytical Models of Trial Decision Making" 3

2 Introduction 3

3 Probability thresholds 4
- 3.1 The basic idea . . . 4
- 3.2 Mixed reactions from legal practitioners . . . 5
- 3.3 Implementation and idealization . . . 6
- 3.4 Minimizing expected costs: higher cost ratio, higher threshold . . . 7
- 3.5 Beyond just costs: the benefits of correct decisions . . . 9
- 3.6 SUGGESTION . . . 12
- 3.7 Minimizing overall errors . . . 12
- 3.8 Interval thresholds (Finkelstein) . . . 17

4 Theoretical challenges - NEW CHAPTER WOULD STATRT HERE 17
- 4.1 The problem of priors . . . 17
- 4.2 Naked statistical evidence . . . 17

5 Conjuction paradox – NEW CHAPTER HERE DEVOTED TO PROBABILITY BASED SOLUTIONS 18
- 5.1 The problem . . . 19
- 5.2 Aggregating hypotheses and evidence . . . 20
- 5.3 Probabilistic (in)dependencies . . . 20
- 5.4 A closer look at the conjunction principle . . . 21
- 5.5 Prior probabilities and evidential support . . . 22
- 5.6 Bayes factor threshold . . . 25
- 5.7 Likelihood ratio threshold . . . 30
- 5.8 Should evidential strength be sensitive to priors? . . . 34
- 5.9 Which Measure of (Combined) Evidential Support? . . . 38
- 5.10 The comparative strateGy . . . 38
- 5.11 Specific Narratives [IDEAS OF A SOLUTION] . . . 40

6 The likelihood strategy 41
- 6.1 Kaplow . . . 42
- 6.2 Dawid . . . 42

6.3 Likelihood and DAC 44
6.4 Kaplow 46

7 Challenges (again) 46
7.1 Likelihood ratio and the problem of the priors 46
7.2 Dawid's likelihood strategy doesn't help 46
7.3 Problem's with Kaplow's stuff 50

8 Probabilistic Thresholds Revised 54
8.1 Likelihood ratios and naked statistical evidence 54
8.2 Conjunction paradox and Bayesian networks 54

9 Conclusions 54

References 55

1 SAMPLE CHAPETR TITLE - "Probability Thresholds as Analytical Models of Trial Decision Making"

2 Introduction

After the evidence has been presented, examined and cross-examined at trial, trained judges or lay jurors must reach a decision. In many countries, 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 meets 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 meet the preponderance standard, but not 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 the murder of 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.

Not sure if it is clear what you mean by this.

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, sometimes conflicting, paraphrases of what these standards mean. The meaning of 'proof beyond a reasonable doubt' is the most controversial. It has been equated with 'moral certainty' or 'abiding conviction' (Commonwealth v. Webster, 59 Mass. 295, 320, 1850) or with '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).

To further complicate things, differences between countries and legal traditions exist. The 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.'

R: check the formulation in Poland

While there are inevitable differences between legal traditions, the question of how strong the

evidence should be to warrant a finding of civil or criminal liability has universal appeal. Any system of adjudication whose decisions are informed by evidence will confront this question in one way or another. Not all legal systems will explicitly formulate standards of proof for trial decisions. Some legal systems may specify rules about how evidence should be weighed without formulating decision criteria such as standards of proof. But even without explicit proof standards, the triers of facts, judges or jurors, will have to decide whether the evidence is sufficient to judge the defendant legally liable.

Need to revise this when the chapter is done.

We will not survey the extensive legal literature and case law about proof standards. We will instead examine 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 challenges 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 literature: 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 trial proceedings.

Here you sound like you're gonna list a bunch of reasons but you give only one. Consider adding reasons or reformulating this bit.

3 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 defendant who faces criminal charges. The prosecution presents evidence to support its accusation, and the defense offers counterevidence. As a trier of fact, you are confronted with the question whether the totality of the evidence presented at trial warrants a conviction. More specifically, the question is whether the evidence as a whole establishes the defendant's guilt beyond a reasonable doubt.

3.1 The basic idea

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, 1979a; Laplace, 1814; Laudan, 2006). On this interpretation, so long as the defendant's guilt 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 the threshold.

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 is tacitly assumed in empirical studies about people's understanding of proof beyond a reasonable doubt (Dhami, Lundrigan, & Mueller-Johnson, 2015). This research examines where decision-makers set the bar for convictions, say at 80% or 90% probability, but does not question the assumption that standards of proof function as probabilistic thresholds 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 their 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 their versions of the facts with, say, 75-80% probability.

3.2 Mixed reactions from legal practitioners

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).

The Anderson decision is all the more interesting as it applies to a civil case. The German court did not think trial decisions could rest on a probability, not even in a civil case. The same conclusion would be less surprising if it applied to a criminal case. For example, Buchak (2014) has argued that an attribution of criminal culpability is an ascription of blame which requires a full belief in someone's guilt, and a proposition that is highly probable on the evidence, no matter how high its probability, cannot amount to a full belief. One is left wondering, however. If a high probability of guilt short of 100% isn't enough and certainty cannot be required either, how else could the standard of proof be met? The question becomes more pressing in civil cases if we replace 'guilt' with 'civil liability'. Anticipating this worry, Germany's Federal Court in the Anderson case endorsed a conception of proof standards that acknowledges the inveitable fallibility of trial decisions while at the same time maintaining 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 dilemmas that bedeviled early theorists of probability and evidence law. When Jacob Bernoulli—one of the pionerres of probability theory—discusses the requirement for a criminal conviction in his *Ars Conjectandi* (1713), 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 with a threshold probability of guilt. A few decades later, the Italian legal penologist Cesare Beccaria in his celebrated treatise *On Crimes and Punishments* (1764) remarks that the certainty needed to convict is 'nothing but a probability, though a probability of such a sort to be called certainty' (chapter 14). This suggestive yet admittedly elusive remark indicates that the standard of decision in criminal trials should be a blend of probability and certainty. But what this blend of probability and certainty should be like is unclear. At best, Beccaria's suggestion brings us back to paraphrases of proof beyond a reasonable doubt such as 'moral certainty' or 'abiding conviction'.

Not all legal practitioners, however, resist a probabilistic interpretation of standards of proof. Some actually find such interpretation plausible, even obvious. For example, Justice Harlan of the United States Supreme Court writes:

> . . . in a judicial proceeding in which there is a dispute about the facts of some earlier event, the factfinder cannot acquire unassailably accurate knowledge of what happened. Instead, all the factfinder can acquire is a belief of what probably happened. The intensity of this belief – the degree to which a factfinder is convinced that a given act actually occurred – can, of course, vary. In this regard, a standard of proof represents an attempt to instruct the factfinder concerning the degree of confidence our society thinks he should have in the correctness of factual conclusions for a particular type of adjudication.[1]

After this methodological premise, Justice Harlan explicitly endorses a probabilistic interpretation of standards of proof, using the expression 'degree of confidence' instead of 'probability':

> Although the phrases 'preponderance of the evidence' and 'proof beyond a reasonable doubt' are quantitatively imprecise, they do communicate to the finder of fact different notions concerning the degree of confidence he is expected to have in the correctness of his

[1] In re Winship, 397 U.S. 358, 370 (1970). This is a landmark decision by the United States Supreme Court establishing that the beyond a reasonable doubt standard must be applied to both adults and juvenile defendants.

factual conclusions.

Justice Newman of United States Court of Appeals for the Second Circuit proposes a similar definition. He worries that words such as 'probability' or 'likelihood' may confuse jurors, as these words often qualify predictions about future events that may or may not occur. Instead, Newman prefers the expression 'degree of certainity'. In chracterizing proof beyond a reasonable doubt, he writes:

> Were I the trier of fact, I would think about my own degree of certainty about the defendant's guilt, and, with a scale of 0 to 100 in mind, not vote to convict unless my degree of certainty exceeded 95 on that scale (p. 269) CITED FROM: Jon O. Newman (2006), Quantifying the standard of proof beyond a reasonable doubt: acomment on three comments. Law, Probability and Risk. 5, pp. 267-269

You only talk about Harlan; it would be nice to have more examples of people embracing probabilistic explications. M: GOOD POINT. ADDED MORE. NEED TO ADD EVEN MORE.

OTHER EXAMPLES TO CHECK. SEE REFERENCES BELOW:

FRANKLIN, J. (2006) Case Comment-United States v. Copeland, 369 F. Supp. 2d 365 (E.D.N.Y. 2005): quantification of the 'proof beyond reasonable doubt' standard. Law, Probability and Risk, 5, 159-165.

TILLERS, P. and GOTTFRIED, J. (2006) Case Comment-United States v. Copeland, 369 F. Supp. 2d 365 (E.D.N.Y. 2005): A Collateral Attack on the Legal Maxim That Proof Beyond a Reasonable Doubt Is Unquantifiable? Law, Probability and Risk, 5, 135-157.

WEINSTEIN, J. B. and DEWSBURY, I. (2006) Comment on the meaning of 'proof beyond a reasonable doubt'. Law, Probability and Risk, 5, 167-173.

3.3 Implementation and idealization

The remarks by Justice Harlan, Newman and others notwithstanding, legal practioners seem in general opposed to quantifying standards of proof probabilistically. This resistance has many causes. One key factor is the conviction that a probabilistic intepretation of legal standards of proof is unrealistic because its implementation would face unsurmountable challenges. How can the relevant probabilities—such as the probability of someone's guilt—be quantified? How will the triers of facts apply probabilistic thresholds? Should the application of the thresholds be automatic—that is, if the evidence meets the threshold, the triers of fact should find against the defendant (say, convict in a criminal trial) and otherwise find in favor of the defendant? The challenge, in general, is to articulate how probabilistic thresholds can be operationalized as part of trial decisions. This is by no means obvious. Judges and jurors do not weigh evidence in an explicitly probabilistic manner. Nor do they explicitly use probability thresholds to guide their decisions. CITE EMPIRICAL EVIDENCE HERE FOR NARRATIVE AND PLAUSIBILITY THEORY AGAINST PROBABILISTIC APPROACH. And even if judges and jurors were to change the way they assess, and reasona about, the evidence presented at trial there would remain the problem of computainal tractability. How could all the variables needed to assess the relevant probabilities be taken into acount in a compution ally tractable way? CITE ALLEN. SAY MORE ABOUT COMPUTIONAL TRACTBILITY EVEN FOR BAYESIAN NETWORK.

To alleviate the force of these worries, the probabilistic interpretion of proof standards can be broken down into two separate claims, what we might call the 'quantification claim' and the 'threshold claim'. In a criminal trial, these claims would look 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 evidence in defense of, the accused).
THRESHOLD CLAIM	an appropriately high threshold guilt probability, say 95%, should be the decision criterion for criminal convictions.

Those worried about implementation might reason thusly. 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 should be rejected.

One way to answer this objection is to bite the bullet. Legal probabilists can admit that probabilistic thresholds constitute a revisionist theory. If they are to be implemented in trial proceedings, they will require changes. Jurors and judges will have to become familiar with probabilistic ideas. They will have to evaluate the strength of the evidence numerically, even for evidence that is not, on its face, quantitative in nature. But this response will simply highten the resistance toward a probabilistic intepretation

of proof standards, or at least, the likelihood of success of such a program of radical reform of trial proceedings is uncertain. But there is a less radical way for legal probabilists to respond, one that admits that legal probabilism tacitly assumes a certain degree of idealization.

Legal probabilists can admit they are not—at least, not yet—engaged with implementation or trial reform. More specifically, the quantification claim can be interpreted in at least two different ways. One interpretation is 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 be understood as an idealization or a regulative ideal. For instance, the authors of a book on probabilistic inference in forensic science write:

> 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. . . (Taroni, Biedermann, Bozza, Garbolino, & Aitken, 2014, (p. xv))

Even from a probabilist standpoint, the quantification of guilt can well be an idealization which has, primarily, a heuristic role. MAYBE ALSO ADD THAT RONALD ALLEN WOULD NOT OBJECT TO THIS, AS HE THINKS THAT PROBABILITY ARE TOOLS IN PLAUSIBILITY REASONING. ADD CITATION.

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, the standard 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 less likely to encounter resistance.

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 under this scenario judges and 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*. (Tribe, 1971)

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

3.4 Minimizing expected costs: higher cost ratio, higher threshold

Let's start with the simplest illustration of how the probabilistic interpretation of proof standards can serve as a theoretical, analytical tool for conceptual clarification. Standards of proof are usually ranked from the least demanding, such as preponderance of the evidence, to the most demanding, such as proof beyond a reasonable doubt, even though this distinction has not always been around.[2] Can we give a principled justification for the use of multiple standards and their ranking? A common argument is that more is at stake in a criminal trial than in a civil trial. A mistaken conviction 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 decisions in civil trials are mostly about imposing monetary compensation. This difference in the stakes warrants different standards of proof, more stringent for criminal than civil cases. This informal argument can be made precise by pairing probability thresholds

[2] See e.g. United States v. Feinberg, 140 F.2d 592 (2d Cir. 1944): " to distinguish between the evidence which should satisfy reasonable men, and the evidence which should satisfy reasonable men beyond a reasonable doubt. While at times it may be practicable to deal with these as separate without unreal refinements, in the long run the line between them is too thin for day to day use" (594).

with expected utility theory, a well-establish paradigm of rational decision-making used in psychology and economic theory.

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. This theory can be applied to a variety of situations, including civil or criminal trials. To see how this works, note that trial decisions can be factually erroneous in two ways. A trial decision can be a false positive—i.e. a decision to hold the defendant liable (to convict, in a criminal case) even though the defedant committed no wrong (or committed no crime). A trial decision can also be a false negative—i.e. a decision not to hold the defendant liable (or to acquit, in a criminal case) even though the defendant did commit the wrong (or committed the crime). Let $\mathsf{cost}(CI)$ and $\mathsf{cost}(AG)$ 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 $\mathsf{P}(G|E)$ and $\mathsf{P}(I|E)$ be the guilt probability and the innocence probability estimated on the basis of the evidence presented at trial. Given a simple decision-theoretic model (Kaplan, 1968), a conviction should be preferred to an acquittal whenever the expected cost resulting from a mistaken conviction—namely, $\mathsf{P}(I|E) \cdot \mathsf{cost}(CI)$—is lower than the expected cost resulting from a mistaken acquittal—namely, $\mathsf{P}(G|E) \cdot \mathsf{cost}(AG)$. That is,

$$\text{convict provided } \frac{\mathsf{cost}(CI)}{\mathsf{cost}(AG)} < \frac{\mathsf{P}(G|E)}{\mathsf{P}(I|E)}.$$ [3]

For the inequality to hold, the ratio of posterior probabilities $\frac{\mathsf{P}(G|E)}{\mathsf{P}(I|E)}$ should exceed the cost ratio $\frac{\mathsf{cost}(CI)}{\mathsf{cost}(AG)}$. So long as the costs can be quantified, the probability threshold can be determined. For example, suppose mistaken conviction is nine times as costly as a mistaken acquittal. The corresponding probability threshold will be 90%. On this reading, in order to meet the standard of proof beyond a reasonable doubt, the prosecution should provide evidence that establishes the defendant's guilt with at least 90% probability, or in formulas, $\mathsf{P}(G|E) > 90\%$. The higher the cost ratio $\frac{\mathsf{cost}(CI)}{\mathsf{cost}(AG)}$, the higher the requisite threshold. The lower the cost ratio, the lower the requiste threshold. For example, if the cost ratio is 99, the threshold would be as high as 99%, but if the cost ratio is 2, the threshold would only be 75%.

The same line of argument applies to civil cases. Let a false attribution of liability FL be a decision to find the defendant liable when the defendant committed no civil wrong (analogous to the conviction of an innocent in a criminal case). Let a false attribution of non-liability FNL be a decision not to find the defendant liable when the defendant did commit the civil wrong (analogous to the acquittal a factually guilty defendant in a criminal case). Let $\mathsf{P}(L|E)$ and $\mathsf{P}(NL|E)$ be the liability probability and the non-liability probability given the evidence presented at trial. So long as the objective is to minimize the costs of erroneous decisions, the rule of decision would be as follows:

$$\text{find the defendant civilly liable provided } \frac{\mathsf{cost}(FL)}{\mathsf{cost}(FNL)} < \frac{\mathsf{P}(L|E)}{\mathsf{P}(NL|E)}.$$ [4]

If the cost ratio $\frac{\mathsf{cost}(FL)}{\mathsf{cost}(FNL)}$ is set to 1, the threshold for liability judgments should equal 50%, a common intepretation of the preponderance standard in civil cases. This means that $\mathsf{P}(L|E)$ should be at least 50% for a defedant to be found civilly liable.

The difference between proof standards in civil and criminal cases lies in the different cost ratios. The cost ratio in civil cases, $\frac{\mathsf{cost}(FL)}{\mathsf{cost}(FNL)}$, is typically lower than the cost ratio in criminal cases, $\frac{\mathsf{cost}(CI)}{\mathsf{cost}(AG)}$, because a false positive in a criminal trial (a mistaken conviction) is considered a more harmful error than a false positive in a civil trial (a mistaken attribution of civil liability). This difference in the cost ratio can have a consequentialist or a retributivist justification (Walen, 2015). From a consequentialist perspective, the loss of personal freedom or even life can be considered a greater loss than being forced to pay an undue monetary compensation. From a retributivist perspective, the moral wrong that results from the mistaken conviction of an innocent person can be regarded as more egregious than the moral wrong that results from the mistaken attribution of civil liability. This difference in consequences or moral wrongs can be captured by positing a higher cost ratio in criminal than civil cases.

Along similar lines, Justice Harlan of the United Supreme Court draws a clear differencce in the cost ratio between criminal and civil litigation:

[3] This follows from $\mathsf{P}(I|E) \cdot \mathsf{cost}(CI) < \mathsf{P}(G|E) \cdot \mathsf{cost}(AG)$.

[4] This follows from $\mathsf{P}(NL|E) \cdot \mathsf{cost}(FL) < \mathsf{P}(L|E) \cdot \mathsf{cost}(FNL)$

> In a civil suit between two private parties for money damages, for example, we view it as no more serious in general for there to be an erroneous verdict in the defendant's favor than for there to be an erroneous verdict in the plaintiff's favor ... In a criminal case, on the other hand, we do not view the social disutility of convicting an innocent man as equivalent to the disutility of acquitting someone who is guilty. In Re Winship (1970), 397 U. S. 358, 371.

To underscore the differences in the cost ratios, Harlan cites an earlier decision of the United States Supreme Court that emphasizes how a defendant's liberty has a transcending value:

> [t]here is always in litigation a margin of error ..., representing error in factfinding, which both parties must take into account ... [w]here one party has at stake an interest of transcending value – as a criminal defendant his liberty – ... this margin of error is *reduced* as to him by the process of placing on the other party [i.e. the prosecutor] the standard of ... persuading the factfinder at the conclusion of the trial of his guilt beyond a reasonable doubt. Speiser v. Randall (1958), 357 U.S. 513, 525-26.

Justice Newman of the United States Court of Appeals for the Second Circuit made a similar point by linking directly the ratio of errors and the degree of certainity required for a conviction:

> ... all must recognize that factfinders are fallible and that any system of adjudicating guilt will inevitably run some risk of both convicting the innocent and acquitting the guilty....Whatever ratio [of false conviction to false acquittals] we find acceptable, one of the major variables in achieving that ratio is the degree of certainty we impose on factfinders. (p. 980) QUOTED FROM: Jon O. Newman (1993), Beyond Reasonable Doubt, New York University Law Review, 68(5), pp. 979-1002

3.5 Beyond just costs: the benefits of correct decisions

The analysis provided so far is limited since it only weighs the costs of mistaken decisions, but leaves out the benefits of correct decisions. A more comprehensive analysis should consider both. Even though the basic idea is the same—that is, trial decision-making is viewed as an instrument for maximizing overall social welfare (Dekay, 1996; Laudan, 2016; Posner, 1973)— a mre comprehensive analysis would afford a more nuanced understanding. It is therefore instructive to explore the implications of weighing the costs of incorrect decisions as well as the benefits of correct decisions.

For simplicity, we will quantify costs and benefits with units of utility using the abbreviation $\mathsf{ut}(...)$. Benefits will correspond to positive numbers and costs to negative numbers. In a criminal trial, the (negative) utility of a mistaken conviction should be weighed together the (negative) utility of an incorrect acquittal: $\mathsf{ut}(CI)$ v. $\mathsf{ut}(AG)$. In addition, the (positive) utility of a correct conviction should be weighed toghether with the (positive) utility of a correct acquittal: $\mathsf{ut}(CG)$ v. $\mathsf{ut}(AI)$. Given this set-up, a conviction would be justified provided its expected utility—that is, $\mathsf{P}(G|E)\cdot\mathsf{ut}(CG)+\mathsf{P}(I|E)\cdot\mathsf{ut}(CI)$—exceeds the expected utility of an acquittal—that is, $\mathsf{P}(G|E)\cdot\mathsf{ut}(AG)+\mathsf{P}(I|E)\cdot\mathsf{ut}(AI)$. By elementary algebraic steps, the threshold is identified by the equation:

$$\mathsf{P}(G|E) > \frac{1}{1+\frac{\mathsf{ut}(CG)-\mathsf{ut}(AG)}{\mathsf{ut}(AI)-\mathsf{ut}(CI)}}.$$ [5]

In a number of cases, this new formula returns the same threshold as the earlier one. If the benefits of convictions and acquittals are zero, this inequality identifies the same threshold as the earlier inequality that only considered costs. For example, if the costs of a mistaken conviction is nine times the cost of a mistaken acquittal, that is, $\frac{\mathsf{ut}(CI)}{\mathsf{ut}(AG)} = 9$, while the benefits are zero, the decision threhshold should be a guilt probability of $\frac{1}{1+(1/9)} = 0.9$, as before. Or suppose the magnitude of the benefits resulting from acquitting an innocent defendant (say +9 units of utility) is the same as the magnitude of the costs resulting from convicting an innocent (-9 units of utility). Similarly, suppose the magnitude of the benefits resulting from convicting a guilty defendant (say +1 unit of utility) is the same as the magnitude of the cost resulting from acquitting a guilty defendant (say -1 unit of utility). Again, the threshold would be $\frac{1}{1+(1+1)/(9+9)} = 0.9$.

