Canadian Journal of Philosophy (2024), 54: 3, 189–206
doi:10.1017/can.2025.6

ARTICLE

Probability of Guilt

Mario Günther

LMU Munich, Munich, Germany and Carnegie Mellon University, Pittsburgh, PA, USA
Email: Mario.Guenther@lmu.de

Abstract
In legal proceedings, a fact-finder needs to decide whether a defendant is guilty, or not, based on probabilistic evidence. We defend the thesis that the defendant should be found guilty just in case it is rational for the fact-finder to believe that the defendant is guilty. We draw on Leitgeb's stability theory for an appropriate notion of rational belief and show how our thesis solves the problem of statistical evidence. Finally, we defend our account of legal proof against challenges from Staffel and compare it to a recent competitor put forth by Moss.

Keywords: legal epistemology; legal proof; statistical evidence; stability theory of belief

1. Introduction

In a criminal trial, the accused is to be found guilty or innocent.[1] The decision is rendered by the fact-finder—a judge or jury—who is governed by a burden of proof and the available and admissible evidence. A criminal conviction, for example, requires the prosecution to prove the defendant's guilt beyond reasonable doubt. This means the evidence presented in court must be enough to remove any reasonable doubt that the accused is guilty of the crime with which they are charged. It is far from clear, however, what the phrase "removing any reasonable doubt" means. A popular way to understand the standard of proof *beyond reasonable doubt* is to require a high level of confidence in the guilt of the defendant.[2] However, consider this case:

> One hundred prisoners are in a yard under the supervision of a guard. At some point, ninety-nine of them collectively kill the guard. Only one prisoner refrains, standing alone in a corner. We know this from a video recording. The video shows that the participation ratio is 99:1, but does not allow for the identification of the ninety-nine killers. There is no other evidence. After the fact, a prisoner is picked at random and tried.[3]

Should the randomly picked prisoner be found guilty? Well, it seems we should be quite confident that the prisoner is guilty. Since 99 out of 100 prisoners killed the guard and the defendant on trial is one of the 100 prisoners, the probability of his guilt is 99%. If we measure confidence in subjective

[1] A notable exception is Scottish law where *not proven* is an available verdict. In most other jurisdictions, the verdict may only be *guilty* or *not guilty*.
[2] In the law, guilt is usually understood to imply an *actus reus* (or objective element of a crime) and a *mens rea* (or criminal intent of a crime). For this article, we put the intricate issue of what constitutes a *mens rea* aside and focus on beliefs about *actus reus*.
[3] The example dates back to Nesson (1979). The wording is taken from Di Bello (2019).

probabilities, also called degrees of belief or credences, the *beyond reasonable doubt* standard seems to amount to the following: the guilt of a defendant is proven if the fact-finder should have a high degree of belief in his guilt. This probabilistic version of the evidential standard is met in the example. And yet, there is something odd about convicting this particular prisoner standing trial. A high probability of guilt seems simply not enough for conviction. However, if a 99% probability of guilt is not enough, what is?[4]

Imagine we had no video recording but an eyewitness who testifies about the randomly picked prisoner: "I saw him killing the guard!" The eyewitness is very reliable but not perfectly reliable. Let us say she raises the probability of the prisoner's guilt to 99%. For many, this eyewitness testimony suffices for a conviction, while the statistics on its own does not. Given that the probability of guilt is the same in the two cases, what makes for the intuitive difference?[5]

Fact-finders have a difficult task. They are forced to decide whether or not a defendant is guilty based on evidence that is probabilistic in one way or another. The video and eyewitness cases illustrate that it is unclear how a high probability of guilt should translate into a binary verdict. Purely statistical evidence seems to differ from individual evidence like the eyewitness testimony in their respective support for a verdict of guilt, even if they make the defendant's guilt equally likely. However, in light of the same probability of guilt, is it not irrational for a fact-finder to judge the two cases differently?

In this article, we defend the thesis that legal proof should be tantamount to rational belief of guilt. A defendant should be found guilty if and only if (iff) it is rational for the fact-finder to believe that the defendant is guilty.[6] We understand rational belief as Leitgeb's (2014b) stably high credence. A high credence of guilt is not sufficient for a stably high credence or rational belief of guilt. We will show that statistical evidence alone does not allow for stably high credence and so is insufficient for rational belief. Hence, a defendant should on our account of legal proof not be found guilty based on statistical evidence alone—in contrast to the verdict of many other probabilistic accounts. Unlike them, our account can trace back the intuitive difference between the cases to what is rational to believe based on the given evidence. And so, it is rational for a fact-finder to judge the two cases differently.

On the thesis we defend, rational belief is governed by norms for full belief and norms for credences. This notion of rational belief implies a probabilistic threshold view that solves the problem posed by statistical evidence.[7] Notably, legal proof requires nothing more than rational belief of guilt. We need not impose any further condition on legal proof—unlike other accounts.[8] Different standards of legal proof will simply turn out to be special cases of rational belief.

We proceed as follows. In Section 2, we show that a high degree of belief in a defendant's guilt is not sufficient for rationally believing that the defendant is guilty. In Section 3, we explain what would be sufficient for a rational belief in guilt and observe that Leitgeb's (2014b) stability theory provides us with an appropriate notion of rational belief. We then apply our account, in Section 4, to the prisoners' example. In Section 5, we defuse Staffel's (2016) challenges against the stability theory

[4]Haack (2014) and Smith (2020) level attacks on the thesis that standards of proof are best understood in terms of probabilities. For a qualified defense of this thesis, see Hedden and Colyvan (2019).

