

# Journal of Applied Logics

The IfCoLog Journal of Logics and their Applications

Volume 6 • Issue 5 • August 2019

## Special Issue Reasoning for Legal AI

**Guest Editors**  
Livio Robaldo  
and Leendert van der Torre

Journal of Applied Logics The IfCoLog Journal of Logics and their Applications

Available online at  
[www.collegepublications.co.uk/journals/ifcolog/](http://www.collegepublications.co.uk/journals/ifcolog/)  
Free open access

### Contents

#### Articles

<b>Introduction to Legal AI</b> Livio Robaldo and Leendert van der Torre	711
<b>I/O Logic in HOL</b> Christoph Benz Müller, Ali Farjami, Paul Meder and Xavier Parent	715
<b>Åqvist's Dyadic Deontic Logic E in HOL</b> Christoph Benz Müller, Ali Farjami and Xavier Parent	733
<b>LEGIS: A Proposal to Handle Legal Normative Exceptions and Leverage Inference Proofs Readability</b> Cleyton Rodrigues, Eunice Palmeira, Fred Freitas, Italo Oliveira and Ivan Varzinczak	757
<b>Principles for a Judgement Editor Based on Binary Decision Diagrams</b> Guillaume Aucher, Jean Berbinau and Marie-Laura Morin	783
<b>Time and Compensation Mechanisms in Checking Legal Compliance</b> Guido Governatori and Antonino Rotolo	817
<b>Rights and Punishment: The Hohfeldian Theory's Applicability and Morals in Understanding Criminal Law</b> Réka Markovich	849
<b>Probabilistic Legal Decision Standards Still Fail</b> Rafal Urbaniak	867
<b>A Deontic Argumentation Framework Towards Doctrine Reification</b> Régis Riveret, Antonino Rotolo and Giovanni Sartor	905
<b>Two Limitations in Legal Knowledge Base Constructing and Formalizing Law</b> Réka Markovich	941
<b>Efficient Full Compliance Checking of Concurrent Components for Business Process Models</b> Silvano Colombo Tosatto, Guido Governatori, Nic Van Beest and Francesco Olivieri	963

Sponsored by



ISBN 978-1-84890-311-1



Published by



7.44 x 9.69  
246 mm x 189 mm

.625  
15.87mm

7.44 x 9.69  
246 mm x 189 mm

Content Type: Black & White  
Paper Type: White  
Page Count: 298  
File Type: InDesign  
Request ID: CSS2689735

---

## CONTENTS

---

### ARTICLES

<b>Introduction to Legal AI</b> . . . . .	<b>711</b>
<i>Livio Robaldo and Leendert van der Torre</i>	
<b>I/O Logic in HOL</b> . . . . .	<b>715</b>
<i>Christoph Benz Müller, Ali Farjami, Paul Meder and Xavier Parent</i>	
<b>Åqvist’s Dyadic Deontic Logic E in HOL</b> . . . . .	<b>733</b>
<i>Christoph Benz Müller, Ali Farjami and Xavier Parent</i>	
<b>LEGIS: A Proposal to Handle Legal Normative Exceptions and Leverage Inference Proofs Readability</b> . . . . .	<b>755</b>
<i>Cleyton Rodrigues, Eunice Palmeira, Fred Freitas, Italo Oliveira and Ivan Varzinczak</i>	
<b>Principles for a Judgement Editor Based on Binary Decision Diagrams</b> . . .	<b>781</b>
<i>Guillaume Aucher, Jean Berbinau and Marie-Laure Morin</i>	
<b>Time and Compensation Mechanisms in Checking Legal Compliance</b> . . . .	<b>815</b>
<i>Guido Governatori and Antonino Rotolo</i>	
<b>Rights and Punishment: The Hohfeldian Theory’s Applicability and Morals in Understanding Criminal Law</b> . . . . .	<b>847</b>
<i>Réka Markovich</i>	

<b>Probabilistic Legal Decision Standards Still Fail . . . . .</b>	<b>865</b>
<i>Rafal Urbaniak</i>	
<b>A Deontic Argumentation Framework Towards Doctrine Reification . . . .</b>	<b>903</b>
<i>Régis Riveret, Antonino Rotolo and Giovanni Sartor</i>	
<b>Two Limitations in Legal Knowledge Base Constructing and Formalizing Law . . . . .</b>	<b>941</b>
<i>Réka Markovich</i>	
<b>Efficient Full Compliance Checking of Concurrent Components for Business Process Models . . . . .</b>	<b>963</b>
<i>Silvano Colombo Tosatto, Guido Governatori, Nic Van Beest and Francesco Olivieri</i>	

---

# PROBABILISTIC LEGAL DECISION STANDARDS

## STILL FAIL

(AT LEAST WHEN IT COMES TO THE DIFFICULTY ABOUT CONJUNCTION AND  
THE GATECRASHER PARADOX)

RAFAL URBANIAK

*University of Gdansk (Poland) & Ghent University (Belgium)*

---

### Abstract

Various probabilistic explications of the phrase *the court's decision regarding a fact, given the evidence, is justified* have been proposed. In this paper I evaluate them against two conceptual challenges: the difficulty about conjunction and the gatecrasher paradox. I argue that despite arguments to the opposite, all proposed models fail to solve these two problems.

## 1 Introduction

Imagine you are a trier of fact in a legal proceeding in which the defendant's guilt is identified as equivalent to a certain factual statement  $G$  and that somehow you succeeded in properly evaluating  $P(G|E)$  – the probability of  $G$  given the total evidence presented to you,  $E$  (and perhaps some other relevant probabilities). For various reasons, some of which will be mentioned soon, this is an idealized situation. One question that arises in such a situation is: *when should you decide against the defendant? when is the evidence good enough?*

What we are after here is a condition  $\Psi$ , formulated in (primarily) probabilistic terms, such that the trier of fact, at least ideally, should accept any relevant claim (including  $G$ ) just in case  $\Psi(A, E)$ . The requirement that the condition should apply to any relevant claim whatsoever (and not just a selected claim, such as  $G$ ) will be called the **equal treatment requirement**.<sup>1</sup>

---

This research has been funded by Narodowe Centrum Nauki (grant no. 2016/22/E/HS1/00304) and Fonds Wetenschappelijk Onderzoek. The author would like to express his gratitude to Marcello Di Bello, Pavel Janda, Alicja Kowalewska, and two anonymous referees for detailed comments on earlier drafts of this paper.

<sup>1</sup>The requirement is not explicitly mentioned in the discussion, but it is tacitly assumed, so it is useful to have a name for it. Moreover, it will turn out crucial when it comes to finding a resolution of the difficulties, but further details need to wait till the last section of this paper.

For instance, one straightforward attempt might be to say: convict if  $P(G|E)$  is above a certain threshold, otherwise acquit. From this perspective, whether assessment of facts leading to conviction is justified is a matter of whether the factual statement corresponding to guilt is sufficiently probable given the evidence.

As it turns out, the idea that such a probabilistic explication  $\Psi$  can be given does not play nicely with some other desiderata that we might want to put forward for what a rational trier of fact should think about facts and evidence.

A large-scale attack on probabilistic approach to legal decisions has been launched quite some time ago by [6], and some of the developments in probabilistic evidence scholarship are to some extent a reaction to some of Cohen's objections. My goal here is to focus on two of them – the **difficulty about conjunction** and the **gatecrasher paradox**. They correspond to two requirements. One, that  $\Psi$  should be such that for any relevant  $A$  and  $B$  there should be no difference between the trier's acceptance  $A$  and  $B$  separately, and her acceptance of their conjunction,  $A \wedge B$ , that is, that  $\Psi(A, E)$  and  $\Psi(B, E)$  just in case  $\Psi(A \wedge B, E)$ . Two, that any such explication should help us make sense of cases in which the probability of guilt given the evidence is high, and yet, conviction is not justified.