But consider now the following utility assignments: $\mathsf{ut}(CI) = -9$, $\mathsf{ut}(AG) = -1$, $\mathsf{ut}(CG) = 5$, and $\mathsf{ut}(AI) = 5$. The correspoding threshhold would be $\frac{1}{1+(5+1)/(5+9)} = 0.7$, significanly below the 90%

[5] SHOW COMPUTATIONS HERE

threshold. If the benefits of correct decisions are further increaesed, say at 7 units of utility each, the threhsold would be lowered further to 66% since $\frac{1}{1+(7+1)/(7+9)} = 0.66$. So the the benefits of correct decisions are not at all inconsequential. In order to keep the threshold for criminal convictions releatively high the benefit of correct conviction would have to be close to zero (as seen before) or alternatively, the benefit of a correct acquittals would have to be significantly higher than the benefit of correct conviction. Say $\mathsf{ut}(CG) = 2$ but $\mathsf{ut}(AI) = 18$, while still $\mathsf{ut}(CI) = -9, \mathsf{ut}(AG) = -1$, the threhhold would be $\frac{1}{1+(2+1)/(18+9)} = 0.9$.

The new inequality shows that the threshold depends on the ratio of the difference between utilities, not so much the cost ratio or the benefit ratio. A given cost and benefit ratio may correspond to different thresholds. In the examples above, even though the cost ratio was fixed at 9:1 and the benefit ratio at 1:1, the threshold was 70% in one case and 66% in the other. The difference in the threshold is due to the difference in the absolute magnitute of the benefits resulting from correct decisions, increased from +5 units of utility to +7 units of utility.

A question suggests itself. How should utilities be assigned? The assignments may be a moral question (what the right conception of justice dictates), an empirical question (what the majority thinks the utilities should be), or a political question (how political ideologies affect utility assignments). For one thing, elected officials should set the appropriaate assignments of utilities, and elected officials should represent the will of the people. On other hand, it is curious that the standard of proof should be allowed to vary depending on the political party who is is charge at the moment. Perhaps, standards of proof should be part of a country's constitution and not vary depending on the political party in power.

We will now explore different strategies for assigning utility to correct and incorretc trial decisions. One stratgy is to identfy the main sources of harm and the main sources of benefits reslting from trial decision. The weight placed on each diferent harms or benefits is likely to be a matter of political idealogy.

Let's start with a politically neutral assignment of utilities. Consider the loss resulting from convicting an innocent defendant. This loss includes: the inappropriate assignment of culpability (-1); damaged reputation (-1); loss of income (-1); severance from family members (-1); putting other citizens at risk of victimization by failing to convict the actual perpetrator (-1); weakening the deterrence function of the trial system by failing to apprehend the perpetrator (-1). This is a total utility loss of -6. What about the correct acquittal of an innocent? There woud be no inappropriate assignment of culpability (0); no reputional damage (0); no loss of income (0); no severance from familiy members (0). The actual perpetrator, however, could still victimize others (-1) and deterrence would be weakened (-1). The innocent defendant who is acquitted could still experience damaged reputation and other harms, but we shall leave these details aside. All in all, trying an innocent defendant, no matter the final decision, carries the baseline costs of failing to identify the true perpetrator (-1) and putting other citizens at risk of victimization (-1). This baseline cost increases when an innocent person is wrongly convicted. So, by adding everything up, $\mathsf{ut}(AI) - \mathsf{ut}(CI) = -2 + 6 = +4$. Next, consider the loss resulting from acquitting a guilty defendant. This will include putting other citizens at risk of victimization by the perpetrator who is not convicted (-1) and possibly weaken the deterrence function of the trial system (-1). There would be no inappropriate assignment of culpability (0); no damaged reputation (0); no loss of income (0); no severance from family members (0). What about the conviction of a guilty defendant? The defendant would still experience damaged reputation (-1); loss of income (-1); severance from family members (-1). However, citizens would enjoy a lower risk of victimization (+1) and the deterrence function of the trial system would be reaffirmed (+1). There would also be no incorrect assertions about the citizen's culpabilty (0). All things considered, $\mathsf{ut}(CG) - \mathsf{ut}(AG) = -1 + 2 = +1$. So, assuming the utilities are correctly assigned, the threshold for a criminal conviction should be $\frac{1}{1+\frac{1}{4}} = 0.8$, lower than the earlier 90% threshold. (Incidentally, if the correct assignment of culpability is counted as a positive benefits (say +1) from a correct conviction, then $\mathsf{ut}(CG) - \mathsf{ut}(AG) = 0 + 2 = +2$ and the threshold would be $\frac{1}{1+\frac{2}{4}} = 0.6666$, an even lower threshold.)

The above analysis is—arguably—politically neutral because it takes into account the costs of a conviction, even for those who are factually guilty, such as loss of income and severance from familiy members. These are issues often emphasized by those on the left who are wary of the costs of criminalization and punitiveness, even for those who, strictly speaking, did committ a crime. At the same time, the above analysis also takes into account the negative consequences that result from failing to apprehend the actual perpetrator, such as hightened crime victimization for others, a point often made by people on the right who are concerned with so-called "law and order".

Suppose someone is extremely concerned about risk of victimization for different reasons, such as more conservative political views or having grown up in a high crime area. Suppose the costs resulting from the risk of victimization resulting from a false acquittal or the trial of an innoocent are increased from -1 to -3. Converserely, the benefits resulting from lower risk of of victimizatio associated with a correct conviction are also increased from +1 to +2. Then, $\mathsf{ut}(AI) - \mathsf{ut}(CI) = +4$, as before, but $\mathsf{ut}(CG) - \mathsf{ut}(AG) = 0 + 3 = +3$. The threshold would be $\frac{1}{1+\frac{3}{4}} = 0.57$, significantly lower than before. Unsurprisingly, someone who is very concerned with the risk of victimization will favor a lower threshold for conviction in criminal cases. (CITE LAUDAN and SAUNDERS HERE).

Let's now explore the view of someone—perhaps ore progressive, liberal and left-leaning—who is more concerned about negative effects of criminalization, such as loss of income and severance from family members. Suppose the utility loss is increased from -1 to -2 for each of these items. Then, $\mathsf{ut}(AI) - \mathsf{ut}(CI) = -2 + 8 = +6$, as before, but $\mathsf{ut}(CG) - \mathsf{ut}(AG) = -3 + 2 = -1$. The threshodl would be $\frac{1}{1+\frac{-1}{6}} = 1.2$. Interestingly, under this assignment of utilities, convciting a defendant never maximizes expected utility unless the evidence establishes guilt with 120% probability. This is clearly impossible. In other words, if the costs of criminalization are so high, convicting anyone is never justified, not even someone whose guillt is 100% probable.

This conclusion could show one of two things. First, it could show that the assignment of utility above is non-sensical, because it leads to the non-sensical conclusion that no one should ever be convicted. Alternatively, those who stand by that assignment of utilities (or one like that) will be committed to say that the practice of convicting people as we know it shoud be abolished, at least until the cost of criminalization become less burdensome. The latter view is by no means a non-starter, as it agrees with radical proposals about prison abolition. (CITE APPROPRIATE REFERENCES ABOUT PRISON ABOLITION).

As the above discussion shows, probabilistic thresholds, when paired with expected utility theory, provide an analytical framework to justify, or at least meaningfully debate, different degrees of stringency necessary for decision criteria—i.e. legal proof standards—in criminal trials. The same discusison could be had for civil trials.

This analytical framework allows for even finer distinctions, not explictily codified in the law. The law typically makes coarse distinctions between standards of proof, such as 'proof beyond a reasonable doubt' for criminal cases, 'preponderance of the evidence' for civil cases and 'clear and convincing evidence' for a narrow subset of civil cases in which the accusation against the defendant is particularly serious. But for rather different crimes, associated with rather different punishments, say murder and grand theft, the same standard of proof is applied for both. It is not obvious why this should be so, except that a finer distinction may cause more confusion than there need be. If the probability required for a conviction or a finding of civil liability against the defendant is a function of weighing the costs and benefits that would result from true and false positives (as well as true and false negatives), the stringency of the threshold should depend on costs and benefits, and thus different cases may require different thresholds. Cases in which the charge is more serious than others—say, murder compared to grand theft—may require higher thresholds so long as the cost of a mistaken decision against the defendant is more significant. In countries that allow for the death penalty or life imprisonment for certain crimes but not others, the cost of a mistaken conviction would be more serious for crimes with harsher punishments, other things being equal. Thus, the threhold should be placed appropriately higher. We could even think that the threhsold should vary across individual cases even for defendants charged with the exact same crime, provided the costs are different for different individuals. However, whether or not standards of proof should vary in this way is debated (Kaplow, 2012; Picinali, 2013). Ultimately, the question is what considerations should be admissible in the calculus of costs and benefits.

The discussion so far might have proceeded from the wrong assumption. To weigh the costs and benefits of convictions and acquittals, correct and incorrect, is one thing. It is another to weigh the costs and benefits of punishment. So perhaps the calculus should only strictly apply to trial decisions, not to what flows from them.

NEEDS MORE EXPLANATION HERE

1) OTHER QUESTION TO ADDREES IS WHAT COST AND BENEFITS? MAYBE ONLY COSTS AND BENEFIT OF A CONVICTION (BLAME ATTRIBUTION), NOT EVERYTHING THAT FOLLOWS FROM A CONVICTION.
2) RELATED POINT. SOME COSTS AND BENEFITS ARE ASSOCIATED WITH LITIGATION PER SE, OTHERS WITH PUNISHMENT, SO IN TALKING ABOUT COSTS AND BENEFITS

OF TRIAL DECISION PERHPAS WE SHOULD TAKE A MORE NARROW APPROACH. WHAT ARE THE UNIQUE COSTS AND BENEFITS OF TRIAL DECISION?

3) SOME COSTS AND BENEFITS ARE SHARED BY MULTIPLE TYPES OF DECISIONS. E.G. CONVICTION ALL DEPRIVE THE DEFENDANT OF INCOME AND FAMILY TIES, WHETHER THE DEFENDANT IS GUILTY OR INNOCENT. DOES THIS MEAN THESE COSTS AND BENEFITS ARE IRRELEVANT FOR THE CALCULUS OF UILITY?
4) WHAT THE ANALYSIS IS MISSING A LARGER LOOK AT THE CONTEXT OF THE TRIAL, PLEAE BARGAINING, THE CRIMINAL JUSTICE SYSTEM AND SOCIETY MORE GENERALLY. MULTI STAGE ANALYSYS. THE MAXIMIZATION OF EXP UTILITY FRAMEOWKR OBSCURES THIS COMPLEXITY.
5) COMPLEXITY PROBLEM. ALLEN.

3.6 SUGGESTION

MARCELLO: IF WE END UP DIVIDING THIS CHAPTER INTO TWO OR THREE SEPARATE CHAPTERS, WE COULD CONTINUE THE DISCUSISON OF THE ANALYTICAL POWER OF THE PROBABILITSTIC APPROACH MORE IN DETAIL HERE, DRAWING ON SOME OF THE MATERIALS ALREADY IN THE LONGER VERSION OF THE SEP ENTRY.

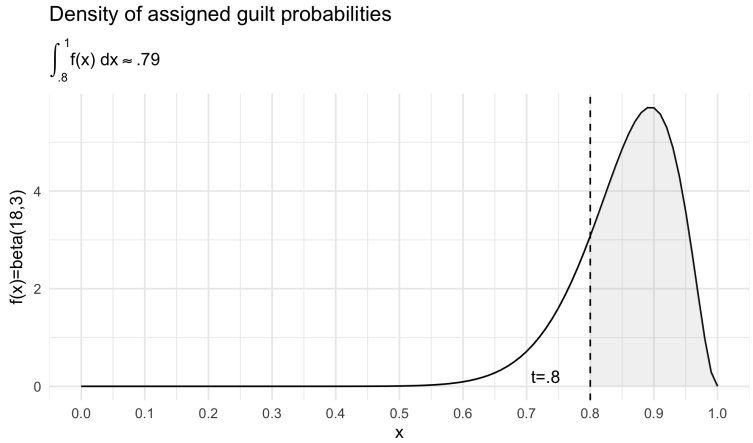
HERE IS A TENTATIVE IDEA OF WHAT TO DISCUSS:

(1) SIMPLE EXPECTED UTILITY MODEL - **DONE, SEE ABOVE**
(2) LAUDAN MODEL, THIS IS A MORE COMPLICATED EXPECTED UTILITY MODEL, PARTLY BORROWED FROM LAPLACE - **YET TO BE DONE**
(3) SIGNAL DETECTION THEORY MODEL - **YET TO BE DONE, ONLY PARTLY DONE**
(4) HAMER MODEL AND KAYE MODEL FOR ERROR MINIMIZATION (DISCUSSED IN THE SEP ENTRY, INTEGRALS, DERIVATIVES, ETC.) - **DONE SEE BELOW**
(5) GOOD AND BAD THINGS ABOUT THESE MODELS, BUT OVERALL THEY SHOW THAT THE PROBABILISTIC FRAMEWORK IS A RICH ANALYTICAL TOOL **YET TO BE DONE**

3.7 Minimizing overall errors

Instead of maximizing expected utility (or minimizing expected costs), standards of proof can be analyzed as decision criteria that have long term effects on the epistemic performance of the trial system. Think about the criminal justice as a whole, making decisions about the guilt and innocence of thousands of defendants facing trial. The system will make a number of decisional errors, committing type I and type II errors. Viewing standards of proof as probability thresholds helps to understand how decisional errors are managed and allocated at this systemic level.

Consider an idealized model of the criminal trial system. Each defendant is assigned a probability x of criminal liability (or guilt) based on the evidence presented at trial. As is customary, this probability ranges between 0 and 1, or 0% and 100%. Since over a period of time many defendants face charges, the guilt probability will have its own distribution. Extreme guilty probabilities set at 0% or 100%, presumably, are assigned rarely in trials if ever, while values between 40% and 80% are more common. A rigorous way to express this distribution is by means of a probability density function, call it $f(x)$. The figure below uses a right skewed distribution, for example, beta(18,3).

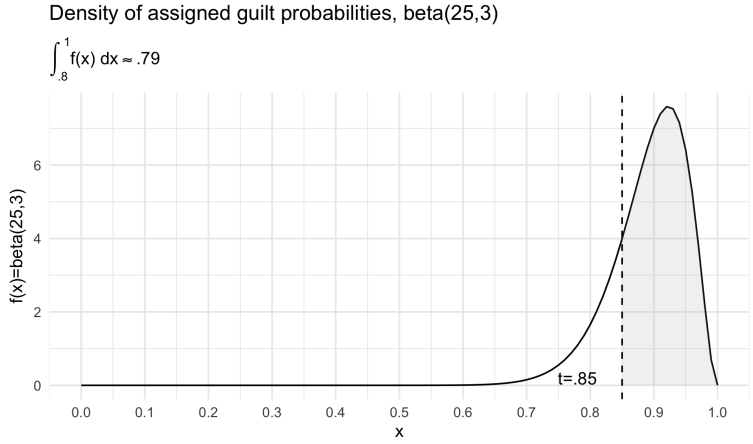


What does the distribution represent? Does it represent the probability of guilt assigned to defendants at the beginning or the end of the trial? There should be a difference between the two—hopefully—or else trial proceedings would be useless. Suppose that the distribution represents the guilt probabilities as they are assigned to defendants at the end of the trial, once all the evidence, counterevidence, arguments and counterarguments have been proferred and weighed appropriately.

The choice of the distribution is for illustrative purposes only. There are no empirical data suggesting this is the right distribution to use. But its choice is not arbitrary either. The right skew of the distribution reflects the assumption that defendants in criminal cases are prosecuted only if the incriminating evidence against them is strong. It should be no surprise that most defendants are assigned a high probability of guilt. This is plausible in principle. For people should not be prosecuted if the evidence against them is weak. The distribution of the probability of liability in civil cases over a period of time might look quite different, perhaps centered around 50% or 60%.

In the figure above, the threshold for conviction is set at > 80%, and the area under the curve to the right of the threshold is about .79. According to this model, 79% of defendants on trial are convicted and 21% acquitted. These figures are close to the rates of conviction and acquittal in many countries (REFERENCES?). Since $f(x)$ is a probability density, the total area under the curve adds up to 1, encompassing all defendants, both convicted and acquitted defendants.

If the threshold becomes more stringent—for example, it moves up to 85%—the rate of conviction would decrease. This holds provided the underlying distribution does not change. But, if the threshold is set higher, those who are prosecuted will tend to face comparatively stronger evidence and thus the distribution will become more skewed toward the right—say beta(25,3). As a consequence, the rate of conviction could still be about 79% even with a more stringent threshold of 85%.



The two graphs above depict the rate of conviction among those who are facing trial, not the rate of conviction in the general population overall. As just shown, the rate of conviction could remain the same even if the probability threhsold is made more stringent. But, the rate of conviction in the general population is likely to diminish so long as higher thresholds, by acting as deterrents against prosecution, make it less likely that people would be prosecuted .

This formal model does not yet make any distinction between factually guilt and factually innocet

defendants. But, presumably, some defendants committed the acts they are accused of and others did not. This is not a clear-cut distinction, however. Some defendants may have committed the acts they are accused of to some extent, but not to the full extent they are accused of, while others may be completely innocent of any crime whatsoever. Leaving this subtlety aside, the formal model can be refined to distinguish between factually innocent and guilty defendants.

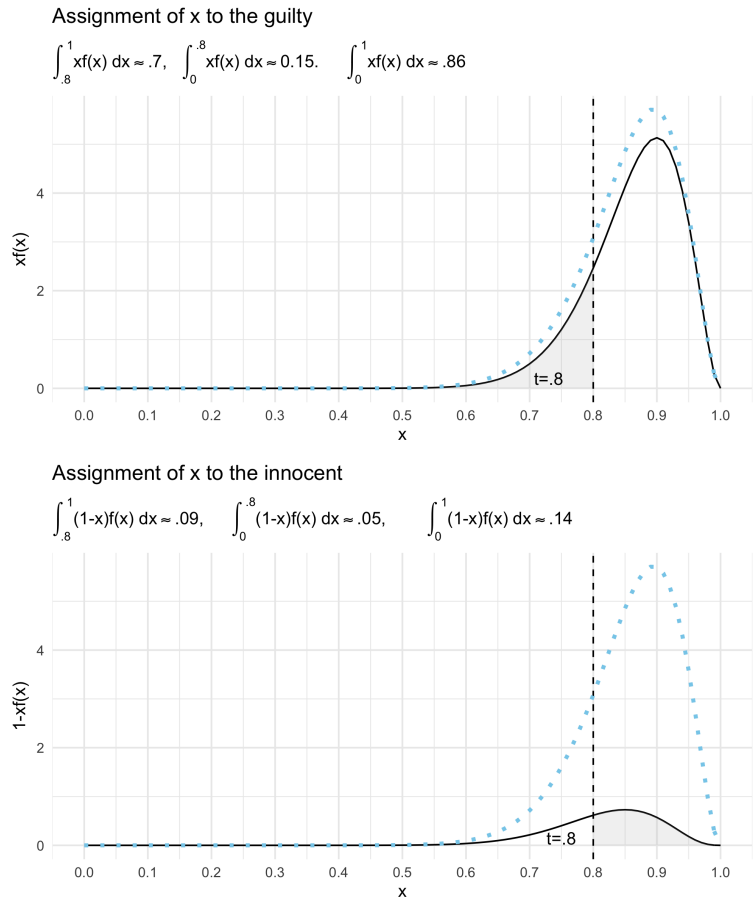
The simplest refinement would create two separate distributions, one distribution for the factually innocent defendants and the other for the factually guilty defendants. The problem with this is that we have little idea about what these distributions should look like in the first place. Hopefully, the innocent distribution will be more left skewed and the guilty distribution more right skewed. Guilty defendants should be assigned, on average, higher guilt probabilities than innocent defendants. The two distributions could still overlap to some extent as some guilt defendants could be assigned as low guilt probabilities as some innocent defendants and conversely some innocent defendants could be assigned as high guilt probabilities as some guilty defendants. This is unfortunate, but also an inevitable consequence of the fallibility of the trial system.

A more principled way to add two separate distributions to the model, one for guilty and another for innocent defendants, would be to derive them from the overall distribution of defendants. This can be done by following the simple principle that, among those defendants who are assigned a probability of, say, 80%, there should be a corresponding proportion of 80% guilty people and 20% innocent people. These are of coruse expected values, not actual values. Say you are throwing a fair six-faced die. In the long run, you would expect that in 1/6 of the throws the die would land, say on "4".

The expected proportion of guilty and innocent defendants on trial, out of all defendants, can be inferred from the density distribution $f(x)$ under certain assumptions. Suppose each defendant is assigned a guilt probability based on the best and most complete evidence. From the perspective of judges and jurors (or anyone who has access to the evidence and evaluates it the same way), x% of defendants who are assigned x% guilt probability are expected to be guilty and $(1-x)$% innocent. For example, 85% of defendants who are assigned a 85% guilt probability are expected to be guilty and 15% innocent; 90% of defendants who are assigned a 90% guilt probability are expected to be guilty and 10% innocent; and so on.

