The Difficulty With Conjunction

Marcello Di Bello and Rafal Urbaniak

1 Introducing the difficulty

According to a common probabilistic interpretation, the standard of proof in legal trials functions as a probability threshold (see Bernoulli, 1713; Dekay, 1996; Kaplan, 1968; Kaye, 1979; Laplace, 1814; Laudan, 2006). In criminal cases, the standard is very stringent, and so the probability threshold should be high, say the defendant's criminal liability should be established with at least .95 probability. In civil cases, the standard is less stringent, so the probability threshold should be lower, say the defendant civil liability should be established with at least .5 probability.

The previous chapter examines a theoretical problem for this probabilistic interpretation: the paradox of naked statistical evidence. This chapter examines another theoretical problem, which we will **the difficulty with conjunction**, also known as **the conjunction paradox**. The chapter is structured as follows. First, we describe the difficulty with conjunction. Next, we explore different strategies that legal probabilists can pursue to respond to this difficulty. These strategies are promising and worth examining. But we show that they are ultimately unsatisfactory. Finally, we articulate our proposal.

Our solution to the difficulty with conjunction complements our solution to the problem of naked statistical evidence. We believe both problems suggest that the claim of criminal or civil liability should not be understood as an abstract, general proposition. It should be understood as a theory or explanation of the facts tailored to the individual defendant on trial. This insight—already formalized using Bayesian networks in chapter XY —contributes to a more adequate understanding of the standard of proof in civil and criminal cases.

REFER TO EARLIER CHAPTER

First formulated by Cohen (1977), the difficulty with conjunction has enjoyed a great deal of scholarly attention every since (Allen, 1986; Allen & Pardo, 2019; Allen & Stein, 2013; Haack, 2014; Schwartz & Sober, 2017; Stein, 2005). This difficulty arises when a claim of wrongdoing, in a civil or criminal proceeding, is broken down into its constituent elements. By the probability calculus, the probability of a conjunction is often lower than the probability of the conjuncts. So, according to the probabilistic interpretation of the standard of proof, it might happen that even when each constituent element (each individual conjunct) is established by the required standard of proof, the overall claim of wrongdoing (the conjunction) fails to meet the required standard. Cohen and others after him believe that this outcome is counter-intuitive and runs contrary to trial practice.

Alicja: M: Check format of references throughout the chapter. We do not want initilas of authors, only their last name.

Alicja: R: fix references to Cohen

Check references. Add reference to "Unraveling the conjunction paradox" by Mark Spottswood in LPR and perhaps others.

1.1 Simple formulation

A simple example will help to fix ideas. Suppose that, in order to prevail in a criminal trial, the prosecution should establish two claims by the required standard: first, that the defendant caused the victim's death; and second, that the defendant's action was premeditated. Cohen (1977) argues that common law systems subscribe to what he calls a **conjunction principle**, according to which if two claims, A and B, are established according to the governing standard of proof, so is their conjunction $A \wedge B$ (and *vice versa*). If the conjunction principle holds, the following must be equivalent, where S is a placeholder for the standard of proof:

Alicja: R: fix ref

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

If we generalize to more than just two constituent claims, the conjunction principle requires that:

$$S[C_1 \wedge C_2 \wedge \cdots \wedge C_k] \Leftrightarrow S[C_1] \wedge S[C_2] \wedge \cdots \wedge S[C_k],$$

where $S[C_i]$ means that claim or hypothesis C_i is established according to standard S. The principle goes in both directions: call the implication from left to right **distribution**, and the one in the opposite direction **aggregation**. Aggregation posits that establishing the individual claims by the requisite standard is enough to establish the conjunction by the same standard. Distribution posits that establishing the conjunction by the requisite standard is enough to establish the individual claims by the same standard. Aggregation and distribution identify properties that the standard of proof should arguably possess. The difficulty with conjunction is traditionally concerned with the failure of aggregation, but we will see later on that on some probabilistic explications of the standard of proof distribution can also fail.

Legal scholars disagree about the tenability of the conjunction principle (more on this toward the end of this chapter). For the time being, however, let us assume the principle correctly captures two desidred features of the standard of proof. The principle has some degree of plausibility and is consistent with 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, 397 U.S. 358, 364, 1970)

A plausible way to interpret this quotation is to assume the following identity: to establish someone's guilt beyond a reasonable doubt *just is* to establish each element of the crime beyond a reasonable doubt (abbreviated as BARD). Thus,

$$\mathsf{BARD}[C_1 \wedge C_2 \wedge \cdots \wedge C_k] \Leftrightarrow \mathsf{BARD}[C_1] \wedge \mathsf{BARD}[C_2] \wedge \cdots \wedge \mathsf{BARD}[C_k],$$

where the conjunction $C_1 \wedge C_2 \cdots \wedge C_k$ comprises all the material facts that, according to the applicable law, constitute the crime with with the accused is charged. A similar argument could be run for the standard of proof in civil cases, preponderance of the evidence or clear and convincing evidence.

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 in a criminal case presents evidence that establishes claims A and B, separately, by the required probability, say about .95 each. Has the prosecution met its burden of proof? If each claim is established by the requisite probability threshold, each claim is established by the requisite standard (assuming the threshold-based interpretation of the standard of proof). And if each claim is established by the requisite standard, then liability as a whole is established by the requisite standard (assuming the conjunction principle). But this cannot be right. Even though each claim is established by the requisite probability threshold, if the claims are independent, the probability of their conjunction will be only $.95 \times .95 \approx .9$, below the required .95 threshold. So liability as a whole is *not* established by the requisite standard (assuming a threshold-based probabilistic interpretation of the standard). This contradicts the conjunction principle.

The difficulty with conjunction shows that a threshold-based interpretation of the standard of proof violates the conjunction principle by violating aggregation. Even though aggregation posits that establishing each conjunct by the required standard of proof is enough to establish the conjunction as a whole, the probability of each conjunct, considered in isolation, can meet the required probability threshold without the conjunction as a whole meeting the threshold.

This difficulty is persistent. It does not subside as the number of constituent claims increases. If anything, the difficulty becomes more apparent. Say the prosecution has established three separate claims by .95 probability. Their conjunction—again if the claims are independent—would be about .85 probable, even further below the .95 probability threshold.

Nor does the difficulty with conjunction generally subside if the claims are no longer regarded as independent. The probability 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, are established with .95 probability, enough for each to meet the threshold, the probability of $A \wedge B$ should still be below the .95 threshold unless $\mathsf{P}(A|B) = 1$. For example, that someone premeditated a harmful act against another (call it premed) makes it more likely that they did cause harm in the end (call it harm). Since $\mathsf{P}(\mathsf{harm}|\mathsf{premed}) > \mathsf{P}(\mathsf{harm})$, the two claims are not independent. Still, premeditation does not always lead to harm, so $\mathsf{P}(\mathsf{harm}|\mathsf{premed})$ will often be below 1. If both claims are established with .95, the probability of the conjunction $\mathsf{harm} \wedge \mathsf{premed}$ should still be below the .95 threshold so long as $\mathsf{P}(\mathsf{harm}|\mathsf{premed})$ is still below 1.

Alicja: R: double-check that abbreviations in formulae are in textsf (outside of math) and mathsf inside math throughout the chapter.

1.2 Adding the evidence

The discussion so far proceeded without mentioning the evidence in support of the claims that constitute the wrongdoing. This is a simplification. As we will see, it is crucial to pay attention to the supporting evidence. With this in mind, for two claims, the conjunction principle can be formulated as follows:

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

where a and b denote the evidence for claims A and B respectively, and S denotes the standard by which the evidence establishes the claim in question. In the case of more than two claims, the formulation of the conjunction principle should be extended accordingly.

Does a threshold-based probabilistic interpretation of the standard of proof also conflict with this revised version of the conjunction principle? The answer is positive, but seeing why requires a bit more work. We should check whether, if both $\mathsf{P}(A|a)$ and $\mathsf{P}(B|b)$ meet the threshold, say .95, then so does $\mathsf{P}(A\wedge B|a\wedge b)$. We are no longer just comparing the probability of $A\wedge B$ to the probability of A and the probability of B as such. Rather, we are comparing the probability of $A\wedge B$ given the combined evidence $a\wedge b$ to the probability of A given evidence a and the probability of B given evidence b.

To fix ideas, consider an example. In an aggravated assault case, the prosecution should establish two claims: first, that the defendant injured the victim; and second, that the defendant knew he was interacting with a public official. Let $\mathsf{witness}$ denote a testimony that the defendant injured the victim, call this claim injury. Let call denote the fact that the defendant made a call to an emergency number. This is evidence that the defendant knew he was dealing with a firefighter, call this claim $\mathsf{firefighther}$. If $\mathsf{P}(\mathsf{injury}|\mathsf{witness})$ and $\mathsf{P}(\mathsf{firefighter}|\mathsf{call})$ both meet the required probability threshold, does $\mathsf{P}(\mathsf{injury}\wedge\mathsf{firefighter}|\mathsf{witness}\wedge\mathsf{call})$ also meet the threshold?

The answer is negative, at least provided two probabilistic independence assumptions hold. The first is that $\mathsf{P}(\mathsf{injury}|\mathsf{witness}) = \mathsf{P}(\mathsf{injury}|\mathsf{witness}\wedge\mathsf{call})$. This assumption is plausible because that the defendant called a firefighter for help, as opposed to someone else, does not make it more (or less) likely that he would cause injury. The second assumption is that $\mathsf{P}(\mathsf{firefighther}|\mathsf{call}) = \mathsf{P}(\mathsf{firefighther}|\mathsf{witness}\wedge\mathsf{call}\wedge\mathsf{injury})$. This assumption is plausible because that the defendant injured the victim and there is a testimony to that effect does not make it more (or less) likely that the victim was a firefighter. To be sure, these informal arguments are not rigorous, a point to which we will soon return. But, assuming for now that the two assumptions hold, it follows that:[1]

$$\mathsf{P}(\mathsf{injury}\wedge\mathsf{firefighther}|\mathsf{witness}\wedge\mathsf{call}) = \mathsf{P}(\mathsf{injury}|\mathsf{witness})\times\mathsf{P}(\mathsf{firefighther}|\mathsf{call}).$$

If the equality holds, even when $\mathsf{P}(\mathsf{injury}|\mathsf{witness})$ and $\mathsf{P}(\mathsf{firefighther}|\mathsf{call})$ meet the required probability threshold, $\mathsf{P}(\mathsf{injury}\wedge\mathsf{firefighther}|\mathsf{witness}\wedge\mathsf{call})$ usually will not.[2] This violates aggregation.