I will argue that even most recent proposals of what such a  $\Psi$  should be have failed to address these difficulties. This, however, does not mean that I side with Cohen and claim that thinking of evidence in legal context in terms of probabilities is doomed. Quite the contrary: probabilistic tools are highly useful, and their utility can be increased (and defended) by addressing Cohen's concerns properly. In this paper, however, I leave this positive task for a later occasion, restricting myself to a negative task of showing that legal probabilism so far has not reached this stage.

The paper is somewhat uneven: parts of it contain mostly philosophical discussion, while some other parts get involved with rather detailed and lengthy mathematical arguments whose philosophical points can be summed up in one or two sentences. To avoid discouraging the philosophically minded reader and disappointing the more mathematically oriented one, let me suggest three different reading strategies.

**A quick philosophical look** You do not want to get bogged down in mathematical details, but you want to find out what the philosophical gist of the paper is. In this case, after reading this section, read *Section 3* to find out what the basic variant of the view under criticism is, *Section 4* for a description of the difficulty about conjunction, *Section 7* - which introduces the gatecrasher paradox, and *Section 14* for an informal survey of the results and a philosophical discussion thereof. This is the shortest strategy.

**The middle way** You want to understand the key mathematical aspects, but you

do not care too much about detailed arguments and digressions that can be skipped. Read the whole paper, except for the parts marked in the following manner.

---

OPTIONAL CONTENT STARTS

---

This is how inessential technicalities and longer side remarks are marked.

---

OPTIONAL CONTENT ENDS

---

**The meticulous reader strategy** Read everything. Or, start with a quick philosophical look as described above, and then read everything.

The structure of this paper is as follows. I start with a more detailed discussion of legal probabilism in *Section 2*. Then, I go through various ways legal probabilism so far has tried to navigate around the difficulties that I am concerned with, each time explaining why it failed. To illustrate the idea of a probabilistic model of a decision standard, and to get considerations of conceptual difficulties started, in *Section 3* I move to the first and most straightforward candidate: *Threshold-based LP*. In *Section 4* I explain how this view falls prey to the difficulty about conjunction. Next, in *Section 5* I look at an important attempt to save threshold-based legal probabilism from the difficulty about conjunction, and in *Section 6* I argue that it fails. In *Section 7* I introduce the other conceptual difficulty that will be used as a measuring stick in the assessment of probabilistic models: the gatecrasher paradox. In *Section 8* I outline Cheng’s Relative LP. In *Section 9* I explain how it is supposed to handle DAC, and in *section 10* I look at Cheng’s attempt to avoid the gatecrasher paradox. In *Section 11* I argue that the approach fails on both counts. In *Section 12* I introduce the last candidate to be discussed, Kaplow’s Decision-Theoretic LP, and in *Section 13* I argue that it also cannot handle the difficulties. *Section 14* contains an informal overview and discussion.

## 2 Legal probabilism and its motivations

In multiple domains of applications, probabilistic methods have been exceedingly successful. Some of the successful applications involve forensic and judiciary contexts.<sup>2</sup> What they seem to have in common is that they pertain to the interpretation or weighing of particular pieces of evidence, or an evaluation of a particular argument involving probabilities or statistics.

What is somehow more contentious is whether a more general and all-encompassing probabilistic model of evidence evaluation and decision about conviction can

---

<sup>2</sup>See for instance [11, 1, 26, 19, 23].

be successfully built: one which would explicate the phrase *given the evidence, the conviction is justified* in probabilistic terms (where the claim under consideration is a factual claim considered as equivalent to guilt according to the law). On one hand, *prima facie*, what the judge or the jury seem to evaluate is the probability of guilt given total available evidence and common sense knowledge, and so, probabilistic tools seem fit to model, at least in abstraction, such phenomena. On the other hand, particular pieces of forensic evidence aside, precise probabilities are hard to come by, and conceptual and practical difficulties with developing such a model abound.

To avoid setting the bar too high, let me get clear on what, on the present approach, a successful probabilistic model is *not* required to achieve. Namely, I am putting aside most of the issues that have to do with practicality.<sup>3</sup> I will not be concerned with the lack of real data to support certain probability assessment, I will not be concerned with people being bad at reasoning about probabilities, etc. Basically, I will not be concerned with those practical issues that would arise if one would like to deploy a probabilistic model directly, by writing down numerical values for all the probabilities relevant in a given case and simply calculating the probability of guilt. I will simply grant that at least for now, successful deployments of this type are not viable.

This, however, does not mean that developing a general probabilistic model is pointless. There are multiple ways in which such a model, even if unfit for direct deployment, could be useful. Once we have a probabilistic model, a vast array of mathematical results pertaining to probability can be used to deepen our understanding of the rationality of legal decisions. If at least in abstraction adequate, the model could be useful for diagnosing various types of biases that humans are susceptible to in such contexts; it could be useful as a measuring stick against which various qualitative inference patterns are assessed, and it could be useful as a source of insights about various aspects of legal decisions and evidence presentation methods. Just as understanding physics might be useful for deepening our understanding of how things work, and for building things or moving them around without performing direct exact calculations, a general probabilistic model – again, if adequate – could help us get better at understanding and making legal decisions without its direct deployment in practice.

Just because I put strong practicality requirements aside, it does not mean that I put no constraints on the probabilistic model to be developed. While *sufficient* conditions of adequacy of such a model are somewhat hard to explicate and I will not get into a deeper discussion thereof, there is at least a fairly clear *necessary* condition.

---

<sup>3</sup>My impression is that with a few exceptions, most of the arguments about legal probabilism in the early stage of the debate were concerned mostly with practicality. See [2, 14, 8, 24, 30, 28, 18, 16, 27].

A successful probabilistic model should either avoid or explain away what seem to be important conceptual difficulties that it runs into. And this is what I will be focusing on in this paper: investigating whether available probabilistic models of legal decision standards avoid or explain away the conceptual difficulties that – it seems – they should be able to handle. In particular, I will focus on two pieces of paradoxical flavor, the **difficulty about conjunction (DAC)** and the **paradox of the gatecrasher**.

One reason to choose these two is that they are easy to explain: and it would be nice if we could handle basic conceptual difficulties before we move to more complex issues. Another reason is that in some variant or another, these have been widely discussed in literature. DAC has a very close cousin named the lottery paradox, which occupied the minds of many, and the gatecrasher paradox and related thought experiments and real cases have been extensively discussed by philosophers trying to identify the factor that makes naked statistical evidence actionable.

Let us start with a very general assumption that all the approaches that will be discussed in what follows share; we will call it *Legal Probabilism* (LP). It is the view that the legal notion of probability is to be governed by the mathematical principles of probability theory, and that the decision process in juridical fact-finding is to be explicated by means of probabilistic tools. LP is fairly general: it does not tell us *how exactly* the decision standards are to be explicated in probabilistic terms, it only tells us that somehow they should.