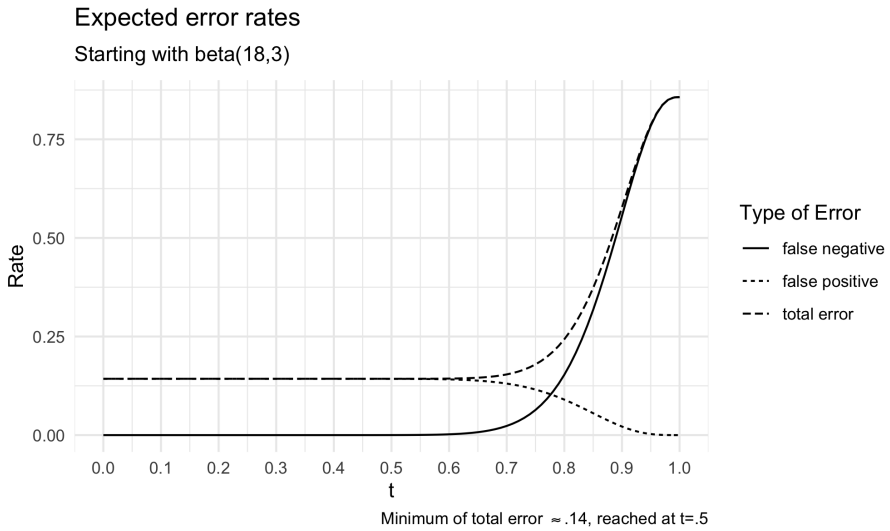
So the expected guilty distribution as a function of x will be $xf(x)$, while the expected innocent distribution will be $(1-x)f(x)$. In other words, the function $xf(x)$ describes the (expected) assignment of guilt probabilities for guilty defendants, and similarly, $(1-x)f(x)$ the (expected) assignment of guilt probabilities for innocent defendants. Neither of these functions is a probability density, since $\%\int_0^1 xf(x)\,\mathrm{d}x = 0.86$ and $\int_0^1 (1-x)f(x)\,\mathrm{d}x = 0.14$. These numbers express the (expected) proportion of guilty and innocent defendants out of all defendants on trial, respectively 86% and 14%.

The rates of incorrect decisions—false convictions and false acquittals or more generally false positives and false negatives—can be inferred from this model as a function of the threshold t (Hamer, 2004, 2014). The integral $\int_0^t xf(x)\,\mathrm{d}x$ equals the expected rate of false acquittals, or in other words, the expected proportion of guilty defendants who fall below threshold t (out of all defendants), and the integral $\int_t^1 (1-x)f(x)\,\mathrm{d}x$ equals the expected rate of false convictions, or in other words, the expected proportion of innocent defendants who fall above threshold t (out of all defendants). The rates of correct decisions—true convictions and true acquittals or more generally true positives and true negatives—can be inferred in a similar manner. The integral $\int_t^1 xf(x)\,\mathrm{d}x$ equals the expected rate of true convictions and $\int_0^t (1-x)f(x)\,\mathrm{d}x$ the expected rate of true acquittals. In the figure below, the regions shaded in gray correspond to false negatives (false acquittals) and false positives (false convictions). The remaining white regions within the solid black curve correspond to true positives (true convictions) and true negatives (true acquittals). Note that the dotted blue curve is the original overall distribution for all defendants.



The size of the grey regions in the figures above—which correspond to false positives and false negatives—is affected by the location of threshold t. As t moves upwards, the rate of false positives decreases but the rate of false negatives increases. Conversely, as t moves downwards, the rate of false positives increases but the rate of false negatives decreases. This trade-off is inescapable so long as the underlying distribution is fixed. We have already remarked on the possibility that the distribution would change in shape as a result of changes in the probability threhsold. We will retunr to this point later in the chapter.

Below are both error rates—false positives and false negatives—and their sum plotted against a choice of t, while holding fixed the density function binom(18,3). The graph shows that any threshold that is no greater than 50% would minimize the total error rate %(comprising false positives and false negatives). A more stringent threshold, say > 90%, would instead significantly reduce the rate of false positives but also significantly increase the rate of false negatives, es expected.



In general, the threshold that minimizes the expected rate of incorrect decisions overall, no matter the underlying distribution, lies at 50%. The claim that setting threshold at $t = .5$ minimizes the expected error rate holds given the distribution $f(x)$ =beta(18,3) as well as any other distribution (**???**). To show this, let $E(t)$ %as a function of threshold t be the sum of rates of false positive and false negative decisions:

$$E(t) = \int_0^t x f(x)\,\mathrm{d}x + \int_t^1 (1-x) f(x)\,\mathrm{d}x.$$

The overall rate of error is minimized when $E(t)$ is the lowest. To determine the value of t for which $E(t)$ is the lowest, set the derivative of $E(t)$ %and $R(t)$ to zero, that is, $\frac{d}{dt}E(t) = 0$. By calculus, $t = 1/2$.[6] This claims holds when the two decisional errors are assigned the same weight, or in other words, the costs of false positives and false negatives are symmetric. The > 50% threshold therefore should be most suitable for civil trials. In criminal trials, however, false convictions are typically considered significantly more costly than false acquittals, say a cost ratio of 9:1 (but see (Epps, 2015)). The sum of the two error rates can be weighted by their respective costs:

$$E(t) = \int_0^t x f(x)\,\mathrm{d}x + 9\int_t^1 (1-x) f(x)\,\mathrm{d}x.$$

Given a cost ratio of 9:1, the optimal threshold that minimizes the (weighted) overall rate of error is no longer 1/2, but rather, $t = 9/10 = 90\%$.[7]

Whenever the decision threshold is more stringent than > 50%, the overall (unweighted) error minimization may be sacrificed to pursue other goals, for example, protecting more innocents against mistaken convictions, even at the cost of making a larger number of mistaken trial decisions overall.

The standard 'proof beyond a reasonable doubt' is often paired with the Blackstone ratio, the principle that it is better that ten guilty defendants go free rather than even just one innocent be convicted. The

[6] Note that $\frac{d}{dt}E(t)$ is the the sum of the derivatives of $\int_0^t x f(x)\,\mathrm{d}x$ and $\int_t^1 (1-x) f(x)\,\mathrm{d}x$, that is,

$$\frac{d}{dt}E(t) = \frac{d}{dt}\int_0^t x f(x)\,\mathrm{d}x + \frac{d}{dt}\int_t^1 (1-x) f(x)\,\mathrm{d}x.$$

By the fundamental theorem of calculus,

$$\frac{d}{dt}\int_0^t x f(x)\,\mathrm{d}x = t f(t) \text{ and } \frac{d}{dt}\int_t^1 (1-x) f(x)\,\mathrm{d}x = -(1-t) f(t).$$

By plugging in the values,

$$\frac{d}{dt}E(t) = t f(t) - (1-t) f(t).$$

Since $\frac{d}{dt}E(t) = 0$, then $t f(t) = (1-t) f(t)$ and thus $t = 1 - t$, so $t = 1/2$ or a > 50% threshold.

[7] The proof is the same as before. Since $t f(t) = 9(1-t) f(t)$, it follows that $t = 9/10$.

exact ratio is a matter of controversy (Volokh, 1997). It is tempting to think that, say, a 99% threshold guarantees a 1:99 ratio between false convictions and false acquittals. But this would be hasty for at least two reasons. First, probabilistic thresholds affect the expected rate of mistaken decisions. The actual rate may deviate from its expected value (**???**-). Second, if the threshold is 99%, *at most* 1% of decision against defendants are expected to be mistaken (false convictions) and *at most* 99% of the decisions in favor of the defendant are expected to be mistaken (false acquittals). The exact ratio will depend on the probabilities assigned to defendants and how they are distributed (**?**). The (expected) rate of false positives and false negatives—and thus their ratio—depend on where the threshold is located but also on the distribution of the liability probability as given by the density function $f(x)$.

3.8 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 - NEW CHAPTER WOULD STATRT HERE

Let's take stock. We briefly examined difficulties in implementation for probabilistic standards of proof and set those aside. We then offered a few illustrations how probabilistic standards can be used as analytical tools to theorize about decision-making at trial. But even if probabilistic thresholds are used solely as analytical tools, legal probabilists are not yet out of the woods. Even if the practical problems can be addressed or set aside, theoretical difficulties remain. We will focus on three in particular: the problem of priors; naked statistical evidence; and the difficulty about conjunction, also called the conjunction paradox. The latter two are difficulties that any theory of the standard of proof – not just a probabilistic theory – should be able to address. The first difficulty is peculiar to the probabilistic interpretation of standards of proof. We will examine each difficulty in turn and then examine a promising line of response within legal probabilism based on likelihood ratios instead of posterior probabilities.

4.1 The problem of priors

4.2 Naked statistical evidence

Suppose one hundred, identically dressed prisoners are out in a yard during recreation. Suddenly, ninety-nine of them assault and kill the guard on duty. We know that this is what happened from a video recording, but we do not know the identity of the ninety-nine killers. After the fact, a prisoner is picked at random and tried. Since he is one of the prisoners who were in the yard, the probability of his guilt would be 99%. But despite the high probability, many have the intuition that this is not enough to establish guilt beyond a reasonable doubt. Hypothetical scenarios of this sort suggest that a high probability of guilt, while perhaps necessary, is not sufficient to establish guilt beyond a reasonable doubt.

Perhpas, the resistance in the prisoner scenario lies in the fact that the prisoner was picked at random, and that any prisoner would be 99% likely to be one of the killers. Since the statistics cannot single out the one innocent prisoner, they are bad evidence. But consider this case. Suppose two people enter a department store. There are no other customers in the store. After they exit the store, a member of the staff finds that an item of merchandise is missing. Since no staff member could be culpable—they are strictly surveilled—the culprit must be one of the customers. One of the customers, John, has scored high in a compulsivity test and has been arrested for stealing in department stores several times in the past. The other customer, Rick, has never been arrested for stealing in a department store and shows no sign of high compulsivity. Statistics show that people with a high degree of compulsivity and who

have stolen merchandise in department stores before are more likely than others to steal merchandise if they are unsupervised. So John is most likely the culprit. Suppose studies show that people like John, when unsupervised, will steal 99 times out of 100 times. Instead, people like Rick, when unsupervised, will only steal 1 time out of 100 times. So John is 99 times more likely than Rick to have stolen the merchandise. Can these statistics be enough to convict John? Again, it seems not. There is no evidence against him specifically, say, no merchandise was found on him that could link him to the crime. Many would feel uneasy about convicting John despite the fact that, between the two suspects, he is the one who is most likely the culprit.

A similar hypothetical can be constructed for civil cases. Suppose a bus company, Blue-Bus, operates 90% of the buses in town on a certain day, while Red-Bus only 10%. That day a bus injures a pedestrian. Although the buses of the two companies can be easily recognized because they are respectively painted blue and red, the pedestrian who was injured cannot remember the color of the bus involved in the accident. No other witness was around. Still, given the statistics about the market shares of the two companies, it is 90% probable that a Blue-Bus bus was involved in the accident. This is a high probability, well above the 50% threshold. Yet the 90% probability that a Blue-Bus bus was involved in the accident would seem—at least intuitively—insufficient for a judgment of liability against Blue-Bus. This intuition challenges the idea that the preponderance standard in civil cases only requires that the plaintiffi establish the facts with a probability greater than 50%.

Confronted with these hyptheticals, legal probabilists could push back. Hypotheticals rely on intuitive judgments, for example, that the high probability of the prisoners's guilt in the scenario above does not amount to proof beyond a reasonable doubt. But suppose we changed the numbers and imagined there were one thousand prisoners of whom nine hundred and ninety-nine killed the guard. The guilt probability of a prisoner picked at random would be 99.9%. Even in this situation, many would insist that guilt has not been proven beyond a reasonable doubt despite the extremely high probability of guilt. But others might say that when the guilt probability reaches such extreme values, values as high as 99.9% or higher, people's intuitive resistance to convicting should subside (Roth, 2010). A more general problem is that intuitions in such hypothetical scenarios are removed from real cases and thus are potentially unreliable as a guide to theorize about standards of proof (Allen & Leiter, 2001; Hedden & Colyvan, 2019; Lempert, 1986).

Another reason to be suspicious of these hypotheticals is that they seem to amplify biases in human reasoning. Say an eyewitness was present during the accident and testified that a Blue-Bus bus was involved. Intuitively, the testimony would be considered enough to rule against Blue-Bus, at least provided the witness survived cross-examination. We exhibit, in other words, an intuitive preference for judgments of liability based on testimonial evidence compared to judgments based on statistical evidence. This preference has been experimentally verified (Arkes, Shoots-Reinhard, & Mayes, 2012; Niedermeier, Kerr, & Messeé, 1999; Wells, 1992) and exists outside the law (Ebert, Smith, & Durbach, 2018; Friedman & Turri, 2015; Sykes & Johnson, 1999). But testimonial evidence is no less prone to error than statistical evidence. In fact, it may well be more prone to error. The unreliability of eyewitness testimony is well-known, especially when the environmental conditions are not optimal (Loftus, 1996). So are we justified in exhibiting an intuitive preference for eyewitness testimony as opposed to statistical evidence, or is this preference a cognitive bias to avoid?

These reservations notwithstanding, the puzzles about naked statistical evidence cannot be easily dismissed. Puzzles about statistical evidence in legal proof have been around for a while (Cohen, 1977; Kaye, 1979b; Nesson, 1979; Thomson, 1986). Philosophers and legal scholars have shown a renewed interest in both criminal and civil cases (Blome-Tillmann, 2017; Bolinger, 2018; Cheng, 2012; Di Bello, 2019b; Enoch, Spectre, & Fisher, 2012; Ho, 2008; Moss, 2018; Nunn, 2015; Pardo, 2018; Pritchard, 2005; Pundik, 2017; Redmayne, 2008; Roth, 2010; Smith, 2018; Stein, 2005; Wasserman, 1991). Given the growing interest in the topic, legal probabilism cannot be a defensible theoretical position without offering a story about naked statistical evidence.

5 Conjuction paradox – NEW CHAPTER HERE DEVOTED TO PROBABILITY BASED SOLUTIONS

Another theoretical difficulty that any theory of the standard of proof should address is the the conjunction paradox or difficulty about conjunction. First formulated by Cohen (1977), the difficulty about

conjunction has enjoyed a great deal of scholarly attention every since (Allen, 1986; Allen & Stein, 2013; Allen & Pardo, 2019; Haack, 2014; Schwartz & Sober, 2017; Stein, 2005). This difficulty arises when an accusation of wrongdoing, in a civil or criminal proceeding, is broken down into its constituent elements. The basic problem is that the probability of a conjuction is often lower than the probability of the conjuncts. Thus, even if each conjunct meets the requisite probability threshold, the conjunction does not. This chapter examines the difficulty about conjuction and how legal probabilists can respond.

5.1 The problem

Suppose that in order to prevail in a criminal trial, the prosecution should establish by the required standard, first, that the defendant caused harm to the victim (call it claim A), and second, that the defendant had premeditated the harmful act (call it claim B). Cohen (1977) argues that common law systems subscribe to a conjunction principle, that is, if A and B are established according to the governing standard of proof, so is their conjunction (and vice versa). If the conjunction principle holds, the following must be equivalent, where S is a placeholder for the standard of proof:

Separate	A is established according to S and B is established according to S
Overall	The conjunction $A \wedge B$ is established according to S

Let $S[X]$ mean that claim or hypothesis X is established according to standard S. Then, in other words, the conjunction principles requires that:

$$S[A \wedge B] \Leftrightarrow S[A] \wedge S[B].$$

The conjunction principle is consistent with—perhaps even required by—the case law. For example, the United States Supreme Court writes that in criminal cases

> the accused [is protected] against conviction except upon proof beyond a reasonable doubt of *every fact* necessary to constitute the crime with which he is charged. In re Winship (1970), 397 U.S. 358, 364.

A plausible way to interpret this quotation is to posit this identity: to establish someone's guilt beyond a reasonable doubt *just is* to establish each element of the crime beyond a reasonable doubt. Thus,

$$\mathsf{BARD}[A_1 \wedge \cdots \wedge A_n] \Leftrightarrow \mathsf{BARD}[A_1] \wedge \cdots \wedge \mathsf{BARD}[A_n],$$

where the conjunction $A_1 \wedge \cdots \wedge A_n$ comprises all the material facts that, according to the applicable law, constitute the crime with with the accused is charged.

The problem for the legal probabilist is that the conjunction principle conflicts with a threshold-based probabilistic interpretation of the standard of proof. For suppose the prosecution presents evidence that establishes claims A and B, separately, to the required probability, say about 95% each. Has the prosecution met the burden of proof? Each claim was established to the requisite probability threshold, and thus it was established to the requisite standard (assuming the threshold-based interpretation of the standard of proof). And if each claim was established to the requisite standard, then (i) guilt as a whole was established to the requisite standard (assuming the conjunction principle). But even though each claim was established to the requisite probability threshold, the probability of their conjunction—assuming the two claims are independent—is only $95\% \times 95\% = 90.25\%$, below the required 95% threshold. So (ii) guilt as a whole was *not* established to the requisite standard (assuming a threshold-based probabilistic interpretation of the standard). Hence, we arrive at two contradictory conclusions: (i) that the prosecution met its burden of proof and (ii) that it did not meet its burden.

The difficulty about conjunction—the fact that a probabilistic interpretation of the standard of proof conflicts with the conjunction principle—does not subside when the number of constituent claims increases. If anything, the difficulty becomes more apparent. Say the prosecution has established three separate claims to 95% probability. Their conjunction—again if the claims are independent—would be about 85% probable, even further below the 95% threshold. Nor does the difficulty about conjunction subside if the claims are no longer regarded as independent. The probabilty of the conjunction $A \wedge B$, without the assumption of independence, equals $\mathsf{P}(A|B) \times \mathsf{P}(B)$. But if claims A and B, separately, have been established to 95% probability, enough for each to meet the threshold, the probability of $A \wedge B$ could still be below the 95% threshold unless $\mathsf{P}(A|B) = 100\%$. For example, that someone premediated a harmful act against another (claim B) makes it more likely that they did cause harm in the end (claim

A). Since $\mathsf{P}(A|B) > \mathsf{P}(A)$, the two claims are not independent. Still, premeditation does not always lead to harm, so $\mathsf{P}(A|B)$ should be below 100%. Consequently, in this case, the probability of the conjunction $A \wedge B$ would be below the 95% threhsold.

False in whole generality, give a counterexample with more specific numbers. M: I changed things a bit. Maybe not it's clear now. The counterexample is basically P(A)=P(B)=0.95, but P(AB)=Pr(A)*P(A|B) and since P(A|B) is below 1, then P(AB) is below 0.95.

5.2 Aggregating hypotheses and evidence

So far the discussion proceeded without mentioning explicitly the evidence proferred in support of the different claims that constitute the allegation of wrongdoing. This is, however, a simplification. Say evidence a establishes claim A and other evidence b establishes a distinct claim B. The question now is whether the combination of a and b establishes the conjunction $A \wedge B$, and viceversa. In other words, the question is whether the following conjunction principle holds:

$$\mathsf{S}[a, A] \text{ and } \mathsf{S}[b, B] \text{ iff } \mathsf{S}[a \wedge b, A \wedge B],$$

where $\mathsf{S}[e, H]$ means that evidence e establishes hypothesis H by standard S. This conjunction principle differs from the earlier one since it mentions explicitly the evidence a and b.

Understood in terms of posterior probability, the conjunction principle above fails in some cases. For suppose that $\mathsf{P}(A|a) > t$ and $\mathsf{P}(B|b) > t$, for a threshold t, or in other words, given the evidence a and b, both A and B are sufficiently probable (for a fixed threshold). It does not generally follow that $A \wedge B$ is sufficienlty probable given the combined evidence $a \wedge b$. By the probability calculus,

$$\begin{aligned}\mathsf{P}(A \wedge B|a \wedge b) &= \mathsf{P}(A|a \wedge b) \times \mathsf{P}(B|a \wedge b \wedge A) \\ &= \mathsf{P}(A|a) \times \mathsf{P}(B|b)\end{aligned}$$

The second equality holds assuming certain relationships of independence, specifically, the independence of A from b given a, and of B from $a \wedge A$ given b. These relationships of independence do not always hold, but they do sometimes. For example, in an aggravated assault case, evidence a could be a witness testimony that the defendant physically injured the victim (claim A), and b evidence that the defendant knew that the victim was a firefighter (claim B), for example, another testimony that the defendant earlier called the firefighther for help. Presumably, $\mathsf{P}(A|a) = \mathsf{P}(A|a \wedge b)$ because the fact that the defendant called a firefighter for help (b) does not make it more (or less) likely that he would physically injure him (A). Further, $\mathsf{P}(B|b) = \mathsf{P}(B|a \wedge b \wedge A)$ because the fact that the defendant injured the victim (A) and there is a testimony to that effect (a) does not make it more (or less) likely that the victim was a firefighter (B). Given these assumptions, if—as is normally the case—neither $\mathsf{P}(A|a)$ nor $\mathsf{P}(B|b)$ equal 1, then

$$\mathsf{P}(A \wedge B|a \wedge b) < \mathsf{P}(A|a) \;\; \& \;\; \mathsf{P}(A \wedge B|a \wedge b) < \mathsf{P}(B|b).$$

This is another manifestation of the difficulty about conjuction. If each piece of evidence a and b establishes claims A and B with 95% probability, the combined evidence $a \wedge b$ need not establish the conjuction $A \wedge B$ with 95% probability. The conjunction principle fails here.

Interestingly, even if the independence assumptions are dropped, the difficulty about conjunction still arises in a number of circumstances. Suppose evidence $a \wedge b$ establishes claim A and also claim B, separately, right above the probability threshold t. Since $\mathsf{P}(A \wedge B|a \wedge b) = \mathsf{P}(A|a \wedge b) \times \mathsf{P}(B|a \wedge b \wedge A)$, it follows that $\mathsf{P}(A \wedge B|a \wedge b)$ would be below t so long as $\mathsf{P}(B|a \wedge b \wedge A)$ is below 100%, which would often be the case since (i) evidence is fallible and (ii) one hypothesis does not usually entail the other. So even though A and B are established to the required probability, the conjunction is not.

5.3 Probabilistic (in)dependencies

The conjunction paradox is a difficult problem, as the vast literature on the topic attests. Before we move on, it is important to become clear about the assumptions underlying the formulation of the paradox, in particular, the assumptions of probabilistic independence. We will then explore a number of probability-based proposals and distinguish the promising ones from those that ultimately fail.