[5]Many proposals for the intuitive difference have been put forth in philosophy and legal theory alike, e.g. Thomson (1986), Redmayne (2008), Smith (2010, 2018), Pardo (2018). For an overview and critical assessment, see Gardiner (2018).

[6]A similar thesis has been suggested by Buchak (2014, see pp. 299–303): a rational agent blames a person just in case she fully believes the person is guilty. This thesis—legal proof is tantamount to justified full belief of guilt—is a "tempting proposal" according to Moss (2018, p. 206). Unlike Buchak, however, we think that belief is rational only if full belief and credences cohere.

[7]This is big news because it is commonly thought that "threshold views of the relationship between *licensed court verdicts* and rational credence are false" (Buchak, 2014, p. 291).

[8]Here is a nonexhaustive list of conditions on legal proof: Enoch et al. (2012) and Enoch and Fisher (2015) suggest to impose sensitivity; Günther (2024) imposes epistemic sensitivity; Pritchard (2015, 2018) imposes safety; and Smith (2010, 2018) normic support. Blome-Tillmann (2017) imposes sufficiently high evidential probability of knowledge. Moss (2021) imposes (probabilistic) knowledge. Steele and Colyvan (2023) impose *some* degree of meta-certainty.

https://doi.org/10.1017/can.2025.6 Published online by Cambridge University Press

and we analyze different standards of proof, *beyond reasonable doubt* and *preponderance of the evidence*. Our account of legal proof compares favorably to Moss's (2021), or so we argue in Section 6.

2. High Credence of Guilt

A fact-finder is often in an unenviable position: a binary decision must be made. A fact-finder usually cannot hedge the decision, unlike private persons (Ross, 2021, p. 14). In particular, a finder must come to a verdict about the defendant's guilt. In addition, this is complicated by the fact that most evidence is probabilistic in one way or another. It is very rare—if not impossible—that a piece of evidence supports a belief without any possibility of revocation.

We said that a defendant should be found guilty just in case it is rational for the fact-finder to believe that the defendant is guilty. Given that evidence comes in probabilistic form and a defendant is either convicted or else acquitted, the finder should have a rational procedure to translate her degrees of belief based on the available evidence into an all-or-nothing belief required for the final verdict.

It is standard to represent a rational agent's credences, or degrees of belief, by a probability distribution that satisfies the standard axioms of probability theory enriched by the ratio definition of conditional probability. On this picture, there is an obvious candidate for bridging credences and qualitative beliefs: an agent should believe a proposition just in case her credence in that proposition exceeds a certain threshold. More formally, let the probability distribution P represent an agent's credences. A simple threshold view then says: a rational agent believes a proposition A iff $P(A) > s$ for some fixed threshold s.

According to the simple threshold view, the fact-finder in our prisoner's example should believe that each prisoner is guilty. To see this clearly, let us formalize the example. Let w_i denote the possible world where prisoner i is innocent and all the other prisoners are guilty. The fact-finder considers the set $W = \{w_1, \ldots, w_{100}\}$ of mutually exclusive and jointly exhaustive possible worlds as relevant possibilities. The finder has maximal credence in the proposition that exactly one of the 100 prisoners is innocent, that is, $P(W) = 1$. She believes that any one of the 100 prisoners may be innocent, and is certain that one is—even though she does not know which one. Having only the $99:1$ participation statistics available, and no further evidence that would make a difference between the prisoners, her credence should be given by the uniform probability distribution P over W:

$$P(\{w_1\}) = P(\{w_2\}) = \ldots = P(\{w_{100}\}) = {}^1\!/_{100}.$$

For $1 \le i \le 100$, $P(\{w_i\}) < s$ for any reasonably high threshold. Indeed, on any reasonable threshold view, the finder should *not* believe that any one of the prisoners is innocent.

Should the fact-finder believe that each prisoner is *not* innocent? To answer this question, let us make explicit a standard propositional framework, which includes a negation. A proposition is a subset of a finite set W of possible worlds. A proposition $A \subseteq W$ is consistent iff $A \neq \emptyset$. A proposition A is consistent with a proposition B iff $A \cap B \neq \emptyset$. A entails B iff $A \subseteq B$. The negation $\neg A$ of a proposition is given by its complement $W \setminus A$, the conjunction $A \wedge B$ of two propositions by their intersection $A \cap B$, and the disjunction $A \vee B$ by their union $A \cup B$. The proposition that prisoner i is not innocent, or equivalently guilty, is then given by the set $W \setminus \{w_i\}$. Prisoner i is guilty in all worlds but w_i. The credence $P(W \setminus \{w_i\}) = .99$ of guilt surpasses a reasonably high threshold.[9] Hence, on the simple threshold view, the fact-finder should believe that each prisoner is guilty. However, the finder is also certain that one of the prisoners is innocent. The threshold view then

[9] Note that the number of prisoners could be increased and so the threshold could be even higher.