There are some loose ends in this discussion, however. Why should one subscribe to the independence assumptions in the aggravated assault case? Further, do these assumptions hold in other cases? The argument should be made both more rigorous and more general. To this end, we will represent formally the relationship between claims A, B and their conjunction $A\wedge B$, as well as the supporting evidence a, b and their conjunction $a\wedge b$. A natural way to do that is to use Bayesian networks. We have already introduced Bayesian networks in Chapter ZZ. Here we will only sketch the ideas necessary to discuss the difficulty with conjunction. REF TO OTHER CHAPETR

1.3 Independent hypotheses

A Bayesian network is a formal model that consists of a graphical part (a directed acyclic graph, DAG) and a numerical part (a probability table). The nodes in the graph represent random variables that can take different values. For ease of exposition, we will use 'nodes' and 'variables' interchangeably. The nodes are connected by directed edges (arrows). No loops are allowed, hence the name acyclic.

[1] By the chain rule and the independence assumptions $\mathsf{P}(A|a) = \mathsf{P}(A|a\wedge b)$ and $\mathsf{P}(B|b) = \mathsf{P}(B|a\wedge b\wedge A)$, the following holds:

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

[2] The only additional assumption to make here is that both $\mathsf{P}(\mathsf{injury}|\mathsf{witness})$ and $\mathsf{P}(\mathsf{firefighther}|\mathsf{call})$ are below 1, as is usually the case given that the evidence offered in a trial is fallible.

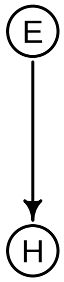

Figure 1: DAG of the simplest evidential relation

The simplest evidential relation, one of evidence bearing on a hypothesis of interest, can be represented by the directed graph displayed in Figure 1. The arrow need not have a causal interpretation. The direction of the arrow indicates which conditional probabilities should be supplied in the probability table. Since the arrow goes from H to E, we should specify the probabilities of the different values of E conditional on the different values of H.

Return now to the difficulty with conjunction. The set up is this. Two items of evidence, a and b, each support their own hypothesis, A and B. This can be represented by two directed graphs visualized in Figure 2.
The two directed graphs should be joined to represent the conjunction $A \wedge B$. How can this be done? Suppose for now that claims A and B are independent. We will relax this assumption later. The conjunction can be represented by adding a conjunction node AB and drawing arrows from nodes A and B into node AB (Figure 3). This arrangement makes it possible to express the meaning of $A \wedge B$ via a probability table (Table 1). This table mirrors the truth table for the conjunction in propositional logic.[3]

Figure 2: DAGs of a supporting A and b supporting B.

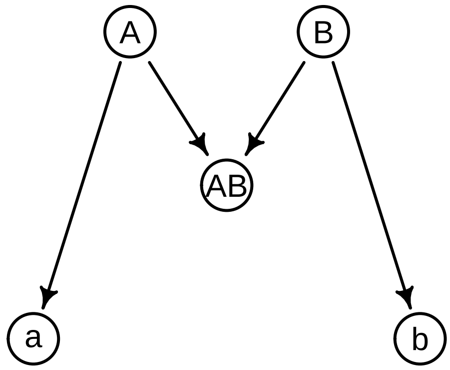

Figure 3: DAG of the conjunction set-up, with the usual independence assumptions built in (DAG 1).

[3]The difference is that the values 1 and 0 stand for two different things depending on where they are in the table. In the columns corresponding to the nodes they represent node states: true and false; in the Pr column they represent the conditional probability of a given state of AB given the states of A and B listed in the same row. For instance, take a look at row two. It says: if A and B are both in states 1, then the probability of AB being in state 0 is 0. In principle we could use 'true' and 'false' instead of 1 and 0 to represent states, but the numeric representation is easier to use in programming, which we do quite a bit in the background, so the reader might as well get used to this harmless ambiguity. For binary nodes, we will consistently use '1' and '0' for the states, it's just probabilities that in this case end up being extreme.

AB	A	B	Pr
1	1	1	1
0	1	1	0
1	0	1	0
0	0	1	1
1	1	0	0
0	1	0	1
1	0	0	0
0	0	0	1

Table 1: Conditional probability table for the conjunction node.

The structure of the directed graph satisfies the desired independence assumptions. First, the two claims A and B are probabilistically independent of one another. Their independence is guaranteed by the fact that the conjunction node AB is a collider and thus no information flows through it.[4] Second, the supporting items of evidence a and b are also probabilistically independent of one another. The reason is the same: node AB blocks any flow of information between the evidence nodes. Notably, the independence of the items of evidence is not always explicitly stated in the formulation of the conjunction paradox. The Bayesian network forces us to make this explicit. This is a good thing.

With this set-up in place, the conjunction paradox arises again, because aggregation is violated. By the theory of Bayesian networks, the directed graph in Figure 1 ensures the following:[5]

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

If, as is normally the case, neither $\mathsf{P}(A|a)$ nor $\mathsf{P}(B|b)$ equal 1, then

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

Thus, even when claims A and B are sufficiently probable given their supporting evidence a and b (for a fixed threshold t)—in symbols, $\mathsf{P}(A|a) > t$ and $\mathsf{P}(B|b) > t$—it does not generally follow that $A \wedge B$ is sufficiently probable given combined evidence $a \wedge b$. Once again, the conjunction principle fails because aggregation fails. The difficulty with conjunction persists.

The argument here goes beyond the specific example about aggravated assault in the previous section. The argument only assumes that the directed graph in Figure 3 is an adequate representation of a situation in which two items of evidence, a and b, support their own hypothesis, A and B. The graph encodes two plausible relations of probabilistic independence: between hypotheses A and B and between items of evidence a and b. The theory of Bayesian networks does the rest of the work.

1.4 Dependent hypotheses

Consider now what happens if the probabilistic independence between claims A and B is dispensed with. To this end, it is enough to draw an arrow between A and B in our directed graph. The result is displayed in Figure 4. Since there is now an open path between nodes A and B, the new graph no longer guarantees the probabilistic independence of A and B, nor the indepedence of evidence nodes a and b. Note however, we still assume there is no *direct* dependence between the items of evidence; this assumption will turn out crucial for the status of aggregation). The items of evidence are still probabilistically independent of one another *conditional* on their respective hypotheses. That is, $\mathsf{P}(a|A) = \mathsf{P}(a|A \wedge b)$ and $\mathsf{P}(b|B) = \mathsf{P}(b|B \wedge a)$. So a and b still counts as independent lines of evidence despite not being (unconditionally) probabilistically independent.[6]

As it turns out, the difficulty with conjunction arises even without the independence of hypotheses A and B, at least in a number of circumstances. Suppose evidence a establishes claim A and b establishes

[4] A more formal treatment of this point is provided in **REFER TO OTHER CHAPTER**.

[5] This is because the only path between A and B goes through AB, which is a collider; as long as we do not condition on it, all paths between A and B remain blocked. See our chapter introducing Bayesian networks for details on this issue. **REFER TO APPROPRIATE CHAPTER**

[6] Here is an illustration of the idea of independent lines of evidence without unconditional independence. Suppose the same phenomenon (say blood pressure) is measured by two instruments. The reading of the two instruments (say 'high' blood pressure) should be *probabilistically dependent* of one another. After all, if the instruments were both infallible and they were measuring the same phenomenon, they should give the exact same reading. On the other hand, the two instruments measuring the same phenomenon should count as *independent lines of evidence*. This fact is rendered in probabilistic terms by means of probabilistic independence conditional on the hypothesis of interest. These ideas can be worked out more systematically in the language of Bayesian networks. Roughly, two variables are probabilistically dependent if there is an open path between them. On the other hand, an open path can be closed by conditioning on one of the variables along the path. For a more rigorous exposition of the notions of open and closed paths, see **CITE EARLIER CHAPTERS**.

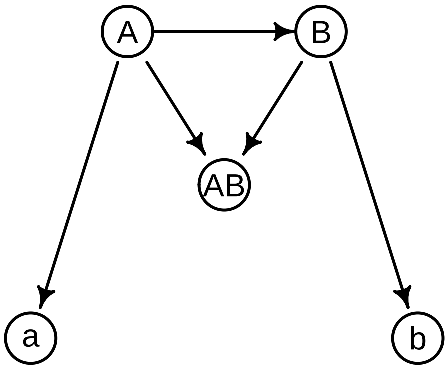

Figure 4: DAG of the conjunction set-up, without independence between A and B (DAG 2).

claim B, separately, right above the probability threshold t. The graph in Figure 4 ensures that:[7]

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

Note that, in a number of cases, $\mathsf{P}(b|B)/\mathsf{P}(b|a)\mathsf{P}(B|A)$ will be less than one, or equivalently:

$$\frac{\mathsf{P}(b|B)}{\mathsf{P}(b|a)} < \frac{1}{\mathsf{P}(B|A)}$$

The expression $\frac{\mathsf{P}(b|B)}{\mathsf{P}(b|a)}$ will typically be greater than one because evidence b should provide positive support for B even in combination with a.[8] The inequality can come out true for a range of probabilities. For instance, if $\mathsf{P}(b|B) = .9$, $\mathsf{P}(b|a) = \mathsf{P}(B|A) = .55$, the left-hand side is ≈ 1.63 and the right-hand side is ≈ 1.81. If $\mathsf{P}(b|B)/\mathsf{P}(b|a)\mathsf{P}(B|A) < 1$, then $\mathsf{P}(A \wedge B|a \wedge b) < \mathsf{P}(A|a)$.[9] There are plenty of such cases. The general pattern is that the higher the right side (for values above one), the lower the probability of $\mathsf{P}(A|A)$ for the inequality to hold. Since, in a number of cases $\mathsf{P}(A \wedge B|a \wedge b)$ is smaller than $\mathsf{P}(A|a)$ and $\mathsf{P}(B|b)$, aggregation does not hold generally even assuming that A and B are probabilistically dependent.

M: Sober and Schwarz claim that, given dependence between A and B, aggregation fails less often, so the paradox is contained. maybe use simulation to check that?

In fact, we can study how often, in principle, the joint posterior is below both of the individual posteriors. To this goal, we simulated 10k random Bayesian networks based on DAG 1 and DAG 2. The joint posterior is lower than both individual posteriors 68% of the time for DAG 1, and around 60% for DAG 2 (see the appendix for details).

2 Evidential Strength

The failures of the conjunction principles so far are failures of aggregation. When the probability of A and the probability of B are both above a given threshold, the probability of the conjunction $A \wedge B$ often is not. This can happen whether or not A and B are probabilistically independent. These failures of aggregation occur if the standard of proof is understood as a posterior probability threshold.

[7] Here is a quick derivation:

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

[8] Because of the relations of independence in Figure 4, $\mathsf{P}(b|B)/\mathsf{P}(b|a) = \mathsf{P}(b|B \wedge a)/\mathsf{P}(b|a)$. We are assuming that $\mathsf{P}(B|a \wedge b) > \mathsf{P}(B|a)$, which is equivalent to $\mathsf{P}(b|B \wedge a)/\mathsf{P}(b|a) > 1$, or in other words, b provides positive supports for B even in light of a. Hence, $\mathsf{P}(b|B)/\mathsf{P}(b|a) > 1$.

[9] By symmetric reasoning, the analogous conclusion that $\mathsf{P}(A \wedge B|a \wedge b) < \mathsf{P}(B|b)$ follows.