One assumption often made in the formulation of the paradox is that claims A and B are probabilistically independent. This is not always the case, but we have seen that the paradox does subside if the two claims are dependent of one another. So the assumption, in formulating the paradox, should be that either the two claims are fully probabilistically independent or positively probabilistically dependent. We exclude cases in which the claim are negatively probabilistically dependent. For it would be odd

if the prosecution or the plaintiff would try to combine two claims where one makes the other less probable.

The other assumption—which is not always stated upfront in the presentation of the paradox—is that the supporting items of evidence are also independent. They are not, however, unconditionally probabilistically independent. They are independent conditionally on the claim they support. This notion of conditional independence captures formally the thought that two or more items of evidence constitute *independent lines of evidence* (SEE DISCUSSION IN EARLIER CHAPTERS).

Bayesian network are useful for representing graphically these relationships of independence. Consider the networks in Figure 1. The structure of the networks is rather natural because each piece of evidence bears on its hypothesis and is probabilistically independent conditional on one of the hypotheses. One might wonder, however, why the arrows go from A and B into the node representing the conjunction $A \wedge B$. This setting can capture the meaning of the conjunction. The constraint that, for the conjunction to be true, both A and B have to be true, can be defined using conditional probability tables. The two networks only differ in one detail, whether or not an arrow exist between claims A and B. If there is no arrow, A and B are probabilistically independent since $A \wedge B$ is a collider node (SEE DISCUSSION IN EARLIER CHAPTERS). To eliminate the independence of the two claims, while holding everything else fixed, it is enough to draw an arrow between A and B. See the network in Figure 1 (bottom). Arrows cab be drawn between the evidence nodes themselves, but this modification would undermine them being independent lines of evidence.

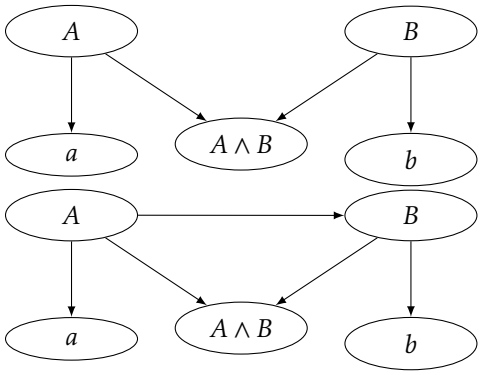


Figure 1: Two Bayesian networks for two pieces of evidence and a composite hypothesis.

The two Bayesian networks above provide a compact representation of the class of models we will be concerned with. We will assume throughout that the items of evidence are independent lines of evidence. Often, we will also assume that the claims themselves are independent, and occasionally consider the situation in which they are positively probabilistically dependent.

5.4 A closer look at the conjunction principle

Let's now turn to possisble solutions to the conjunction paradox within the probabilistic framework. The first thing to note is that the paradox would not arise without the conjunction principle. So could legal probabilists reject this principle and let the paradox disappear? In current discussions in epistemology, an analogous principle about knowledge or justification has been contested because it appears to deny the fact that risks of error accumulate (CITE). If one is justifiably sure about the truth of each claim considered seperataly, one should not be equally sure of their conjunction. You have checked each page of a book and found no error. So, for each page, you are nearly sure there is no error. Having checked each page and found no error, can you be sure that the book as a whole contains no error? Not really. As the number of pages grow, it becomes virtually certain that there is at least one error in the book you have overlooked, although for each page you are nearly sure there is no error. (ADD CITATION ABOUT PREFACE PARADOX) The same applies to other contexts, say product quality control. You may be sure, for each product you checked, that it is free from defects. But you cannot, on this basis alone, be sure that all products you checked are free from defects. Since the risks of error accumulate, you must have missed at least one defective product.

Hey, we can quote a paper that's out by Alicja!:) Also, I guess you want me to find the right refs? M: Yes, if you can

Suppose the legal probabilist does away with the conjunction principle. Now what? How should they define standards of proof? Two immediate options come to mind, but neither is without problems.

One option stipulates that, in order to establish the defendant's guilt beyond a reasonable doubt (or civil liability by preponderance of the evidence), the party making the accusation should establish each claim, separately, to the requisite probability, say at least 95%, without needing to establish the conjunction to the requisite probability. Call this the *atomistic account*. On this view, the prosecution could be in a position to establish guilt beyond a reasonable doubt without estalishing the conjunction of different claims with a sufficiently high probability. This account would allow convictions in cases in which the probability of the defendant's guilt, call it G, is low, just because G is a conjunction of several independent claims that separately satisfy the standard of proof. For example, if each constituent claim is established with 95% probability, the composite claim—assuming, as usual, probabilistic independence between individual claims—would only be established with 59% probability, a far cry from proof beyond a reasonable doubt. This is counterintuitve as it would allow convictions when the defendant is most likely innocent. Under the atomistic account, the composite claim representing the case as a whole would often be established with a probability below the required threshold.

The other option is to require that the prosecution in a criminal case (or the plantiff in a civil case) establish the accusation as a whole—say the cojunction of A and B—to the requisite probability. Call this the *holistic account*. This account is not without problems either. The proof of $A \wedge B$ would impose a higher requirement on the separate probabilities of the conjuncts. If the conjunction $A \wedge B$ is to be proven with at least 95% probability, the individual conjuncts should be established with probability higher than the 95% threshold. So the more conjuncts, the higher their required probability. The more constituent claims, the higher the posterior probability for each claim needed to meet the requisite probability threhsold. Assume, for the sake of illustration, the independence and equiprobability of the constituent claims. If a composite claim consists of k individual claims, these individual claims will have to be established with proabilility of at least $t^{1/k}$, where t is the threhsold applicable to the composite claim.[8] For example, if there are ten constituent claims, they will have to be proven with $0.5^{1/10} = 0.93$ even if the standard of proof is only > 0.5. If the standard is more stringent, as is appropriate in criminal cases, say > 0.95, each individual claim will have to be proven with near certainity, which would make the task extremely demanding on the prosecution. For example, if there are ten constituent claims, they will have to be proven with $0.95^{1/10} = 0.995$.[9]

```
## [1] 0.933033
## [1] 0.9948838
## [1] 0.5987369
```

We have a dilemma here: either (under the holistic approach) the standard is too demanding on the prosecution (or the plantiff) because it would require the individual claims to be established to extremely high probabilities, or (under the atomistic approach) the standard is too lax because it would allow findings of liability when the defendant most likely committed no wrong. Denying the conjunction principle, then, is not without difficulties of its own. Absent the conjuction principle, legal probabilists should still explain how individual claims relate to larger claims in the process of legal proof.

5.5 Prior probabilities and evidential support

It is worth examining the holistic account more closely, focusing in particular on the role of prior probabilities, an aspect that has gone unnoticed so far. The problem with the holistic approach is that it would require, especially in criminal cases, individual claims to be established with a very high probability, often making the task unsurmountable for the prosecution. Or so it would seem. But a composite claim such as $A \wedge B$ will have, other things being equal, a lower prior probability than any individual claim A or B. In general, a composite claim consists of k individual claims. If the composite claim has a prior probability of π, each constituent claim, assuming they are independent and equiprobable, will have a prior probability of $\pi^{1/n}$. The prior probability of the individual claims will

[8] Let p the probability of each consitutuent claim. To meet threshold t, the probability of the composite claim, p^k, should satisfy the constraint $p^k > t$, or in other words, $p > t^{1/k}$.

[9] Moreover, the standard that applies to one of the conjuncts would depend on what has been achieved for the other conjuncts. For instance, assuming independence, if $\mathrm{P}(A)$ is 96%, then $\mathrm{P}(B)$ must be at least 99% so that $\mathrm{P}(A \wedge B)$ is above a 95% threshold. But if $\mathrm{P}(A)$ is 99.99%, then $\mathrm{P}(B)$ must only be greater than 95% to reach the same threshold. Thus, the holistic account would require that the elements of an accusation be proven to different probabilities depeding on how well other claims have been established.

approach one as the number of constituent claims increases.

Is this enough to eliminate the conjunction paradox? Perhpas so. Cohen worried that, as the number of constituent claims increases, the prosection or the plaintiff would see their case against the defendant become progressively weaker and it would become impossisble for them to establish liability. But this worry is an exaggeration. The paradox, as is commonly formulated, starts by assuming that the constituent claims are established by the required probability threshold and then shows that the probability of the conjuction may fall below the threshold. However, following the holistic approach, the order of presentation can be reversed. Start by assuming that the composite claim is established by the required probability threshold. No doubt the individual claims will have to be established with a higher probability, a violation of the conjuction principle. Yet, this violation is not as counterintuitive as it might first appear for two reasons. First, since risks aggregate, it is natural that the probability of a conjunction would be lower than the probability of the conjunct. Second, the prior probabilities of the conjuncts will be higher than the prior probability of the conjunction. Thus, establishing the conjuncts with a higher probability will not be exceedingly demanding.

Along this lines, Dawid (1987), in one of the earliest attempts to solve the conjunction paradox from a probabilistic perspective, wrote:

> ... it is not asking too much of the plaintiff to establish the case as a whole with a posterior probability exceeding one half, even though this means that the several component issues must be established with much larger posterior probabilities; for the *prior* probabilities of the components will also be correspondingly larger, compared with that of their conjunction. The overall effect of subdiving a case into more and more component issues ... [gives] an advantage to the plaintiff, even though he has to establish each with a high probability. (p. 97)

The price of this strategy, however, is the denial of the conjunction principle, the very motivation behind the conjunction paradox. Cohen could insist that this solution amounts to denying the paradox itself. To satisfy Cohen, legal probabilists should offer a justification of the conjunction principle in probabilistic terms, something that Cohen maintains cannot be done. Or can it be done?

Dawid observed that the prior probabilities of the conjuncts are correspondingly higher than the prior probability of the conjunction. The conjunction principle, instead, ignores the role of prior probabilities and treat the conjuncts and the conjuction only in relation to the evidence, irrespective of the prior probabilities. So, in order to capture the conjunction principle, legal probabilists should rely on probabilistic measures that are not heavily depend on prior probabilities.

To this end, it is instructive to compare the impact of one item of evidence on the probability of an individual claim, for different level of sensitivity and specificity of the evidence, with the impact of k items of evidence on the probability of a composite claim that consists of k individual claims, again for different levels of specificity and sensitivity of the evidence. More specifically, let c_i be a constituent claim and e_i its supporting evidence. The comparison is between prior probabilities $\mathsf{P}(C_i)$ and $\mathsf{P}(C_1 \wedge C_2 \wedge \cdots \wedge C_k)$, contrasted with the posterior probabilities $\mathsf{P}(C_i|E_i)$ and $\mathsf{P}(C_1 \wedge C_2 \wedge \cdots \wedge C_k|E_1 \wedge E_2 \wedge \cdots \wedge E_k)$. Figure 2 (top) compares one item of evidence supporting an individual claim and five items of evidence supporting a composite claim consisting of five claims. As is customary, the items of evidence and constituent claims are independent. In addition, for the sake of simplicity, the prior probabilities of the constituent claims are assumed to be the same. There is a significant difference in posterior probabilities, as expected, but there is a also a significant difference in prior probabilities. Since the composite claim starts out less likely than any individual claim, it is natural—other things being equal—that its posterior probability would be correspondingly lower.

What happens if we make the same comparison between individual and composite claims by equalizing their prior probability? If the claims are independent and equiprobable, let x be the prior probability of an individual claim (when it is considered in isolation) and let $x^{1/k}$ the prior probability of the same individual claim when it is part of a composite claim that consists of k claims. In this way, the prior probability of the composite claim equals the prior probability of the individual claim since $(x^{1/k})^k = x$, as desired. Figure 2 (bottom) shows the result of this process of equalization. Here we are explicitly factoring out the role of prior probabilities.

We note two things. First, the difference in posterior probability, though still present, is less significant, especially for values above the 50% threshold or even more clearly above the 95% threshold. Second, whatever remaining difference in posterior probability is now reversed, that is, a composite claim supported by several items of evidence has a higher posterior probability compared to an individual

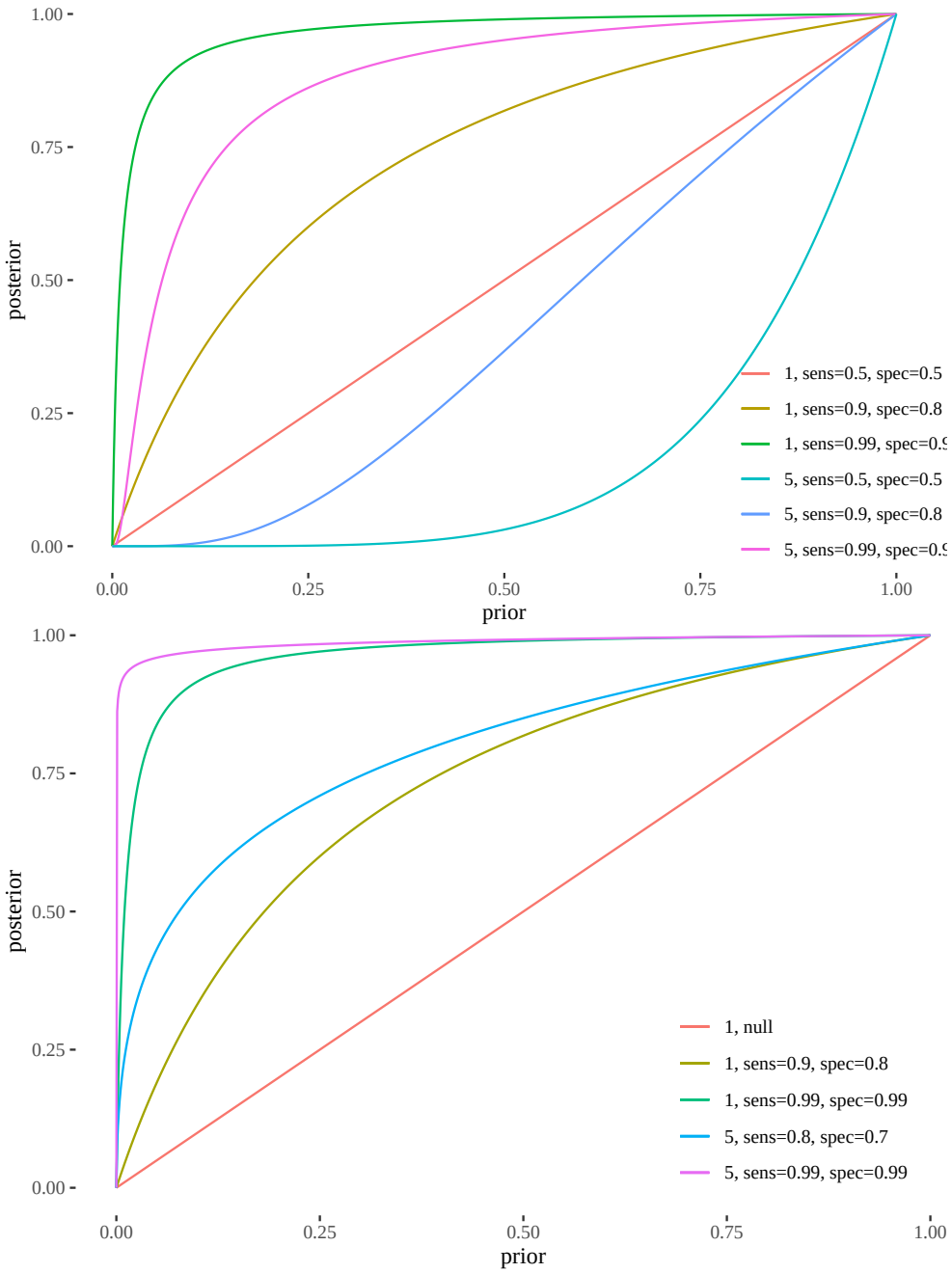


Figure 2: The comparison here is betwen individual support and joint support. Top graph: The null line for joint support ($y = x * x$) is much below the null line for individual support ($y = x$). Bottom graph: the two null lines are equalized and the posterior lines adjusted accordingly. The posterior lines for individual and joint support get much closer especially for high posterior probability values.

claim supported by one item of evidence. These obervations establish that, by factorig out prior probabilities and under certain independence assumptions, whenever the individual claims meet the applicable threhsold, so does the composite claim, and conversely, whenever the composite claim, say $A \wedge B$, meets the appliacable threshold, so do the individual claims (at least for relatively high values above 50% or higher). Both directions of the conjunction principle have been, to some approximation, justified using Figure 2 (bottom).

This is a promising strategy for the legal probabilist to develop. Failure of the conjunction principle could be an artifact caused by the fact that the prior probabilities of the individual claims differ from the prior probability of the composite claim. So, once the prior probabilities are factored out, the conjunction might be validated in the probabilistic framework, and with that, the conjunction paradox would go away. For this strategy to succeed, however, the standard of proof should be understood in a manner that is not prior sensitive. This is quite natural. The standard of proof is a sufficiency criterion for how strong the evidence should be in order to justify a finding of criminal or civil liability against the defendant. The strength of the evidence in favor of a hypothesis—say the defendant is guilty of murder—need not depend on the prior probability of the hypothesis. In fact, it should solely depend on the quality of the evidence supporting the hypothesis, or else very good evidence in favor of an unlikely hypothesis could be regarded as worse evidence than otherwise poor evidence in favor of likely hypothesis.

As we will see in detail in what follows, this strategy cannot succeed. To preview, the basic argument is this. Suppose the standard of proof is no longer understood as a threshold on the posterior probability given the evidence, but rather, as threshold on evidential strength. Two common probabilistic measures of evidential strength are the Bayes factor or the likelihood ratio. We discussed this topic in earlier chapters (REFER TO EARLIER CHAPTERS). Under plausible assumptions, these measures of evidential strength validate one direction of the conjunction principle, what we call aggregation. If a is sufficiently strong evidence in favor of A and b is sufficiently strong evidence in favor of B, then $a \wedge b$ is sufficiently strong evidence in favor of the conjunction $A \wedge B$. In fact, the evidential support for the conjunction will often exceed that for the individual claims, a point already made by Dawid (1987):

> suitably measured, the support supplied by the conjunction of several independent testimonies exceeds that supplied by any of its constituents.

Dawid thought this fact was enough for the conjuction paradox to 'evaporate'. To some extent, this is true since we are vindicating one direction of the conjunction principle. However, the other direction does not hold, what we call distribution. If $a \wedge b$ is sufficiently strong evidence in favor of $A \wedge B$, it does not follow that a is sufficiently strong evidence in favor of A or b sufficiently strong evidence in favor of B. It is not even true that, if $a \wedge b$ is sufficiently strong evidence in favor of $A \wedge B$, then $a \wedge b$ is sufficiently strong evidence in favor of A or B. This is odd. It would mean that, given a body of evidence, one can establish beyond a reasonable doubt that $A \wedge B$ (say the defendant killed the victim *and* did so intentionally) while failing to establish by the same standard one of the conjuncts.

Interestingly, if instead of probabilistic measures of evidential strength such as the Bayes factor or the likelihood ratio, we use posterior probabilities, the other direction of conjuction principle—aggregation, not distribition—fails. So we are in a dilemma. If the standard of proof is understood as a threshold relative to the posterior probability, the conjunction principle fails becaue aggregation fails while distribution succeeds. If, on the other hand, the standard of proof is understood as a threshold relative to measures of evidential strength, the conjuction principle fails because distribution fails while aggregation succeeds. From a probabilistic perspective, it seems impossible to capture both directions of the conjunction principle into one unified account.

The arguemnt taht follows is somewhat tedious. The reader can take our word for it and jump ahead or dive into it at their own peril! In what follows, by using the theory of Bayesian networks, we will first verify under what conditions Dawid's claim that 'suitably measured, the support supplied by the conjunction of several independent testimonies exceeds that supplied by any of its constituents' holds. Next, we will show—perhpas surprisingly—that the difficulty about conjunction does not subside even after switching from posterior probabilities to measures of evidential support.

5.6 Bayes factor threshold

A common probabilistic measure of the support of E in favor of H is the Bayes factor $\mathsf{P}(E|H)/\mathsf{P}(E)$. Since by Bayes' theorem

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

the Bayes factor measures the extent to which a piece of evidence increases the probability of a hypothesis. The greater the Bayes factor (for values above one), the stronger the support of E in favor of H. Putting aside reservations about this measure of evidential support (discussed earlier in Chapter CROSSREF), the Bayes factor $\mathsf{P}(E|H)/\mathsf{P}(E)$, unlike the conditional probability $\mathsf{P}(H|E)$, offers a potential way to overcome the difficulty about conjunction.

Say a and b, separately, support A and B to degree s_A and s_B respectively, that is, $\mathsf{P}(a|A)/\mathsf{P}(a) = s_A$ and $\mathsf{P}(b|B)/\mathsf{P}(b) = s_B$, where both s_A and s_B are greater than one. Does the combined evidence $a \wedge b$ provide at least as much support in favor of the combined claim $A \wedge B$ as the individual support by a and b in favor of A and B considered separately? The combined support should be measured by the combined Bayes factor $\mathsf{P}(a \wedge b|A \wedge B)/\mathsf{P}(a \wedge b)$. The latter, under suitable independence assumptions, equals the product of the individual supports s_A and s_B.[10] That is,

$$\frac{\mathsf{P}(a \wedge b|A \wedge B)}{\mathsf{P}(a \wedge b)} = \frac{\mathsf{P}(a|A)}{\mathsf{P}(a)} \times \frac{\mathsf{P}(b|B)}{\mathsf{P}(b)}$$
$$s_{AB} = s_A \times s_B.$$

Thus, the combined support s_{AB} will always be higher than the individual support so long as s_A and s_B are greater than one. This result can be generalized beyond two pieces of evidence and holds generally. Figure REFERENCE TO FIGURE BELOW compares the Bayes factor of one item of evidence, say $\frac{\mathsf{P}(a|A)}{\mathsf{P}(a)}$ with the combined Bayes factor for five item of evidence, say $\frac{\mathsf{P}(a_1 \wedge ... a_5|A_1 \wedge ... A_5)}{\mathsf{P}(a_1 \wedge ... a_5)}$, for dfferent values of sensitivity and specificity of the evidence. The latter always exceeds the fomer.