https://doi.org/10.1017/can.2025.6 Published online by Cambridge University Press

implies that the fact-finder should believe both at the same time: (i) each prisoner is guilty and (ii) one prisoner is not. However, this implication casts doubt on the rationality of the simple threshold view. Here, (ii) contradicts (i) if belief is closed under conjunction. And so, it seems to be irrational if one and the same agent believes that Prisoner 1 is guilty, and Prisoner 2, and …, and Prisoner 100, but also that one of those 100 prisoners is innocent. A simple threshold view demands of the fact-finder to have all-or-nothing beliefs that are inconsistent.[10]

Imagine a fact-finder who believes that only one person killed John, but she also believes that Jim alone killed John and that Mary alone killed John. Such a finder seems rather irrational because she violates a rationality norm of qualitative belief: beliefs ought to be closed under conjunction. If a rational agent believes A and B, she should also believe $A \wedge B$. We would be uneasy if our imagined fact-finder was to find guilty Jim *and* Mary based on her qualitatively irrational beliefs. As long as the finder believes that only one person killed John, the unease persists. Only if the finder gives up one of her qualitative beliefs so that her beliefs are consistent can we understand and accept her verdict. Her all-or-nothing beliefs should therefore be at least logically consistent to warrant a verdict.[11]

We represent—as is standard for qualitative models of belief—a rational agent's beliefs by a non-empty set $Bel \subseteq \wp(W)$ of propositions the agent believes. A rational agent's all-or-nothing beliefs are consistent and closed under logical consequence. We express closure under logical consequence as follows: for all $A, B \subseteq W$, (i) if $A \in Bel$ and $A \subseteq B$, then $B \in Bel$, and (ii) if $A \in Bel$ and $B \in Bel$, then $A \cap B \in Bel$. Hence, the conjunction B_W of all believed propositions is also in Bel. Thus, Bel uniquely determines B_W and vice versa:

$$A \in Bel \text{ iff } B_W \subseteq A.$$

To sum up, our agent should, according to the simple threshold view, believe that each prisoner is guilty, and that there is one innocent prisoner. As it seems rational to close qualitative belief under conjunction, we obtain a contradiction. Our agent believes $W \setminus \{w_i\}$ for $1 \leq i \leq 100$. By closing her beliefs under conjunction, she believes:

$$(W \setminus \{w_1\}) \cap (W \setminus \{w_2\}) \cap \ldots \cap (W \setminus \{w_{100}\}) = \varnothing.$$

It is logically inconsistent to believe at the same time that all prisoners are guilty and one is innocent.[12] To believe that each prisoner is guilty is irrational, even though the finder should have a high credence in the guilt of any individual prisoner. A high credence in guilt is thus not sufficient for rational belief of guilt.

We have said that a defendant should be found guilty iff it is rational for the fact-finder to believe that the defendant is guilty. We have seen that a fact-finder needs to convert her credences based on the available evidence into an all-or-nothing belief. It seems plausible that the finder should believe a defendant is guilty if her credence in the defendant's guilt exceeds a certain threshold. On the simple threshold view, the finder should believe that each prisoner in our example is guilty, and that one of

[10]This is well known from the literature on the Lottery Paradox. See Kyburg (1961), Hempel (1962), and the more recent review by Wheeler (2007). In light of the paradox, Kyburg and Christensen (2004, Chs. 3&4) among many others reject the closure of rational belief under conjunction.

[11]The consistency of rational belief and its closure under conjunction has been defended among others by Levi (1967), Leitgeb (2014a), and Smith (2022).

[12]The simple threshold fact-finder violates the closure of rational belief under conjunction and so is not qualitatively rational. As $P(\varnothing) = 0$, it is not rational for her to believe that *all* prisoners are guilty even though it is rational for her to believe of *each* individual prisoner that he is guilty. She believes that Prisoner 1 is guilty, that Prisoner 2 is guilty, …, and that Prisoner 100 is guilty, and yet she does not believe that all prisoners are guilty. This is odd because it is not clear how the first 100 beliefs taken together differ from her belief that all prisoners are guilty (Douven, 2002, p. 395).

https://doi.org/10.1017/can.2025.6 Published online by Cambridge University Press

them is innocent. Since the verdict of the finder is binary, the rationality norms of qualitative belief should apply. However, once we close the finder's beliefs under conjunction—which is a rationality norm of qualitative belief—we obtain a contradiction: the finder would be required to believe that *all* prisoners are guilty and *one* is innocent. However, this does not answer the question whether the fact-finder should believe that the particular prisoner standing trial is guilty. We will provide an answer in the next section.

3. Leitgeb's Stability Theory of Rational Belief

In the last section, we have observed that a fact-finder should be *doubly rational.* She should have rational credences, because evidence comes in probabilistic form and she should have rational binary beliefs because her verdict must be either "guilty" or "not guilty," as opposed to, let us say, "guilty to 82%." In our prisoner's example, the fact-finder's beliefs become logically inconsistent and so irrational once her beliefs are governed by the standard rationality norms for credences, those for all-or-nothing belief, and a threshold view bridging her credences and her all-or-nothing beliefs. The finder has no doubly rational belief about whether the prisoner standing trial is innocent or guilty.

