VARIETIES OF LEGAL PROBABILISM: A SURVEY[1]

Małgorzata Stefaniak*
University of Gdansk

Rafal Urbaniak**
University of Gdansk

Abstract: *Legal Probabilism is the view that mathematics, and probability theory in particular, can be used to explicate the standard of legal decisions. While probabilistic tools are sometimes used in courtrooms, the construction of a general model of evidence evaluation remains a challenge. Conceptual difficulties facing Legal Probabilism include the difficulty about conjunction, the difficulty about corroboration and the gatecrasher paradox. These problems need to be addressed before we construct a general model. In this survey we discuss the three difficulties and present some theories proposed as their solutions.*

Key words: *legal probabilism, Bayesian epistemology, legal decision standards.*

ODMIANY PROBABILIZMU PRAWNEGO – PRZEGLĄD

Streszczenie: *Probabilizm prawny to pogląd, że teoria prawdopodobieństwa może być użyta do eksplikacji standardów decyzji prawnych dotyczących faktów. O ile narzędzia probabilistyczne są czasami używane w sądzie, to skonstruowanie ogólnego modelu, który pozwoliłby na ocenę dowodów, stanowi wyzwanie. Probabilizm prawny zmaga się z takimi trudnościami jak problem koniunkcji, problem dotyczący koroboracji czy paradoks widza bez biletu (gatecrasher paradox). Problemy te należy rozwiązać zanim przystąpimy do tworzenia ogólnego modelu. W niniejszym przeglądzie omawiamy wspomniane trudności oraz proponowane w literaturze sposoby ich rozwiązania.*

Słowa kluczowe: *probabilizm prawny, epistemologia bayesiańska, standardy decyzji prawnych.*

1. BAYESIAN EPISTEMOLOGY

From the perspective of **Bayesian epistemology**, degrees of beliefs, called **credences**, are represented by real numbers from the [0,1] interval and are assumed to satisfy the axioms of standard probability theory.[2] This part of the view is called **probabilism**. Probabilism allows Bayesian philosophers to use established results of probability theory to address various philosophical problems.[3]

According to probabilism, credences should satisfy the standard axioms of probability: probability should take values between 0 and 1 inclusive, logically impossible events have probability 0, logically certain events have probability 1, and the probability of the union of finitely many disjoint events is the sum of their individual probabilities (this last condition is called **finite additivity**).[4]

If the agent's credence in given evidence E is greater than 0, we can talk about the conditional probability of hypothesis H given this evidence, $P(H|E)$, which is defined by:

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

Probabilism, together with the standard axioms of probability imply that an ideal agent's credences satisfy the synchronic Bayes' Theorem which tells us how $P_t(E|H)$, the conditional credence in the evidence E given the hypothesis H at a time t, is related to $P_t(H|E)$, the conditional credence in the hypothesis given the evidence at the same time t.

$$\text{(Bayes)} \qquad P_t(H|E) = \frac{P_t(E|H)P_t(H)}{P_t(E)}.$$

We say that Bayes' theorem is synchronic because it only tells us something about the relation between various credences at the same moment in time. On the other hand, **Bayesian updating rule** – another component of Bayesianism – is a *diachronic* rule that tells us how our credences should be revised over time as we obtain new evidence.[4]

---

[2] There are non-standard theories of probability which, for instance, differ in their approach to additivity or their use conditional probability as a primitive.

[3] See, for instance, (De Finetti, 1937), (Ramsey, 1978), (Bovens and Hartmann, 2004), (Bradley, 2015).

[4] Sometimes additivity is assumed to hold for countable unions of events. This, however, is irrelevant for this survey which deals with finite number of events only.

The Bayesian updating rule requires that we update by **conditionalization**: once we find out that E holds (and this is the only thing we find out), our new credence in H, $P_E(H)$, should be equal to our earlier conditional credence in H given E, $P(H|E)$:

(Conditionalization) $P_E(H) = P(H|E)$.

Bayesian epistemology is an attractive approach that has been applied to various problems in epistemology. It has been used to analyze problems and paradoxes related to induction, to study ways to establish agreement in groups, to investigate the concept of justification, to reason about knowledge from (expert) testimony, to develop a confirmation theory, and to further investigate some other philosophical issues pertaining to decision theory and game theory.

Moreover, Bayesian methods have been influential outside of philosophy. Most notably, they have been used to improve statistical methods and inferences in any domain that relies on the use of statistics (Lindley, 1970; Bolstad and Curran, 2016; Savchuk and Tsokos, 2011).

In addition to many useful applications there are various more direct arguments for Bayesianism. The main ones fall into three categories: the Dutch Book Arguments (DBA), Epistemic Utility Arguments and Representation Arguments. Let us go over them briefly.