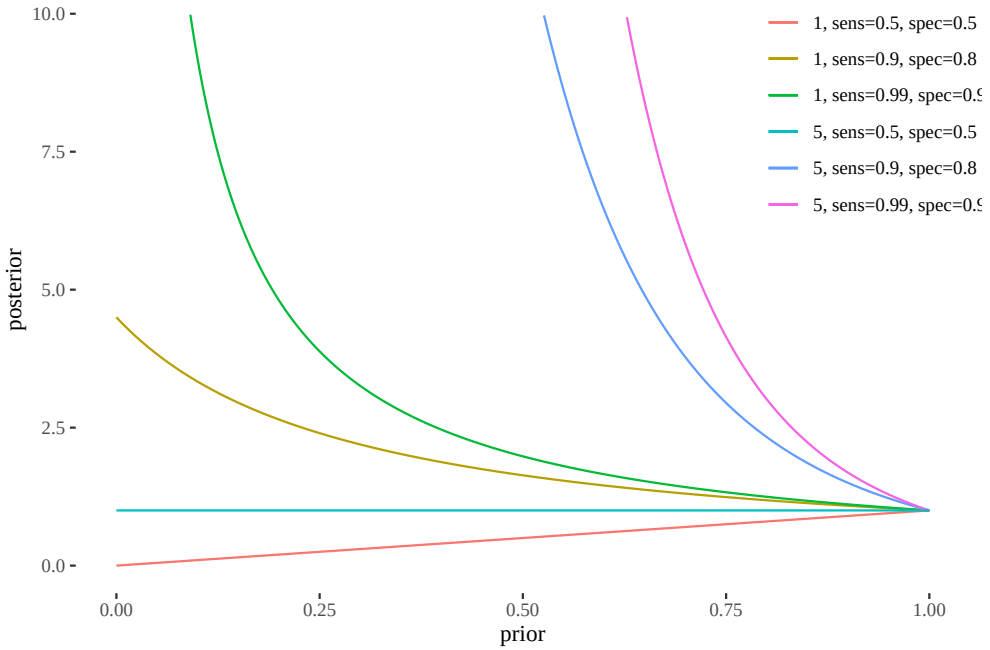
\begin{figure}

```
## Warning: Removed 91 rows containing missing values (geom_path).
## Warning: Removed 526 rows containing missing values (geom_path).
## Warning: Removed 628 rows containing missing values (geom_path).
```

[10] By the probability calculus,

$$\begin{aligned}
\frac{\mathsf{P}(a \wedge b|A \wedge B)}{\mathsf{P}(a \wedge b)} &= \frac{\mathsf{P}(A \wedge B|a \wedge b)}{\mathsf{P}(A \wedge B)} \\
&= \frac{\frac{\mathsf{P}(A \wedge B \wedge a \wedge b)}{\mathsf{P}(a \wedge b)}}{\mathsf{P}(A \wedge B)} \\
&= \frac{\frac{\mathsf{P}(A) \times \mathsf{P}(B|A) \times \mathsf{P}(a|A \wedge B) \times \mathsf{P}(b|A \wedge B \wedge a)}{\mathsf{P}(a) \times \mathsf{P}(b|a)}}{\mathsf{P}(A \wedge B)} \\
&=^{*} \frac{\frac{\mathsf{P}(A) \times \mathsf{P}(B) \times \mathsf{P}(a|A) \times \mathsf{P}(b|B)}{\mathsf{P}(a) \times \mathsf{P}(b)}}{\mathsf{P}(A) \times \mathsf{P}(B)} \\
&= \frac{\mathsf{P}(a|A)}{\mathsf{P}(a)} \times \frac{\mathsf{P}(b|B)}{\mathsf{P}(b)} \\
s_{AB} &= s_A \times s_B
\end{aligned}$$

The step marked by the asterisk rests on the independence assumptions codified in the Bayesian network in Figure 1 (top).



\end{figure}

Dawid's claim that 'the support supplied by the conjunction of several independent testimonies exceeds that supplied by any of its constituents' is therefore vindicated. The claim holds in the class of cases characterized by the relationships of probabilistic independence encoded in the Bayesian network in Figure 1 (top). As noted before, this network ensures that claim A and B are probabilistically independent, as well as that items of evidence a and b are conditionally independent (which captures formally the fact that they are independent lines of evidence).

What happens if A and B are not necessarily probabilistically independent as in the Bayesian network in Figure 1 (bottom)? Given this network, the following holds:

$$\frac{\mathsf{P}(a \wedge b | A \wedge B)}{\mathsf{P}(a \wedge b)} = \frac{\mathsf{P}(a|A)}{\mathsf{P}(a)} \times \frac{\mathsf{P}(b|B)}{\mathsf{P}(b|a)}$$
$$s_{AB}^{'} = s_A \times s_B^{'}$$

Note that factor $s_B = \frac{\mathsf{P}(b|B)}{\mathsf{P}(b)}$ was replaced by $s_B^{'} = \frac{\mathsf{P}(b|B)}{\mathsf{P}(b|a)}$.[11] Now, $s_B^{'}$ is usually lower than s_B because $\mathsf{P}(b|a) > \mathsf{P}(b)$ (assuming, at least, a and b are convergent pieces of evidence; SEE DISCUSSION IN EARLIER CHAPTERS). At the same time, $s_B^{'}$ should still be greater than one since b, by assumption, positively supports B even in combination with a.[12] Hence, $s_{AB}^{'}$ should still be greater than one provided s_A and S_B are both greater than one, but it might not always exceed the support supplied by s_A and s_B individually. For suppose $s_A = 2$ and $s_B = 3$, but $s_B^{'} = 1.2$. Then, $s_{AB}^{'} = 2 \times 1.2 = 2.4$, which is below the individual support s_B, but still above s_A. For the case in which A and B are probabilistically dependent, Dawid's claim should be amended as follows. Even though the support supplied by the conjunction of several independent testimonies need not always exceed that supplied by any of its constituents, it is always at least as great as the smallest support supplied by its constituents.[13]

[11] Given the second Bayesian network, b need not be probabilistically independent of a, and thus there is no guarantee that $\mathsf{P}(b|a) = \mathsf{P}(a)$.

[12] Note that $\frac{\mathsf{P}(b|B)}{\mathsf{P}(b|a)} = \frac{\mathsf{P}(b|B \wedge a)}{\mathsf{P}(b|a)}$ (by the probabilistic independence of a and b given B). So the claim that $\frac{\mathsf{P}(b|B)}{\mathsf{P}(b|a)} > 1$ is equivalent $\mathsf{P}(B|b \wedge a) > \mathsf{P}(B|a)$ since by Bayes' theorem $\mathsf{P}(B|b \wedge a) = \frac{\mathsf{P}(b|B \wedge a)}{\mathsf{P}(b|a)} \times \mathsf{P}(B|a)$. Presumably, evidence b should still raise the probability of B even in cojunction with a, or else b would be useless evidence. **M: THIS CLAIM NEEDS TO BE PROVEN MORE RIGOROUSLY BUT I THINK IT'S CORRECT. BASIC INTUITION IS THAT EVIDENCE a RAISES THE PROBABILITY OF CLAIM A (OR B) AND THEN EVIDENCE b FURTHER RAISES THE PROBABILITY OF CLAIM A (OR B). THIS HAPPENS WHEN WE HAVE CONVERGENT EVIDENCE.**

[13] DO WE NEED A PROOF OF THIS? THIS SHOULD BE CLEAR BY PLOTTING. IN PLITTING THIS, WE SHOULD ENSURE THAT BOTH BF ARE GREATER THAN ONE, THEN THE COMBINED ONE WILL BE GREATER THAN THE SMALLEST EVEN WHEN A AND B ARE DEPENDENT.

If the combined support equals $s_A \times s_B$ or $s_A \times s_B^{'}$, does the difficulty about conjunction evaporate, as Dawid thought? One hurdle here is that the standard of proof would no longer be formalized as a posterior probability threshold, but instead as a threshold about the Bayes factor. The threshold would no longer be a probability between 0% and 100%, but rather a number somewhere above 1. The greater this number, the more stringent the standard of proof, for any value above one. In criminal trials, for example, the rule of decision would be: guilt is proven beyond a reasonable doubt if and only if the evidential support in favor of G—as measured by the Bayes factor $\frac{\mathsf{P}(E|G)}{\mathsf{P}(E)}$—meets a suitably high threshold t_{BF}. The obvious question at this point is, how do we identify the appropriate threshold?

good question, will think about it, will need to take a look at "Bayesian Choice", also need to think about a counterexample

One strategy is to derive the Bayes factor threshold, call it t_{BF}, from the posterior threshold t. Since $posterior = Bayes\,factor \times prior$, the Bayes factor threshold can be determined as follows:

$$\frac{t}{prior} = t_{BF}$$

The higher the prior probability, the lower t_{BF}. Whether this is a desirable property for a decision threshold can be questioned, but the same can be said about the posterior threshold t. The higher the prior probability, the easier to meet the posterior threshold.

Presumably, the threshold t_{BF} should be applied to individual as well as composite claims. Since the threshold varies depending on the priors, the thresholds for the individual claims A and B, denoted by t_{BF}^{A} and t_{BF}^{B}, will differ from the threshold for the composite claim $A \wedge B$, denoted by $t_{BF}^{A \wedge B}$. At issue here is whether the conjunction principle can be formalized in a plausible manner with the Bayes factor. Unfortunately, the answer is negative. To see why, first recall the conjunction principle:

$$\mathrm{S}[a, A] \text{ and } \mathrm{S}[b, B] \text{ iff } \mathrm{S}[a \wedge b, A \wedge B],$$

where $\mathrm{S}[E, H]$ means that evidence E supports hypothesis H by standard S. If the standard of proof is formalized using the Bayes factor, the conjunction principle would boil down to:

$$\frac{\mathsf{P}(a|A)}{\mathsf{P}(a)} > t_{BF}^{A} \text{ and } \frac{\mathsf{P}(b|B)}{\mathsf{P}(b)} > t_{BF}^{B} \text{ iff } \frac{\mathsf{P}(a \wedge b|A \wedge B)}{\mathsf{P}(a \wedge b)} > t_{BF}^{A \wedge B}$$

Consider a posterior threshold $t = 0.95$, as might be appropriate in a criminal case. If A and B both have a prior probability of 10%, the threshold $t_{BF}^{A} = t_{BF}^{B} = 0.95/0.1 = 9.5$ for A or B individually. Assuming independence of A and B, the composite claim $A \wedge B$ will be associated with the threshold $t_{BF}^{A \wedge B} = 0.95/(0.1 * 0.1) = 95$, a much higher value. But if each individual claim meets its Bayes factor threshold of 9.5 and the two claims are independent, the joint Bayes factor would equal the multiplication of the individual Bayer factors, that is, $9.5 * 9.5 = 90.25$. This is not quite enough to meet $t_{BF}^{A \wedge B} = 95$, but is fairly close. The difference in absolute terms grows as the prior probability of the individual claims becomes lower, but the combined Bayes factor remains only 5% below the value needed to meet $t_{BF}^{A \wedge B}$.[14] Perhaps this is a good enough approximation. However, as the number of constituent claims grows, the difference becomes larger.[15] In addition, the difference becomes larger with a lower posterior probability threshold, say 0.5. Even with just two claims, $t_{BF}^{A \wedge B} = 0.5/(0.1 * 0.1) = 50$, but $t_{BF}^{A} * t_{BF}^{B} = (0.5/0.1) * (0.5/0.1) = 25$, only half the required value. The conjunction principle therefore fails in a large number of cases even using the Bayes factor threshold.

```
## [1] 9.5
## [1] 95
## [1] 90.25
## [1] 95
## [1] 9500
## [1] 9025
## [1] 950000
## [1] 902500
## [1] 5
```

[14] The difference at here is between $t_{BF}^{A \wedge B} = 0.95/p^2$ and $t_{BF}^{A} * t_{BF}^{B} = (0.5/p)^2$. Note that $\frac{0.95/p^2 - (0.5/p)^2}{0.95/p^2} = 5\%$, for any value of the prior p.

[15] Given five constituent claims, $\frac{0.95/p^5 - (0.95/p)^5}{0.95/p^5} = 18\%$.

```
## [1] 50
## [1] 25
```

The alternative here is to fix the Bays factor threshold regardless of the prior probability of the claim of interest. This raises the difficult question of how to fix the Bayesian factor threshold irrespective of the priors. Standard decision theory can no longer be used. As it turns out, even if the question can be satisfactorily answered, the fixed threshold approach gives rise to a complication that proves fatal. If the standard of proof is formalized using a fixed Bayes factor threshold t_{BF}, the conjunction principle would boil down to:

$$\frac{\mathsf{P}(a|A)}{\mathsf{P}(a)} > t_{BF} \text{ and } \frac{\mathsf{P}(b|B)}{\mathsf{P}(b)} > t_{BF} \text{ iff } \frac{\mathsf{P}(a \wedge b|A \wedge B)}{\mathsf{P}(a \wedge b)} > t_{BF}$$

The left-to-right direction—call it aggregation—is likely to hold for any threshold t_{BF} greater than one. As shown earlier, the combined evidential support is greater than the individual evidential support if A and B are independent, or greater than the smallest individual support if A and B are dependent. Aggregation could not be justified using posterior probabilities $\mathsf{P}(A|a)$ and $\mathsf{P}(B|b)$ nor could it be justified generally using a variable Bayesian factor threshold. So it is an advantage of the fixed Bayes factor threshold that it can justify this direction of the conjunction principle.

However, the right-to-left direction—call it distribution—has now become problematic. For suppose the combined evidential support, $\frac{\mathsf{P}(a\wedge b|A\wedge B)}{\mathsf{P}(a\wedge b)}$, barely meets the threshold. This implies that the individual support, say $\frac{\mathsf{P}(a|A)}{\mathsf{P}(a)}$, could be be below the threshold unless $\frac{\mathsf{P}(b|B)}{\mathsf{P}(b)} = 1$ (which should not happen if b positively supports B). So, curiously, there would be cases in which, even though the conjunction $A \wedge B$ is established to the desired standard of proof, one of the individual claims fails to meet the standard. This is odd. More specifically, the following distribution principle fails in some cases:

Not always true; just give a specific numerical counterexample. M: I added 'often'. Is this enough?

$$\text{If } \mathrm{S}[a \wedge b, A \wedge B], \text{ then } \mathrm{S}[a, A] \text{ and } \mathrm{S}[b, B]. \tag{DIS1}$$

Could this principle be rejected? Perhaps, it is not as essential as we thought at first. Since the evidence is not held constant, the support supplied by $a \wedge b$ could be stronger than that supplied by a and b individually. So even when the conjunction $A \wedge B$ is established to the requisite standard given evidence $a \wedge b$, it might still be that A does not meet the requisite standard (given a) nor does B (given b).

But consider a less controversial version, holding the evidence constant:

$$\text{If } \mathrm{S}[a \wedge b, A \wedge B], \text{ then } \mathrm{S}[a \wedge b, A] \text{ and } \mathrm{S}[a \wedge b, B]. \tag{DIS2}$$

This principle is harder to deny. That is, one would not want to claim that, holding fixed evidence $a \wedge b$, establishing the conjunction might not be enough for establishing one of the conjuncts. One cannot be willing to assent to the conjunction without being willing to assent to one of the conjuncts against a fixed body of evidence. Certainly any formalization of the standard of proof should obey (DIS2). And yet, it is this very principle that we should deny if we understand the standard of proof using the Bayes factor. [16] In case A and B are probabilistically independent, (DIS1) and (DIS2) are in fact equivalent, so rejectign one requires rejecting the other.[17] The intuitive reason for this— perhpas surprising—result is that $A \wedge B$ has a much lower prior probability than A (or B) considered separately. Thus, the same body of evidence is going to have a larger impact on a hypothesis with a lower prior probability, other things being equal. This larger impact on the prior is reflected in a larger Bayes factor.

[16] To show that (DIS2) fails, it is enough to show that $\frac{\mathsf{P}(B|a\wedge b\wedge A)}{\mathsf{P}(B|A)} > 1$ because $S[a \wedge b, A \wedge B] > S[a \wedge b, A]$ iff $\frac{\mathsf{P}(a\wedge b|A\wedge B)}{\mathsf{P}(a\wedge b)} > \frac{\mathsf{P}(a\wedge b|A)}{\mathsf{P}(a\wedge b)}$ iff $\frac{\mathsf{P}(A\wedge B|a\wedge b)}{\mathsf{P}(A\wedge B)} > \frac{\mathsf{P}(A|a\wedge b)}{\mathsf{P}(A)}$ iff $\frac{\mathsf{P}(A|a\wedge b)\times\mathsf{P}(B|a\wedge b\wedge A)}{\mathsf{P}(A)\times\mathsf{P}(B|A)} > \frac{\mathsf{P}(A|a\wedge b)}{\mathsf{P}(A)}$. Now, $\frac{\mathsf{P}(B|a\wedge b\wedge A)}{\mathsf{P}(B|A)} > 1$ so long as $a \wedge b$ positively supports B even under the assumption of A, unless A entailed B. We naturally exclude the situation in which one claim entails the other because otherwise there would be no need to establish the two claims. Establishing one claim alone would suffice. To see why $a \wedge b$ positively supports B (or A), note that $S[a, A] = \frac{\mathsf{P}(a|A)}{\mathsf{P}(a)} \leq \frac{\mathsf{P}(a|A)\times\mathsf{P}(b|A\wedge a)}{\mathsf{P}(a)\times\mathsf{P}(b|a)} = \frac{\mathsf{P}(a\wedge b|A)}{\mathsf{P}(a\wedge b)} = S[a \wedge b, A]$. The key step here is $\frac{\mathsf{P}(a|A)}{\mathsf{P}(a)} \leq \frac{\mathsf{P}(a|A)\times\mathsf{P}(b|A\wedge a)}{\mathsf{P}(a)\times\mathsf{P}(b|a)}$. The latter holds because $\frac{\mathsf{P}(b|A\wedge a)}{\mathsf{P}(b|a)} \geq 1$. It is useful to distinguish two cases. First, if A and B are probabilistically independent, as in the Bayesian network in Figure 1 (top), then $\frac{\mathsf{P}(b|A\wedge a)}{\mathsf{P}(b|a)} = \frac{\mathsf{P}(b)}{\mathsf{P}(b)} = 1$. Second, if A and B are probabilistically dependent, as in the Bayesian network in Figure 1 (bottom), evidence b positively supports claim A (even conditional on a) so long as b positively supports B. The assumption is that claim A and B are positively correlated, and thus, any evidence that supports one of the claims is going to support the other claim, as well. **SEE EARLIER CHAPTERS FOR A MORE RIGOROUS PROOF OF THIS LAST POINT.**

[17] $\frac{\mathsf{P}(a\wedge b|A)}{\mathsf{P}(a\wedge b)} = \frac{\mathsf{P}(a|A)\mathsf{P}(b|A)}{\mathsf{P}(a)\mathsf{P}(b)} = \frac{\mathsf{P}(a|A)}{\mathsf{P}(a)}$. In other words, $S[a \wedge b, A] = S[a, A]$.

All in all, using Bayes factor to understad the standard of proof has counterintuitive consequences, what we will call the distribution paradox. For suppose the prosecution provided evidence for claim A, but this evidence still falls short of the threshold t (a certain number above 1). Just by tagging an additional claim B and without doing any further evidentiary work, the prosecution could provide sufficiently strong evidence (which meets the threshold t) in favor of claim $A \wedge B$. So, it could well happen that, while the prosecution failed to prove beyond a reasonable doubt that the defendant injured the victim, the prosecution could nevertheless prove beyond a reasonable doubt that the defendant injured the victim and did so intentionally. This is odd.

5.7 Likelihood ratio threshold

Let's now replace the Bayes factor with the likelihood ratio, another probabilistic measure of evidential support. As we shall understand it for now, the likelihood ratio compares the probability of the evidence on the assumption that a hypothesis of interest is true and the probability of the evidence on the assumption that the negation of the hypothesis is true, that is, $\frac{\mathsf{P}(E|H)}{\mathsf{P}(E|\neg H)}$. The greater the likelihood ratio (for values above one), the stronger the evidential support in favor of the hypothesis (as contrasted to the its negation). We discussed extensively the advantages and limitations of this account in **REFERENCE TO EARLIER CHAPTER**.

We can think of the the likelihood rato as the following:

$$\frac{\textit{sensitivity}}{\textit{1- specificity}}$$

Unlike the Bayes factor, the likleihood ratio is not sensitive to the priors so long as sensitivity and specificity are not. In this sense, it is a more suitable measure if we want to factor out the effects that priors may have on the assessment of evidential strength. But this advantage is short lived. Even the likelihood ratio is sensitive to priors when one considers a composite claim instead of an individual claim.

To see why, consider the combined ikelihood ratio $\frac{\mathsf{P}(a \wedge b|A \wedge B)}{\mathsf{P}(a \wedge b|\neg(A \wedge B))}$. The numerator can be computed easily:[18]

$$\mathsf{P}(a \wedge b|A \wedge B) = \mathsf{P}(a|A) \times \mathsf{P}(b|B)$$

The equality requires the independence assumptions codified in the Bayesian networks in Figure 1. That is, the two items of evidence should be independent of one another conditional on the hypothesis they support. The numerator does not depend on the priors associated with $A \wedge B$. Call it *combined sensitivity*, simply resulting from multiplying the sensitivity of the individual items of evidence, a and b, relative to their respective hyptheses, A and B.