Is doubly rational belief in the guilt of a defendant simply asking too much for finding a defendant guilty? We would not say so. Each set of rationality norms is independently plausible for a notion of rational belief. In fact, we have identified three desiderata for rational belief, which come into conflict in the prisoner's example:

(i) All-or-nothing belief should be consistent and closed under logical consequence.
(ii) Credences should obey the axioms of probability theory and the ratio definition of conditional probability.
(iii) An agent should believe a proposition just in case her credence in the proposition is high enough.

It has long been thought that (i)–(iii) cannot be true together, and so at least one of them must be given up. Then, Leitgeb (2014b, 2015, 2017) has shown how the three desiderata give rise to a theory of rational belief. The rough idea is that a rational agent believes A, iff she still assigns A a high enough credence when she conditions on any proposition she considers possible. In brief, a rational agent believes A, iff her credence in A is *stably* high. What gives rise to Leitgeb's theory is (i) relative to a fixed partition, (ii), and a specification of the desideratum (iii), where high enough credence is understood as stably high credence. In this section, we sketch this stability theory of belief and apply it in the next section to the prisoner's example.

What does it exactly mean for a credence in A to be stably high? Recall that an agent's credences are represented by a probability distribution P over a finite set W of worlds. A proposition A is stably high with respect to a credence function P, or simply A is P-stable, iff $P(A|B) > 1/2$ for all $B \subseteq W$ such that $A \cap B \neq \emptyset$ and $P(B) > 0$. P-stability demands that the probability of A remains higher than the probability of its negation when conditioning on each proposition B that is consistent with A and the conditional probability is defined. An agent's credence in A is P-stable only if she considers A to be more likely than not given each proposition consistent with it. It is noteworthy that the P-stability of a non-empty proposition A entails that $P(A|W) = P(A) > 1/2$. For $A \neq \emptyset$ is consistent with W and $P(W) = 1$. Furthermore, any proposition of probability 1 is P-stable.

Leitgeb's theory allows us to determine what a doubly rational agent believes given the agent's credence function P and a consistent P-stable proposition. That is, given an agent's credences P and a non-empty P-stable proposition, one can determine Bel such that the given P-stable proposition is the strongest believed proposition B_W. In fact, Leitgeb has shown that the other

https://doi.org/10.1017/can.2025.6 Published online by Cambridge University Press

direction holds as well.[13] This gives us the core of the stability theory: an agent believes a proposition A just in case her strongest and consistent belief B_W entails it and is P-stable. In symbols,

$$A \in Bel \text{ iff } B_W \subseteq A \text{ and } B_W \text{ is P-stable.} \tag{1}$$

The P-stable proposition B_W is a non-empty subset of any proposition A the rational agent believes. Hence, the rational agent assigns any proposition she believes a credence at least as great as the credence in B_W. In addition, each proposition A she assigns an equal or greater credence than B_W is a superset of B_W due to P-stability. Where B_W is a P-stable proposition, Leitgeb's theory thus entails a *Lockean Thesis* with threshold $P(B_W)$:[14]

$$A \in Bel \text{ iff } P(A) \geq P(B_W), \text{for all A} \subseteq \text{W}.$$

What an agent believes thus depends on her credence function P and the choice of the strongest believed and P-stable proposition B_W.

Here is a reading of Leitgeb's theory geared toward our purposes. A rational agent comes equipped with a belief set Bel and a credal distribution P. At any given time, her all-or-nothing beliefs must cohere with her credences according to Equation (1). This synchronic constraint on her beliefs determines at each moment in time which propositions are consistent with her qualitative beliefs. A proposition A is consistent with her beliefs iff she does not believe $\neg A$.[15] If so, let us say she considers A to be possible, or equivalently A is epistemically possible; if not, not. Now, she believes A just in case she still assigns A a credence over $1/2$ when she conditions on any proposition she considers possible.

Titelbaum (2020) suggested that we may conceive of the propositions she considers possible as propositions she believes she might learn later. Thus, she has an all-or-nothing belief in A only if, for any proposition she believes she might learn later, she still would assign A a high credence if she were to learn that proposition.[16] On this reading, a rational agent believes now a proposition if she is confident in it now and anticipates continued confidence in the future.

Titelbaum's suggestion may be adapted to a preliminary notion of *belief beyond reasonable doubt*. A rational agent believes a proposition A beyond reasonable doubt when she is confident in it now and she anticipates no relevant possibility that would lower her confidence below $1/2$. We will come back to this notion later. But first, we apply our reading of Leitgeb's theory to the prisoner's example.

4. The Prisoner's Example Revisited

In the prisoner's example, we have the evidence that 99 out of 100 prisoners kill a prison guard. This 99:1 statistics says that each of the 100 prisoners may be innocent and we have no reason to believe that any one is more or less likely to be guilty than any of the others. That is, all 100 prisoners are serious candidates for killing the guard and the probabilistic evidence does not discriminate between those candidates. Crucially, we have a reference class of 100 prisoners whose members are all on a par concerning their probability of guilt. Hence, we represented the fact-finder's

[13] See Leitgeb (2013, fn. 26).

[14] We could also choose a threshold slightly below $P(B_W)$.

[15] A proposition A is consistent with her beliefs just in case $B_W \cap A \neq \emptyset$. This formula is equivalent to $B_W \not\subseteq W \setminus A$. As all believed propositions must be supersets of B_W, $\neg A \notin Bel$ iff $B_W \cap A \neq \emptyset$.