Some have regarded these failures as strong enough ground to advocate a different conception of probability, along the lines of so-called Baconian probability or fuzzy logic, or reject legal probabilism altogether. We do not rehearse these lines of argument here. Our focus is on *probabilistic* strategies for capturing the principle of aggregation, and more generally, for handling the difficulty with conjunction. Our working hypothesis is that probability theory need not be revised. Probability theory is sufficiently well-established that we should not reject it unless extremely strong reasons mandate it. Since we have not explored any alternative strategies, within the probabilistic approach, the failures of aggregation are not yet strong reasons to reject probability theory in this context.

R:Add footnote about what these strategies are like

Now, as we have seen, aggregation fails in a large class of cases if the standard of proof is understood as a posterior probability threshold. As an alternative, legal probabilists can think of proof standards as decision criteria formulated in terms of evidential strength. This is the strategy we examine in this section. Instead of a posterior probability threshold, the standard of proof can be modeled using a probabilistic measure of evidential strength. As we shall see, this alternative way of modeling proof standards succeeds in capturing aggregation (at least, given some additional assumptions), but is still not entirely satisfactory.

2.1 A dilemma

We begin by providing an outline of the argument. Two common probabilistic measures of evidential strength are the Bayes factor and the likelihood ratio. We discussed them earlier in Chapter XX. We will show that, under plausible assumptions, these measures of evidential strength validate one direction of the conjunction principle: 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):

REFER TO EARLIER CHAPTERS

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

That is, probability theory justifies this claim: if distinct items of evidence a and b constitute sufficiently strong evidence for claims A and B, so does the conjunction $a \wedge b$ for the composite claim $A \wedge B$ (although, there are some caveats and extra assumptions for this to hold, and *contra* Dawid, it is false that joint likelihood ratio is always higher than each of the individual likelihood ratios; more on this later).

Dawid thought that vindicating aggregation was enough for the conjunction paradox to 'evaporate.' Unfortunately, we will show that on the evidential strength interpretation of the standard of proof, the other direction of the conjunction principle, distribution, does not hold. 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* acted intentionally) while failing to establish one of the conjuncts.

We face a dilemma. If the standard of proof is understood as a posterior probability threshold, the conjunction principle fails because aggregation fails while distribution succeeds. If, on the other hand, the standard of proof is understood as a threshold relative to evidential strength, the conjunction principle fails because distribution fails while aggregation succeeds. From a probabilistic perspective, it seems impossible to capture both directions of the conjunction principle.

In what follows, we develop more precisely the argument that, on the evidential strength approach, (a) aggregation succeeds but (b) distribution fails. The argument for these two claims is tedious. The reader should arm themselves with patience or take our word for it and jump ahead. On the other hand, a more curious reader is welcome to read the appendix for the technical details related to the issues we discuss further on.

2.2 Combined support: Bayes factor

The first step in the argument shows that the combined support supplied by multiple pieces of evidence (e.g. $a \wedge b$) for a conjunctive claim (e.g. $A \wedge B$) typically exceeds the individual support supplied by individual pieces of evidence for individual claims. This claim holds for the Bayes factor and to some extent for the likelihood ratio. We start with the Bayes factor $\mathsf{P}(E|H)/\mathsf{P}(E)$ as our measure of the support of E in favor of H. 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, as compared to its prior probability. 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 (see Chapter XX), 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 with conjunction by vindicating aggregation.

REFER TO EARLIER CHAPTER

2.2.1 Independent hypotheses

Suppose items of evidence a and b positively support A and B, separately. In other words, both Bayes factors ${}^{\mathsf{P}(a|A)}/_{\mathsf{P}(a)}$ (abbreviated BF_A) and ${}^{\mathsf{P}(b|B)}/_{\mathsf{P}(b)}$ (abbreviated BF_B) are greater than one. Does the combined evidence $a \wedge b$ provide at least as much support in favor of the joint claim $A \wedge B$ as the individual support by a and b in favor of A and B considered separately? The combined support here is:

$$\frac{\mathsf{P}(a \wedge b|A \wedge B)}{\mathsf{P}(a \wedge b)} = \frac{\mathsf{P}(a|A)}{\mathsf{P}(a)} \times \frac{\mathsf{P}(b|B)}{\mathsf{P}(b)}$$
$$BF_{AB} = BF_A \times BF_B$$

(See the appendix for a proof.) This claim holds assuming (roughly) that hypotheses A and B are independent and that items of evidence a and b are independent. These assumptions are plausible insofar as the Bayesian network in Figure 3 is a plausible representation of the situation at hand. Thus, the combined support BF_{AB} will always be higher than the individual support so long as BF_A and BF_B are greater than one, that is, if the individual piece of evidence positively support their respective hypotheses.

This result generalizes beyond two pieces of evidence. Figure 5 compares the Bayes factor of one item of evidence, say ${}^{\mathsf{P}(a|A)}/_{\mathsf{P}(a)}$ with the combined Bayes factor for five items of evidence, say ${}^{\mathsf{P}(a_1 \wedge \dots \wedge a_5|A_1 \wedge \dots \wedge A_5)}/_{\mathsf{P}(a_1 \wedge \dots \wedge a_5)}$, for different values of sensitivity and specificity of the evidence.[10] The combined Bayes factor always exceeds the individual Bayes factors provided, as usual, the individual pieces of evidence positively support the individual hypotheses.[11] Under these conditions, Dawid's claim that 'the support supplied by the conjunction of several independent testimonies exceeds that supplied by any of its constituents' (if support is to be measured in terms of Bayes factor) is vindicated.

Alicja: R: recode lty scale so that the order is five-two-one in the legend.

2.2.2 Dependent hypotheses

If A and B are not necessarily probabilistically independent as in the Bayesian network in Figure 4, the combined Bayes factor BF_{AB} is still greater than both the individual Bayes factor BF_A and BF_B if the probabilistic measure fits DAG 2. To see why, first note that the following holds:

$$\frac{\mathsf{P}(a \wedge b|A \wedge B)}{\mathsf{P}(a \wedge b)} = \frac{\mathsf{P}(a|A)}{\mathsf{P}(a)} \times \frac{\mathsf{P}(b|B)}{\mathsf{P}(b|a)}$$
$$BF_{AB} = BF_A \times BF_B' \tag{1}$$

(See the appendix for a proof.) The difference from the case of independent hypotheses is that $BF_B = {}^{\mathsf{P}(b|B)}/_{\mathsf{P}(b)}$ was replaced by $BF_B' = {}^{\mathsf{P}(b|B)}/_{\mathsf{P}(b|a)}$. Since b need not be probabilistically independent of a, there is no guarantee that $\mathsf{P}(b|a) = \mathsf{P}(b)$. However, if the probabilistic measure fits DAG 2, if BF_B is greater than 1, then so is BF_B'. In such circumstances, (1) entails that the joint Bayes factor, BF_{AB}, will be greater than any of the individual Bayes factors> Interestingly, if the underlying DAG allows for a direct dependence between the items of evidence, the claim fails, and the joint Bayes factor can be lower than either of the individual Bayes factors (see the appendix for proof).

R: I though hard about including the graphics here, but at this point I don't think this would contribute to clarity, give it a thought.

[10] The **sensitivity** of a piece of evidence E relative to a hypothesis H is $\mathsf{P}(E|H)$, while its **specificity** is $\mathsf{P}(\neg E|\neg H)$.

[11] The order is reversed if the items of evidence oppose the individual hypotheses. Neutral evidence results in a combined Bayes factor of 1, no matter the prior or the number of items of evidence

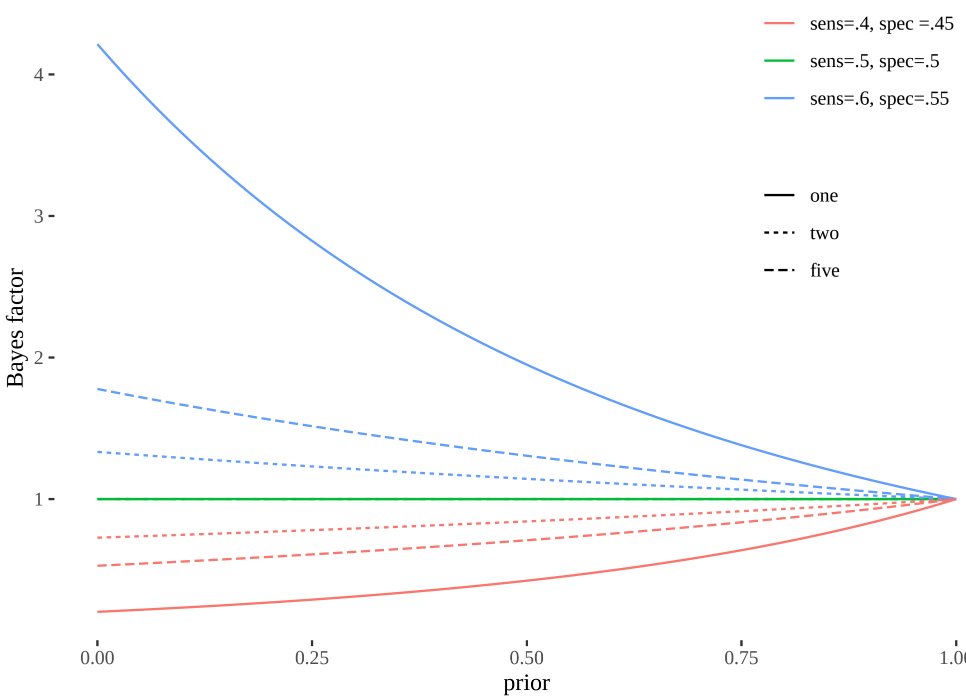

Figure 5: Bayes factor for one, two and five items of evidence and the corresponding claims, given different degrees of specificity and sensitivity of the evidence. The independence assumptions in Figure 3 hold.

2.3 Combined support: likelihood ratio

The likelihood ratio is another probabilistic measure of evidential support, extensively discussed in Chapter XX. The likelihood ratio compares the probability of the evidence on the assumption that a hypothesis of interest is true (*sensitivity*) and the probability of the evidence on the assumption that the negation of the hypothesis is true (*1- specificity*). That is,

REFER TO EARLIER CHAPTER

$$\frac{\mathsf{P}(E|H)}{\mathsf{P}(E|\neg H)} = \frac{\textit{sensitivity}}{\textit{1- specificity}}$$

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). Unlike the Bayes factor, the likelihood ratio does not vary depending on the prior probability of the hypothesis so long as sensitivity and specificity do not.[12]

The question to examine is whether the combined support measured by the combined likelihood ratio

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

exceeds the individual support measured by the individual likelihood ratios ${}^{\mathsf{P}(a|A)}/_{\mathsf{P}(a|\neg A)}$ and ${}^{\mathsf{P}(b|B)}/_{\mathsf{P}(b|\neg B)}$. Under suitable assumptions, the answer is positive. So, details aside, Bayes factor and likelihood ratio agree here. The argument for the likelihood ratio, however, is more laborious.