The denominator is more involved:[19]

18

$$\begin{aligned}
\mathsf{P}(a \wedge b|A \wedge B) &= \frac{\mathsf{P}(A \wedge B \wedge a \wedge b)}{\mathsf{P}(A \wedge B)} \\
&= \frac{\mathsf{P}(A) \times \mathsf{P}(B|A) \times \mathsf{P}(a|A \wedge B) \times \mathsf{P}(b|A \wedge B \wedge a)}{\mathsf{P}(A) \times \mathsf{P}(B|A)} \\
&=^{*} \frac{\mathsf{P}(A) \times \mathsf{P}(B|A) \times \mathsf{P}(a|A) \times \mathsf{P}(b|B)}{\mathsf{P}(A) \times \mathsf{P}(B|A)} \\
&= \mathsf{P}(a|A) \times \mathsf{P}(b|B)
\end{aligned}$$

The asterisk marks the step that requires the indepedence assumptions in Figure 1.

19

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

The asterisk marks the step that requires the indepedence assumptions in Figure 1.

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

The same independence assumptions invoked before are needed here. Unlike the numerator, the denominator—call it *combined specificity*—depends on the priors of A and A and thus on the priors of $A \wedge B$. Putting numerator and denominator together yields the formula for the combined likelihood ratio.

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

As with the Bayes factor, under suitable independence assumptions, the combined likelihood ratio exceeds the individual likelihood ratio so logn as the two pieces of evidence have the same sensitivity and specificity. If they have differnt levels of sensitivity and specificity, the combined likelihood ratio never goes below the lowest of the two individual likelihood ratios.

Becasue of the many variables at play, it is not easy to compare the combined evidential support and the individual support supplied by a and b towards A and B, as measured by the individual and combined likelihood ratio. To circumvent this difficulty, we make three simplifying assumptions. First, the sensitivity of a piece of evidence, say $\mathsf{P}(a|A)$, is the same as its specificity, $\mathsf{P}(\neg a|\neg A)$. Let $\mathsf{P}(a|A) = x$ and $\mathsf{P}(b|B) = y$. So $\mathsf{P}(a|\neg A) = 1 - x$ and $\mathsf{P}(b|\neg B) = 1 - y$. Finally, the sensitivity (and thus the specificity) of the two pieces of evidence is the same, that is, $\mathsf{P}(a|A) = x = \mathsf{P}(b|B) = y$. Finally, as is customary, claims A and B are independent of one another. The combined likelihood ratio therefore reduces to the following, where $\mathsf{P}(A) = k$ and $\mathsf{P}(B) = t$:

$$\frac{\mathsf{P}(a \wedge b | A \wedge B)}{\mathsf{P}(a \wedge b | \neg(A \wedge B))} = \frac{xx}{\frac{(1-k)t(1-x)x + k(1-t)x(1-x) + (1-k)(1-t)(1-x)(1-x)}{(1-k)t + (1-t)k + (1-k)(1-t)}}$$

The graph of the combined likelihood ratio can now be easily plotted against the single likelihood ratios. As Figure 3 shows, the combined likelihood ratio varies depeding on the prior probabilities $\mathsf{P}(A)$ and $\mathsf{P}(B)$, as expected, but always exeeeds the individual likelihood ratios whenever they are greater than one (that is, the two piece of evidence provides positive support for their respective hypothesis)

What happens if we relax the three simplifying assumptions? Suppose the sensitivity and specificity of the two pieces of evidence are not the same. Their likelihood ratios will then also be different. In this case, the combined likelihood ratio is not always greater than the individual ratios, but it is always greater than the smallest of the two provided the individual likelihood ratios are greater than one. **M: what about dropping other assumptions?**

M: Need to add simuation results to make this argument fully general and drop all the simplifying assumptions. For example, what if the two hyptheses are not independent?

The argument, once again, verified Dawid's claim that 'the support supplied by the conjunction of several independent testimonies exceeds that supplied by any of its constituents'. One caveat is that, if evidential support is measured by the likelihood ratio, the support supplied by the conjunction of different independent pieces of evidence always exceeds the smallest of the support supplied by its consitituents, but there are cases in which it does not exceed the support supplied by some of its constituents.

Like the Bayes factor, the likelihood ratio can be used to formalize the standard of proof by equating the standard to a threshold t above one. The greater the threshold, the more stringent the standard. In criminal trials, for example, the rule of decision would be: guilt is proven beyond a reasonable doubt if and only if the evidential support in favor of G—as measured by the likelihood ratio $\frac{\mathsf{P}(E|G)}{\mathsf{P}(E|\neg G)}$—meets a suitably high threshold t above one.

By the ratio version of Bayes' theorem,

$$\textit{posterior ratio} = \textit{likelihood ratio} \times \textit{prior ratio},$$

and thus

$$\frac{\textit{posterior ratio}}{\textit{prior ratio}} = \textit{likelihood ratio}.$$

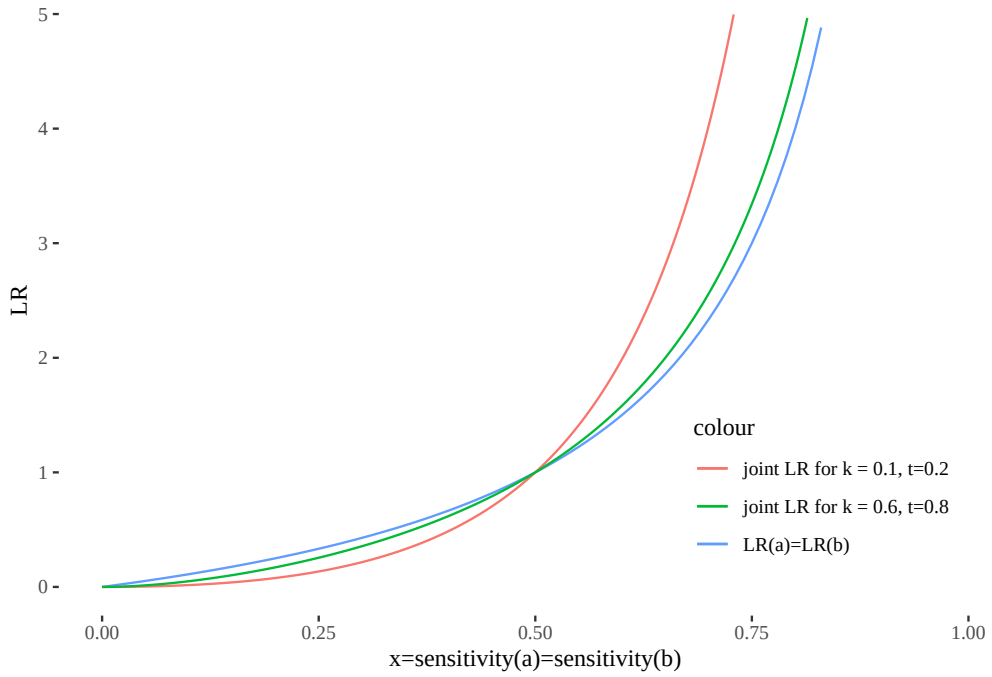


Figure 3: Combined likelihood ratios exceeds individual Likelihood ratios. Changes in the prior probabilities t and k do not invalidate this result.

If the posterior ratio is fixed at, say $t/1-t$, the threshold on the likelihood ratio, call it t_{LR} is defined as:

$$\frac{t/1-t}{\textit{prior ratio}} = t_{LR}.$$

The likelihood ratio threshold will vary depending on the prior. The higher the prior, the lower the likelihood ratio threshold. So

Consider the individual claim A and B and compared them with the composite claim $A \wedge B$. Suppose the two claims are independent. Consider a posterior threshold of 95% as might be appropriate in a criminal case. Assuming A and B have a prior probability of 20% and 30% respectively, the likelihood ratio threshold for A and B will be $t^A_{LR} \approx 76$ and $t^B_{LR}a \approx 44$. The likelihood ratio threshold for the composite claim $A \wedge B$ will be $t^{A\wedge B}_{LR} \approx 297$.

```
## [1] 76
## [1] 44.33333
## [1] 297.6667
```

Consider a posterior threshold of 50% as might be appropriate in a civil case. Assuming A and B have a prior probability of 20% and 30% respectively, the likelihood ratio threshold for A and B will be $t^A_{LR} \approx 4$ and $t^B_{LR}a \approx 2$. The likelihood ratio threshold for the composite claim $A \wedge B$ will be $t^{A\wedge B}_{LR} \approx 15$.

```
## [1] 4
## [1] 2.333333
## [1] 15.66667
```

As expected, other things being equal, the likelihood ratio threshold is lower for civil than criminal cases. The threshold is variable and depends on then prior, so claim that have higher priors are asociated with lower likelihood ratio threshold, such as A and B, than claims associated with lower priors such as $A \wedge B$.

Now suppoose the individual likelihood ratios meet the threshold t^A_{LR} and t^B_{LR} give a posterior threshold of 0.95. To ensure that t^A_{LR} is met, evidence a should have a sensitivity of at least (roughly) 0.99 (and a specificity of 1-sensitivity). To ensure that t^B_{LR} is met, evidence b should have a sensitivity of at least (roughly) 0.98 (and a specificity of 1-sensitivity). Holding fixed the values for sensitivity and specificity, does the combined likelihood ratio meet the threshold $t^{A\wedge B}_{LR}$? Not quite. The combined likelihood ratio

equals about 145, far short that what the threshold $t_{LR}^{A \wedge B}$ requires, namely a likelihood ratio of 297

```
## [1] 76
## [1] 75.92308
## [1] 44.33333
## [1] 44.45455
## [1] 145.1446
```

Things do not look any better for a lower threshold. Now suppoose the individual likelihood ratios meet the threshold t_{LR}^{A} and t_{LR}^{B} give a posterior threshold of 0.5. To ensure that t_{LR}^{A} is met, evidence a should have a sensitivity of at least 0.8 (and a specificity of 1-sensitivity). To ensure that t_{LR}^{B} is met, evidence b should have a sensitivity of at least 0.7 (and a specificity of 1-sensitivity). To ensure that t_{LR}^{B} is met, evidence b should have a sensitivity of at least 0.7 (and a specificity of 1-sensitivity). Holding fixed the values for sensitivity and specificity, does the combined likelihood ratio meet the threshold $t_{LR}^{A \wedge B}$? Note quite. The combined likelihood ratio equals about 5, far short that what the threshold $t_{LR}^{A \wedge B}$ requires, namely a likelihood ratio of 15.

```
## [1] 4
## [1] 4
## [1] 2.333333
## [1] 2.333333
## [1] 5.222222
```

By using a likelihood ratio threshold that is prior dependent, the conjunction principle–in particular, what we called aggregation—fails. The alternative is to fix a likelihood ratio threshold irrespective of the prior. Setting aside the problem of how ot identify the appropriate threshold, we will see that this approach is able to justify one direction of the conjunction principle—what we called aggregation—but still fails to justify the other direction—what we called distribution So the distribution paradox arises here again.

Consider aggregation first. Say both individual likelihood ratios $\frac{\mathsf{P}(a|A)}{\mathsf{P}(a|\neg A)}$ and $\frac{\mathsf{P}(b|B)}{\mathsf{P}(b|\neg B)}$ are above the requisite threshold t for meeting the standard of proof. Will the combined likelihood ratio $\frac{\mathsf{P}(a \wedge b|A \wedge B)}{\mathsf{P}(a \wedge b|\neg(A \wedge B))}$ also be above the threshold? The answer is affirmative. As we argued earlier, the combined likelihood ratio is never below the lowest of the individual likelihood ratio. If both individual likelihood ratios meet the threhsold, so does the combined likelihood ratio. But this approach faces another problem, the same problem that plagues the Bayes factor. That is, likelihood ratios still fail to capture the other direction of the conjunction principle, what we called distribution.

For suppose evidence $a \wedge b$ supports $A \wedge B$ to the required threshold t. If evidential support is measured by the likelihood ratio, the threshold in this case should be some order of magnitude greater than one. If the combined likelihood ratio meets the threshold t_{LR}, one of the individual likelihood ratios may well be below t_{LR} So—if the standard of proof is interpreted using evidential support measured by the likelihood ratio—even though the conjuction $A \wedge B$ was proven according to the desired standard, one of individual claims might not. The right-to-left direction of the conjunction principle—that is, if $S[a \wedge b, A \wedge B]$, then $S[a, A] \wedge S[b, B]$, what we called earlier the distribution principle (DIS1)—fails.

Consider now the second and less objectionale extrapolation principle discussed earlier. This is the principle that holds the evidence fixed throughout, repeated below for convenience:

$$\text{If } \mathrm{S}[a \wedge b, A \wedge B], \text{ then } \mathrm{S}[a \wedge b, A] \text{ and } \mathrm{S}[b \wedge b, B]. \tag{DIS2}$$

Here again, the support of $a \wedge b$ in favor of $A \wedge B$ could exceed that of $a \wedge b$ in favor of A alone (or B alone) if evidential support is measured using likelihood ratios.[20] Even this second, seemingly unobjectionable version of extrapolation fails. The same counterintuitive consequences that arose with

[20]Note that, assuming either of the Baysian networks in Figure 1, $S[a \wedge b, A] = \frac{\mathsf{P}(a \wedge b|A)}{\mathsf{P}(a \wedge b|\neg A)} = \frac{\mathsf{P}(a|A)}{\mathsf{P}(a|\neg A)} \times \frac{\mathsf{P}(b|A)}{\mathsf{P}(b|\neg A)}$, where

$$\frac{\mathsf{P}(b|A)}{\mathsf{P}(b|\neg A)} = \frac{\mathsf{P}(B|A) \times \frac{\mathsf{P}(b|B)}{\mathsf{P}(b|\neg B)} + (1 - \mathsf{P}(B|A))}{\mathsf{P}(B|\neg A) \times \frac{\mathsf{P}(b|B)}{\mathsf{P}(b|\neg B)} + (1 - \mathsf{P}(B|\neg A))}.$$

SEE PROOF IN EARLIER CHAPTERS. If A and B are assumed to be probabilistically independent, the numerator and the denominator will be the same, so $\frac{\mathsf{P}(b|A)}{\mathsf{P}(b|\neg A)} = 1$. Thus, $S[a \wedge b, A] = \frac{\mathsf{P}(a|A)}{\mathsf{P}(a|\neg A)} = S[a, A]$. Since $S[a, A] < S[a \wedge b, A \wedge B]$ (see Figure 3), it follows $S[a \wedge b, A] < S[a \wedge b, A \wedge B]$. **What if A and B are dependent? Need simulation data here.**

Bayes factor manifest themselves here. The distribution paradox persists.

M: Might be goot to emphasize how devastaing this finding is for legal probabilists who endorsed likelihood ratios as the solution to many problems with legal probabilism.

M: This whole section should be generalized to the case in which A and B are not independent using the simulation data.

5.8 Should evidential strength be sensitive to priors?

Neither the Bayes factor nor the likelihood ratio managed to fully justify both directions of the conjunction principle. One direction, aggregation, was justified. So the original concern that was driving Cohen's formulation of the conjuction paradox was addressed. But the other direction, distribution, failed. The failure of distribution creates a paradox of its own, what we called distribution paradox. It is odd that one could have sufficiently strong evidence in support of $A \wedge B$, while not having sufficiently strong evidence for A or B. This occurs even when A and B are probabilistically independent. If they were dependent of one another–say A and B were mutually reinforcing—it is possible the evidence would strongly support the conjunction, but not one of the conjuncts in isolation (becuase the additional support from the other claim, A or B, would be missing). But the failure of distribution manifests itself even when A and A are independent. What should we make of this? This problem exists for both the Bayes factor and the likelihood ratio.

Start by noticing that the larger Bayes factor associated with the composite claim, holding the evidence fixed, need not be a sign of stronger evidence, but merely an artifact of the lower prior probability of the composite claim. The same can be said for the combined likelihood ratio. Holding fixed the sensitivty and specificity of a and b, the combined likelihood ratio can be changed by varying the priors of A and B. The lower the priors, the stronger the likelihood ratio. So, arguably, the same body of evidence may strongly suport the composite hypothesis $A \wedge B$, while failing to stronggly support A or B simply because $A \wedge B$ has a lower prior probability and this lower prior probability, everything else being equal, inflates the likelihood ratio or the Bayes factor qua measures of evidential strength.

To circumvent the phenomenon of prior sensitivity, evidential strength can be thought as a general measure of the relationship between prior and posterior. The graph in Figure 4 below represents to what extent the evidence changes the prior probability of a select hypothesis for any value of the prior. The graph compares the 'base line' (representing no the change from prior to posterior) and the 'posterior line' (representing the posterior as a function of the prior for a given assignment of sensitivity and specificity of the evidence). Roughly, the larger the area between the two base line and the posterior line, the stronger the evidence. Crucially, this area does not depend on the prior probability of the hypothesis, but solely on the sensitivity and specificity of the evidence. As expected, any improvement in sensitivity or specificity will increase the area between the base line and the posterior line. To be precise, what matters is the ratio of sensitivity to to 1-specificity, not their abosolute values. So evidence with sensitivity and specificity at 0.9 and 0.9 would the same as evidence evidence with sensitivity and specificity at 0.09 and 0.09 becaue $0.9/(1-0.9) = 0.09/(1-0.99)$.

The same approch can model the joint evidential strength of two items of evidence, $a \wedge b$, relative to the combined hypothesis, $A \wedge B$. For simplicity, assume a and b are independent lines of evidence supporting their respective hypothesis A and B probabilistically independent of the other, as in the Bayesian network in Figure 1 (top). The graph in Figure shows how the prior probabilities are impacted by one piece of evidence in support of a single hypothesis—say a supports A—versus two pieces of evidence in support of a joint hypthesis—say $a \wedge b$ supports $A \wedge B$. The base line is lower in in the latter case than in the former case, simply because the prior probability of $A \wedge B$ is lower than the prior probability of A. The prior of A equals x and the prior of $A \wedge B$ equals x^2. The assumption here is that A and B have the same prior probability, and as noted before, are probabilistically independent of one another.

Note the difference between the graph on top and the one at the bottom where the base line for the individual and the composite case are equalized. TWe should ensure that a composiite claim such as $A \wedge B$ is given the same probability as an indiividual claim such as A or B. If the claims are independent and equiprobable, let x be the prior probability of an individual claim (when it is considered in isolation) and let $x^{1/k}$ the prior probability of the same individual claim when it is part of a composite claim that consists of k claims. In this way, the prior probability of the composite claim equals the prior probability of the individual claim since $(x^{1/k})^k = x$, as desired. These different claims are then plotted having the same priors. We can then see the impact the evidence has on the individual claim versus the impact it has on the composite claim. The impact is roughly the same for high values of sensitivity and specificity, but still different for lower values, especially if the priors are below 50%. The same equalization can be

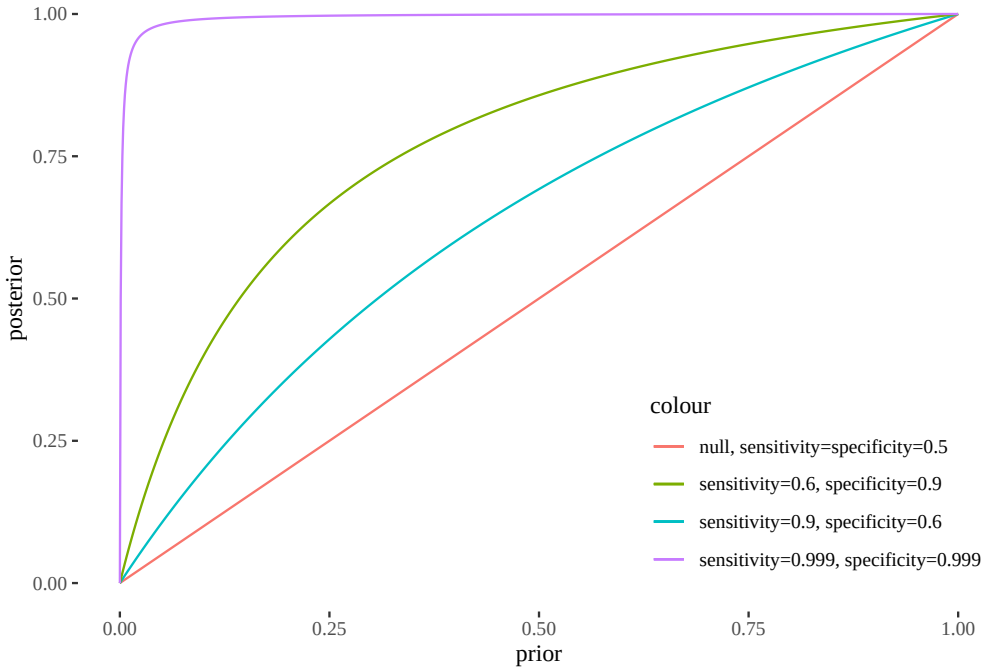
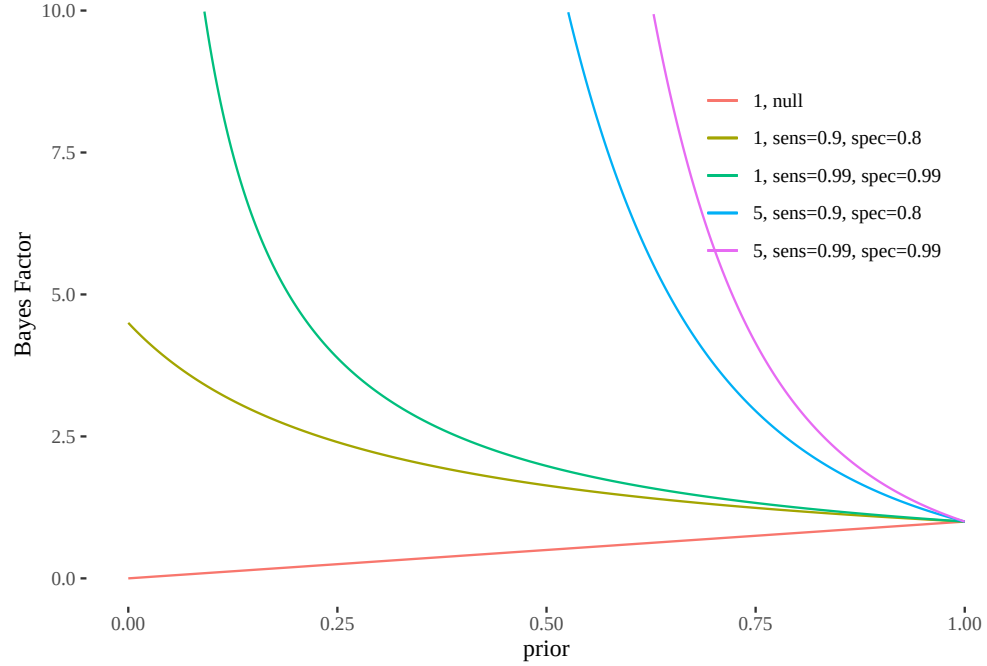