A Dutch Book is a series of bets in which a player is bound to lose money regardless of what happens. The core of **Dutch Book Arguments** is that sure loss can happen to anyone who violates the rules of probabilism, and cannot happen to anyone who does not. Thus, according to DBA, given that it is irrational to be subjected to sure-loss bets, it is irrational to diverge from probabilism.

The underlying idea behind the **Epistemic Utility Arguments** is that we want our credences to be accurate — as close to what holds in reality as they can be. Suppose we have a measure of accuracy that quantifies the distances between what holds in reality and credence functions defined over multiple propositions. If we take two credence functions P_1 and P_2, we say that P_1 dominates P_2 if and only if P_1 is more accurate than P_2, no matter what the reality looks like.

The Accuracy Theorems (various variants arise from differences among possible assumptions) (Pettigrew, 2011, 2016) say, roughly, that if a credence function P_1 violates the rules of probability, it is dominated by a credence function P_2 which satisfies the rules of probability.

Moreover, no credence function that satisfies the rules of probability is dominated by another credence function. Epistemic Utility Arguments rely on this theorem by

---

5   So, Bayesian updating rule does not follow from Bayes' Theorem and needs a separate justification.

insisting that since it is irrational to have a credence function that can be dominated, it is irrational to have a non-probabilistic credence function.

Representation Theorems, roughly speaking, say that if our preferences between outcomes satisfy certain reasonable conditions (such as *Transitivity*),[6] they can be represented as resulting from beliefs that satisfy the rules of probability (Resnik, 1987). Hence, even if we do not have any numerical credence in mind, the very existence of the correspondence between our qualitative choices and their numerical representation means that we can make inferences using numerical representation and interpret them in the domain of qualitative choices.

So, it seems that we have a decent case for at least trying to think about epistemological issues from the Bayesian perspective. Of course, some degree of idealization is required — after all, we are prone to arithmetical errors and we are not too good at specifying numerical values of our credences. Our hope is that Bayesian reasoning, assuming it applies to our decision problem, will allow us to understand the problem better and will tell us how an ideal rational agent would solve it.

In particular, perhaps, the Bayesian approach can be helpful in our thinking about the epistemic standards of judiciary fact-finding. Once we develop a fairly adequate Bayesian model of this phenomenon, the hope is that we can use Bayesian methods to learn how we should think about various inferences in such contexts.

In *Section* 1 we have introduced Bayesian epistemology and related concepts: Bayes' Theorem, conditionalization, Dutch Book Arguments, and two epistemic utility arguments based on Accuracy Theorems and Representation Theorems.

In *Section* 2 we will introduce Legal Probabilism (LP) and its shapes: Classical Legal Probabilism (CLP) and Threshold-based Legal Probabilism (TLP). *Section* 3 will describe some practical challenges to LP. In *Section* 4 we will turn to conceptual challenges to LP: the so-called *difficulty about corroboration* (with Cohen's result about it described in *Subsection* 4.2), the so-called *difficulty about conjunction*, and the gatecrasher paradox. Next, we will discuss the existing attempts at addressing the last two issues. First, we describe Cheng's Relative Legal Probabilism (RLP) and second, Kaplow's Decision-Theoretic Legal Probabilism (DTLP). In the last step we introduce David Miller's concept of contrapositive probability.

---

[6] Various representation theorems arise from considering different assumptions about agent's preferences and qualitative comparative credences.

2. Legal probabilism

At least some aspects of thinking about evidence in trials are amenable to Bayesian analysis. This approach not only helps avoiding various mistakes, but also provides us with tools to assess the weight of scientific evidence presented in court.

Indeed, successful applications of probabilistic methods in forensic and judiciary contexts abound.[7] They usually pertain to the interpretation or weighing of particular pieces of evidence, or an evaluation of a particular argument involving probabilities or statistics.

To get a taste of the utility of Bayesian method in the courtroom, let us have a quick look at one example. In 1996 Sally Clark gave birth to her first son, who died of Sudden Infant Death Syndrome (SIDS). One year later, her second child also died of SIDS. Sally Clark was accused of murdering her children. An expert, Sir Roy Meadow, testified that the probability of one SIDS death in such a family was 1/8500, and therefore the probability of two SIDS deaths in the family was $(1/8500)^2$. Based on this testimony, Sally Clark was convicted.