### 3 Threshold-based legal probabilism (TLP)

LP comes in various shapes. It is one thing to say that the standards of juridical proof are to be explicated in probabilistic terms, it is another to provide such an explication. The threshold-based legal probabilism has it that once the probability of guilt (or, to be more precise, the factual statement that according to law is equivalent to guilt) given the total evidence available is assessed, conviction is justified just in case this probability is above a certain threshold.<sup>4</sup>

*Classical Legal Probabilism* (CLP), stemming from [4], keeps the threshold constant.<sup>5</sup>

---

<sup>4</sup>In the Anglo-Saxon tradition there is a distinction between decision standards in civil and in criminal cases. In the former, decision is to be made on the preponderance of probability, and in criminal cases, the guilt statement is supposed to be beyond reasonable doubt. Assuming these are to be modeled by probability thresholds different from 1, there is no essential difference here as far as the conceptual difficulties to be discussed in this paper are involved.

<sup>5</sup>Again, we are going to ignore the difference between civil and criminal litigation here. If one wants to keep this distinction in mind, CLP can be easily revised by positing one threshold for



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

A slightly weaker (and perhaps more common among evidence scholars) variant of this view, let us call it the *Sensitive Legal Probabilism* (SLP), also embraces the idea that what is to be evaluated is the probability of guilt given the evidence, but abandons the requirement that there should be a single threshold for all cases; rather, SLP suggests that the context of each particular case will determine which threshold is appropriate for it.

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

Before we move to the discussion of the two key difficulties that we will be interested in, let me briefly mention one issue about TLP that I will not be concerned with. A careful reader might already have the following complaint: *if you are saying that there is a conviction probability threshold, what exactly is it and why?* And indeed, it seems quite difficult to point to any particular choice of value and argue that the choice is not to a large extent arbitrary.

One reason why I will not be concerned with this problem is that this is an issue that seems to pertain to TLP only, while I would like to focus on problems that seem to be more general.

Another reason is that once decision-theoretic tools are allowed, there might be reasons to think that the choice of threshold is not that arbitrary [17]. Say the probability of guilt (or responsibility) is  $p$ , the disutility of acquitting a guilty person is  $d_g$ , and the disutility of convicting an innocent person is  $d_i$ . From the perspective of minimalization of expected disutility, we would like to convict, or find for the plaintiff, just in case the expected disutility of a mistaken acquittal is greater than the expected disutility of an incorrect conviction:

$$pd_g > (1 - p)d_i$$

---

criminal cases, and one for civil ones.

Now, solving for  $p$  gives us:

$$\begin{aligned} pd_g &> d_i - pd_i \\ pd_g + pd_i &> d_i \\ p(d_g + d_i) &> d_i \\ p &> \frac{d_i}{d_g + d_i} \end{aligned}$$

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

Finally, as practice (such as conviction decisions based on DNA identification) indicates, there are probabilities of guilt clearly considered high enough for conviction, and there are ones which clearly are not high enough for conviction. Perhaps, there are some borderline cases, but these are not too many. From this perspective, the phrase *probability of guilt sufficient for conviction* can be argued to be vague, but the vagueness does not seem too damaging in practice (at least, not more than the vagueness that is already there, even without probabilistic tools). Moreover, it is a rather common practice to theorize about notions which in practice are somewhat vague using idealized mathematical tools which do not tolerate vagueness. As long as the results obtained hold independently of any particular choice of the precisification of a given vague notion, the initial vagueness is not a deep obstacle to the utility of these theoretical considerations.

Now that we know what the first explication is, let us move to the first of the two conceptual difficulties that we actually will be concerned with – the difficulty about conjunction.

## 4 The difficulty about conjunction

The *Difficulty About Conjunction* (DAC) proceeds as follows. Say we focus on a civil suit where a plaintiff is required to prove their case on the balance of probability, which for the sake of argument we construe as passing the 0.5 probability threshold.<sup>6</sup> Suppose the plaintiff's claim to be proven based on total evidence  $E$  is composed of two elements,  $A$  and  $B$ , independent conditionally on  $E$ .<sup>7</sup> The question is, what exactly is the plaintiff supposed to establish? It seems we have two possible readings:

---

<sup>6</sup>This is a natural choice given that the plaintiff is supposed to show that their claim is more probable than the defendant's. The assumption is not essential. DAC can be deployed against any  $\neq 1$  guilt probability threshold.

<sup>7</sup>These assumptions, again, are not too essential. In fact, the difficulties become more severe as the number of elements grows, and, extreme cases aside, do not tend to disappear if the elements

---

**Requirement 1**  $P(A \wedge B|E) > 0.5$

**Requirement 2**  $P(A|E) > 0.5$  and  $P(B|E) > 0.5$

---

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

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

Should we then abandon **Requirement 1** and remain content with **Requirement 2**? Cohen [6, 66] convincingly argues that we should not. Not evaluating a complex civil case as a whole is the opposite of what the courts themselves normally do. There are good reasons to think that every common law system subscribes to a sort of conjunction principle, which states that if  $A$  and  $B$  are established on the balance of probabilities, then so is  $A \wedge B$ .

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

---

are dependent.

## 5 Dawid’s likelihood strategy

One well-known attempt to handle DAC from the probabilistic perspective without any drastic changes to the probabilistic model is due to Dawid [9]. Here is how it proceeds (the considerations that follow apply to other sorts of uncertain evidence; we’ll focus on witnesses for the sake of simplicity). Imagine the plaintiff produces two independent witnesses:  $W_A$  attesting to  $A$ , and  $W_B$  attesting to  $B$ . Say the witnesses are regarded as 70% reliable and  $A$  and  $B$  are probabilistically independent, so we infer  $P(A) = P(B) = 0.7$  and  $P(A \wedge B) = 0.7^2 = 0.49$ .

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

---

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

---

Intuitively, after hearing them, we would still take  $P(A_1)$  to be close to 1 and  $P(A_2)$  to be close to 0, because we already have fairly strong convictions about the issues at hand. In general, how we should revise our beliefs in light of a testimony depends not only on the reliability of the witness, but also on our prior convictions.<sup>8</sup> And this is as it should be: as indicated by Bayes’ Theorem, one and the same testimony with different priors might lead to different posterior probabilities.

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

$$\begin{aligned} O(A) &= \frac{P(A)}{P(\neg A)} \\ O(A|a) &= \frac{P(A|a)}{P(\neg A|a)} \\ LR(a|A) &= \frac{P(a|A)}{P(a|\neg A)}. \end{aligned}$$

---

<sup>8</sup>An issue that Dawid does not bring up is the interplay between our priors and our assessment of the reliability of the witnesses. Clearly, our posterior assessment of the credibility of the witness who testified  $A_2$  will be lower than that of the other witness. But a deeper discussion goes beyond the scope of this paper.

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

$$\frac{P_{posterior}(A)}{P_{posterior}(\neg A)} = \frac{P_{prior}(A|a)}{P_{prior}(\neg A|a)} = \frac{P_{prior}(a|A)}{P_{prior}(a|\neg A)} \times \frac{P_{prior}(A)}{P_{prior}(\neg A)}$$

that is,

$$O_{posterior}(A) = O_{prior}(A|a) = \underbrace{LR_{prior}(a|A)}_{\text{conditional likelihood ratio}} \times O_{prior}(A) \quad (1)$$

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

---

OPTIONAL CONTENT STARTS

---

Dawid gives no general argument, but it is not too hard to give one. Let  $rel(a) = P(a|A) = P(a|\neg A)$ . We have in the background  $P(a|\neg A) = 1 - P(a|A) = 1 - rel(a)$ .