We start with observing that, in a large class of cases captured by the Bayesian network in Figure 3 or Figure 4, the following holds:

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

The equality is sufficiently general and it holds whether or not hypotheses A and B are probabilistically independent. (See the appendix for a proof.)

[12] Whether the sensitivity and specificity of the evidence depends on the prior probability of the hypothesis is debated in the literature **CITE**. Further, we will see that specificity does depend on the prior probability in the case of conjunctive hypotheses such as $A \wedge B$.

2.3.1 Same sensitivity and specificity, and independence

For illustrative purposes, we first consider a simplified picture. Let the sensitivity and specificity of the items of evidence be the same and equal x. Let also A and B be probabilistically independent in agreement with Figure 3. The combined likelihood ratio can now be plotted as a function of x.[13] Figure 6 shows that the combined likelihood ratio always exceeds the individual likelihood ratios whenever they are greater than one (or in other words, as is usually assumed, the two pieces of evidence provide positive support for their respective hypotheses). Interestingly, the combined likelihood ratio varies depending on the prior probabilities $\mathsf{P}(A)$ and $\mathsf{P}(B)$.

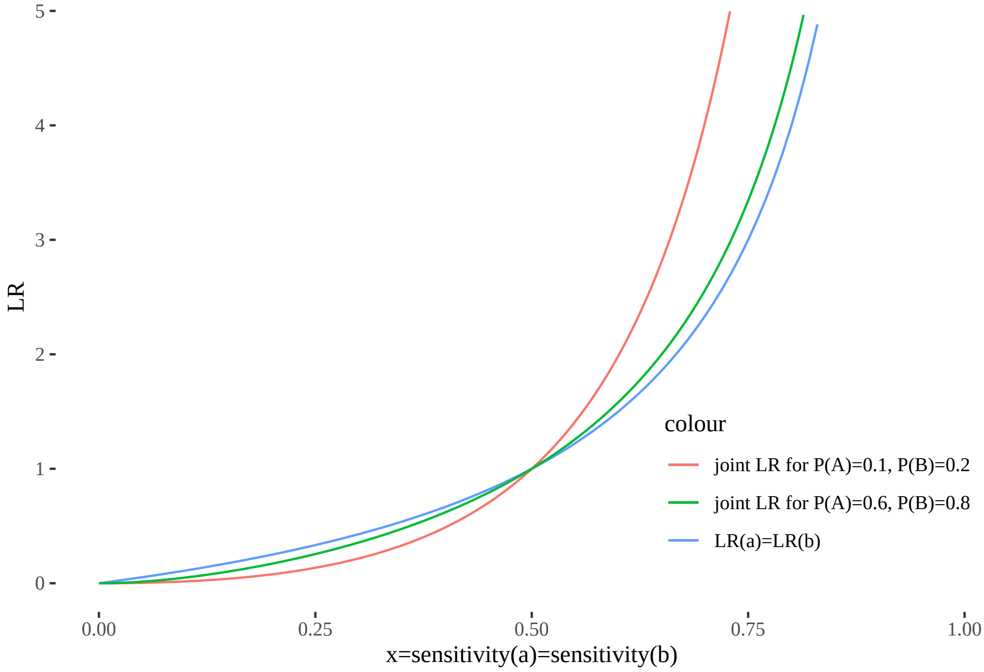

Figure 6: Combined likelihood ratios exceeds individual Likelihood ratios as soon as sensitivity is above .5. Changes in the prior probabilities t and k do not invalidate this result.

As with the Bayes factor, the combined likelihood ratio exceeds the individual likelihood ratios. But the graph only covers cases in which the two pieces of evidence have the same sensitivity and specificity and hypotheses A and B are independent. What happens if these assumptions are relaxed?

2.3.2 Different sensitivity and specificity

If the items of evidence have different levels of sensitivity and specificity, the combined likelihood ratio never goes below the lower of the two individual likelihood ratios, but can be lower than the higher individual likelihood ratio. We established this claim by means of a computer simulation (see the appendix for details). This holds if the probabilistic measure fits DAG 1 and DAG 2, but fails if there is direct dependence between the pieces of evidence. The simplified set-up in the previous situation does not contradict this claim but follows from it. In the simplified set-up, both individual likelihood ratios were the same, so whenever the joint likelihood was higher than the minimum of the individual likelihood ratios, it was higher than the both of them. In this sense, the joint likelihood ratio behaves differently than the joint Bayes factor in that it is greater than the lower of the individual likelihood ratios, rather than being greater than both of them. Here Dawid's claim that 'the support supplied by the conjunction of several independent testimonies exceeds that supplied by any of its constituents' should

[13] In this simplified set-up, the combined likelihood ratio reduces to the following, where $\mathsf{P}(A) = k$ and $\mathsf{P}(B) = t$:

$$\frac{\mathsf{P}(a \wedge b | A \wedge B)}{\mathsf{P}(a \wedge b | \neg(A \wedge B))} = \frac{x^2}{\frac{(1-k)t(1-x)x + k(1-t)x(1-x) + (1-k)(1-t)(1-x)(1-x)}{(1-k)t + (1-t)k + (1-k)(1-t)}}$$

therefore be weakened. That is, the evidential support (as measured by the likelihood ratio) supplied by the conjunction of several independent items of evidence exceeds the support supplied by at least one individual item of evidence but possibly not all, and only if there is no direct dependence between the items of evidence.

2.4 Vindicating aggregation

We have seen that under suitable assumptions the combined evidential support for the conjunction exceeds the individual support for at least one of the individual claims. More precisely:

$$STR[a \wedge b, A \wedge A] \geq min(STR[a,A], STR[b,B]),$$

where $STR[...]$ stands for the strength of the evidential support of a piece of evidence toward a hypothesis, measured by the Bayes factor or the likelihood ratio. This fact can be used to justify aggregation. By the principle above, the following implication holds:

$$STR[a,A] > t_{STR} \wedge STR[b,B] > t_{STR} \Rightarrow STR[a \wedge b, A \wedge B] > t_{STR},$$

where t_{STR} is a threshold on the strength of evidential support. This implication resembles the aggregation principle, repeated below for convenience:

$$S[a,A] \wedge S[b,B] \Rightarrow S[a \wedge b, A \wedge B],$$

where S stands for the standard of proof by which a claim must be established based on the evidence.

The argument for justifying the principle of aggregation, however, is not immediate. Aggregation is a principle about the standard of proof. The Bayes factor and the likelihood ratio are instead measures of evidential strength. The standard of proof must be connected to evidential strength. Naturally enough, the decision criterion should be formulated in terms of evidential strength rather than posterior probabilities. In criminal trials, for example, the rule of decision could be: guilt is proven beyond a reasonable doubt if and only if the evidential support in favor of H—as measured by the Bayes factor $\mathsf{P}(E|H)/\mathsf{P}(E)$ or the likelihood ratio $\mathsf{P}(E|H)/\mathsf{P}(E|\neg H)$—meets a suitably high threshold t_{BF} or t_{LR}. The new threshold is no longer be a probability between 0 and 1, but a number somewhere above one. The greater this number, the more stringent the standard of proof, for any value above one. The question at this point is, how to identify the appropriate evidential strength threshold? The answer to this question is not obvious. Below we examine two possible approaches.

2.4.1 Variable threhsold

The first approach we consider will soon turn out to be inadequate. It is still useful to understand why this approach does not work to formulate one that is more promising. A threshold on evidential strength can be derived from the threshold on posterior probability. The advantage of the posterior probability threshold is that its stringency can be determined in a decision-theoretic manner via the minimization of expected costs (see Chapter XX). The threshold for the Bayes factor and the likelihood ratio, call them REFER TO ER-ALIER CHAPPER
t_{BF} or t_{LR}, can be derived from the threshold t for the posterior probability by manipulating a simple equation.

Consider t_{BF} first. Since

$$\text{Bayes factor} = \frac{\text{posterior}}{\text{prior}},$$

the Bayes factor threshold can be defined as follows:

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

Note that t_{BF} will depend on the prior probability of the hypothesis of interest. The higher its prior probability, the lower t_{BF}. Whether this is a desirable property for a decision threshold can be questioned, but a similar point holds about the posterior threshold t: the higher the prior probability, the easier to meet the threshold. The same strategy works for the threshold t_{LR}. By the odds version of Bayes' theorem,

$$\text{likelihood ratio} = \frac{\text{posterior odds}}{\text{prior odds}}.$$

If the posterior ratio is fixed at, say $t/1-t$, t_{LR} can be obtained as follows:

$$t_{LR} = \frac{t/1-t}{\text{prior odds}}.$$

Once again, the higher the prior, the lower the likelihood ratio threshold.

Unfortunately, this approach incurs a major shortcoming: aggregation still fails. There will be cases in which the conjuncts taken separately satisfy the decision standard t_{BF} or t_{LR}, while the conjunction does not. The culprit is the fact that t_{BF} and t_{LR} have different absolute values when applied to individual claims A and B compared to the composite claim $A \wedge B$. To illustrate this point, consider the principle of aggregation formulated in terms of the Bayes factor threshold:

Question: how does Kaplow think about it? Does he only derive threshold for the ultimate claim? I don't remember now.

$$\frac{\mathsf{P}(a|A)}{\mathsf{P}(a)} > t^A_{BF} \text{ and } \frac{\mathsf{P}(b|B)}{\mathsf{P}(b)} > t^B_{BF} \Rightarrow \frac{\mathsf{P}(a \wedge b|A \wedge B)}{\mathsf{P}(a \wedge b)} > t^{A \wedge B}_{BF}$$

Note the superscripts A, B or $A \wedge B$. The Bayes factor threshold t_{BF} is now indexed by the superscript to the claim of interest because the threshold T_{BF} is prior-dependent and thus claim-dependent (since different claims have different prior probabilities). Now, for simplicity, suppose the individual claims A and B are probabilistically independent. Consider a posterior threshold of .95 as might be appropriate in a criminal case. If A and B both have a prior probability of, say .1, the threshold $t^A_{BF} = t^B_{BF} = .95/.1 = 9.5$ for A or B individually. The composite claim $A \wedge B$ will be associated with the threshold $t^{A \wedge B}_{BF} = .95/(.1 \times .1) = 95$, 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 result from the multiplication of the individual Bayes factors, that is, $9.5 \times 9.5 = 90.25$. This is not quite enough to meet $t^{A \wedge B}_{BF} = 95$. So aggregation fails. An analogous point holds for the likelihood ratio threshold.[14]

Perhaps, this is not the path that the proponent of the evidential strength approach would take anyway. The variable threshold for Bayes factor or the likelihood ratio is parasitic on the posterior probability threshold. It should not be surprising that, if there are reasons to reject the posterior probability threshold, these reasons also apply to other thresholds that are parasitic to it.

2.4.2 Fixed threshold

The second approach we consider consists in fixing the evidential strength threshold regardless of the prior probability of the hypothesis of interest. This approach raises the question of how the threshold should be fixed irrespective of the priors. Standard decision theory can no longer help here. But assuming the question can be answered, it can be shown that the fixed threshold approach vindicates aggregation.