What went wrong? One problem was that the expert confused the directions of conditional probabilities (this mistake is called the *prosecutor's fallacy*). The expert testified that P(*Evidence*|*Innocent*), P($E|I$) in short, is low (($1/8500)^2 \approx 1{=}73$ mln.), but what matters for us is P($I|E$), the probability of innocence, given the evidence. And calculating this, according to Bayes' Theorem, requires not only the previously mentioned conditional probability, but also the prior probability of innocence:

$$P(I|E) = \frac{P(E|I)P(I)}{P(E)}.$$

So, while the expert testified that P($E|I$) is very low, no one considered that P(I) is very high, which might result in P($I|E$) still being quite high.[8]

Misleading presentation of the weight of probabilistic evidence led to the wrongful conviction. Despite Clark's acquittal on the second appeal she never recovered from the trauma of the conviction and died a few years after release. Proper reflection on Bayes' Theorem in odds form reveals the appropriate way of presenting evidence of this sort (Aitken et al., 2010).

---

[7]   See, for instance, (Finkelstein and Levin, 2001), (Aitken and Taroni, 2004), (Taroni et al., 2006), (Lucy, 2013), (Robertson et al., 2016).

[8]   The expert made yet another mistake: to obtain the probability of a conjunction of two events he multiplied the probabilities of the conjunctions. This is legitimate only if the conjunctions are independent, which sibling SIDS deaths in a single household might not be.

$$odds(I) = \frac{P(I)}{P(\neg I)}$$

$$odds(I|E) = \frac{P(I|E)}{P(\neg I|E)}$$

$$likelihood(E|I) = \frac{P(E|I)}{P(E|\neg I)}.$$

In the odds form, Bayes' Theorem says:[9]

$$odds(I|E) = likelihood(E|I) \times odds(I)$$

Notably, the prior odds of innocence, while not being part of evidence, are an essential factor in establishing the posterior odds of innocence. The evidence strength, measured by the conditional likelihood, is only part of the equation.

Leaving aside human tragedies caused by incompetent experts, we will turn now to a broader issue. The challenge is to construct a general probabilistic model of evaluating evidence and making decision about conviction. In other words, how should we explicate the phrase *given the evidence, the factual claim considered sufficient for conviction is justified* in probabilistic terms?

Legal Probabilism (LP) is the view that this challenge can be met, that the legal notion of probability is to be governed by the mathematical principles of standard probability theory, and that the decision criterion in juridical fact-finding is to be modeled with probabilistic tools.

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 a proper explication. One example of an explication is the **Classical Legal Probabilism** (CLP) (Bernoulli, 1713), according to which the decision rule is:[10]

(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.

---

[9] To see how it follows, let us start with the usual formulation:

$$P(I|E) = \frac{P(E|I)P(I)}{P(E)}$$

Divide both sides by $P(\neg I|E)$:

$$\frac{P(I|E)}{P(\neg I|E)} = \frac{P(E|I)P(I)}{P(E)P(\neg I|E)}$$

Next, by Bayes' Theorem, substitute $P(\neg I|E)$ with $\frac{P(E|\neg I)P(\neg I)}{P(E)}$. $P(E)$ cancels out, and we are left with the formula in question.

[10] For the sake of simplicity we talk throughout the paper about the probability of guilt. This may sound too careless for a lawyer. For such a reader, let us emphasize that what we mean by "guilt" is *questio facti*, the probability of the defendant's having committed a certain act.

A slightly weaker (and more common among evidence scholars) variant of LP, let us call it the **Threshold-based Legal Probabilism** (TLP), 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, TLP suggests that the context of each particular case will determine the appropriate threshold.

(TLP) 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.

The debate about LP started in the Sixties,[11] continued for quite a few years[12] and led to a careful level of acceptance of some probabilistic methods in the court of law.[13] In the process, however, many have argued that the models offered by LP are either inadequate or unhelpful.[14]

Some concerns dealt with the practicality of the approach. Even if the Bayesian model is theoretically adequate, applying it in practice may still not be a good idea. This is an important concern, and we will discuss it briefly in *Section 3* indicating what roles, in our opinion, a probabilistic model should and should not play.

Other concerns were more conceptual. These include paradoxes that seem to arise when one accepts CLP or TLP (put forward mostly by Cohen (1977)), and difficulties focused on the fact that LP seems to be blind to various phenomena that an adequate account of legal fact-finding should explain. For instance, in legal proceedings arguments go back-and-forth between opposing parties, cross-examination is crucial, and judiciary fact-finding involves the so-called inference to the best explanation and requires reasoning not only evidence-to-hypothesis, but also hypotheses-to-evidence. Yet, seemingly, LP takes no notice of these aspects (Wells, 1992; Stein, 2005; Allen and Pardo, 2007; Dant, 1988).

We have discussed Legal Probabilism and the basic forms it takes. Now we will move to practical, and then conceptual challenges to it.