We want to find the condition under which  $P(A|a) = P(a|A)$ . Set  $P(A) = p$  and start with Bayes' Theorem and the law of total probability, and go from there:

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

First we multiplied both sides by the denominator. Then we divided both sides by  $P(a|A)$  and multiplied on the right side. Then we used our background notation and information. Next, we manipulated the right-hand side algebraically and

moved  $-p$  to the left-hand side. Move  $2rel(a)p$  to the left and manipulate the result algebraically to get to the last line last line (naturally, we assume  $rel(a) \neq 1$ ).

---

OPTIONAL CONTENT ENDS

---

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

A sensible way to probabilistically interpret the 70% reliability of a witness who testifies that  $A$  is to take it to consist in the fact that the probability of a positive testimony if  $A$  is the case, just as the probability of a negative testimony (that is, testimony that  $A$  is false) if  $A$  isn't the case, is 0.7:<sup>9</sup>

$$P_{prior}(a|A) = P_{prior}(\neg a|\neg A) = 0.7.$$

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

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

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

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

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

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

Assuming the independence of the witnesses, this will hold in non-degenerate cases that do not involve extreme probabilities, on the assumption of independence of  $a$  and  $b$  conditional on all combinations:  $A \wedge B$ ,  $A \wedge \neg B$ ,  $\neg A \wedge B$  and  $\neg A \wedge \neg B$ .<sup>10, 11</sup>

---

<sup>9</sup>In general setting, these are called the *sensitivity* and *specificity* of a test (respectively), and they don't have to be equal. For instance, a degenerate test for an illness which always responds positively, diagnoses everyone as ill, and so has sensitivity 1, but specificity 0.

<sup>10</sup>Dawid only talks about the independence of witnesses without reference to conditional independence. Conditional independence does not follow from independence, and it is the former that is needed here (also, four non-equivalent different versions of it).

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

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

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

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

By the law of total probability we have:

$$\begin{aligned} P(a \wedge b) &= P(a \wedge b | A \wedge B)P(A \wedge B) + \\ &\quad + P(a \wedge b | A \wedge \neg B)P(A \wedge \neg B) \\ &\quad + P(a \wedge b | \neg A \wedge B)P(\neg A \wedge B) + \\ &\quad + P(a \wedge b | \neg A \wedge \neg B)P(\neg A \wedge \neg B) \end{aligned} \tag{2}$$

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

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

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

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

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

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

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

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

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

---

OPTIONAL CONTENT ENDS

---

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

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

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

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

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

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

---

OPTIONAL CONTENT STARTS

---

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

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

Our independence assumption means:

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

Abbreviate the probabilities the way we already did:



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

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

$$p_{11} + p_{10} + p_{01} + p_{00} = 1 \quad (3)$$

So, we can use this with (2) to get:

$$\begin{aligned} P(a \wedge b) &= \mathbf{ab}p_{11} + \mathbf{a(1-b)}p_{10} + (\mathbf{1-a})\mathbf{b}p_{01} + (\mathbf{1-a})(\mathbf{1-b})p_{00} \\ &= p_{11}\mathbf{ab} + p_{10}(\mathbf{a-ab}) + p_{01}(\mathbf{b-ab}) + p_{00}(\mathbf{1-b-a+ab}) \end{aligned} \quad (4)$$

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

$$\begin{aligned} P(A \wedge B | a \wedge b) &= \frac{P(a \wedge b | A \wedge B)P(A \wedge B)}{P(a \wedge b)} \\ &= \frac{\mathbf{ab}p_{11}}{p_{11}\mathbf{ab} + p_{10}(\mathbf{a-ab}) + p_{01}(\mathbf{b-ab}) + p_{00}(\mathbf{1-b-a+ab})} \end{aligned} \quad (5)$$

To answer our question we therefore have to compare the content of (5) to  $p_{11}$  and our claim holds just in case:

$$\begin{aligned} &\frac{\mathbf{ab}p_{11}}{p_{11}\mathbf{ab} + p_{10}(\mathbf{a-ab}) + p_{01}(\mathbf{b-ab}) + p_{00}(\mathbf{1-b-a+ab})} > p_{11} \\ &\frac{\mathbf{ab}}{p_{11}\mathbf{ab} + p_{10}(\mathbf{a-ab}) + p_{01}(\mathbf{b-ab}) + p_{00}(\mathbf{1-b-a+ab})} > 1 \\ &p_{11}\mathbf{ab} + p_{10}(\mathbf{a-ab}) + p_{01}(\mathbf{b-ab}) + p_{00}(\mathbf{1-b-a+ab}) < \mathbf{ab} \end{aligned} \quad (6)$$

Proving (6) is therefore our goal for now. This is achieved by the following reasoning:<sup>12</sup>

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

---

<sup>12</sup>Thanks to Pawel Pawlowski for working on this proof with me.

The last line is what we have been after.

---

OPTIONAL CONTENT ENDS

---

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

## 6 Troubles with the likelihood strategy

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

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

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

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

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

When the problem is analysed carefully, the ‘paradox’ evaporates: suitably measured, the support supplied by the conjunction of several independent testimonies exceeds that supplied by any of its constituents.

[p. 97]

This is quite a different claim from the content of Fact 1, because previously the joint probability was claimed only to increase as compared to the prior, and here it

is claimed to increase above the level of the separate increases provided by separate testimonies. Regarding this issue Dawid elaborates (we still use the  $p_{ij}$ -notation that we've already introduced):

“More generally, let  $P(a|A)/P(a|\neg A) = \lambda$ ,  $P(b|B)/P(b|\neg B) = \mu$ , with  $\lambda, \mu > 0.7$ , as might arise, for example, when there are several available testimonies. If the witnesses are independent, then

$$P(A \wedge B|a \wedge b) = \lambda\mu p_{11}/(\lambda\mu p_{11} + \lambda p_{10} + \mu p_{01} + p_{00})$$

which increases with each of  $\lambda$  and  $\mu$ , and is never less than the larger of  $\lambda p_{11}/(1 - p_{11} + \lambda p_{11})$ ,  $\mu p_{11}/(1 - p_{11} + \mu p_{11})$ , the posterior probabilities appropriate to the individual testimonies.” [p. 95]

This claim, however, is false.

---

OPTIONAL CONTENT STARTS

---

Let us see why. The quoted passage is a bit dense. It contains four claims for which no arguments are given in the paper. The first three are listed below as (7), the fourth is that if the conditions in (7) hold,  $P(A \wedge B|a \wedge b) > \max(P(A|a), P(B|b))$ . Notice that  $\lambda = LR(a|A)$  and  $\mu = LR(b|B)$ . Suppose the first three claims hold, that is:

$$\begin{aligned} P(A \wedge B|a \wedge b) &= \lambda\mu p_{11}/(\lambda\mu p_{11} + \lambda p_{10} + \mu p_{01} + p_{00}) \\ P(A|a) &= \frac{\lambda p_{11}}{1 - p_{11} + \lambda p_{11}} \\ P(B|b) &= \frac{\mu p_{11}}{1 - p_{11} + \mu p_{11}} \end{aligned} \tag{7}$$

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

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

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

So what went wrong in Dawid's calculations in (7)? Well, the first formula is correct. However, let us take a look at what the second one says (the problem with the third one is pretty much the same):

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

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

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

Now, on the assumption that witness' sensitivity is equal to their specificity, we have

$P(a|\neg A) = P(\neg a|A)$  and can substitute this in the denominator:

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

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

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

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

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

---

OPTIONAL CONTENT ENDS

---

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

---

OPTIONAL CONTENT STARTS

---

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

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

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

With the joint likelihood, the reasoning is just a bit more tricky. We will need to know what  $P(a \wedge b | \neg(A \wedge B))$  is. There are three disjoint possible conditions in

which the condition holds:  $A \wedge \neg B$ ,  $\neg A \wedge B$ , and  $\neg A \wedge \neg B$ . The probabilities of  $a \wedge b$  in these three scenarios are respectively  $\mathbf{a}(1 - \mathbf{b})$ ,  $(1 - \mathbf{a})\mathbf{b}$ ,  $(1 - \mathbf{a})(1 - \mathbf{b})$  (again, the assumption of independence is important), and so on the assumption  $\neg(A \wedge B)$  the probability of  $a \wedge b$  is:

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

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

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

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

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

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

---

OPTIONAL CONTENT ENDS

---

Fact 1 is so far the most optimistic reading of the claim that if witnesses are independent and fairly reliable, their testimonies are going to provide positive support for the conjunction,<sup>13</sup> and any stronger reading of Dawid's suggestions fails. But Fact 1 is not too exciting when it comes to answering the original DAC. The original question focused on the adjudication model according to which the deciding agents are to evaluate the posterior probability of the whole case conditional on all evidence, and to convict if it is above a certain threshold. The problem, generally, is that it might be the case that the pieces of evidence for particular elements of the claim can have high likelihood and posterior probabilities of particular elements can be above the threshold while the posterior joint probability will still fail to meet the threshold. The fact that the joint posterior will be higher than the joint prior does not help much. For instance, if  $\mathbf{a} = \mathbf{b} = 0.7$ ,  $pr = \langle 0.1, 0.5, 0.3, 0.1 \rangle$ , the posterior of

---

<sup>13</sup>And this is the reading that Dawid in passing suggests: "the combined support is always positive, in the sense that the posterior probability of the case always exceeds its prior probability." [9, 95]

$A$  is  $\approx 0.777$ , the posterior of  $B$  is  $\approx 0.608$  and the joint posterior is  $\approx 0.216$  (yes, it is higher than the joint prior  $= 0.1$ , but this does not help the conjunction to satisfy the decision standard).

To see the extent to which Dawid's strategy is helpful here, perhaps the following analogy might be useful. Imagine it is winter, the heating does not work in my office and I am quite cold. I pick up the phone and call maintenance. A rather cheerful fellow picks up the phone. I tell him what my problem is, and he reacts:

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

Dawid's discussion is in the vein of the above conversation. In response to a problem with the adjudication model under consideration Dawid simply invites us to abandon thinking in terms of it and to abandon requirements crucial for the model. Instead, he puts forward a fairly weak notion of support (analogous to a fairly weak sense of the building providing improvement), according to which, assuming witnesses are fairly reliable, if separate fairly reliable witnesses provide positive support to the conjuncts, then their joint testimony provides positive support for the conjunction.

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

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

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

---

OPTIONAL CONTENT STARTS

---

In fact, Cohen replied to Dawid's paper [7]. His reply, however, does not have

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

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

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

---

OPTIONAL CONTENT ENDS

---

## 7 The gatecrasher paradox

Here’s another problem with TLP, the *paradox of the gatecrasher* [6, 21]. A variant of the paradox goes as follows:

Suppose our guilt threshold is high, say at 0.99. Consider the situation in which 1000 fans enter a football stadium, and 991 of them avoid paying for their tickets. A random spectator is tried for not paying. The probability that the spectator under trial did not pay exceeds 0.99. Yet, intuitively, a spectator cannot be considered guilty on the sole basis of the number of people who did and did not pay.<sup>14</sup>

---

<sup>14</sup>The thought experiment that in the absence of any other evidence, the only source of proba-



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

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

The gatecrasher paradox can be considered (or at least has been considered by some scholars) illustrative of a wider phenomenon. According to at least some approaches, there is an important distinction between *naked statistical evidence*, such as the evidence involved in the Gatecrasher Paradox, and *individualized evidence* (such as, say, eyewitness testimony) [13]. Seemingly, judges and human subjects are less willing to convict based on naked statistical evidence than when individualized evidence is available, despite the subjective probability of guilt being the same [31].

Philosophers accepting this distinction have proposed many different explications of what this supposed difference consists in exactly, without much agreement being reached.<sup>15</sup> However, the underdevelopment of philosophical theories aside, as the gatecrasher paradox and some real cases based solely on DNA cold hits that got thrown away indicate, there are at least some cases in which the probability of guilt given the evidence might be high, and the conviction still is not justified. Arguably, a probabilistic explication of judiciary decision standard should at least allow for this possibility and specify the conditions under which this might happen.

## 8 Cheng's Relative Legal Probabilism (RLP)

Let us think about juridical decisions in analogy to statistical hypothesis testing. We have two hypotheses under consideration: defendant's  $H_\Delta$  and plaintiff's  $H_\Pi$ , and we are to pick one:  $D_\Delta$  stands for the decision for  $H_\Delta$  and  $D_\Pi$  is the decision that  $H_\Pi$ . If we are right, no costs result, but incorrect decisions have their price. Let us say that if the defendant is right and we find against them, the cost is  $c_1$ ,

---

bilistic information is the statistics, and so that the probability of guilt corresponds to the frequency of unpaid admissions. If the reader does not agree, I ask her to play along, and to notice that in such a case a principled story of what the probability of guilt is and why is needed.

<sup>15</sup>See [22] for a critical survey and [10] and [25] for more recent proposals.

and if the plaintiff is right and we find against them, the cost is  $c_2$ :

		Decision	
		$D_\Delta$	$D_\Pi$
Truth	$H_\Delta$	0	$c_1$
	$H_\Pi$	$c_2$	0

Arguably, we need a decision rule which minimizes the expected cost. Say that given our total evidence  $E$  we have the corresponding probabilities:

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

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

where  $\mathbf{P}$  stands for the prior probability (this will be the case throughout our discussion of Cheng). The expected costs for deciding that  $H_\Delta$  and  $H_\Pi$ , respectively, are:

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

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

so, assuming that we are minimizing expected cost, we would like to choose  $H_\Pi$  just in case  $E(D_\Pi) < E(D_\Delta)$ . This condition is equivalent to:

$$\begin{aligned} c_1 p_\Delta &< c_2 p_\Pi \\ c_1 &< \frac{c_2 p_\Pi}{p_\Delta} \\ \frac{c_1}{c_2} &< \frac{p_\Pi}{p_\Delta} \end{aligned} \tag{8}$$

[5, 1261] insists:

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

If we grant this assumption, (8) reduces to:

$$\begin{aligned} 1 &< \frac{p_\Pi}{p_\Delta} \\ p_\Pi &> p_\Delta \end{aligned} \tag{9}$$

That is, in standard civil litigation we are to find for the plaintiff just in case  $H_{\Pi}$  is more probable given the evidence than  $H_{\Delta}$ , which doesn't seem like an insane conclusion.<sup>16</sup>

So on this approach, rather than directly evaluating the probability of  $H_{\Pi}$  given the evidence and comparing it to a threshold, we compare the support that the evidence provides for alternative hypotheses  $H_{\Pi}$  and  $H_{\Delta}$  (where, let's emphasize again, the latter doesn't have to be the negation of the former), and decide for the better supported one. Let's call this decision standard **Relative Legal Probabilism (RLP)**.<sup>17</sup>

## 9 RLP vs. DAC

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

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

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

$$\begin{aligned} \frac{P(S \wedge C|E)}{P(S \wedge \neg C|E)} &= \frac{P(S|E)P(C|E)}{P(S|E)P(\neg C|E)} = \frac{P(C|E)}{P(\neg C|E)} > 1 \\ \frac{P(S \wedge C|E)}{P(\neg S \wedge C|E)} &= \frac{P(S|E)P(C|E)}{P(\neg S|E)P(C|E)} = \frac{P(S|E)}{P(\neg S|E)} > 1 \\ \frac{P(S \wedge C|E)}{P(\neg S \wedge \neg C|E)} &= \frac{P(S|E)P(C|E)}{P(\neg S|E)P(\neg C|E)} > 1 \end{aligned} \tag{10}$$

---

<sup>16</sup>Notice that this instruction is somewhat more general than the usual suggestion of the preponderance standard in civil litigation, according to which the court should find for the plaintiff just in case  $P(H_{\Pi}|E) > 0.5$ . This threshold, however, results from (9) if it so happens that  $H_{\Delta}$  is  $\neg H_{\Pi}$ , that is, if the defendant's claim is simply the negation of the plaintiff's thesis. By no means, Cheng argues, this is always the case: often the defendant offers a story which is much more than simply the denial of what the opposite side said.