If the standard of proof is formalized using a fixed threshold for the Bayes factor or the likelihood ratio, the conjunction principle boils down to one of these:

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

$$\frac{\mathsf{P}(a|A)}{\mathsf{P}(a|\neg A)} > t_{LR} \text{ and } \frac{\mathsf{P}(b|B)}{\mathsf{P}(b|B)} > t_{LR} \Leftrightarrow \frac{\mathsf{P}(a \wedge b|A \wedge B)}{\mathsf{P}(a \wedge b|\neg(A \wedge B))} > t_{LR}$$

The superscripts have been dropped since the threshold is the same for individual and conjunctive claims. We now show that the left-to-right direction—the principle of aggregation—holds for any thresholds

[14] Say A and B have prior probabilities of .2 and .3 respectively. On this approach, 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 $t^{A \wedge B}_{LR} \approx 297$. Now suppose the individual likelihood ratios meet their threshold and respective sensitivity and specificity are identical. For t^A_{LR} to be met, evidence a should have sensitivity of at least 0.988. For t^B_{LR} to be met, evidence b should have sensitivity 0.978. Now with these separate sensitivies, The combined likelihood ratio equals about 145, far short that what the threshold $t^{A \wedge B}_{LR}$ requires, namely a likelihood ratio of 297. Once again, aggregation fails.

t_{BF} or t_{LR} greater than one. As seen previously, the combined evidential support is usually greater than at least one, if not all, individual evidential supports, whether measured by the Bayes factor or the likelihood ratio. So whenever BF_A and BF_B meet the threshold t_{BF}, then usually also BF_{AB} meets t_{BF}. The same applies for the likelihood ratio threshold. Whenever LR_A and LR_B meet the threshold t_{LR}, then usually also LR_{AB} meets t_{LR}. Success. Aggregation is finally vindicated. Since aggregation could not be justified using posterior probabilities $\mathsf{P}(A|a)$ and $\mathsf{P}(B|b)$ or using a variable evidential strength threshold, this result militates in favor of the fixed evidential strength threshold.

Unfortunately, the right-to-left direction—the seemingly uncontroversial principle of distribution—has now become problematic. Suppose the combined Bayes factor, $\mathsf{P}(a \wedge b|A \wedge B)/\mathsf{P}(a \wedge b)$, barely meets the threshold. The individual support, say $\mathsf{P}(a|A)/\mathsf{P}(a)$, could be still below the threshold unless $\mathsf{P}(b|B)/\mathsf{P}(b) = 1$ (which should not happen if b positively supports B). The problem for likelihood ratio is analogous. For suppose evidence $a \wedge b$ supports $A \wedge B$ to the required threshold t. 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.

To be specific, we have just shown that the following distribution principles fails:

$$S[a \wedge b, A \wedge B] \Rightarrow S[a, A] \wedge S[b, B], \tag{DIS1}$$

where S is a placeholder for the standard of proof. Perhaps, the problem is with principle (DIS1). Should it be rejected? It might not be 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 by 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). Or at least one might argue this way.

To accommodate this line of reasoning, consider this other distribution principle:

$$S[a \wedge b, A \wedge B] \Rightarrow S[a \wedge b, A] \wedge S[a \wedge b, B]. \tag{DIS2}$$

Principle (DIS2) is less controversial because it holds the evidence constant. This principle is harder to deny: one would not want to claim that, while holding fixed evidence $a \wedge b$, establishing the conjunction might not be enough for establishing one of the conjuncts. It seems that any formalization of the standard of proof should obey (DIS2). Yet, (DIS2) also fails both for Bayes factor and likelihood ratio threshold. After all, if A and B are probabilistically independent, (DIS1) and (DIS2) are in fact equivalent principles so long as the standard of proof is interpreted as a fixed evidential strength threshold. So the counterexamples to (DIS1) also work against (DIS2).

M: Need proof of this. Appendix? What about if they are not independent. Need proof. Maybe simulation?

R: should we get into this?

Curiously, on the fixed evidential strength threshold approach, no matter whether one uses Bayes factor or likelihood ratio, there would be cases in which, even though the conjunction $A \wedge B$ is established by the desired standard of proof, one of the individual claims fails to meet the standard (see the appendix). This is odd.

3 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 Pardo & Allen (2008). Say the prosecutor or the plaintiff puts forward 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, if one thinks of the defense hypothesis H_d as a substantial account of what has happened. The defense may elect to challenge the hypothesis put forward 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.

3.1 Comparing posteriors

Setting aside this controversy for the time being, we first work out the comparative strategy using posterior probabilities. More specifically, the standard of proof is understood comparatively as follows: 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 ${}^{\mathsf{P}(H_p|E)}/{}_{\mathsf{P}(H_d|E)}$ should be above a suitable threshold. Presumably, the ratio threshold should be higher for criminal than civil cases. In fact, in civil cases it seems enough to require that the ratio ${}^{\mathsf{P}(H_p|E)}/{}_{\mathsf{P}(H_d|E)}$ be above 1, or equivalently, that $\mathsf{P}(H_p|E)$ 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. But, if two hypotheses are exclusive and exhaustive, ${}^{\mathsf{P}(H_p|E)}/{}_{\mathsf{P}(H_d|E)} > 1$ implies that $\mathsf{P}(H_p|E) > .5$, the standard probabilistic interpretation of the preponderance standard.

One advantage of the comparative aapproach—as Cheng (2012) 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. The costs of a false positive is c_{FP} and that of a false negative is C_{FN}, both greater than zero. REFERENCE TO EARLIER CHAPTER FOR MORE COMPLEX COST STRUCTURE 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 positives to false negatives equals one. So the rule of decision would be: find against the defendant 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 appropriately greater than $\mathsf{P}(H_d|E)$.

Does the comparative strategy just outlined solve the difficulty with conjunction? 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 ${}^{\mathsf{P}(S \wedge C|E)}/{}_{\mathsf{P}(H_d|E)} > 1$. To answer it, we have to decide what the defense hypothesis H_d should be. 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

$$\begin{aligned}
\frac{\mathsf{P}(S \wedge C|E)}{\mathsf{P}(S \wedge \neg C|E)} &= \frac{\mathsf{P}(S|E)\mathsf{P}(C|E)}{\mathsf{P}(S|E)\mathsf{P}(\neg C|E)} = \frac{\mathsf{P}(C|E)}{\mathsf{P}(\neg C|E)} > 1 \\
\frac{\mathsf{P}(S \wedge C|E)}{\mathsf{P}(\neg S \wedge C|E)} &= \frac{\mathsf{P}(S|E)\mathsf{P}(C|E)}{\mathsf{P}(\neg S|E)\mathsf{P}(C|E)} = \frac{\mathsf{P}(S|E)}{\mathsf{P}(\neg S|E)} > 1 \\
\frac{\mathsf{P}(S \wedge C|E)}{\mathsf{P}(\neg S \wedge \neg C|E)} &= \frac{\mathsf{P}(S|E)\mathsf{P}(C|E)}{\mathsf{P}(\neg S|E)\mathsf{P}(\neg C|E)} > 1
\end{aligned} \tag{2}$$

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 vindicated, at least for simple cases involving independence. Success.

What about the opposite direction, distribution? Distribution is not generally satisfied. Suppose ${\mathsf{P}(S\wedge C|E)}/{\mathsf{P}(H_d|E)} > 1$, or in other words, the combined hypothesis $S \wedge C$ has been established by preponderance of the evidence. The question is whether the individual hypotheses have been established by the same standard, specifically, whether $\frac{\mathsf{P}(C|E)}{\mathsf{P}(\neg C|E)} > 1$ and $\frac{\mathsf{P}(S|E)}{\mathsf{P}(\neg S|E)} > 1$. If ${\mathsf{P}(S\wedge C|E)}/{\mathsf{P}(H_d|E)} > 1$, the combined hypothesis is assumed to be more probable than any of the competing hypotheses, in particular, ${\mathsf{P}(S\wedge C|E)}/{\mathsf{P}(\neg S\wedge C|E)} > 1$, ${\mathsf{P}(S\wedge C|E)}/{\mathsf{P}(S\wedge \neg C|E)} > 1$ and ${\mathsf{P}(S\wedge C|E)}/{\mathsf{P}(\neg S\wedge \neg C|E)} > 1$. We have:

$$
\begin{aligned}
1 &< \frac{\mathsf{P}(S\wedge C|E)}{\mathsf{P}(S\wedge \neg C|E)} = \frac{\mathsf{P}(S|E)\mathsf{P}(C|E)}{\mathsf{P}(S|E)\mathsf{P}(\neg C|E)} = \frac{\mathsf{P}(C|E)}{\mathsf{P}(\neg C|E)} \\
1 &< \frac{\mathsf{P}(S\wedge C|E)}{\mathsf{P}(\neg S\wedge C|E)} = \frac{\mathsf{P}(S|E)\mathsf{P}(C|E)}{\mathsf{P}(\neg S|E)\mathsf{P}(C|E)} = \frac{\mathsf{P}(S|E)}{\mathsf{P}(\neg S|E)} \\
1 &< \frac{\mathsf{P}(S\wedge C|E)}{\mathsf{P}(\neg S\wedge \neg C|E)} = \frac{\mathsf{P}(S|E)\mathsf{P}(C|E)}{\mathsf{P}(\neg S|E)\mathsf{P}(\neg C|E)}
\end{aligned}
\tag{3}
$$

In the first two cases, clearly, if the composite hypothesis meets the threshold, so do the individual claims. But consider the third case. ${\mathsf{P}(S|E)\mathsf{P}(C|E)}/{\mathsf{P}(\neg S|E)\mathsf{P}(\neg C|E)}$ might be strictly greater than ${\mathsf{P}(C|E)}/{\mathsf{P}(\neg C|E)}$ or ${\mathsf{P}(S|E)}/{\mathsf{P}(\neg S|E)}$. It is possible that ${\mathsf{P}(S|E)\mathsf{P}(C|E)}/{\mathsf{P}(\neg S|E)\mathsf{P}(\neg C|E)}$ is greater than one, while either${\mathsf{P}(C|E)}/{\mathsf{P}(\neg C|E)}$ or ${\mathsf{P}(S|E)}/{\mathsf{P}(\neg S|E)}$ are not, say when they are 3 and 0.5, respectively. Distribution fails. And the same problem would arise with a more stringent threshold as might be appropriate in criminal cases.

There is a more general problem with Cheng's comparative approach. Much of the heavy lifting here is done by the strategic splitting of the defense line into multiple scenarios. Suppose, for illustrative purposes, $\mathsf{P}(H_p|E) = 0.37$ 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 on this approach 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, the more alternative scenarios into which $\neg H_p$ needs to be divided. This normally would lead to lowering even further the probability of each of them (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 a disadvantage the defense is. This seems undesirable.