[16] Note that the fact-finder need not actually learn a proposition of which she now believes that she might learn it later. We are silent in this article on the related problem Bayesians face with learning conditional information (Günther & Trpin, 2023). For a possible solution, see Günther (2017, 2018, 2022) and Günther and Sisti (2022).

https://doi.org/10.1017/can.2025.6 Published online by Cambridge University Press

credences by a uniform distribution over the 100 mutually exclusive and jointly exhaustive possibilities of innocence.

On Leitgeb's theory, rational belief depends on how the relevant epistemic possibilities are partitioned. A partition Π on W is a set of pairwise disjoint non-empty subsets u_i of W such that $\cup u_i = W$. Propositions are understood relative to a partition Π: any proposition is a subset of Π. It is rational to believe a proposition relative to a given partition. As we will see shortly, rational belief is partition-sensitive, and so rational beliefs are not deductively closed across partitions. The logical closure of rational belief is valid only with respect to the same partition.

One can define a new probability distribution P_Π in terms of our uniform distribution P. The probabilities of the partition cells and of their unions are determined by P that is—unlike P_Π—defined for all subsets of W. Of course, given the 99:1 statistics, it is justified to choose the partition Π of relevant possibilities that correspond to W. The difference between this most fine-grained partition Π and W is negligible in the prisoner's example:

$$\Pi = \{\{w_1\}, \{w_2\}, \ldots, \{w_{100}\}\}.$$

Now, the only P_Π-stable proposition is W, and so $B_W = W$. The belief, for instance, that Prisoner 1 is guilty is not P_Π-stable. To see this, consider the probability that Prisoner 1 is guilty given that Prisoner 1 or Prisoner 2 is innocent:

$$P_\Pi(W \setminus \{w_1\} | \{w_1, w_2\}) = \frac{P(\{w_2\})}{P(\{w_1, w_2\})} = 1/2.$$

P_Π-stability requires that the probability of the proposition under consideration remains strictly over $1/2$. However, on our reading of Leitgeb's theory, if the fact-finder considers the possibility that Prisoner 1 or Prisoner 2 is innocent, her confidence that Prisoner 1 is guilty does not surpass $1/2$. A similar argument applies to each prisoner i for $1 \le i \le 100$. Hence, the finder should not believe that any prisoner is guilty.[17]

Unlike the simple threshold view discussed above, Leitgeb's theory says that the fact-finder should not believe that the randomly picked prisoner is guilty. More precisely, the fact-finder should believe that exactly one of the 100 prisoners is innocent and she should not believe that any prisoner i is guilty relative to the partition Π. The fact-finder's credence $P_\Pi(W \setminus \{w_i\}) = \frac{99}{100}$ of guilt is high enough for any prisoner, but not stably high. She is therefore not allowed to have a rational belief in the guilt of any particular prisoner.

Now, imagine Prisoner 1 is standing trial and we have no statistical evidence. Instead, an eyewitness comes forward and testifies that she saw Prisoner 1 joining in to kill the guard. She is so reliable, let us say, that her testimony makes it 99% likely that Prisoner 1 killed the guard. Unlike the 99:1 statistics, the eyewitness testimony is only about Prisoner 1—it is silent on the other 99 prisoners. This eyewitness evidence shifts the perspective. While the 99:1 statistics answers the question how likely it is that a randomly picked prisoner is guilty, the testimony answers the question whether or not Prisoner 1 killed the guard. For the latter question, there are only two relevant possibilities: Prisoner 1 killed the guard or else he did not. The testimony is either correct or not. This suggests that the testimony partitions all underlying possibilities in just two cells:

$$\Pi' = \{\{w_1\}, \{w_2, \ldots, w_{100}\}\}.$$

[17] In fact, only the belief in W is P_Π-stable. To see this, suppose for reduction that the fact-finder believes the proposition $\{w_1, \ldots, w_i\}$ for $1 \le i < 100$. The finder thus considers $\{w_i, \ldots, w_{100}\}$ to be possible. Now, the probability of $\{w_1, \ldots, w_i\}$ conditional on $\{w_i, \ldots, w_{100}\}$ is $1/(100-(i-1))$. The latter term is $\le 1/2$. However, by P_Π-stability, the fraction should be strictly $> 1/2$. Contradiction.

https://doi.org/10.1017/can.2025.6 Published online by Cambridge University Press

$\{w_2, \ldots, w_{100}\}$ represents that the witness's testimony is correct and Prisoner 1 killed the guard, while $\{w_1\}$ represents that the testimony is incorrect and Prisoner 1 is innocent. Now, there are two $P_{\Pi'}$-stable propositions, W and $\{w_2, \ldots, w_{100}\}$. To see why the latter is $P_{\Pi'}$-stable, consider the probability that Prisoner 1 is guilty given any proposition relative to Π':

$$P_{\Pi'}(W \setminus \{w_1\}|B) > 1/2 \text{ for all } \mathrm{B} \subseteq \Pi'$$
$$\text{such that } (\mathrm{W} \setminus \{\mathrm{w}_1\} \cap \mathrm{B}) \neq \varnothing \text{ and } \mathrm{P}(\mathrm{B}) > 0.$$