<sup>17</sup>I was not aware of any particular name for Cheng's model so we came up with this one. We're not particularly attached to it, and it is not standard terminology.

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

## 10 RLP vs. the Gatecrasher Paradox

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

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

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

[5, 1267] insists that in civil trials the prior probabilities should be equal. Granted this assumption, prior odds are 1, and we have:

$$\frac{P(H_{\Pi}|E)}{P(H_{\Delta}|E)} = \frac{P(E|H_{\Pi})}{P(E|H_{\Delta})} \tag{11}$$

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

$$P(E|H_{\Pi}) > P(E|H_{\Delta}) \tag{12}$$

Thus, (12) tells us to convict just in case  $LR(E) > 1$ .

In the case of the gatecrasher paradox, our evidence is statistical. In our variant  $E$ ="991 out of 1000 spectators gatecrashed". Now pick a random spectator, call him Tom, and let  $H_{\Pi}$ ="Tom gatecrashed." [5, 1270] insists:

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

So, on his view, in such a case,  $P(E|H_{\Pi}) = P(E|H_{\Delta})$ , the posterior odds are, by (11), equal to 1, and conviction is unjustified.

## 11 Troubles with RLP

There are various issues with how RLP has been deployed to resolve the difficulties that CLP and TLP run into. First of all, to move from (8) to (9), Cheng assumes that the costs of wrongful decision is the same, be it conviction or acquittal. This is by no means obvious. If a poor elderly lady sues a large company for serious health damage that it supposedly caused, leaving her penniless if the company is liable is definitely not on a par with mistakenly making the company lose a small percent of their funds. Even in cases where such costs are equal, careful consideration and separate argument is needed. If, for instance,  $c_1 = 5c_2$ , we are to convict just in case  $5 < \frac{p_{II}}{p_{\Delta}}$ . This limits the applicability of Cheng's reasoning about DAC, because his reasoning, if correct (and I will argue that it is not correct later on), yields only the result that the relevant posterior odds are greater than 1, not that they are greater than 5. The difficulty, however, will not have much impact on Cheng's solution of the gatecrasher paradox, as long as  $c_1 \leq c_2$ . This is because his reasoning, if correct, establishes that the relevant posterior odds are below 1, and so below any higher threshold as well.

Secondly, Cheng's resolution of DAC uses another suspicious assumption. For (10) to be acceptable we need to assume that the following pairs of events are independent conditionally on  $E$ :  $\langle S, C \rangle$ ,  $\langle S, \neg C \rangle$ ,  $\langle \neg S, C \rangle$ ,  $\langle \neg S, \neg C \rangle$ . Otherwise, Cheng would not be able to replace conditional probabilities of corresponding conjunctions with the result of multiplication of conditional probabilities of the conjuncts. But it is far from obvious that speeding and neck injury are independent. If, for instance, the evidence makes it certain that if the car was not speeding, the neck injury was not caused by the accident,  $P(\neg S \wedge C|E) = 0$ , despite the fact that  $P(\neg S|E)P(C|E)$  does not have to be 0!

Without independence, the best that we can get, say for the first line of (10), is:

$$\begin{aligned} P(S \wedge C|E) &= P(C|E)P(S|C \wedge E) \\ P(S \wedge \neg C|E) &= P(\neg C|E)P(S|\neg C \wedge E) \end{aligned}$$

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

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

are we to convict once we notice that given the evidence, the accusation is rather false, because  $P(\neg H_{\Pi}|E) = 0.63$ ?

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

In the process of solving the gatecrasher paradox, to reach (11), Cheng makes another controversial assumption: that the prior odds should be one, that is, that before any evidence specific to the case is obtained,  $P(H_{\Pi}) = P(H_{\Delta})$ . One problem with this assumption is that it is not clear how to square this with how Cheng handles DAC. For there, he insisted we need to consider *three different* defense scenarios, which we marked as  $H_{\Delta_1}$ ,  $H_{\Delta_2}$  and  $H_{\Delta_3}$ . Now, do we take Cheng’s suggestion to be that we should have

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

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

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

Finally, let us play along and assume that in the gatecrasher scenario the conviction is justified just in case (12) holds. Cheng insists that it does not, because  $P(E|H_{\Pi}) = P(E|H_{\Delta})$ . This supposedly captures the intuition that whether Tom paid has no impact on the statistics that we have.

But this is not obvious. Here is one way to think about this. Tom either paid the entrance fee or did not. Consider these two options, assuming nothing else about the case changes. If he did pay, then he is among the 9 innocent spectators. But this means that if he had not paid, there would have been 992 gatecrashers, and so

$E$  would be false (because it says there was 991 of them). If, on the other hand, Tom in reality did not pay (and so is among the 991 gatecrashers), then had he paid, there would have been only 990 gatecrashers and  $E$  would have been false, again!

So whether conviction is justified and what the relevant ratios are depends on whether Tom really paid. Cheng's criterion (12) results in the conclusion that Tom should be penalized if and only if he did not pay. But this does not help us much when it comes to handling the paradox, because the reason why we needed to rely on  $E$  was exactly that we did not know whether Tom paid.

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

$$\begin{aligned} P(E|H_{\Pi}) &= \frac{P(H_{\Pi}|E)e}{\pi} \\ P(E|H_{\Delta}) &= \frac{P(H_{\Delta}|E)e}{1 - \pi} \end{aligned}$$

Assuming our posteriors are taken from the statistical evidence, we have  $P(H_{\Pi}|E) = 0.991$  and  $P(H_{\Delta}|E) = 0.009$ . So we have:

$$\begin{aligned} LR(E) &= \frac{P(H_{\Pi}|E)e}{\pi} \times \frac{1 - \pi}{P(H_{\Delta}|E)e} \\ &= \frac{P(H_{\Pi}|E) - P(H_{\Pi}|E)\pi}{P(H_{\Delta}|E)\pi} \\ &= \frac{0.991 - 0.991\pi}{0.009\pi} \end{aligned} \tag{13}$$

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

## 12 Kaplow's Decision-Theoretic LP (DTLP)

On RLP, at least in certain cases, the decision rule leads us to (12), which tells us to decide the case based on whether the likelihood ratio is greater than 1. Quite independently, Kaplow [15] suggested another approach to juridical decisions which focuses on likelihood ratios, of which Cheng's suggestion is only a particular case.<sup>18</sup>

---

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

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

Let  $LR(E) = P(E|H_{\Pi})/P(E|H_{\Delta})$ . In whole generality, DTLP invites us to convict just in case  $LR(E) > LR^*$ , where  $LR^*$  is some critical value of the likelihood ratio.