3.2 Comparing likelihoods

Instead of posterior probabilities, likelihoods can also be compared. The standard of proof would then be as follows: the ratio between likelihoods ${\mathsf{P}(E|H_p)}/{\mathsf{P}(E|H_d)}$ should be above a suitable threshold. Note that the posterior ratio ${\mathsf{P}(H_p|E)}/{\mathsf{P}(H_d|E)}$ from before was replaced by the likelihood ratio ${\mathsf{P}(E|H_p)}/{\mathsf{P}(E|H_d)}$ where H_d and H_p, as before, need not be exhaustive hypotheses. In civil cases, the likelihood ratio should perhaps be just be above 1, meaning that the evidence supports H_p more strongly than it supports H_d. In criminal cases, the ratio should be several orders of magnitude above one. This approach runs into the same problem as Cheng's. It cannot justify distribution.

We do not provide all the details of the argument. The reasoning is analogous. Consider the car crash example from before, where S stands for the defendant's speeding, C stands for the statement that the crash caused neck injury, and E stands for the total evidence. The plaintiff's hypothesis H_p is $S \wedge C$. Suppose ${\mathsf{P}(E|S\wedge C)}/{\mathsf{P}(E|H_d)} > 1$, or in other words, the combined hypothesis $S \wedge C$ has been established by preponderance of the evidence. The question is whether the individual hypotheses have been established by the same standard, specifically, whether $\frac{\mathsf{P}(E|C)}{\mathsf{P}(E|\neg C)} > 1$ and $\frac{\mathsf{P}(E|S)}{\mathsf{P}(E|\neg S)} > 1$. Focusing on a specific defense hypothesis, $\neg S \wedge \neg C$, the following holds:[15]

$$
1 < \frac{\mathsf{P}(E|S\wedge C)}{\mathsf{P}(E|\neg S\wedge \neg C)} = \frac{\mathsf{P}(E|S)\mathsf{P}(E|C)}{\mathsf{P}(E|\neg S)\mathsf{P}(E|\neg C)} \tag{4}
$$

Note that ${\mathsf{P}(E|S)\mathsf{P}(E|C)}/{\mathsf{P}(E|\neg S)\mathsf{P}(E|\neg C)}$ might be strictly greater than ${\mathsf{P}(E|C)}/{\mathsf{P}(E|\neg C)}$ or ${\mathsf{P}(E|S)}/{\mathsf{P}(E|\neg S)}$. It is possible that ${\mathsf{P}(E|S\wedge C)}/{\mathsf{P}(E|H_d)}$ is greater than one, while either $\frac{\mathsf{P}(E|C)}{\mathsf{P}(E|\neg C)}$ and $\frac{\mathsf{P}(E|S)}{\mathsf{P}(E|\neg S)}$ are not, say when they are 3 and 0.5, respectively. Once again, distribution fails.

R: take a look at this additional argument, is this good enough?

A general difficulty remains, related to how the comparative likelihood strategy is sensitive to the choice of the hypotheses. In many plausible situations it might be the case there might be hypotheses that

[15] NEED PROOF

one wishers to compare H_1, H_2 such that $\mathsf{P}(E|H_1)$ is much (say, at least a few times) larger than $\mathsf{P}(E|H_2)$ while $\mathsf{P}(E|H_1)$ is still smaller than $\mathsf{P}(E|\neg H_1)$ in such circumstances, the comparative likelihood strategy seems to be recommending the acceptance of H_1, even though, intuitively, the evidence seems to support $\neg H_1$ to a larger extent.

4 Rejecting the conjunction principle?

A number of strategies that legal probabilists can pursue to address the difficulty with conjunction have proven problematic. Perhaps, a different perspective should be taken here. Observe that the problem would not arise without the conjunction principle. Should legal probabilists simply reject this principle? So far we have not challenged it, but it is time to scrutinize it more closely. In this section, we provide an epistemic argument and a legal argument to question the conjunction principle. We also caution that merely rejecting the conjunction principle will not dissolve the difficulty with conjunction. More work needs to be done. We take it on in the final section.

4.1 The legal argument

Before moving further, it is worth asking what the law says about the conjunction principle. The answer, perhaps unsurprisingly is, not very much. We have been assuming that the law agrees with the conjunction principle. At least, this is what Cohen thought. Matters, however, are not so clear-cut. Looking at legal practice, the conjunction principle is an uncertain principle at best.

The best place to look is how jury instructions are formulated. Do they obey the conjunction principle? To some extent, they do. For example, here are jury instructions about negligence claims in civil cases:

> A negligence claim has three elements:
>
> 1. [Defendant] did not use ordinary care;
> 2. [Defendant's] failure to use ordinary care caused [Plaintiff's] harm; and
> 3. [Plaintiff] is entitled to damages as compensation for that harm.
>
> [Plaintiff] must prove each element by a preponderance of the evidence—that each element is more likely so than not so. If [Plaintiff] proves each element, your verdict must be for [Plaintiff]. If [Plaintiff] does not prove each element, your verdict must be for [Defendant].[16]

The elements are explicitly separated and the standard of proof is applied to each element separately. This seems to confirm the conjunction principle. Other jury instructions are more ambiguous:

> In order to find that the plaintiff is entitled to recover, you must decide it is more likely true than not true that:
>
> 1. the defendant was negligent;
> 2. the plaintiff was harmed; and
> 3. the defendant's negligence was a substantial factor in causing the plaintiff's harm.[17]

The elements are still separated, but the standard of proof ('more likely than not') applies to the conjunction as a whole, not the individual claims. At least, these jury instruction are at best ambiguous between an atomistic reading (the standard applies to each claim separately) and a holistic reading of the standard of proof (the standard applies to the conjunction). Only the atomistic reading would justify the conjunction principle.

This quick survey of jury instructions gives some reassurance that, should we decide to reject the conjunction principle, we would not violate a well-entrenched, indispensable legal principle.

CITE SOBER AND SCHAWRTZ WHI DIUD THE EMPIRICAL ANALYSIS OF JURY INSTRUCTIONS

4.2 Risk accumulation

Beside legal uncertainty about the conjunction principle, there are also independent theoretical reasons to question the principle. In current discussions in epistemology about knowledge or justification, a principle similar to the conjunction principle has been contested because it appears to deny the fact that risks of error accumulate (Kowalewska, 2021). If one is reasonably sure about the truth of each claim

[16] Standardized Civil Jury Instructions for the District of Columbia (Civil Jury Instructions, revised edition 2017), Sec. 5.01.
[17] Alaska Civil Pattern Jury Instructions, Sec. 3.01 (Civil Pattern Jury Instructions 2017).)

considered separately, one should not be equally reasonably sure of their conjunction. You have checked each page of a book and found no error. So, for each page, you are reasonably sure there is no error. Having checked each page and found no error, can you be equally reasonably 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 reasonably sure there is no error (Makinson, 1965). A reasonable doubt about the existence of an error, in one page or another, creeps up as one considers more and more pages. The same observation applies to other contexts, say product quality control. You may be reasonably sure, for each product you checked, that it is free from defects. But you cannot, on this basis alone, be equally reasonably 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.

Risk accumulation challenges aggregation: even if the probability of several claims, considered individually, is above a threshold t, their conjunction need not be above t. It does not, however, challenge distribution. If, all risks considered, you have good reasons to accept a conjunction, no further risk is involved in accepting any of the conjuncts separately. This is also mirrored by what happens with probabilities. If the probability of the conjunction of several claims is above t, so is the probability of each individual claim.

R: What's our take on risk accummulation?

The standard of proof in criminal or civil cases can be understood as a criterion concerning the degree of risk that judicial decisions should not exceed. If this understanding of the standard of proof is correct, the phenomenon of risk accumulation would invalidate the conjunction principle, specifically, it would invalidate aggregation. It would longer be correct to assume that, if each element is proven according to the applicable standard, the case as a whole is proven according to the same standard. And, in turn, if the conjunction principle no longer holds, the conjunction paradox will disappear. Or will it?

4.3 Atomistic and holistic approches

Matters are not so straightfoward, however. Suppose legal probabilists do 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.

Let's stipulate that, in order to establish the defendant's guilt beyond a reasonable doubt (or civil liability by preponderance of the evidence or clear and convincing evidence), the party making the accusation should establish each claim, separately, to the requisite probability, say at least .95 (or .5 in a civil case), 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 establishing 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 is relatively low, just because guilt 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, a composite claim consisting of five subclaims—assuming, as usual, probabilistic independence between the subclaims—would only be established with probability equal to .77, a far cry from proof beyond a reasonable doubt. This is counterintuitve, as it would allow convictions when the defendant is not very likely to have committed the crime. A similar argument can be run for the civil standard of proof 'preponderance of the evidence.' 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 atomistic approach is a non-starter.

Another option is to require that the prosecution in a criminal case (or the plaintiff in a civil case) establish the accusation as a whole—say the conjunction of A and B—to the requisite probability. Call this the *holistic account*. This account is not without problems either.

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 $\mathsf{P}(A)$ is .96, then $\mathsf{P}(B)$ must be at least .99 so that $\mathsf{P}(A \wedge B)$ is above a .95 threshold. But if $\mathsf{P}(A)$ is .9999, then $\mathsf{P}(B)$ must only be slightly greater than .95 to reach the same threshold. Thus, the holistic account might require that the elements of an accusation be proven to different probabilities—and thus different standards—depending on how well other claims have been established. This result runs counter to the tacit assumption that each element should be established to the same standard of proof.

M: Cite Ubaniak's paper on this point.

Fortunately, this challenge can be addressed. It is true that different elements will be established with different probabilities, depending on the probabilities of the other elements. But this follows from the fact that the prosecution or the plaintiff may choose different strategies to argue their case. They may decide that, since they have strong evidence for one element and weaker evidence for the other, one element should be established with a higher probability than the other. What matters is that the case as a whole meets the required threshold, and this objective can be achieved via different means. What will never happen is that, while the case as a whole meets the threshold, one of the constituent elements does not. As seen earlier, the probability of the conjunction never exceeds the probability of one of the conjunct, or in other words, distribution is never violated.

A more difficult challenge is the observation that 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 .95. So the more constituent claims, the higher the posterior probability for each claim needed for the conjunction to meet the requisite probability threshold.

This difficulty is best appreciated by running some numbers. 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 probability of at least $t^{1/k}$, where t is the threshold to be applied to the composite claim.[18] For example, if there are ten constituent claims, they will have to be proven with $.5^{1/10} = .93$ even if the probability threshold is only .5. If the threshold is more stringent, as is appropriate in criminal cases, say .95, each individual claim will have to be proven with near certainty. This would make the task extremely demanding on the prosecution, if not downright impossible. If there are ten constituent claims, they will have to be proven with $.95^{1/10} = .995$. So the plaintiff or the prosecution would face the demanding task of establishing each element of the accusation beyond what the standard of proof would seem to require.

We reached an impasse. Under the atomistic approach, the standard is too lax because it allows for findings of liability when the defendant quite likely committed no wrong. Under the holistic approach, the standard is too demanding on the prosecution (or the plaintiff) because it requires the individual claims to be established with extremely high probabilities. Bummer.