There are just two strict subsets B of Π'. $\{w_1\}$ is inconsistent with $W \setminus \{w_1\}$ and $P(W \setminus \{w_1\}|W \setminus \{w_1\}) = 1$. On Leitgeb's theory, it is permissible to pick $B_W = \{w_2, \ldots, w_{100}\}$ and so $P_{\Pi'}(B_W) = .99$ as a threshold for all-or-nothing belief. Hence, it is rationally permissible to believe that Prisoner 1 is guilty.[18] Based on the eyewitness evidence, the fact-finder may rationally believe that Prisoner 1 is guilty.[19] Rational belief in Prisoner 1's guilt is permissible relative to the eyewitness partition Π' but not relative to the partition Π induced by the statistical evidence. This partition sensitivity of rational belief shows that rational beliefs are not logically closed across partitions: a rational belief in Prisoner 1's guilt relative to Π' does not imply a rational belief in Prisoner 1's guilt relative to Π.

We have seen that different kinds of evidence may point to different partitions of the underlying set of possibilities. The uniform probability measure P over the fine-grained partition of the possibilities W induces a symmetry between the prisoners: each prisoner is just as likely as any other to be innocent. The 99:1 statistics does not probabilistically discriminate between this or that possibility. The innocence of each prisoner is a relevant epistemic possibility and the finder's uniform credences do not break the symmetry. We suggest it is this symmetry why it feels so random to convict one of the prisoners based on statistical evidence alone: looking at the probability values, it could likewise have been any other prisoner.[20]

The eyewitness testimony, by contrast, biases the fact-finder's credences toward Prisoner 1 being guilty. The reason is simply that the very coarse-grained partition allowing only the two possibilities that Prisoner 1 killed the guard, or else he did not, gives a rather strong indication of what to believe about Prisoner 1. The eyewitness evidence dissipates the air of randomness: the two possibilities are far from being equally likely.

The prisoner's example suggests a general distinction between *statistical* and *individual* evidence. Statistical evidence assigns uniform probabilities to a certain partition of possibilities. The more uniform the probability distribution is, the more statistical the evidence. A completely uniform distribution is purely statistical. Individual evidence counteracts the uniformity of statistical evidence. It may do so by partitioning the underlying possibilities such that the uniformity is broken. Thereby, it induces a probabilistic difference between partition cells. In brief, statistical evidence is uniform over certain possibilities, and individual evidence probabilistically discriminates between those possibilities.

Purely statistical evidence has it that all possibilities are on a par: no possibility is more likely than any other. This is behind the sentiment that all possibilities might be actual. Moreover, according to Leitgeb's theory, purely statistical evidence alone should never give rise to qualitative beliefs.

[18]Leitgeb (2013) recommends picking the strongest B_W. On this recommendation, it is rational simpliciter to believe that Prisoner 1 is guilty.

[19]If the fact-finder has the 99:1 statistics as additional evidence, the finder may believe that Prisoner 1 is guilty and must believe that one of the Prisoners 2–100 is innocent.

[20]Pritchard (2018) explains the feeling of randomness thus: it is an *easy* possibility that the prisoner standing trial is innocent. He casts easy possibilities in terms of closeness of possible worlds and avoids probabilities. We may stipulate the term for our account by using probabilities: a possibility is easy if it is at least as likely as any other relevant possibility. On this stipulation, it is an easy possibility that the prisoner standing trial is innocent given only the 99:1 statistics. Given only the eyewitness evidence, however, the prisoner being innocent is not an easy possibility.

Our simple account of probabilistic evidence explains the intuition that statistical evidence gives only general information about the members of a reference class and does not single out any member of the class as special. The 99:1 statistics, for instance, does not only pertain to the particular defendant standing trial but in the same way to any other prisoner in the yard. And so, it would be a distortion if no individual would be represented by its own partition cell. According to our simple account, individual evidence may break this uniformity over a partition. It may shift the focus on only a subset or even just one member against its wider reference class. This coarse-graining of the underlying possibilities may induce a probabilistic differential between the new partition cells. As a result of the coarse-graining, the probability distribution over the new partition may well become discriminatory. And so, qualitative belief may become rationally permissible.

5. Staffel's Uneven Statistics, Standards of Proof, and Fine-Graining

Staffel (2016) challenges Leitgeb's stability theory of belief. She claims that it "is irrational to hold outright beliefs based on purely statistical evidence" and observes that Leitgeb's theory does not rule out rational belief based on *uneven statistics* (p. 1725). According to her, a statistics is even if the probability is uniformly distributed over the considered possibilities, and uneven if the distribution is not uniform. The 99:1 participation statistics of the prisoner's example is even: it is equally likely that each prisoner is innocent. Of course, a statistics need not be even. Consider, for example, the following credences P over only four relevant possibilities:

$$P(\{w_1\}) = .6, \quad P(\{w_2\}) = .3, \quad P(\{w_3\}) = .09, \quad P(\{w_4\}) = .01.$$

The proposition $\{w_1\}$ is P-stable, and so it is rationally permissible to believe it on Leitgeb's theory. Staffel's observation is correct: stably high credences can be based on uneven statistics.