3. PRACTICAL CHALLENGES TO LP

Let us consider what legal probabilism would look like, if we were to deploy it directly in the courtroom, actually making all judicial decisions using the probabilistic

---

[11] See (Ball, 1960), (Kaplan, 1968), (Cullison, 1969), (Simon and Mahan, 1970), (Lempert, 1977), (Kaye, 1979).
[12] See, for instance, (Tillers and Green, 1988).
[13] (Stein, 2005), (Ho, 2008), (Aitken et al., 2010).
[14] See, for instance, (Tribe, 1971a), (Tribe, 1971b), (Cohen, 1977), (Underwood, 1977), (Nesson, 1979), (Cohen, 1981), (Dant, 1988), (Wells, 1992), (Stein, 2005), (Allen and Pardo, 2007), (Ho, 2008), (Haack, 2014b).

model. First, we should write down all relevant probabilities, then calculate the probability of guilt given the total evidence available, and convict or acquit depending on whether this probability is above or below the threshold. This approach seems to run into the following problems:

- In different cases we seem to have different standards. For instance, the standard is higher in a murder case than, say, in a case of shoplifting. On the one hand, the fact that the threshold depends on what case we are dealing with means that it is relative, which creates a problem for CLP. On the other hand, even if we endorse TLP instead, the question remains: how do we know what threshold value we should employ?

- Trying to standardize the numerical value of the threshold is a hopeless endeavor that would only lead to confusion and mere appearance of clarity and precision.

- The institutions of law would lose their authority if they were to admit that certain mistakes (wrongful conviction and acquittals) are acceptable, and it seems that accepting a guilt probability threshold less than 1 implicitly involves such an admission.[15]

- Not all cases are clear enough for the jury to assign precise numerical probabilities to all pieces of evidence or to determine precise prior conditional probabilities.

- Haack (2014b), Enoch and Fisher (2015), and Smith (2017) suggest that it is a mistake to convict only based on so-called "naked statistical evidence."[16] From this perspective, when we have only statistical evidence, we do not have any story, and nothing substantial which would ensure the validity of the verdict.

- Numerical probabilities might draw attention to quantified evidence at the cost of either ignoring important evidence that is hard to quantify or at the cost of ignoring the question of admissibility of the statistical evidence. In some cases, a proportion in a reference class containing the defendant constitutes an admissible evidence; this is the case with the character evidence, for instance. Yet, in other cases the reference class information would not be admissible; such is the case with race or a social group. Focusing on numerical values might make the audience less sensitive to such admissibility issues.

---

[15] Some people may not consider this to be a problem but some certainly would (cf. Wasserman, 1991).

[16] For instance, in United States vs. Shonubi the average amounts of heroin found on Nigerian drug smugglers caught at JFK airport was used as evidence in determining the total amount smuggled by Shonubi. A natural question is: why is this the right reference class for such considerations? See (Colyvan et al., 2001), (Colyvan and Regan, 2007), (Cheng, 2009), and (Franklin, 2010) for a discussion.

These problems arise when LP is seen as an attempt to quantify everything without a proper reflection and when mathematics is thought of as the only solution to all real-life problems. But this is clearly impossible – in real life, not all relevant probabilities can be estimated in a reasonable way.

This, however, does not mean that LP is a senseless view. Rather, the goal of LP is to formulate a theoretical ideal. This ideal – without direct deployment – can still be used to evaluate or explain the status of different forms of reasoning, qualitative or quantitative, or to obtain more general and abstract insights into the properties of our decision-making.

4. CONCEPTUAL DIFFICULTIES WITH LEGAL PROBABILISM

As we have seen, there are some pretty serious practical reasons why LP might not be considered the best tool to use in the courtroom. Most of them, however, are the result of an extreme interpretation of Legal Probabilism as a desire to unreflectively use numbers to make legal decisions.

Beyond these issues, there are also conceptual challenges to LP. In this section we will describe three that we consider to be the most important ones: the *difficulty about corroboration, the difficulty about conjunction* and *the gatecrasher paradox*.

4.1. THE DIFFICULTY ABOUT CORROBORATION

Corroboration by independent witnesses and convergence of independent items of circumstantial evidence are often encountered in legal proceedings. They are not easy to explicate, however, in proper probabilistic terms. Since both have identical underlying probabilistic structure, in what follows we will talk about corroboration, assuming that what is said applies, *mutatis mutandis*, to convergence.

We can talk about corroboration when two independent witnesses agree in their testimonies. Intuitively, such corroboration sharply increases the probability of their testimonies being true. Suppose two witnesses, *A* and *B*, are fairly reliable. Under normal circumstances, the fact that *A* testified that *S* increases the probability of *S*, and so does the fact that *B* testified that *S*. Intuitively, these two facts, taken together, should increase the probability of *S* much more than each of them taken separately. (Of course, the independence condition is crucial: if *A* and *B* can influence each other, all bets are off.) Would this intuition be reflected in the corresponding probabilities once we represent this reasoning in Bayesian terms?