4.4 Not asking too much

Consider again the holistic approach. It is true that the individual elements (the individual conjuncts) should be established with a higher probability than the case as a whole (the conjunction). This would seem to impose an unreasonably stringent burden of proof on the prosecution or the plaintiff. But the burden might not be as unreasonable as it appears at first. As Dawid (1987) pointed out, in one of the earliest attempts to solve the conjunction paradox from a probabilistic perspective, the prior probabilities of the conjuncts will also be higher than the prior probability of their conjunction:

> . . . 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 [p. 97].

Dawid's proposal is compelling. Still, why exactly is it 'not asking too much' to establish the individual conjuncts by a higher threshold than the case as a whole? The prior probabilities of the conjuncts are surely higher than the prior probability of the conjunction. But what is the notion of 'not asking too much' at work here? Dawid might be recommending—as the rest of his paper suggests—that the standard of proof no longer be understood in terms of just posterior probabilities. Measures of how strongly each claim is being supported by the evidence, such as the Bayes' factor of the likelihood ratio, account for the difference between prior and posterior probabilities. So, presumably, Dawid is recommending these measures as better suited to formalize the standard of proof.

Now, as the reader will have realized, we have pursued Dawid's strategy already. This strategy can justify, on purely probabilistic grounds, one direction of the conjunction principle: aggregation. The evidential support—measured by the Bayes' factor or the likelihood ratio—for the conjunction often exceeds the individual support for (at least one of) the individual claims. This is a success, especially because the failure of aggregation motivated Cohen's formulation of the conjunction paradox.

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

Unfortunately, we have seen that this strategy invalidates a previously unchallenged direction of the conjunction principle: distribution.

5 The proposal

Here is where we have gotten so far. There might be good reasons to reject the conjunction principle, but rejecting it does not automatically solve the difficulty with conjunction. We still need a theory that explains how individual claims are combined, together with the available evidence, to form more complex claims, say the claim that the defendant committed the crime for which they were charged. The conjunction principle provides a recipe—a very simple one at that—to combine individual claims and form conjunctive claims. That recipe might not be right. If it is not, a good theory of the standard of proof should still provide an alternative recipe for combing individual claims.

Our proposal is inspired by the story model of adjudication (Pennington & Hastie, 1993; Wagenaar, Van Koppen, & Crombag, 1993) and the relative plausibility theory (Allen & Pardo, 2019; Pardo & Allen, 2008). It posits that prosecutors and plaintiffs should aim to establish a unified narrative of what happened or explanation of the evidence, not establish each individual element of wrongdoing separately. As we shall see, any attempt to proceed in a piecemeal manner implicitly requires, sooner or later, to weave the different elements together into a unified whole.

Our argument consists of two parts. First, the guilt or civil liability of a defendant on trial cannot be equated with a generic claim of guilt or civil liability as defined in the law. The claim against a defendant facing trial should always be grounded in specific details. Call this the specificity argument. Second, it is erroneous to think of someone's guilt or civil liability as the mere conjunction of separate claims. The separate claims must be unified, not just added up in a conjunction. Call this the unity argument.

5.1 The specificity argument

We start with the specificity argument. The probabilistic interpretation of proof standards usually posits a threshold that applies to the posterior probability of a *generic* hypothesis, such as the defendant is guilty of a crime, call it G, or civilly liable, call it L. In criminal cases, the requirement is formulated as follows: the evidence E presented at trial establishes guilt beyond a reasonable doubt provided $\mathsf{P}(G|E)$ is above a suitable threshold, say .95. The threshold is lower in civil trials. Civil liability is proven by preponderance provided $\mathsf{P}(L|E)$ is above a suitable threshold, say .5.

This formulation conflates two things. The wrongdoing as defined in the applicable law is one thing. The way in which the wrongdoing is established in court is another thing. The wrongdoing is defined in a generic manner and is applicable across a class of situations, whereas the way the wrongdoing is established in court is specific to a situation and tailored to the individual defendant. A prosecutor in a criminal case does not just establish that the defendant assaulted the victim in one way or another, but rather, that the defendant behaved in such and such a manner in this time and place, and that the behavior in question fulfills the legal definition of assault. The requirement of specificity is, for one thing, a consequence of the fact that defendants have a right to be informed with sufficient degree of detail, and also that they should be in a position to prepare a defense.[19]

If this is correct, the probabilistic interpretation of proof standards should be revised. The generic claim that the defendant is guilty or civilly liable should be replaced by a more fine-grained hypothesis, call it H_p, the hypothesis put forward by the prosecutor (or the plaintiff in a civil case), for example, that the defendant, given reasonably well-specified circumstances, approached the victim, pushed and kicked the victim to the ground, and then run away. Hypothesis H_p is a more precise description of what happened that entails the defendant committed the wrong. In defining proof standards, instead of saying—generically—that $\mathsf{P}(G|E)$ or $\mathsf{P}(L|E)$ should be above a suitable threshold, a probabilistic interpretation should read: civil or criminal liability is proven by the applicable standard provided $\Pr(H_p|E)$ is above a suitable threshold, where $H_p$is a reasonably specific description of what happened according to the prosecutor or the plaintiff.

This revision of the probabilistic interpretation of standards of proof may appear inconsequential, but it is not. It is the revision we invoked to address the puzzles of naked statistical evidence in Chapter XX. Recall the gist of the argument. Consider the prisoner hypothetical, a standard example of naked

REFERENCE TO EARLIER CHAPTER

[19] ADD CASE LAW REFERENCES TO BUTRESS THIS POINT.

statistical evidence. The naked statistics E_s make the prisoner on trial .99 likely to be guilty, that is, $\mathsf{P}(G|E_s) = .99$. It is .99 likely that the prisoner on trial is one of those who attacked and killed the guard. This is a generic claim. It merely asserts that the prisoner was – with very high probability – one of those who killed the guard, without specifying what he did, how he partook in the killing, what role he played in the attack, etc. If the prosecution offered 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 $\mathsf{P}(G|E_s) = .99$. That the prisoner on trial is most likely guilty is an artifact of the choice of a generic hypothesis G. When this hypothesis is made more specific—as should be—this probability drops significantly. And the puzzle of naked statistical evidence disappears.

5.2 The unity argument

The specificity argument addresses the problem of naked statistical evidence, but also provides the necessary background for addressing the difficultly about conjunction. In the traditional formulation, not only are G and L understood as generic claims. They are also understood as *mere conjunctions* of simpler claims that correspond to the elements of wrongdoing in the applicable law. Since the probability of a conjunction is often lower than the probability of its conjuncts, the individual claims can be established with a suitably high probability that meets the required threshold even though the conjunction as a whole fails to meet the same threshold. This mismatch gives rise to the difficulty with conjunction.

But someone's guilt (and the same applies to civil liability) cannot be the mere conjunction of the claims corresponding to the elements of wrongdoing as defined in the law. Someone's guilt is a state of affairs that is described by a well-specified series of events that possess a coherent, structured unity. These events, taken as a whole, can be subsumed under the legal definition that consists of several discrete elements. The conjunction paradox assumes that criminal or civil wrongdoings are the mere collection of separate elements. The law is more complicated. Legal definitions often impose a structure on how the different elements relate to one another. They are not merely separate items that should be proven one by one. They often form a structured unity.[20]

But what if the law does not impose any structure among the different elements of wrongdoing? Consider this case in which only two elements must be proven. Element 1: the defendant's conduct caused a bodily injury to victim. Element 2: the defendant's conduct consisted in reckless driving. Call this criminal offense "vehicular assault."[21] The two elements are not independent, but they each add novel information. It could be that the defendant's driving caused an injury to victim, but the driving was not reckless, or the driving was reckless, but no injury ensued. Neither element is presupposed by the other. The law does not impose any specific structure between the elements. But the same unity argument would still apply at a conceptual level.

Consider how one could go about establishing the claim of reckless driving that caused injury to the victim. One option is to offer a detailed reconstruction of what happened. The reconstruction could

[20]Consider, for example, the allegation of negligent misrepresentation to be established by clear and convincing evidence. As an illustration, we follow the jury instructions of the State of Washington. (EXACT REFERENCE) These are the elements a plaintiff should establish: (E1) defendant supplied false information to guide plaintiff in their business transaction; (E2) defendant knew the false information was supplied to guide plaintiff; (E3) defendant was negligent in obtaining or communicating the false information; (E4) plaintiff relied on the false information; (E5) plaintiff's reliance on the false information was reasonable; and (E6) the false information proximately caused damages to the plaintiff. For the plaintiff to prevail in this type of case, they should prove each element by the required standard. What does that actually require? The plaintiff should first establish, at a minimum, that the defendant supplied false information for the guidance of the plaintiff in the course of a business transaction. Say the defendant, a owner of a vacation resort, told plaintiff, a travel agent, that the resort included amenities that were not actually there, and the plaintiff decided to book several clients at the defendant's resort instead of other resorts. The plaintiff could offer copies of emails communication or screenshots from the resort's website. This would take care of the first element. The second element qualifies the first, in that the defendant *knew* they were supplying false information to guide the business transaction. Clearly, establishing the second element presupposes having already established the first since the second element is a qualification of the first. The truth of the second element entails the truth of the first. If the defendant knowingly supplied false information (second element), they did supply false information (first element). The third element requires to show that the defendant was negligent in obtaining or communicating the false information. This is another qualification that applies to the first element and cannot be proven without having already proven the first element. The fourth and fifth element should be understood together. The plaintiff relied on the false information (fourth element) and such reliance was reasonable (fifth element). In turn, establishing the fourth and fifth element presupposes that the plaintiff did supply false information to begin with (first element). Finally, the sixth element concerns causation of harm. This is again a qualification of the first element and cannot be established without having already established the first element.

[21]CITE RELATED LAWS

go something like this. The defendant was driving above the speed limit, veering left and right. The defendant's reached a school crosswalk when children were getting out of school. The defendant hit a child on the crosswalk who was then pushed against a light pole on the sidewalk incurring a head injury. This story is supported by plenty of evidence: other children, people standing around, police officers, paramedics. There is plenty of supporting evidence as the incident occurred in the middle of the day. Taken at face value, this story does establish both elements: reckless driving and cause of injury. Parts of the story are relevant for element 1 (reckless driving) and others are relevant for element 1 (cause of injury). The two cannot be neatly separated, however. Still, what is crucial is that the different parts of the story are part of the same episode, the same unit of wrongdoing.

Could the prosecutor prove vehicular assault in a piecemeal manner? Suppose the prosecutor attempted to do that, by establishing, first, that the defendant drove recklessly, and second—*separately from the first element*—that the defendant's action caused injury. As noted before in the specificity argument, it is not enough to establish that the defendant drove recklessly at some point in time somewhere. Nor is it enough to establish that the defendant's action caused injury. The prosecutor should offer a specific story detailing what happened, a story relevant for the first element and a story relevant for the second element. Say this expectation of specificity is met. Suppose the prosecutor did not simply establish generic element 1 and generic element 2 of the charge, but rather, a reasonably detailed story which, if true, would establish element 1 and a story which, if true, would establish element 2. Wouldn't that be enough? It wouldn't. Even if each element—more precisely, each story associated with each element—was established by the required standard, there would be something missing here.