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

$$\begin{aligned} \frac{P(H_{\Pi}|E)}{P(H_{\Delta}|E)} &= \frac{P(H_{\Pi})}{P(H_{\Delta})} \times \frac{P(E|H_{\Pi})}{P(E|H_{\Delta})} > 1 \Leftrightarrow \\ &\Leftrightarrow \frac{P(E|H_{\Pi})}{P(E|H_{\Delta})} > \frac{P(H_{\Delta})}{P(H_{\Pi})} \end{aligned}$$

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

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

$$P(H_{\Pi}|E) \times G > P(H_{\Delta}|E) \times L \tag{14}$$

Now, (14) is equivalent to:

$$\begin{aligned} \frac{P(H_{\Pi}|E)}{P(H_{\Delta}|E)} &> \frac{L}{G} \\ \frac{P(H_{\Pi})}{P(H_{\Delta})} \times \frac{P(E|H_{\Pi})}{P(E|H_{\Delta})} &> \frac{L}{G} \\ \frac{P(E|H_{\Pi})}{P(E|H_{\Delta})} &> \frac{P(H_{\Delta})}{P(H_{\Pi})} \times \frac{L}{G} \\ LR(E) &> \frac{P(H_{\Delta})}{P(H_{\Pi})} \times \frac{L}{G} \end{aligned} \tag{15}$$

This is the general format of Kaplow's decision standard. Now, let us see how it fares when it comes to DAC and the gatecrasher paradox.



### 13 Troubles with DTLP

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

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

$$\begin{aligned} \frac{P(E|A)}{P(E|\neg A)} &> \frac{P(\neg A)}{P(A)} \times \frac{L}{G} \\ \frac{P(E|B)}{P(E|\neg B)} &> \frac{P(\neg B)}{P(B)} \times \frac{L}{G} \end{aligned}$$

entail

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

Alas, the answer is negative.

---

OPTIONAL CONTENT STARTS

---

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

---

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

---

Suppose that  $L = G$  (this is for simplicity only — nothing hinges on this, counterexamples for when this condition fails are analogous). First, notice that  $A$  and  $B$  taken separately satisfy (15).  $P(A) = P(\neg A) = 0.5$ ,  $P(\neg A)/P(A) = 1$   $P(E|A) = 1$ ,  $P(E|\neg A) = 0.4$ . (15) tells us to check:

$$\begin{aligned} \frac{P(E|A)}{P(E|\neg A)} &> \frac{L}{G} \times \frac{P(\neg A)}{P(A)} \\ \frac{1}{0.4} &> 1 \end{aligned}$$

so, following DTLP, we should accept  $A$ . For analogous reasons, we should also accept  $B$ .  $P(B) = P(\neg B) = 0.5$ ,  $P(\neg B)/P(B) = 1$   $P(E|B) = 0.8$ ,  $P(E|\neg B) = 0.6$ , so

we need to check that indeed:

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

$$\frac{0.8}{0.6} > 1$$

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

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

$$\frac{7}{4} > \frac{7}{3}$$

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

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

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

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

---

 OPTIONAL CONTENT ENDS
 

---

Let us turn to the gatecrasher paradox.

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

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

---

<sup>19</sup>Thanks to Alicja Kowalewska for pressing me on this.

From (13) we already now that

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

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

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

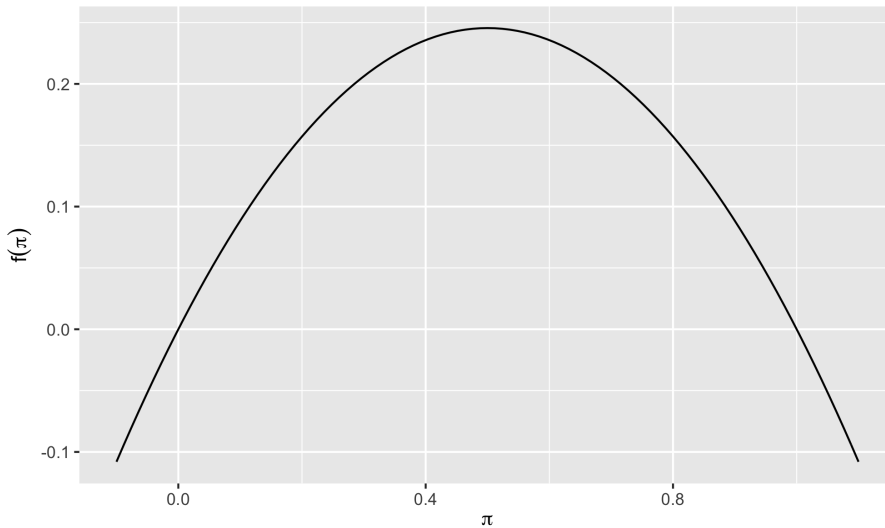
Multiply both sides first by  $009\pi$  and then by  $\pi$ :

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

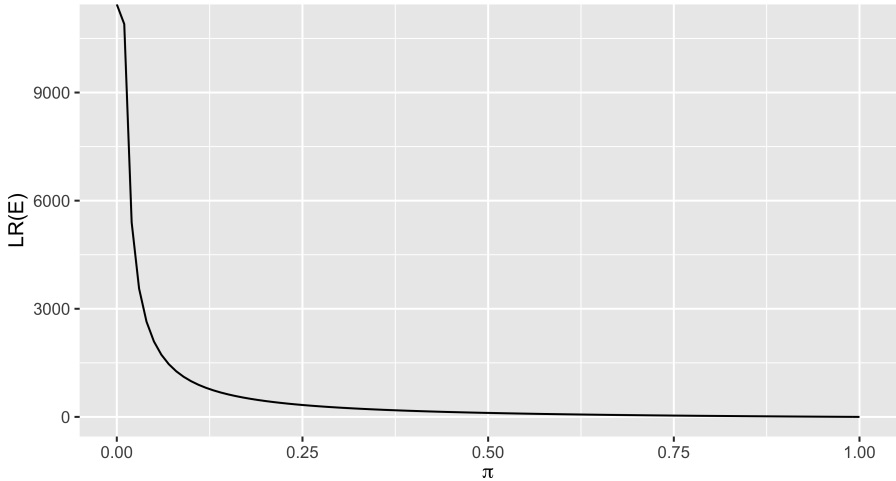
Simplify and call the resulting function  $f$ :

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

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



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



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

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

---

OPTIONAL CONTENT STARTS

---

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

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

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

the general decision standard, plugging in our calculations for  $LR(E)$ :

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

---

OPTIONAL CONTENT ENDS

---

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

---

OPTIONAL CONTENT STARTS

---

There is another, recent approach due to Miller [20].<sup>20</sup> Instead of using  $P(H|E)$ , he introduces a new function,  $Q$ , which he calls contrapositive probability, and defines it as:

$$Q(H|E) = P(\neg E|\neg H)$$

According to a theorem that Miller stated without a proof, if these assumptions hold

$$\begin{aligned}
 Q(H_1|E) &> Q(\neg H_1|E) \\
 Q(H_2|E) &> Q(\neg H_2|E)
 \end{aligned}$$

---

<sup>20</sup>The idea is not developed in any of his papers. What follows is an account based on his lecture at the UNILOG '18 conference.

then it follows that:

$$Q(H_1 \wedge H_2|E) > Q(\neg(H_1 \wedge H_2)|E).$$

Full assessment of this approach will have to wait for a more complete development of the strategy. Note however, that it is far from clear that the above theorem solves the issue. It only applies to cases in which the threshold is 0.5 and says that if conjuncts are above it, then so is the conjunction. It still might be the case that the conjunction has lower “score” than any of the conjuncts, and if so, shifting the threshold might not preserve the value of the theorem.