Take S_A to mean "witness A says that S," and S_B to mean "witness B says that S." The argument structure seems to be as follows: if S_A and S_B, taken separately, increase the probability of S, and S_A and S_B are unrelated (except for being related simply by the truth of S), then the two taken together increase the probability of S much more than each of them separately.

Intuitively, we often have no problem agreeing with arguments of this form. When we find out about some event from two separate, independent sources, our credence in this event usually increases quite a lot. But when we try to model this argument in probabilistic terms, things get tricky.

Multiple proposals as to how to handle such form of reasoning from Bayesian perspective have been put forward. Each has its difficulties. The most important result in this area is due to Cohen (1977, 101-107), who has proved that, as intuitively expected, axioms of probability theory imply increase in the probability of S provided we take into account prior probabilities and assume that some other conditions are satisfied.

4.2. COHEN'S RESULT

Let us take a closer look at the assumptions first.

- Our first two premises say that the testimonies, taken separately, increase the probability of S:

  (1) $\qquad\qquad\qquad P(S|S_A) > P(S)$

  (2) $\qquad\qquad\qquad P(S|S_B) > P(S)$

- The next condition is that the testimonies should be independent, in the following distinct way:

  – First, if B is to corroborate A's testimony, B cannot be more willing to put forward a false testimony when A did so (i.e. S_A assuming $\neg S$ holds). In symbols:

  (3) $\qquad P(S_B|\neg S) \geq P(S_B|S_A \wedge \neg S)$

  – Secondly, if B is to corroborate A's testimony, B's will to testify truthfully cannot be reduced, when A's testimony is true (i.e. B cannot be willing to disagree with A assuming that S holds). So, we have:

  (4) $\qquad P(S_B|S) \leq P(S_B|S_A \wedge S)$

- Next, we assume that $S_A \wedge S_B$ does not have zero probability:

  (5) $\qquad P(S_A \wedge S_B) > 0$

- Finally, we assume that after A's testimony, S still has room to increase:

$$(6) \qquad P(S|S_A) < 1$$

Fact 1. *Conditions* (1), (2), (3), (4), (5) *and* (6) *imply that:*[17]

$$(7) \qquad P(S|S_B \wedge S_A) > P(S|S_A)$$

The theorem tells us that indeed the probability of S increases, but it does not tell us how much. This is a problem for LP for the following reason. If we want to prove that a probabilistic model of legal decision standards is adequate, the properties of this model should correspond to our intuitions about the types of reasoning that we consider to be strongly compelling.

So far, our intuition about the probability of the conclusion increasing a lot upon corroboration does not follow from any straightforward properties of the probabilistic model. On one hand, this lack of a probabilistic model of corroboration can be used as an argument against LP. On the other hand, if a probabilistic model for corroboration can be found, this would count as an argument for Legal Probabilism.

4.3. THE DIFFICULTY ABOUT CONJUNCTION

The difficulty about conjunction (DAC) can be described as follows. Suppose in a civil suit a plaintiff is required to prove the case on the balance of probability and let's assume, for the sake of the argument, that the probability has to be higher than 0.5.[18] Suppose the plaintiff's claim, based on total evidence E, is composed of two elements, A and B, which are independent conditionally on E.[19] The question is, what exactly is the plaintiff supposed to establish? It seems we have two possible interpretations:

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

Requirement 1 states that the plaintiff should show that their claim, defined as the conjunction of A and B, is more likely than its negation. There is a strong intuition that this is what the plaintiff should do. The problem is that this requirement is not equivalent to **Requirement 2**. In fact, if we want to satisfy $P(A \wedge B|E) = P(A|E) \times P(B|E) > 0.5$, satisfying **Requirement 2** will not suffice. 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

---

[17] For the complete proof see (Cohen, 1977, 104-107).

[18] 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 used with any probability threshold which is less than 1.

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

of probability puts a significantly higher requirement on the separate probabilities of the conjuncts.

Moreover, what is required for one of them depends on what has been achieved for the other. If I have already established that $P(A|E) = 0.8$, then I merely 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.

Should we abandon **Requirement 1** and remain content with **Requirement 2**? Cohen (1977, 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 $A \wedge B$.

We have described two of the three conceptual difficulties with LP we have planned to cover. The third and the last one is the gatecrasher paradox.