The prosecutor should establish that the two elements—reckless driving and injury, or the two stories associated with the two elements—are part of the same unity of wrongdoing. It must me *this* reckless driving that caused *this* injury. So, under the piecemeal approach, the prosecutor would be tasked with establishing three claims: (1) the defendant, in some well-specified circumstances, was driving reckless; (2) the defendant, in some well-specified circumstances, caused injury to the victim; and (3) the well-specified circumstances in (1) and (2) are part of the same episode. But once (3) is established, the prosecutor would have effectively established the charge by the required standard in accordance with the holistic approach. The prosecutor did not only establish each separate element (two separate stories) but also combine the two elements (the two stories) together. Once the piecemeal approach is pursued to its logical conclusion, it coincides with the holistic approach.[22]

Let's summarize the unity argument in schematic form. If the prosecutor or the plaintiff is expected to establish claim A and B by the required standard, what the law actually requires—even in terms of the piecemeal approach—is (1) to establish A; (2) to establish B; and (3) to establish A and B are part of the same unit of wrongdoing by the required standard. Item (3) is often implicit, which leaves the impression that the law only requires to establish (1) and (2) separately. Interestingly, (3) entails (1) and (2). In fact, (3) amounts to establishing a unified story, narrative or theory about what happened. Such a narrative should be subsumed under the different elements of wrongdoing as defined in the law. The piecemeal approach and the revised holistic approach, therefore, converge.

To be sure, not all wrongful acts, in civil or criminal cases, require the prosecutor or the plaintiff to establish a unified *spatio-temporal* narrative. It might not be necessary to show that all elements of an offense occurred at the same point in time or in close succession one after the other. Some wrongful acts may consist of a pattern of acts that stretches for several days, months or even years. There may be temporal and spatial gaps that cannot not be filled. We consider several of these examples in our discussion of naked statistical evidence in Chapter XX. Be that as it may, an accusation of wrongdoing SEE PREVIOUS CHAPTER
in a criminal or civil case should still have a degree of cohesive unity. The acts and occurrences that constitute the wrongdoing should belong to the same wrongful act. It is this unity which the plaintiff and the prosecution must establish when they make their case. One way to establish this unity is by providing a unifying narrative, but this need not be the only way. Perhaps the expressions 'theory' or 'explanation' are more apt than 'narrative' or 'story.'

[22] We should be clear that it is not enough for the prosecutor or plaintiff to provide well-specified narrative in support of their allegations, even when they are well-supported by the evidence. When the two narratives are combined into one narrative, its probability could well be below the threshold. If we only require that each element-specific narrative be proven, a defendant could be found criminally or civilly liable even though it is unlikely that they committed the alleged wrongful act. This counter-intuitive result is similar to the one that arose with the atomistic approach.

5.3 Probability, specificity and completeness

We emphasize the distinction between a narrative (or theory, story, explanation, account) and a mere conjunctions of elements of wrongdoing $E_1 \wedge E_1 \wedge \cdots \wedge E_k$. The narrative describes one way among many of instantiating the conjunction. This distinction is important. The claims that constitute a narrative or unified explanation need not map neatly onto the elements of the wrongdoing. The narrative will comprise claims about the evidence itself and how the evidence supports other claims in the narrative, say that witnesses were standing around when the defendant's car hit the child. The narrative or explanation will not only comprise a description of what happened but also of how we know this is what happened.

The distinction between narrative and the mere conjunction of elements matters for how we should understand the standard of proof. Other things being equal, the conjunction is more probable on the evidence than the narrative, and each conjunct even more probable. But this does not mean that the mere conjunction is established by a higher standard of proof than the narrative. As we argued in Chapter XX on naked statistical evidence, a highly probable narrative that nevertheless lacks the desired degree of specificity will fail to meet the standard of proof. By contrast, a more specific narrative that is otherwise less probable than the mere conjunction might well meet the standard. On this account, the standard of proof consists of two criteria: (1) the posterior probability of the proposed narrative (or theory, story, explanation) given the evidence presented at trial; and (2) the degree of specificity, coherence and unity of the narrative or explanation.

REFERENCE TO EARLIER CHAPTERS

R: remember to add your paper to the folder.

R: added this brief survey, take a look!

Are we giving up on legal probabilism, then? We are giving up on *traditional* legal probabilism. Even though ideas such as specificity, coherence and unity cannot be captured by the posterior probability alone, they can be formalized as properties of Bayesian networks. Early development of this approach can be found in (Urbaniak, 2018). While we will develop this approach further later on , the general idea is this. Once we represent the relevant claims and pieces of evidence and their interaction as a Bayesian network (or a set thereof), some of the binary nodes corresponding to various propositions are marked as evidence nodes, and some qualify as narration nodes—these are the nodes that various sides or proposed scenarios disagree about. Within this setup various intuitively needed notions can be explicated. For instance, an accusing narration should "make sense" of evidence in the following sense: any item of evidence presented, should have sufficiently high posterior probability in the network updated with the total evidence obtained. In contrast, a defending narration is supposed to explain evidence in a different sense: if the defense story is rather minimal and mostly constitutes in rebutting the accusations, it isn't reasonable to expect the defense to explain all pieces of evidence, and it is not reasonable to expect the defense to provide a story explaining how each piece of evidence came into existence. Rather, the defense should argue that the probability of the evidence being as it is while the defense's narration is true isn't below a rejection threshold. As another example, a piece of evidence is missing if there is an evidential proposition such that the probability of it being instantiated (given what is known about cases of a given type and about the particularities of a given case) is non-negligible, but it is not included in the evidence. Yet another example: a narration contains gaps just in case there are claims that the narration should choose from (nodes that a narration should consider instantiated), but it does not do so. This, of course, is very hand-wavy at this stage, but at this point we just want to leave the reader with the impression that more can be explicated with Bayesian networks than one might initially expect, leaving detailed development to a different chapter.

REF CHAPTER

REF TO APPROPRIATE CHAPTER

Further, this analysis of the standard of proof—which combines two criteria, posterior probability and specificity—can be evaluated using concepts from probability theory that are not posterior probabilities. To illustrate, compare a trial system that convicts defendants on the basis of claims that generic but highly probable, as opposed to a trial system that convicts defendants on the basis of claims that are more specific but less probable. A natural question to ask at this point is, which trial system will make fewer mistakes—fewer false convictions and false acquittals—in the long run? The answer is not obvious. But the question can be made precise in the language of probability. The question concerns the diagnostic properties of the two trial systems, such as their rate of false positives and false negatives. We examine this question in Chapter YY, including a simulationist study of the impact of various features of narrations in their expected accuracy. To anticipate, we argue that more specific claims are liable to more extensive *direct* adversarial scrutiny than generic claims. The more specific someone's claim, the more liable to be attacked. At the same time, if a specific claim resists adversarial scrutiny, it becomes more firmly established than a less specific claim that survived scrutiny by evading the questions. So specificity plays an accuracy-conducive role even though more specific claims are, other

REFERENCE TO LATER CHAPTER

things being equal, less probable than more generic claims.[23] That is why specificity should be an important ingredient in any theory of the standard of proof.

Another ingredient worth adding to posterior probability and specificity is the completeness of the evidence presented at trial. Could the probability of someone's guilt be extremely high just because the evidence presented is one-sided and missing crucial pieces of information? It certainly can. If the probability of guilt is high because the evidence is partial, guilt was not proven beyond a reasonable doubt. It is a matter of dispute whether knowledge about the partiality of the evidence should affect the posterior probability. After all, if we know that evidence about a hypothesis is missing, shouldn't we revise the assessment of the posterior probability of the hypothesis? This may be true, but the problem is that the content of the missing evidence is unknown. The missing evidence might increase or decrease this probability. We cannot know that without knowing what the evidence turns out to be. If we knew how the missing evidence would affect our judgment about the defendant's guilt, the evidence would no longer be—strictly speaking—missing.

R: do you still want to add these?

THIGNS TO ADD: 1. ROLE OF COMPLETENESS OF EVIDENCE, SEE OREGON DNA CASE, MEANT TO SHOW THAT HOLISTIC AOPPROACH IS NOT AD HOC, WEIGHT, RESILIENCE IMPORTANT FOR ASESSING STRENGHT OF EVIDENCE AND STANDART OF PROOF, GUILT COULD BE HIGHLY PROBABLE GIVEN AVAILABLE EVIDEMCE, BUT THIS NEED NOT BE ENOUGH IF EVIDENCE THAT SHOULD BE THERE IS MISSING

5.4 The conjuction principle revised

What does this discussion tell us about the conjunction paradox? Say we take seriously the idea that the standard of proof is a legal device to optimize the satisfaction of the following three criteria:

1. The defendant's civil or criminal liability must be sufficiently high
2. The narrative, story, theory, explanation that details the defendant's civil or criminal liability should be sufficiently (reasonably) specific
3. The supporting evidence should be sufficiently (reasonably) complete

The conjunction paradox no longer arises given this conception of the standards of proof. That prosecutors and plaintiffs should aim to establish a well-specified, unified account of the wrongdoing would trivialize the conjunction principle and dissolve the difficulty about conjunction. Suppose the prosecutor established a narrative N by a very high probability, say above the required threshold for proof beyond a reasonable doubt. Denote the elements of wrongdoing by $E_1, E_2, \ldots$. Then,

$$\mathsf{P}(E_1 \wedge E_1 \wedge \cdots \wedge E_k | N) = \mathsf{P}(C_i | N) = 1 \text{ for any } i = \{1, 2, ..., k\}.$$

Both directions of the conjunction principle, aggregation and distribution, are now trivially satisfied. Once we condition on the narrative N, each individual claim has a probability of one and thus their conjunction also has a probability of one. The conjunction principle is reduced to a deductive check that the elements of the wrongdoing follow deductively from the narrative put forward. The narrative, however, has a probability short of one, up to whatever value is required to meet the governing standard of proof. The standard applies to the narrative as a whole, and only indirectly—via a deductive check—to the individual elements. This trivialization of the conjunction principle is unsurprising and desirable given that no lawyer has ever been concerned with the reliability of conjunction elimination or introduction.

References

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

Allen, R. J., & 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

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

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

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

[23] CITE POPPER HERE

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

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.

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

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

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

Kowalewska, A. (2021). Reasoning without the conjunction closure. *Episteme*, 1–14. https://doi.org/10.1017/epi.2020.53

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

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

Makinson, D. C. (1965). The paradox of the preface. *Analysis*, *25*(6), 205–207.

Pardo, M. S., & Allen, R. J. (2008). Judicial proof and the best explanation. *Law and Philosophy*, *27*(3), 223–268.

Pennington, N., & Hastie, R. (1993). Reasoning in explanation-based decision making. *Cognition*, *49*, 123–163.

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

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

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

Wagenaar, W., Van Koppen, P., & Crombag, H. (1993). *Anchored narratives: The psychology of criminal evidence.* St Martin's Press.