Figure 4: Strength of evidence as a relationship between prior and posterior. The further away the posterior line from the null line, the stronger the evidence irrespective of the prior probability of the hypothesis.

done with the Bayes factor, yielding similar results. Compare the difference betwen teh graph (top) in which teh Bayes factor has not been equalized relative to the priors and the graph (bottom) in which it has. In the latter case, individual Bayes factor and joint Bayes factor tend to converge especially for relatively high values of sensitivity and specificity SEE FIGURE BELOW.

\begin{figure}

```
## Warning: Removed 91 rows containing missing values (geom_path).
## Warning: Removed 526 rows containing missing values (geom_path).
## Warning: Removed 628 rows containing missing values (geom_path).
```




```
## Warning: Removed 91 rows containing missing values (geom_path).
```

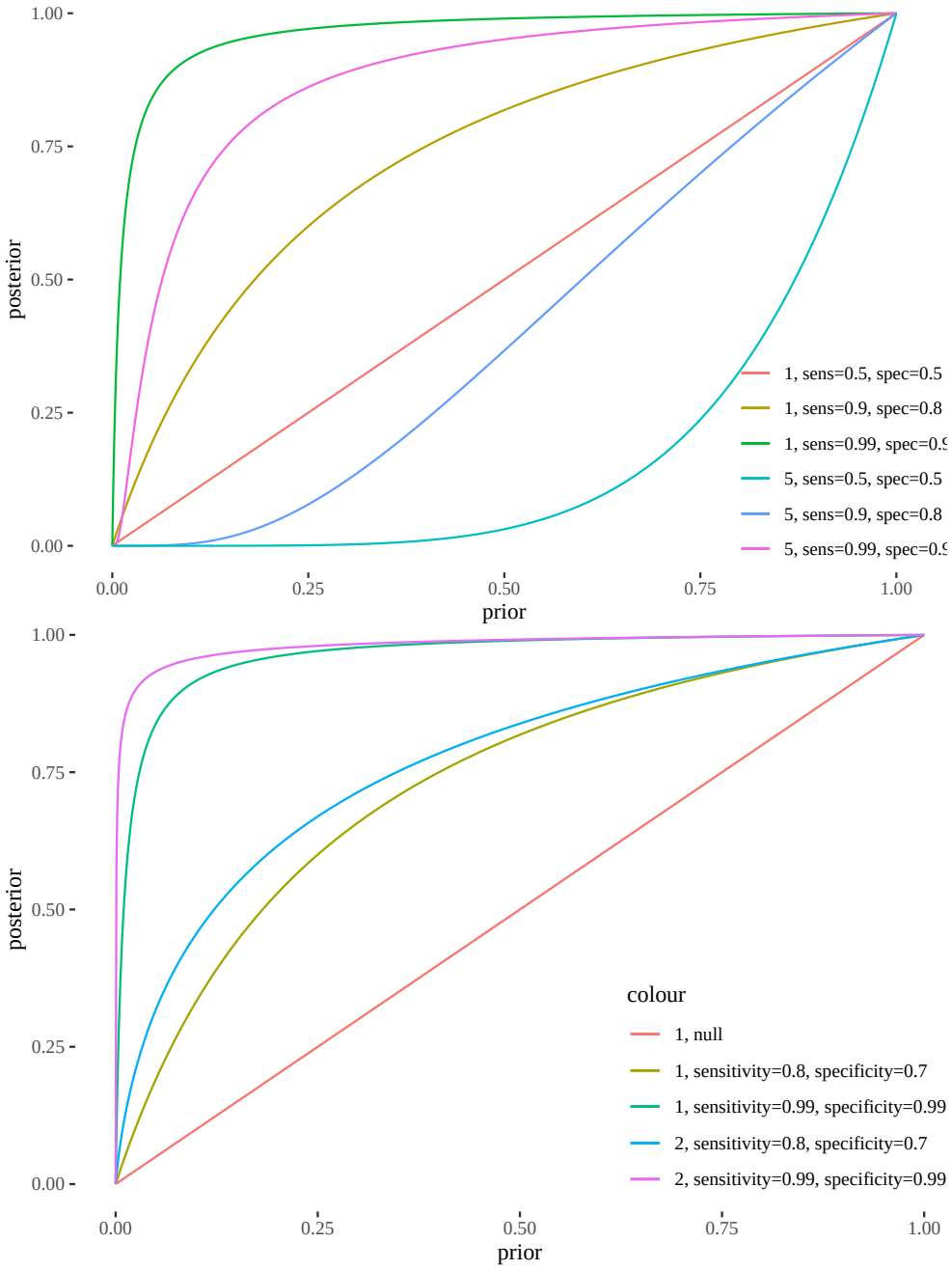
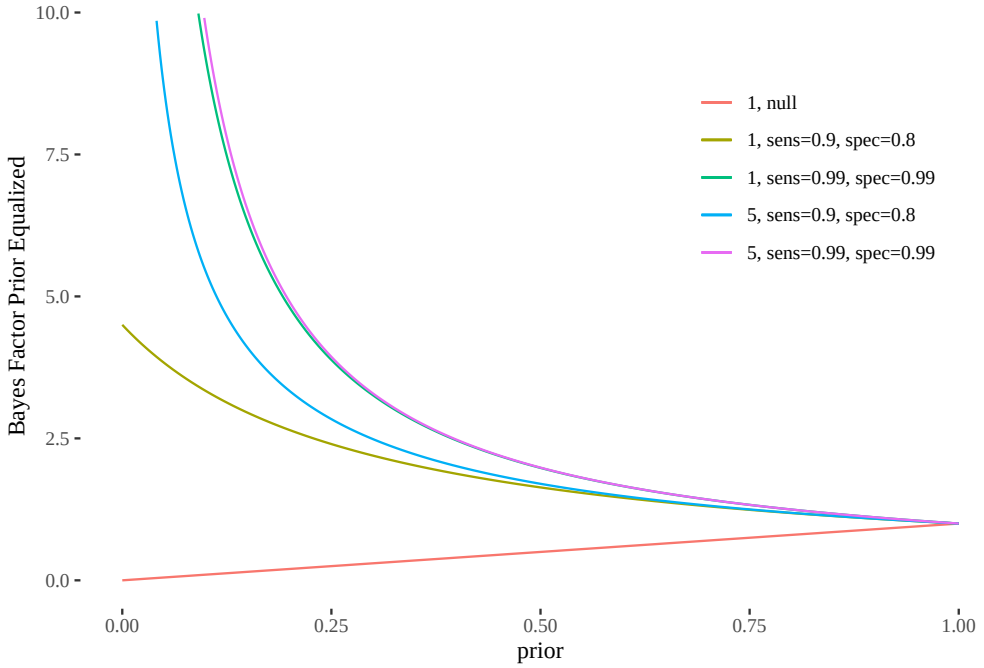


Figure 5: Strenght of evidence as a general relationship between prior and posterior. The comparison here is betwen individual support and joint support. Top graph: The null line for joint support ($y = x * x$) is much below the null line for individual support ($y = x$). Bottom graph: the two null lines are equalized and the posterior lines adjusted accordingly. The posterior lines for individual and joint support get much closer especially for high posterior probability values.

```
## Warning: Removed 41 rows containing missing values (geom_path).
## Warning: Removed 98 rows containing missing values (geom_path).
```



\end{figure}

The approach we have just dscribed—equalizing the prior probabilities across different hypotheses—does not entirely eliminate the distribution paradox but would reduce its scope significantly. There would still be cases in which a composite hypothesis, say $A \wedge B$, receives stronger support than an individual hypothesis, given the body of evidence. So factoring out prior reduces the incidence of the distribition paradox but does not eliminate it entirely. Sensitivity to priors does play a role in the paradox, but it seems that it is not the only factor at play? What else is responsible for the distribution paradox?

NOT SURE HOW TO ANSWER THIS QUESTION

A more general question here is whether we should seek a measure of the strength of evidence—especially, combined evidence for a composite hypothesis—that is not sensitive to priors? Intuitively, the strength of the evidence should not depend on the prior probability of the hypothesis, but solely on the quality of the evidence itself. The prior probability of the hypothesis seems extrinsic to the quality of the evidence since the latter should solely depend on the sensitivity and specificity of the evidence relative to the hypothesis of interest. Strength of evidence determines how much the evidence changes, upwards or downwards, the probability of a hypothesis. However, as the prior probability increases, the smaller the impact that the evidence will have on the probability of the hypothesis. If the prior is close to one, the evidence would have marginal if not null impact. But this does not mean that the evidence weakens as the prior probability of the hypothesis goes up. For consider the same hypothesis which in one context has a very high prior probability and in another has a moderate prior probability (say a disease is common in a population but rare in another). The outcome of the same diagnostic test (say a positive test result) performed on two people, each drawn from two populations, should not count as stronger evidence in one case than in the other. After all, it is the same test that was performed and thus the quality of the evidence should be the same.

This intuition turns out to be incorrect empirically, however. One study in the medical literature has shown, surprisingly, that the sensitivity of a diagnostic test is independent of the prior of the hypothesis beign tested—say whether the patient has a medical condition. However, specificity is dependent on the prior of the hypothesis:

> Overall, specificity tended to be lower with higher disease prevalence; there was no such systematic effect for sensitivity (page E537). Source: Variation of a test's sensitivity and specificity with disease prevalence Mariska M.G. Leeflang, Anne W.S. Rutjes, Johannes B. Reitsma, Lotty Hooft and Patrick M.M. Bossuyt CMAJ August 06, 2013 185 (11) E537-E544; DOI: https://doi.org/10.1503/cmaj.121286

The authors of the study, however, caution that

> Because sensitivity is estimated in people with the disease of interest and specificity in people without the disease of interest, changing the relative number of people with and without the disease of interest should not introduce systematic differences. Therefore, the effects that we found may be generated by other mechanisms that affect both prevalence and accuracy.

So, according to the authors, changes in prevalence need not directly affect specificity since variations in prevalence and variation in specificity may have a common cause. Our earlier calculations about combined specificity and sensitivity agree with experimental results, namely, only specificity depends on the priors. Our calculations, in fact, show that different priors for the individual claim do affect specificity. The variation of specificity in the result of of splitting the negation of the composite hypthesis $\neg(A \wedge B)$ into three further scenarios, $\neg A \wedge B$, $B \wedge \neg A$ and $\neg A \wedge negB$. This prior sensitivity, of course, only applies to composite hyotheses, but to some extent, any hypothesis can be analyzed as a composite hypothesis. The claim that the defedant was running down 5th avnue can be broken down in the conjuction that the defendant was running and that the defendant was at 5th avenue. Any claim, under some level of description, is a composite hypothesis. So, perhpas, the quality or strength of the evidence should depend on the the priors whether the hypothesis is composite or not. Is this another example of base rate neglect?

M: Might be good to have a simulation here that makes vidid why combined specificity is in fact dependent on the priors. Maybe it is, after all, a fallacy to think that the quality/strength of the evidence should be independent of the priors.

Let's grant that the quality of the evidence should depend, contrary to our initial intuition, on the prior probability of the hypotheses. If that is so, it would not be natural to see that evidence – the same evidence – strongly favors $A \wedge B$ without strongly favoring A or B. Perhaps we can make sense of this if we keep in mind the comparison between hypothesis we are making here.

NOT SURE HOW TO CONTIBUE HERE THOUGH!

TO DO:

1. NOTE THAT EVEN BY EQUALIZING PRIORS, TE DISTRBUTION PARADOX DOES NOT GO AWAY. SO WHAT ELSE IS GOING ON HERE? NEED TO MAKE COMPATISON BETWEEN HYPOTHESES. NEED TO FIGURE THIS OUT!
2. TRY TO MAKE SENSE OF THIS, IT IS INTITVELY ACCEPTABLE THAT SUPPORT FOR COMBINED CLAIM, EVEN HOLDING FIXED THE SAME EVIDENCE, SHOULD BE STRONGER THEN SUPPORT FOR INDIVIDUAL CLAIM? THAT IS CLEARLY ODD AND GOES AGAINST COMMON ASSUMPTIONS.

5.9 Which Measure of (Combined) Evidential Support?

THIGNS TO ADD:

1. THE MIN SEEMS TO BE THE MEASURE FOR COMPSUTE CLAIMS TAHT CAPTURE AGGREGATION AND DISRIBUTION BEST. SO THE QUESTION IS WHAT PROBABILISTIC MEASURE CAPTURES MIN?
2. CAN USE LR SEEMS IT IS INDEFFERENCT TO PRIORS, BUT THIS IS NOT THE CASE FOR THE COMPOSITE CLAIM. SO SAME PROBLEM AS WITH BF.
3. IS DEPEDENCY ON PRIOR ANOTHER EXAMPLE OF BASE RATE NEGLECT. WE NEGLECT BAE RATE IN CALCULATING POSTERIOR BUT ALSO IN CALCULATING STRENGHT OF EVIDENCE? WE JUST NEED TO LIVE WITH THE FACT THAT WE HAVE A POOR UNDERSTAND OF EVIDENE THOUGHT PERHPAS GOOD ENOUGH TO GET BY IN THE WORLD. CONNECT TO POINT 1 AND MIN FUNCTION.
4. RUN SIMULATION TO SHOW THAT, ACTUALLY, ONE TEST FOR COMPOSITE CLAIM WOULD PERFORM BETTER (BETTER LR) THEN TEH SAME TEST FOR INDIVIDUAL CLAIM (WORSE LR). SO TEST COUDL PASS QUALITY CRONTROL IN ONE CASE BUT NOT IN THE OTHER.
5. ANY DIFFERENCE BETWEEN AREA UNDER TEH CURVE REPRESENTATION OF EVIDENCE STRENGHT AND LR. WHAT IF WE DECRESES SENSITIVITY AND SPECIFICITY. WHAT WOULD HAPPEN TO THE TWO MEASURES.

5.10 The comparative strateGy

Instead of thinking in terms of absolute thresholds—whether relative to posterior probabilities, the Bayes factor or the likelihood ratio—the standard of proof can be understood comparatively. This suggestion

has been advanced by Cheng (2012) following the theory of relative plausibility by **REFERENCE TO ALLEN AND PARDO HERE**. Say the prosecutor or the plaintiff puts foward a hypothesis H_p about what happened. The defense offers an alternative hypothesis, call it H_d. On this approach, rather than directly evaluating the support of H_p given the evidence and comparing it to a threshold, we compare the support that the evidence provides for two competing hypotheses H_p and H_d, and decide for the one for which the evidence provides better support.

It is controversial whether this is what happens in all trial proceedings, especially in criminal trials. The defense may elect to challenge the hypothesis put foward by the other party without proposing one of its own. For example, in the O.J. Simpson trial the defense did not advance its own story about what happened, but simply argued that the evidence provided by the prosecution, while significant on its face to establish OJ's guilt, was riddled with problems and deficencies. This defense strategy was enough to secure an acquittal. So, in order to create a reasonable doubt about guilt, the defense does not always provide a full-fledged alternative hypothesis. The supporters of the comparative approach, however, will respond that this could happen in a small number of cases, even though in general—especially for tactical reasons—the defense will provide an alternative hypothesis. After all, not to provide one would usually amount to an admission of criminal or civil liability.

Setting aside this controversy for the time being, let's first work out the comparative strategy using posterior probabilities. More specifically, given a body of evidence E and two competing hypotheses H_p and H_d, the probability $\mathsf{P}(H_p|E)$ should be suitably higher than $\mathsf{P}(H_d|E)$, or in other words, the ratio $\frac{\Pr H_p|E}{\Pr H_d|E}$ should be above a suitable threshold. Presumably, the ratio threshold shoud be higher for criminal than civil cases. In fact, in civil cases it seems enough to require that the ratio $\frac{\Pr H_p|E}{\Pr H_d|E}$ be avove 1, or in other words, $\mathsf{P}(H_p|E)$ should be higher than $\mathsf{P}(H_d|E)$. Note that H_p and H_d need not be one the negation of the other. Whenever two hypotheses are one the negation of the other, $\frac{\mathsf{P}(H_p|E)}{\mathsf{P}(H_d|E)} > 1$ implies that $\mathsf{P}(H_p|E) > 50\%$, the standard probabilistic intepretation of the preponderemca standard.

One advantage of this approach—as Cheng shows—is that expected utility theory can set the appropriate comparative threshold t as a function of the costs and benefits of trial decisions. For simplicity, suppose that if the decision is correct, no costs result, but incorrect decisions have their price (**REFERENCE TO EARLIER CHAPTER FOR MORE COMPLEX COST STRUCTURE**). The costs of a false positive is c_{FP} and false negative is C_{FN}, both greater than zero. Intuitively, the decision rule should minimize the expected costs. That is, a finding against the defendant would be acceptable whenever its expected costs—$\mathsf{P}(H_d|E) \times c_{FP}$—are smaller than the expected costs of an acquittal—$\mathsf{P}(H_p|E) \times c_{FN}$— or in other words:

$$\frac{\mathsf{P}(H_p|E)}{\mathsf{P}(H_d|E)} > \frac{c_{FP}}{c_{FN}}.$$

In civil cases, it is customary to assume the costs ratio of false postives to false negatives equals one. So the rule of decision would be: Find against the defedant whenever $\frac{\mathsf{P}(H_p|E)}{\mathsf{P}(H_d|E)} > 1$ or in other words $\mathsf{P}(H_p|E)$ is greater than $\mathsf{P}(H_d|E)$. In criminal trials, the costs ratio is usually considered higher, since convicting an innocent (false positive) should be more harmful or morally objectionable than acquitting a guilty defendant (false negative). Thus, the rule of decision in criminal proceedings would be: Convict whenever $\mathsf{P}(H_p|E)$ is significantly greater than $\mathsf{P}(H_d|E)$.

Does the comparative strategy just outlined solve the difficulty about conjuction? We will work through a stylized case used by Cheng himself. Suppose, in a civil case, the plaintiff claims that the defendant was speeding (S) and that the crash caused her neck injury (C). Thus, the plaintiff's hypothesis H_p is $S \wedge C$. Given the total evidence E, the conjuncts, taken separately, meet the decision threshold:

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

The question is whether $\frac{\mathsf{P}(S \wedge C|E)}{\mathsf{P}(H_d|E)} > 1$. To answer it, we have to decide what the defense hypothesis H_d should. Cheng reasons that there are three alternative defense scenarios: $H_{d_1} = S \wedge \neg C$, $H_{d_2} = \neg S \wedge C$, and $H_{d_3} = \neg S \wedge \neg C$. How does the hypothesis H_p compare to each of them? Assuming independence between C and S, we have

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

So, whatever the defense hypothesis, the plaintiff's hypothesis is more probable. At least in this case, whenever the elements of a plaintiff's claim satisfy the decision threshold, so does their conjunction. The left-to-right direction of the conjunction principle—what we called aggregation—has been vidicated. But what about the opposite direction, what we called extrapolation? Interestingly, if the threshold is just 1—as might be appropriate in civil cases—extrapolation would be satisfied. Even if $\frac{\mathsf{P}(S|E)\mathsf{P}(C|E)}{\mathsf{P}(\neg S|E)\mathsf{P}(\neg C|E)}$ might be strictly greater than $\frac{\mathsf{P}(C|E)}{\mathsf{P}(\neg C|E)}$ or $\frac{\mathsf{P}(S|E)}{\mathsf{P}(\neg S|E)}$, wehenver the former is greater than one the latter must be greater than one. However, suppose the threshold is more stringent than one, as might be appropriate for criminal cases. For some constituent claims A and B in a criminal case, whenever $\frac{\mathsf{P}(A|E)\mathsf{P}(B|E)}{\mathsf{P}(\neg A|E)\mathsf{P}(\neg B|E)}$ barely meet the threhold t, $\frac{\mathsf{P}(A|E)}{\mathsf{P}(\neg A|E)}$ or $\frac{\mathsf{P}(B|E)}{\mathsf{P}(\neg B|E)}$ could be below t. Since the evidence is held fixed throughout, this would be a violation of the extrapolation principle (EXT2).

Another problem with this approach is that much of the heavy lifting here is done by the strategic splitting of the defense line into multiple scenarios. Now suppose $\mathsf{P}(H_p|E) = 0.37$ and the probability of each of the defense lines given E is 0.21. This means that H_p wins with each of the scenarios, so we should find against the defendant. But should we? Given the evidence, the accusation is very likely to be false, because $\mathsf{P}(\neg H_p|E) = 0.63$? 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 $\neg H_p$ needs to be divided. This normally would lead to the probability of each of them being even lower (because now $\mathsf{P}(\neg H_p)$ needs to be "split" between more scenarios). So, if we take this approach seriously, the more elements a case has, the more at disadvantage the defense is. This seems undesirable.

REPEAT SAME ARGUMENT USING LR AND COMPARATIVE STRATEGY

5.11 Specific Narratives [IDEAS OF A SOLUTION]

So far we have assumed the most natural probabilistic interpretation of proof standards, one that posits a threshold on the posterior probabilities of a generic hypothesis such as guilt or civil liability. In criminal cases, the requirement is formulated as follows: guilt is proven beyond a reasonable doubt provided $\Pr(G|E)$ is above a suitable threshold, say 95%. The threshold is lower in civil trials. Civil liability is proven by preponderance provided $\Pr(L|E)$ is above a suitable threshold, say 50%. The claim that the defendant is guilty or civilly liable can be replaced by a more fine-grained hypothesis, call it H_p, the hypothesis put foward by the prosecutor (or the plaintiff in a civil case), for example, the hypothesis that the defendant killed the victim with a firearm while bulglarizing the victim's apartment. H_p can be any hypothesis which, if true, would entail the defendanat is civilly or criminally liable (according to the governing law). Hypothesis H_p is a more precise description of what happened that establishes, if true, the defendant's guilt or civil liability. In defining proof standards, instead of saying – somewhat generically – that $\mathsf{P}(G|E)$ or $\mathsf{P}(L|E)$ should be above a suitable threshold, a probabilistic interpretation could read: civil or criminal liabbility is proven beyond a reasonable doubt provided $\Pr(H_d|E)$ is above a suitable threshold.