4.4. The gatecrasher paradox

An important paradox that was developed to show that a high probability of guilt alone is not sufficient for a conviction is the paradox of the gatecrasher (Cohen, 1977; Nesson, 1979). A variant of the paradox goes as follows:

> Suppose our guilt threshold is high, say 0.99. Consider a case in which 1000 fans enter a football stadium and 991 of them enter without paying. A random spectator is tried for not paying. The probability that a fan on 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.

This thought experiment can be adapted to match any particular threshold that a proponent of CLP might suggest, as long as it is < 1. For any choice of a threshold, we can think of a situation in which the probability of guilt given the evidence is higher than the threshold. Yet, conviction in all such cases seems unjustified.

The problem is not only that CLP leads to a conviction that might be wrong and feels intuitively dubious. Since our evidence about each spectator is exactly the same, CLP seems to imply that all of them should be punished, including the nine that have actually paid, as long as we cannot tell them apart. And arguably, there is something disturbing in the idea of a system of justice which explicitly admits that some innocent people should be punished.

Examples above indicate that LP is not as obvious of a tool to use in legal reasoning as it might have initially seemed.

However, attempts have been made to resolve these problems. In section 4.2 we have already discussed a proposal to resolve the difficulty about corroboration. Now we will discuss attempts to resolve the problems with conjunction and with the gatecrasher paradox. We will first look at Chengs's Relative Legal Probabilism. Next we will cover a more general idea, Kaplow's Decision-Theoretic Legal Probabilism. Finally, in our last section we will look at David Miller's concept of contrapositive probability.

5. CHENG'S RELATIVE LEGAL PROBABILISM (RLP)

Let us think about juridical decisions as analogous to statistical hypothesis testing. We have two hypotheses under consideration: defendant's H_Δ and plaintiff's H_Π, and we are to pick one: D_Δ stands for the decision for H_Δ and D_Π is the decision that H_Π. If we are right, our decision does not incur any costs. Incorrect decisions, however, come at a price. Let's assume that if the defendant is right and we find against him, the cost is c_1; if the plaintiff is right and we find against him, the cost is c_2:

		Decision	
		D_Δ	D_Π
Truth	H_Δ	0	c_1
	H_Π	c_2	0

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

$$p_\Delta = P(H_\Delta|E)$$

$$p_\Pi = P(H_\Pi|E)$$

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

$$E(D_\Delta) = p_\Delta 0 + p_\Pi c2 = c_2 p_\Pi$$

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

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

$$c_1 p_\Delta < c_2 p_\Pi$$

$$c_1 < \frac{c_2 p_\Pi}{p_\Delta}$$

$$(8) \qquad \frac{c_1}{c_2} < \frac{p_\Pi}{p_\Delta}$$

Cheng (2012, 1261) insists:

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

Under this assumption, (8) reduces to:

$$1 < \frac{p_\Pi}{p_\Delta}$$

$$(9) \qquad p_\Pi > p_\Delta$$

That means that in a standard civil litigation we are to find for the plaintiff just in case H_Π is more probable given the evidence than H_Δ, which does not seem like an unreasonable conclusion.

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 $P(H|E) > 0.5$. This threshold, however, results from (9) if the defendant's claim is simply the negation of the plaintiff's, that is, if H_Δ is $\neg H_\Pi$, that is, if the defendant's claim is simply the negation of the plaintiff's. 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 has said.

With this approach, instead of directly evaluating the probability of H given the evidence and comparing it to a threshold, we compare the support that the evidence provides for alternative hypotheses H_Π and H_Δ (where, let us emphasize again, the latter does not have to be the negation of the former), and decide for the one that is better supported. Let us call this decision standard **Relative Legal Probabilism** (RLP).[20]

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

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

---

[20] Since we were not aware of any name commonly used for Cheng's model, we have created a label of our own. We should note that this label is not a part of any standard terminology.

The question, clearly, is whether $\frac{P(S \wedge C \,|\, E)}{P(H_\Delta \,|\, E)} > 1$. But to answer it, we have to specify H_Δ. This is the point at which Cheng's remark that H_Δ need not be simply $\neg H_\Pi$ becomes important. In a case like that, 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 compare to each of them? Cheng argues:

$$(10) \qquad \frac{P(S \wedge C \,|\, E)}{P(S \wedge \neg C \,|\, E)} = \frac{P(S \,|\, E)P(C \,|\, E)}{P(S \,|\, E)P(\neg C \,|\, E)} = \frac{P(C \,|\, E)}{P(\neg C \,|\, E)} > 1$$

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

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

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

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

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

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

Cheng (2012, 1267) insists that in civil trials the prior probabilities should be equal. Under assumption, prior odds are 1, and we have:

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

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 in this case tells us to convict just in case:

$$(12) \qquad P(E|H_\Pi) > P(E|H_\Delta)$$

let us denote the likelihood ratio: $LR(E) = P(E|H_\Pi)/P(E|H_\Delta)$. Thus, (12) tells us to convict just in case $LR(E) > 1$.