The uneven statistics can be interpreted as a four-ticket lottery. Only one of the four tickets wins and the chances of winning are given by P. In world w_i, ticket i wins for $1 \leq i \leq 4$. If we pick the strongest P-stable proposition $B_W = \{w_1\}$, it is rational to believe that Ticket 1 will win and the other tickets will lose. If we pick the next strongest P-stable proposition $B_W = \{w_1, w_2\}$, it is rational to believe that Ticket 1 or Ticket 2 will win, and the other tickets will lose, and so on. Rational belief in an "uneven lottery proposition" such as "Ticket 1 will win" is not excluded on Leitgeb's theory. And as "it is irrational to have outright beliefs in lottery propositions," she thinks beliefs in uneven lottery propositions are irrational (p. 1725). Hence, the stability theory of belief does not provide a theory of rational belief, or so argues Staffel.

Staffel (2016, p. 1729) tailors an alleged case of statistical evidence based on an uneven statistics. Let us follow her and tailor a case of statistical evidence that fits the probability distribution P. You face four people who have attended a football game. You are certain that one of the four people gatecrashed to watch the game, while the other three paid for their tickets. The only evidence you have is as follows: Person 1 sat in Section 1, where 60% of the visitors were gatecrashers; Person 2 sat in Section 2, where 30% of the visitors were gatecrashers; and so on. As the case is structurally indistinguishable from the uneven lottery, the stability theory allows us to rationally believe that Person 1 gatecrashed and the others did not. Staffel (2016, p. 1729) counters:

> [I]t would be irrational for you to form any outright beliefs about which person is or is not the fence-jumper, since the only available evidence is statistical evidence about the percentage of fence-jumpers in the section in which each person sat.

As a consequence, Staffel concludes that the stability theory of belief, at least on its own, cannot explain why it is irrational to believe propositions "based on purely statistical evidence."

We must wonder, however, whether there is anything wrong with rational belief based on uneven statistics. For Staffel, an uneven statistics provides only "purely statistical evidence."

https://doi.org/10.1017/can.2025.6 Published online by Cambridge University Press

However, an uneven statistics does not assign a uniform probability distribution to a certain partition of possibilities. So, on our account of evidence, an uneven statistics may count as individual evidence. To bring this point home, consider the following very simple and very uneven probability distribution P' over only two relevant possibilities:

$$P'(\{w_1\}) = .99, \quad P'(\{w_2\}) = .01$$

In the abstract, it does not seem irrational to believe the P'-stable proposition $\{w_1\}$. If we interpret w_i as the world where ticket i wins, you may believe of Ticket 1 that it will win and of Ticket 2 that it will lose. In fact, we already considered a structurally indistinguishable case. The 99%-reliable eyewitness in the prisoner's example induced a coarse-grained partition consisting of only two cells to which the same probability values were assigned. It is commonly agreed that this eyewitness evidence is individual.

Similarly, imagine you face two people who attended a soccer game and you believe for sure that one and only one of them gatecrashed. Person 1 sat in Section 1, where 99% of the people were gatecrashers; Person 2 sat in Section 2, where 1% of the people were gatecrashers. Would we still say that it is irrational to believe that Person 1 gatecrashed because the belief is "based on purely statistical evidence"?

As is typical in the literature, Staffel gives no criterion for distinguishing between purely statistical and individual evidence (Enoch & Spectre, 2019, pp. 183–184). Purely statistical evidence is introduced by pointing to cases like the prisoner's example. She cites no reason why her uneven statistics is "purely statistical" and why our uneven two-ticket lottery is or is not. In the latter case, it seems rationally permissible to believe the proposition that Ticket 1 wins (which is quite compatible with the stability theory). We do not share Staffel's conviction that it is *always* irrational to have all-or-nothing beliefs based on uneven statistics. In fact, we would deny that the "two-ticket lottery" P' represents purely statistical evidence. And if the uneven distribution P' is not purely statistical, in which sense can we say that the uneven distribution P in our four-ticket lottery is purely statistical?

On our simple account of evidence, both probability distributions, the "four-ticket lottery" P and the "two-ticket lottery" P', reflect individual evidence. They clearly discriminate between the probability of different possibilities and thereby violate our uniformity criterion of purely statistical evidence. With respect to P, the stability theory permits us to believe that Person 1 gatecrashed. But should you believe it? Or should you rather believe that Person 1 or 2 gatecrashed? Or should you only believe that Person 1 or 2 or 3 gatecrashed? All of these beliefs are rationally permissible. In terms of the stability theory, the question is what is the strongest P-stable proposition B_W you should pick. Moreover, the choice of B_W depends on which threshold is appropriate in certain contexts.

Gatecrashing is a matter of civil law, where the burden of proof is the evidential standard known as *preponderance of evidence*. It is typically understood as follows: a plaintiff's claim counts as proven in court just in case the claim is established to be more likely than not. A civil court should thus find a defendant liable if the probability that the defendant is guilty surpasses $1/2$ given the available and admissible evidence. Given the threshold $P(Guilt) > .5$ and that $Guilt$ is P-stable, it is rational to believe that the gatecrasher is guilty. Hence, the fact-finder should believe that Person 1 is guilty of gatecrashing. The question is, of course, whether the threshold of $1/2$ is really appropriate for the situation.