This variation may appear inconsequential. But we argue – perhpas surprisingly – it can address the naked statistical evidence problem and the difficulty about conjunction. Consider the prisoner hypothetical. It is true that the naked statistics make him 99% likely to be guilty, that is, $\mathsf{P}(G|E_s)$. It is 99% likely that he is one the prisoners who attacked and killed the guard. Notice that this a generic claim. It is odd for the prosecution to simply assert that the prisoner was one of those who killed the guard, without saying what he did, how he partook in the killing, what role he played in the attack, etc. If the prosecution offerred a more specific incriminating hypothesis, call it H_p, the probability $\mathsf{P}(H_p|E_s)$ of this hypothesis based on the naked statistical evidence E_s would be well below 99%, even though $\mathsf{P}(G|E_s) = 99\%$. The fact the prisoner on trial is most likely guilty is an artifact of the choice of

a generic hypothesis G. When this hypthesis is made more specific – as it should be – this probability drops significantly. A more detailed defense of this argument is provided in CHAPTER XYZ.

Consider now the difficulty about conjunction, focusing again on criminal cases for the sake of concreteness. This difficulty assumes that prosecutors should establish each element of a crime in isolation. If they manage to prove each element to the desired standard, they have meet their burden. This is an artificial view of legal proof. Consider a Washington statute about negligent driving:

> (1)(a) A person is guilty of negligent driving in the first degree if he or she operates a motor vehicle in a manner that is both negligent and endangers or is likely to endanger any person or property, and exhibits the effects of having consumed liquor or marijuana or any drug or exhibits the effects of having inhaled or ingested any chemical, whether or not a legal substance, for its intoxicating or hallucinatory effects. RCW 46.61.5249

In other words, a prosecutor who wishes to establish beyond a reasonable doubt that the defendant is guilty of negligent driving should establish:

> (a) the the defendant operated a vehicle (b) that, in operating a vehicle, the defendant did so in negligente manner (c) that, in operating a vechicle, the defendant did so in a manner likely to endanger a person pr property (d) that the defendant – presumably, immediately after the incident – exihibited the signs of intoxication by liquor or drugs

These four claims form a common narratives about what happenned. NEED TO COMPLETe THIS. BASIC IDEA IS THAT, FIRST, YOU ESTABLISH THE NARTRATIVES, AND, SECOND, THE NARRATIVE IF TRUE PROVES EACH ELEMENT. SO TO PROVE EACH ELEMENT SIMPLY MEANS TO PROVE A NARRARTIVE FROM WHICH ALL ELEMENTS FOLLOW DEDUCTIVELY. THERE IS NO PINT IN THIKING ABOUT WHETHER EACH ELEMENT HAS BEEN PROVEN.

6 The likelihood strategy

Focusing on posterior probabilities is not the only approach that legal probabilists can pursue. By Bayes' theorem, the following holds, using G and I as competing hypotheses:

$$\frac{\Pr(G|E)}{\Pr(I|E)} = \frac{\Pr(E|G)}{\Pr(E|I)} \times \frac{\Pr(G)}{\Pr(I)},$$

or using H_p and H_d as competing hypotheses,

$$\frac{\Pr(H_p|E)}{\Pr(H_d|E)} = \frac{\Pr(E|H_p)}{\Pr(E|H_d)} \times \frac{\Pr(H_p)}{\Pr(H_d)},$$

or in words

$$\textit{posterior odds} = \textit{likelihood ratio} \times \textit{prior odds}.$$

A difficult problem is to assign numbers to the prior probabiliteis such as $\Pr(G)$ or $\Pr(H_p)$, or priors odds such as $\frac{\Pr(G)}{\Pr(I)}$ or $\frac{\Pr(H_p)}{\Pr(H_d)}$.

DISCUSS DIFFICULTIES ABOUT ASSIGNING PRIORS! WHERE? CAN WE USE IMPRECISE PROBABILKITIES T TALK ABOUT PRIORS – I.E. LOW PRIORS = TOTAL IGNORANCE = VERY IMPRECISE (LARGE INTERVAL) PRIORS? THE PROBLME WITH THIS WOULD BE THAT THERE IS NO UPDSATING POSSIBLE. ALL UPDATING WOULD STILL GET BACK TO THE STARTING POINT. DO YOU HAVE AN ANSWER TO THAT? WOULD BE INTERETSING TO DISCUSS THIS!

Given these difficulties, both practical and theoretical, one option is to dispense with priors altogether. This is not implausible. Legal disputes in both criminal and civil trials should be decided on the basis of the evidence presented by the litigants. But it is the likelihood ratio – not the prior ratio – that offers the best measure of the overall strength of the evidece presented. So it is all too natural to focus on likekihood ratios and leave the priors out of the picture. If this is the right, the question is, how would a probabilistic interpretation of standards of proof based on the likelihood rato look like? At its simplest, this stratgey will look as follows. Recall our discussion of expected utility theory:

$$\text{convict provided} \frac{cost(CI)}{cost(AG)} < \frac{\Pr(H_p|E)}{\Pr(H_d|E)},$$

which is equivalent to

$$\text{convict provided} \frac{cost(CI)}{cost(AG)} < \frac{\Pr(E|H_p)}{\Pr(E|H_d)} \times \frac{\Pr(H_p)}{\Pr(H_d)}.$$

By rearraing the terms,

$$\text{convict provided} \frac{\Pr(E|H_p)}{\Pr(E|H_d)} > \frac{\Pr(H_d)}{\Pr(H_p)} \times \frac{cost(CI)}{cost(AG)}.$$

Then, on this intepretation, the likelihood ratio should be above a suitable threshold that is a function of the cost ratio and the prior ratio. The outstanding question is how this threshold is to be determined.

6.1 Kaplow

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:

$$\mathsf{P}(H_\Pi|E) \times G > \mathsf{P}(H_\Delta|E) \times L \tag{2}$$

Now, (2) is equivalent to:

$$\frac{\mathsf{P}(H_\Pi|E)}{\mathsf{P}(H_\Delta|E)} > \frac{L}{G}$$
$$\frac{\mathsf{P}(H_\Pi)}{\mathsf{P}(H_\Delta)} \times \frac{\mathsf{P}(E|H_\Pi)}{\mathsf{P}(E|H_\Delta)} > \frac{L}{G}$$
$$\frac{\mathsf{P}(E|H_\Pi)}{\mathsf{P}(E|H_\Delta)} > \frac{\mathsf{P}(H_\Delta)}{\mathsf{P}(H_\Pi)} \times \frac{L}{G}$$
$$LR(E) > \frac{\mathsf{P}(H_\Delta)}{\mathsf{P}(H_\Pi)} \times \frac{L}{G} \tag{3}$$

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

6.2 Dawid

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.[21] 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,

$$O_{posterior}(A) = O_{prior}(A|a) = \underbrace{LR_{prior}(a|A)}_{\text{conditional likelihood ratio}} \times O_{prior}(A) \tag{4}$$

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 $1/2$.[22]

[21] 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.

[22] 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:

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

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.

6.3 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.

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:[23]

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

$\mathsf{P}_{prior}(a|\neg A) = 1 - \mathsf{P}_{prior}(\neg a|\neg A) = 0.3$, 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 (4), 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 $\neg A \wedge \neg B$.[24,~25]

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:

$$\begin{aligned}
\mathsf{P}(a \wedge b) = {} & \mathsf{P}(a \wedge b|A \wedge B)\mathsf{P}(A \wedge B) + \\
& + \mathsf{P}(a \wedge b|A \wedge \neg B)\mathsf{P}(A \wedge \neg B) \\
& + \mathsf{P}(a \wedge b|\neg A \wedge B)\mathsf{P}(\neg A \wedge B) + \\
& + \mathsf{P}(a \wedge b|\neg A \wedge \neg B)\mathsf{P}(\neg A \wedge \neg B)
\end{aligned} \tag{5}$$

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

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

[23] 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.

[24] 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).

[25] 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.

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.

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

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).

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

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

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

that is, whether $28 > 28p_{11} - 12p_{00}$ (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:

$$p_{11} + p_{10} + p_{01} + p_{00} = 1 \tag{6}$$

So, we can use this with (5) to get:

$$\begin{aligned} \mathsf{P}(a \wedge b) &= \mathbf{ab}p_{11} + \mathbf{a}(\mathbf{1} - \mathbf{b})p_{10} + (\mathbf{1} - \mathbf{a})\mathbf{b}p_{01} + (\mathbf{1} - \mathbf{a})(\mathbf{1} - \mathbf{b})p_{00} \\ &= p_{11}\mathbf{ab} + p_{10}(\mathbf{a} - \mathbf{ab}) + p_{01}(\mathbf{b} - \mathbf{ab}) + p_{00}(1 - \mathbf{b} - \mathbf{a} + \mathbf{ab}) \end{aligned} \tag{7}$$

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

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

To answer our question we therefore have to compare the content of (8) to p_{11} and our claim holds just in case:

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

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

$$p_{11}\mathbf{ab} + p_{10}(\mathbf{a} - \mathbf{ab}) + p_{01}(\mathbf{b} - \mathbf{ab}) + p_{00}(1 - \mathbf{b} - \mathbf{a} + \mathbf{ab}) < \mathbf{ab} \tag{9}$$

Proving (9) is therefore our goal for now. This is achieved by the following reasoning:[26]

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.	$p_{10}2\mathbf{ab} > p_{10}\mathbf{a},\quad p_{01}2\mathbf{ab} > p_{01}\mathbf{b}$	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 (6)
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.

6.4 Kaplow

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

[27] 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.

7 Challenges (again)

7.1 Likelihood ratio and the problem of the priors

7.2 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

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

[27] 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.

decision standard, and so addresses quite a different question, and so it is not clear that 'theparadox' 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:

- 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 (10), the fourth is that if the conditions in (10) 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:

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

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:

$$\mathsf{P}(A|a) = \frac{\mathsf{P}(a|A)\mathsf{P}(A)}{\mathsf{P}(a)} = \frac{0.6 \times 0.8}{0.6 \times 0.8 + 0.4 \times 0.2} \approx 0.8571$$

$$\mathsf{P}(B|b) = \frac{\mathsf{P}(b|B)\mathsf{P}(B)}{\mathsf{P}(b)} = \frac{0.6 \times 0.2}{0.6 \times 0.2 + 0.4 \times 0.8} \approx 0.272$$

$$\begin{aligned}\mathsf{P}(A \wedge B|a \wedge b) &= \frac{\mathsf{P}(a \wedge b|A \wedge B)\mathsf{P}(A \wedge B)}{\mathsf{P}(a \wedge b|A \wedge B)\mathsf{P}(A \wedge B) + \mathsf{P}(a \wedge b|A \wedge \neg B)\mathsf{P}(A \wedge \neg B) + \\ &\quad + \mathsf{P}(a \wedge b|\neg A \wedge B)\mathsf{P}(\neg A \wedge B) + \mathsf{P}(a \wedge b|\neg A \wedge \neg B)\mathsf{P}(\neg A \wedge \neg B)} \\ &= \frac{\mathbf{ab}p_{11}}{\mathbf{ab}p_{11} + \mathbf{a}(1-\mathbf{b})p_{10} + (1-\mathbf{a})\mathbf{b}p_{01} + (1-\mathbf{a})(1-\mathbf{b})p_{00}} \approx 0.147\end{aligned}$$

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 (10)? 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.

7.3 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 (3). $\mathsf{P}(A) = \mathsf{P}(\neg A) = 0.5$, $\mathsf{P}(\neg A)/\mathsf{P}(A) = 1$ $\mathsf{P}(E|A) = 1$, $\mathsf{P}(E|\neg A) = 0.4$. (3) 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?[28]

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 $A \wedge B$

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

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 (**??**) 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

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

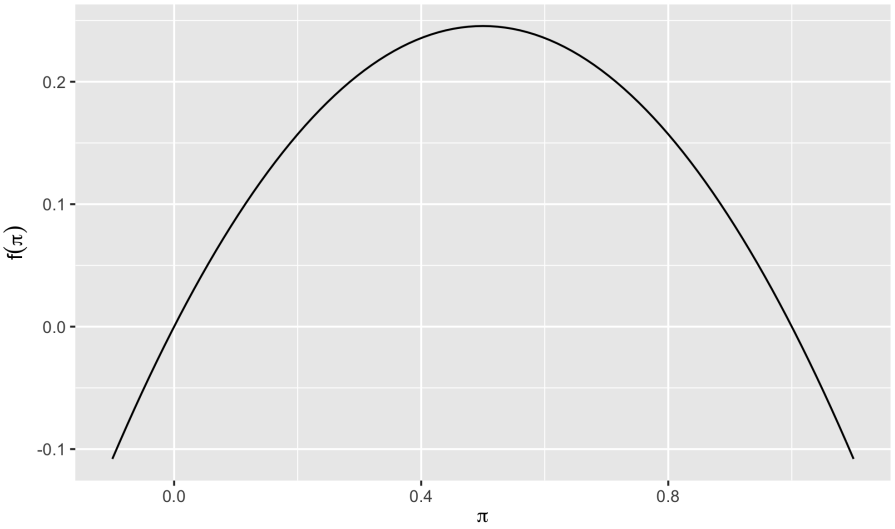
Multiply both sides first by 009π and then by π:

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

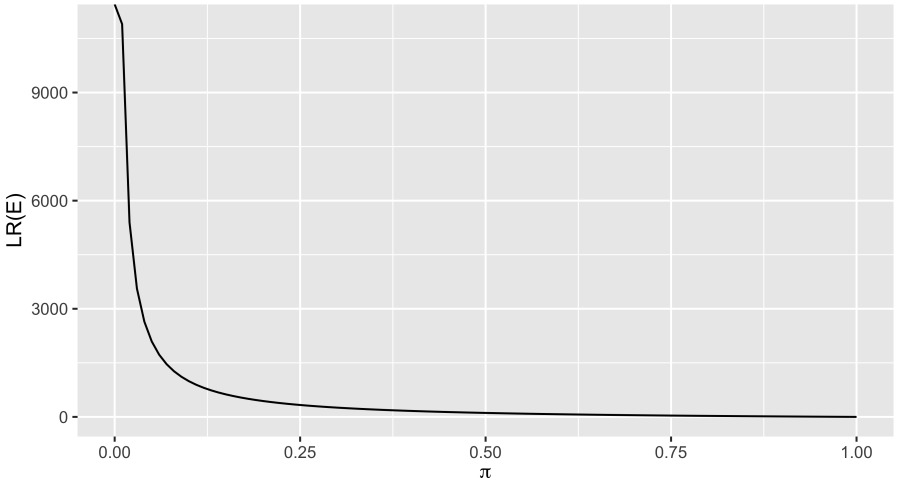
Simplify and call the resulting function f:

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

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:



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.

8 Probabilistic Thresholds Revised

8.1 Likelihood ratios and naked statistical evidence

8.2 Conjunction paradox and Bayesian networks

9 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 $\frac{\mathsf{P}(H\|E)}{\mathsf{P}(H\|\neg E)}$	- 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, 2019a), 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.

References

Allen, R. J. (1986). A reconceptualization of civil trials. *Boston University Law Review*, *66*, 401–437.

Allen, R. J., & Leiter, B. (2001). Naturalized epistemology and the law of evidence. *Virginia Law Review*, *87*(8), 1491–1550.

Allen, R. J., & Stein, A. (2013). Evidence, probability and the burden of proof. *Arizona Law Journal*, *55*, 557–602.

Allen, R., & Pardo, M. (2019). Relative plausibility and its critics. *The International Journal of Evidence & Proof*, *23*(1-2), 5–59. https://doi.org/10.1177/1365712718813781

Arkes, H. R., Shoots-Reinhard, B. L., & Mayes, R. S. (2012). Disjunction between probability and verdict in juror decision making. *Journal of Behavioral Decision Making*, *25*(3), 276–294.

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

Blome-Tillmann, M. (2017). "More likely than not" — Knowledge first and the role of bare statistical evidence in courts of law. In A. Carter, E. Gordon, & B. Jarvi (Eds.), *Knowledge first—approaches in epistemology and mind* (pp. 278–292). Oxford University Press. https://doi.org/10.1093/oso/9780198716310.003.0014

Bolinger, R. (2018). The rational impermissibility of accepting (some) racial generalizations. *Synthese*, 1–17.

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.

Cohen, J. (1977). *The probable and the provable*. Oxford University Press. https://doi.org/10.2307/2219193

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

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

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. (2019a). Probability and plausibility in juridical proof. *International Journal of Evidence and Proof*.

Di Bello, M. (2019b). Trial by statistics: Is a high probability of guilt enough to convict? *Mind*.

Ebert, P. A., Smith, M., & Durbach, I. (2018). Lottery judgments: A philosophical and experimental study. *Philosophical Psychology*, *31*(1), 110–138.

Enoch, D., Spectre, L., & Fisher, T. (2012). Statistical evidence, sensitivity, and the legal value of knowledge. *Philosophy and Public Affairs*, *40*(3), 197–224.

Epps, D. (2015). The consequences of error in criminal justice. *Harvard Law Review*, *128*(4), 1065–1151.

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

Friedman, O., & Turri, J. (2015). Is probabilistic evidence a source of knowledge? *Cognitive Science*, *39*(5), 1062–1080.

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).

Hamer, D. (2004). Probabilistic standards of proof, their complements and the errors that are expected to flow from them. *University of New England Law Journal*, *1*(1), 71–107.

Hamer, D. (2014). Presumptions, standards and burdens: Managing the cost of error. *Law, Probability and Risk*, *13*, 221–242.

Hedden, B., & Colyvan, M. (2019). Legal probabilism: A qualified defence. *Journal of Political Philosophy*, *27*(4), 448–468. https://doi.org/10.1111/jopp.12180

Ho, H. L. (2008). *A philosophy of evidence law: Justice in the search for truth*. Oxford University Press.

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

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

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

Kaye, D. H. (1979b). The paradox of the Gatecrasher and other stories. *The Arizona State Law Journal*, 101–110.

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

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

Laudan, L. (2016). *The law's flaws: Rethinking trials and errors?* College Publications. Retrieved from https://books.google.pl/books?id=MvkWvgAACAAJ

Lempert, R. O. (1986). The new evidence scholarship: Analysing the process of proof. *Boston University Law Review*, *66*, 439–477.

Loftus, E. F. (1996). *Eyewitness testimony (revised edition)*. Harvard University Press.

Moss, S. (2018). *Probabilistic knowledge*. Oxford University Press.

Nesson, C. R. (1979). Reasonable doubt and permissive inferences: The value of complexity. *Harvard Law Review*, *92*(6), 1187–1225. https://doi.org/10.2307/1340444

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

Niedermeier, K. E., Kerr, N. L., & Messeé, L. A. (1999). Jurors' use of naked statistical evidence: Exploring bases and implications of the Wells effect. *Journal of Personality and Social Psychology*, *76*(4), 533–542.

Nunn, A. G. (2015). The incompatibility of due process and naked statistical evidence. *Vanderbilt Law Review*, *68*(5), 1407–1433.

Pardo, M. S. (2018). Safety vs. Sensitivity: Possible worlds and the law of evidence. *Legal Theory*, *24*(1), 50–75.

Picinali, F. (2013). Two meanings of "reasonableness": Dispelling the "floating" reasonable doubt. *Modern Law Review*, *76*(5), 845–875.

Posner, R. (1973). *The economic analysis of law*. Brown & Company.

Pritchard, D. (2005). *Epistemic luck*. Clarendon Press.

Pundik, A. (2017). Freedom and generalisation. *Oxford Journal of Legal Studies*, *37*(1), 189–216.

Redmayne, M. (2008). Exploring the proof paradoxes. *Legal Theory*, *14*(4), 281–309.

Roth, A. (2010). Safety in numbers? Deciding when DNA alone is enough to convict. *New York University Law Review*, *85*(4), 1130–1185.

Schwartz, D. S., & Sober, E. R. (2017). The Conjunction Problem and the Logic of Jury Findings. *William & Mary Law Review*, *59*(2), 619–692.

Smith, M. (2018). When does evidence suffice for conviction? *Mind*, *127*(508), 1193–1218.

Stein, A. (2005). *Foundations of evidence law*. Oxford University Press.

Sykes, D. L., & Johnson, J. T. (1999). Probabilistic evidence versus the representation of an event: The curious case of Mrs. Prob's dog. *Basic and Applied Social Psychology*, *21*(3), 199–212.

Taroni, F., Biedermann, A., Bozza, S., Garbolino, P., & Aitken, C. (2014). *Bayesian networks for probabilistic inference and decision analysis in forensic science* (2nd ed.). John Wiley & Sons.

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

Tribe, L. H. (1971). Trial by mathematics: Precision and ritual in the legal process. *Harvard Law Review*, *84*(6), 1329–1393.

Urbaniak, R. (2018). Narration in judiciary fact-finding: A probabilistic explication. *Artificial*

Intelligence and Law, 1–32. https://doi.org/10.1007/s10506-018-9219-z

Volokh, A. (1997). N guilty men. *University of Pennsylvania Law Review*, *146*(2), 173–216.

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

Wasserman, D. T. (1991). The morality of statistical proof and the risk of mistaken liability. *Cardozo L. Rev.*, *13*, 935.

Wells, G. (1992). Naked statistical evidence of liability: Is subjective probability enough? *Journal of Personality and Social Psychology*, *62*(5), 739–752. https://doi.org/10.1037/0022-3514.62.5.739