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

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

So, on his view, in gatecrasher's paradox, $P(E|H_\Pi) = P(E|H_\Delta)$, which means that the posterior odds are, by (11), equal to 1, and hence the conviction is unjustified.[21]

6. KAPLOW'S DECISION-THEORETIC LEGAL PROBABILISM (DTLP)

With RLP the decision rule leads us, in some cases, to (12), and (12) tells us to decide the case based on whether the likelihood ratio is greater than 1. Quite independently, Kaplow (2014) suggested another approach to juridical decisions which focuses on likelihood ratios, of which Cheng's proposal is only a particular case. We will call this approach Decision-Theoretic Legal Probabilism.[22]

Let $LR(E) = P(E|H_\Pi)/P(E|H_\Delta)$. In general, DTLP prescribes conviction 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 if $P(H_\Pi|E) > 0.5$. From Bayes' Theorem we have:

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

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

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 the potential gain from correct conviction, and L stand for the potential loss resulting

---

[21] Cheng's approach is not uncontroversial. For instance, in (10) speeding and neck injury are taken to be independent conditionally on E, which is far from obvious. If so, Cheng cannot replace conditional probabilities of corresponding conjunctions with the result of multiplication of conditional probabilities of the conjuncts. A deeper discussion of other issues goes beyond the scope of this paper.

[22] Again, the name of the view is by no means standard, it is just a term we coined to refer to various types of Legal Probabilism in a fairly uniform manner.

from mistaken conviction. Taking them into account, Kaplow suggests, we should convict if and only if:

(13) $P(H_\Pi|E) \times G > P(H_\Delta|E) \times L$

Now, (13) is equivalent to:

$$\frac{P(H_\Pi|E)}{P(H_\Delta|E)} > \frac{L}{G}$$

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

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

(14) $$LR\,(E) > \frac{P(H_\Delta)}{P(H_\Pi)} \times \frac{L}{G}$$

So, using this rule, we first have to set the ratio of potential loss to potential gain. This ratio should be then multiplied by the prior odds of innocence, to give us a threshold such that we should convict if the posterior likelihood ratio is above it and acquit otherwise.

While Kaplow does not discuss the difficulties that Cheng tried to resolve, nothing indicates that the strategy of approaching them would be any different in DTLP as compared to RLP.

6.1. Miller's contrapositive probability

Another approach to DAC is due to Miller (2018).[23] 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 we have:

$Q(H_1|E) > Q(\neg H_1|E)$

$Q(H_2|E) > Q(\neg H_2|E)$

then it follows that:

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

is true.

---

[23] The idea is not developed in any of his papers. What follows is an account based on his lecture at the UNI-LOG '18 conference.

Full assessment of this approach will have to wait for a more complete development of the strategy. Note however, that it is not 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. A more elaborate discussion of the strategy is beyond the scope of this survey.

7. Summary

Probabilism is a convenient tool for Bayesian philosophers – it allows them to use established mathematical results to approach various philosophical problems. But things get tricky when we try to use probabilism in a courtroom. Different versions of Legal Probabilism should, ideally, help us avoid making mistakes and should assist us in assessing the weight of scientific evidence presented in court. Unfortunately, if we take a closer look at LP, we will find both practical and conceptual problems. In this survey we presented four variants of probabilism meant to solve these problems. None of them, however, can be said to be fully successful. Whether this means that LP should be abandoned, or rather revised, remains to be seen.

References

Aitken, C., Roberts, P., Jackson, G. (2010). *Fundamentals of Probability and Statistical Evidence in Criminal Proceedings (Practitioner Guide No. 1)*. Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses. Royal Statistical Society's Working Group on Statistics and the Law.

Aitken, C. , Taroni, F. (2004). *Statistics and the evaluation of evidence for forensic scientists*. Wiley Online Library.

Allen, R., Pardo, M. (2007). The problematic value of mathematical models of evidence. *The Journal of Legal Studies, 36*(1),107–140.

Ball, V.C. (1960). The moment of truth: probability theory and standards of proof. *Vanderbilt Law Review, 14*, 807–830.

Bernoulli, J. (1713). *Ars conjectandi, opus posthumum. Accedit Tractatus de seriebus infinitis, et epistola gallicé scripta de ludo pilae reticularis*. Bazylea: Thurneysen Brothers.

Bolstad, W.M., Curran, J.M. (2016). *Introduction to Bayesian statistics*. John Wiley & Sons.