Moreover, the measure has very unintuitive properties. Consider a single toss of a die and its result.  $P(< 2 | < 3)$  (that is, the probability that the result is 1 given it is less than 3) is  $1/2$ , and so  $Q(\geq 3 | \geq 2) = 1/2$ . So, on this approach, the evidence that the result is one of 2, 3, 4, 5, 6 supports the claim that it is one of 3, 4, 5, 6 only at the level of  $1/2$ , despite the corresponding probability being  $4/5$ . Similarly,  $P(< 3 | < 5) = 1/2$  and so  $Q(\geq 5 | \geq 3) = 1/2$  (the same level as before), despite the corresponding probability this time being  $1/4$ .

---

OPTIONAL CONTENT ENDS

---

## 14 Informal overview

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

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

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

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

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

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

View	Convict iff	DAC	Gatecrasher
Threshold-based LP (TLP)	Probability of guilt given the evidence is above a certain threshold	fails	fails
Dawid's likelihood strategy	No condition given, focus on $\frac{P(H E)}{P(H \neg E)}$	<ul style="list-style-type: none"> <li>- If evidence is fairly reliable, the posterior of <math>A \wedge B</math> will be greater than the prior.</li> <li>- The posterior of <math>A \wedge B</math> can still be lower than the posterior of any of <math>A</math> and <math>B</math>.</li> <li>- Joint likelihood, contrary to Dawid's claim, can also be lower than any of the individual likelihoods.</li> </ul>	fails
Cheng's relative LP (RLP)	Posterior of guilt higher than the posterior of any of the defending narrations	The solution assumes equal costs of errors and independence of $A$ and $B$ conditional on $E$ . It also relies on there being multiple defending scenarios individualized in terms of combinations of literals involving $A$ and $B$ .	Assumes that the prior odds of guilt are 1, and that the statistics is not sensitive to guilt (which is dubious). If the latter fails, tells to convict as long as the prior of guilt $< 0.991$ .
Kaplow's decision-theoretic LP (DTLP)	The likelihood of the evidence is higher than the odds of innocence multiplied by the cost of error ratio	fails	convict if cost ratio $< 110.1111$

Thus, each account either simply fails to satisfy the desiderata, or succeeds on rather unrealistic assumptions. Does this mean that a probabilistic approach to legal evidence evaluation should be abandoned? No. This only means that if we are to develop a general probabilistic model of legal decision standards, we have to do better. One promising direction is to go back to Cohen's pressure against **Requirement 1** and push against it. A brief paper suggesting this direction is [3], where the idea is that the probabilistic standard (be it a threshold or a comparative standard wrt. defending narrations) should be applied to the whole claim put forward by the plaintiff, and not to its elements. In such a context, DAC does not arise, but **equal treatment** is violated. Perhaps, there are independent reasons to abandon it, but the issue deserves further discussion. Another strategy might be to go in the direction of employing probabilistic methods to explicate the narration theory of

legal decision standards [29], but a discussion of how this approach relates to DAC and the gatecrasher paradox lies beyond the scope of this paper.

## References

- [1] Aitken, C. and Taroni, F. (2004). *Statistics and the evaluation of evidence for forensic scientists*, volume 16. Wiley Online Library.
- [2] Ball, V. C. (1960). The moment of truth: probability theory and standards of proof. *Vanderbilt Law Review*, 14:807–830.
- [3] Bello, M. D. (2019). Plausibility and probability in juridical proof. *The International Journal of Evidence & Proof*, doi 10.1177/1365712718815355.
- [4] Bernoulli, J. (1713). *Ars conjectandi*.
- [5] Cheng, E. (2012). Reconceptualizing the burden of proof. *Yale LJ*, 122:1254.
- [6] Cohen, J. (1977). *The probable and the provable*. Oxford University Press.
- [7] Cohen, J. (1988). The difficulty about conjunction in forensic proof. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 37(4-5):415–416
- [8] Cullison, A. D. (1969). Probability analysis of judicial fact-finding: A preliminary outline of the subjective approach. *Toledo Law Review*, 1:538–598.
- [9] Dawid, A. P. (1987). The difficulty about conjunction. *The Statistician*, pages 91–97.
- [10] Enoch, D. and Fisher, T. (2015). Sense and sensitivity: Epistemic and instrumental approaches to statistical evidence. *Stan. L. Rev.*, 67:557.
- [11] Finkelstein, M. O. and Levin, B. (2001). *Statistics for lawyers*. Springer.
- [12] Haack, S. (2014a). *Evidence Matters: Science, Proof, and Truth in the Law*. Cambridge University Press.
- [13] Haack, S. (2014b). Legal probabilism: an epistemological dissent. In [12], pages 47–77.
- [14] Kaplan, J. (1968). Decision theory and the factfinding process. *Stanford Law Review*, 20:1065–1092.
- [15] Kaplow, L. (2014). Likelihood ratio tests and legal decision rules. *American Law and Economics Review*, 16(1):1–39.
- [16] Kaye, D. (1979). The paradox of the gatecrasher and other stories. *Arizona State Law Journal*, pages 101–110.
- [17] Kaye, D. H. (1986). Do we need a calculus of weight to understand proof beyond a reasonable doubt? *Boston University Law Review*, 66(3-4).
- [18] Lempert, R. O. (1977). Modeling relevance. *Michigan Law Review*, 75:1021–1057.
- [19] Lucy, D. (2013). *Introduction to statistics for forensic scientists*. John Wiley & Sons.
- [20] Miller, D. (2018). Cohen’s criticisms of the use of probability in the law. Lecture at the Sixth World Congress on Universal Logic, Vichy.
- [21] Nesson, C. R. (1979). Reasonable doubt and permissive inferences: The value of complexity. *Harvard Law Review*, 92(6):1187–1225.
- [22] Redmayne, M. (2008). Exploring the proof paradoxes. *Legal Theory*, 14(4):281–309.



- [23] Robertson, B., Vignaux, G., and Berger, C. (2016). *Interpreting evidence: evaluating forensic science in the courtroom*. John Wiley & Sons.
- [24] Simon, R. J. and Mahan, L. (1970). Quantifying burdens of proof—a view from the bench, the jury, and the classroom. *Law and Society Review*, 5(3):319–330.
- [25] Smith, M. (2017). When does evidence suffice for conviction? *Mind*.
- [26] Taroni, F., Biedermann, A., Bozza, S., Garbolino, P., and Aitken, C. (2006). *Bayesian networks for probabilistic inference and decision analysis in forensic science*. John Wiley & Sons.
- [27] Tillers, P. and Green, E. D., editors (1988). *Probability and Inference in the Law of Evidence. The Uses and Limits of Bayesianism*, volume 109 of *Boston studies in the philosophy of science*. Springer.
- [28] Tribe, L. H. (1971a). A further critique of mathematical proof. *Harvard Law Review*, 84:1810–1820.
- [29] Urbaniak, R. (2018). Narration in judiciary fact-finding: a probabilistic explication. *Artificial Intelligence and Law*, 26(4):345–376.
- [30] Tribe, L. H. (1971b). Trial by mathematics: Precision and ritual in the legal process. *Harvard Law Review*, 84(6):1329–1393.
- [31] Wells, G. (1992). Naked statistical evidence of liability: Is subjective probability enough? *Journal of Personality and Social Psychology*, 62(5):739–752.