Bovens, L., Hartmann, S. (2004). *Bayesian epistemology*. Oxford University Press.

Bradley, D. (2015). *A Critical Introduction to Formal Epistemology*. Bloomsbury Publishing.

Cheng, E. (2012). Reconceptualizing the burden of proof. *Yale Law Journal, 122*(5), 1254–1273.

Cheng, E.K. (2009). A practical solution to the reference class problem. *Columbia Law Review, 109,* 2081–2103.

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

Cohen, L.J. (1981). Subjective probability and the paradox of the gatecrasher. *Arizona State Law Journal, 2*(2), 627–634.

Colyvan, M. , Regan, H.M. (2007). Legal decisions and the reference class problem. *The International Journal of Evidence and Proof, 11*(4), 274–285.

Colyvan, M., Regan, H.M., Ferson, S. (2001). Is it a crime to belong to a reference class. *Journal of Political Philosophy, 9*(2), 168–181.

Cullison, A.D. (1969). Probability analysis of judicial fact-finding: A preliminary outline of the subjective approach. *Toledo Law Review, 1,* 538–598.

Dant, M. (1988). Gambling on the truth: the use of purely statistical evidence as a basis for civil liability. *Columbia Journal of Law and Social Problems, 22,* 31–70.

De Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré, 7,* 1–68. (translated as "Foresight: Its Logical Laws, Its Subjective Sources" in (Kyburg, 1964)).

Enoch, D., Fisher, T. (2015). Sense and sensitivity: Epistemic and instrumental approaches to statistical evidence. *Stanford Law Review, 67,* 557.

Finkelstein, M.O., Levin, B. (2001). *Statistics for Lawyers*. Springer.

Franklin, J. (2010). Feature selection methods for solving the reference class problem: Comment on Edward K. Cheng, a practical solution to the reference class problem. *Columbia Law Review Sidebar, 110,* 12-23.

Haack, S. (2014a). *Evidence Matters: Science, Proof, and Truth in the Law*. Cambridge University Press.

Haack, S. (2014b). Legal probabilism: an epistemological dissent. In: Haack (2014a), 47–77.

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

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

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

Kaye, D. (1979). The paradox of the gatecrasher and other stories. *Arizona State Law Journal, 1033,* 101–110.

Kyburg, H.E. (1964). *Studies in subjective probability*. John Wiley & Sons.

Lempert, R.O. (1977). Modeling relevance. *Michigan Law Review, 75,* 1021–1057.

Lindley, D.V. (1970). *Introduction to probability and statistics from a Bayesian viewpoint*. Cambridge University Press.

Lucy, D. (2013). *Introduction to statistics for forensic scientists*. John Wiley & Sons.

Miller, D. (2018). *Cohen's criticisms of the use of probability in the law*. Lecture at the Sixth World Congress on Universal Logic, Vichy.

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

Pettigrew, R. (2011). *Epistemic utility arguments for probabilism*. Stanford Encyclopedia of Philosophy.

Pettigrew, R. (2016). *Accuracy and the Laws of Credence*. Oxford University Press.

Ramsey, F. (1978). Truth and probability. In: Mellor, D. H., editor, Foundations: *Essays in Philosophy, Logic, Mathematics and Economics*, 58–100. Routledge. [originally published in 1926].

Resnik, M. (1987). *Choices: An Introduction to Decision Theory*. University of Minnesota Press.

Robertson, B., Vignaux, G., Berger, C. (2016). *Interpreting evidence: evaluating forensic science in the courtroom*. John Wiley & Sons.

Savchuk, V., Tsokos, C.P. (2011). *Bayesian theory and methods with applications*. Atlantis Studies in Probability and Statistics (Book 1), Atlantis Press.

Simon, R.J., 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.

Smith, M. (2017). When does evidence suffice for conviction? *Mind, 508*(127), 1193–1218.

Stein, A. (2005). *Foundations of Evidence Law*. Oxford University Press.

Taroni, F., Biedermann, A., Bozza, S., Garbolino, P., Aitken, C. (2006). *Bayesian networks for probabilistic inference and decision analysis in forensic science*. John Wiley & Sons.

Tillers, P., Green, E.D., editors (1988). *Probability and Inference in the Law of Evidence. The Uses and Limits of Bayesianism*. Boston Studies in the Philosophy and History of Science, volume 109. Springer Netherlands.

Tribe, L.H. (1971a). A further critique of mathematical proof. *Harvard Law Review, 84*, 1810–1820.

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

Underwood, B.D. (1977). The thumb on the scale of justice: Burdens of persuasion in criminal cases. *Yale Law Journal, 86*(7), 1299–1348.

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

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