Murder is a matter of criminal law, where the burden of proof is the evidential standard known as *beyond reasonable doubt*. Let us say this standard requires the threshold $P(Guilt) > .9$ and that $Guilt$ is P-stable. Moreover, let us suppose we know that there is exactly one murderer and the probability of guilt for the four suspects is given by P. Well then, it is not rational to believe that Person 1 is the murderer. In the context of murder, our evidential standard is too cautious for such an outright belief. Given the threshold, we can only say that it is rational to believe that Person 1 or

https://doi.org/10.1017/can.2025.6 Published online by Cambridge University Press

2 or 3 is guilty of murder. This context sensitivity between gatecrashing and murder seems far from implausible.[21] We have just analyzed two standards of legal proof in terms of stably high credence.

Staffel (2016, p. 1731) has another challenge for Leitgeb's theory. The stability theory of belief is partition-sensitive. In fact, we have used this partition sensitivity to explain why it is not rational to believe a certain proposition based on our notion of statistical evidence, but it is rational to believe this very proposition based on what we call individual evidence. Staffel tries to turn the partition sensitivity against the stability theory. Recall that $\{w_1\}$ is P-stable: it is rationally permissible to believe that Person 1 gatecrashed. But now, suppose that our fact-finder considers that the flip of a fair coin landed heads (h) or tails (t). As the coin flip is irrelevant to the gatecrashing, it should not make any difference to what a rational agent believes, or so argues Staffel (p. 1732).

The coin flip is—by assumption—irrelevant to the fact-finder's beliefs and so independent of her credences in the other propositions. Still, merely considering the coin flip results in a new, more fine-grained partition:

$$P(\{w_{1h}\}) = .3, \quad P(\{w_{2h}\}) = .15, \quad P(\{w_{3h}\}) = .045, \quad P(\{w_{4h}\}) = .005$$
$$P(\{w_{1t}\}) = .3, \quad P(\{w_{2t}\}) = .15, \quad P(\{w_{3t}\}) = .045, \quad P(\{w_{4t}\}) = .005$$

$P(\{w_{1h}\}) = .3$, for instance, denotes the probability that Person 1 gatecrashed and the coin landed heads.

The fine-graining results in a loss of rational belief. While $\{w_1\}$ is P-stable, $\{w_{1h}, w_{1t}\}$ is not. To see this, consider the conditional probability that Person 1 gatecrashed, given that Person 1 did not gatecrash *or* the coin landed tails:

$$P(\{w_{1h}, w_{1t}\} | W \setminus \{w_{1h}\}) = \frac{P(\{w_{1t}\})}{P(W \setminus \{w_{1h}\})} = \frac{3}{7} < 1/2.$$

Considering the coin flip makes a difference: it is not rationally permissible anymore to believe that Person 1 gatecrashed. And so, "the stability theory must reject the intuition that considering irrelevant propositions should not change our rational beliefs," as Staffel (2016, p. 1732) points out.

From a purely formal perspective, Staffel's challenge of fine-graining stands. The stability theory of belief is prone to a loss of rationally permissible belief when moving to a more fine-grained partition—even if the fine-graining is merely a result of considering irrelevant propositions. On the other hand, it is intuitively questionable why a judge or jury would consider the possibility that Person 1 did not gatecrash *or* the coin landed tails. After all, the coin toss is assumed to be irrelevant for the beliefs about gatecrashing. This suggests the following fix for the example at hand: coarse-grain the underlying partition such that we are de facto back to the "four-ticket lottery."[22]

In general, we must wonder whether a purely formal notion of rationality is enough in our legal cases. We have just seen that the two sets of rationality norms can and should be complemented by a probabilistic threshold that is appropriate to the more specific legal context. Perhaps, we need to go beyond the notion of being doubly rational to a more substantive notion of rationality.

On a more substantive notion of rationality, it becomes questionable why a rational agent would consider possibilities (and propositions) irrelevant for the issue at hand. Would we really consider a fact-finder—who includes an irrelevant fair coin flip in her deliberations about guilt—to be

[21] There is a well-known argument that delivers the optimal threshold for belief depending on the stakes involved. See Kaplan (1968), Cheng (2013, pp. 1259–1261 and 1275–1278), and Günther (2024).

[22] Here is a tentative idea of how one could obtain the coarse-grained partition of relevant propositions: include only the propositions that are direct reasons for believing some target proposition. This idea could be implemented by representing the fact-finder's credence function by a Bayesian network. You then believe a proposition A iff A is entailed by your strongest belief that is P-stable with respect to the partition determined by A and its parent variables in the network. The Markov condition ensures that the other propositions are irrelevant.

substantively rational? We do not think so. It seems almost to be a conceptual truth that a rational fact-finder should only consider propositions that are relevant for the current deliberation. In fact, it seems rational in a substantive sense to abstract away from any irrelevant propositions or possibilities. Typically, it is rational to deliberate on the most coarse-grained level, which does not omit any relevant possibility. Moreover, some possibilities are relevant for a question under discussion—their truth or falsity impacts the probability of the answer—and yet, it is unreasonable to consider them. A case in point is the skeptical possibility that I am a brain in a vat. If true, I did not shoot Jill—simply because I do not have any hands or the like to carry out the deed. Being substantively rational means to abstract away from such unreasonable possibilities as well.

Finally, it is prima facie substantively rational to compartmentalize a more complex deliberation into many sub-deliberations, which involve fewer propositions. Otherwise, it is hard to see how we can, based on probabilistic evidence, hold outright beliefs that, in turn, figure as explicit premises in the argument following the current one. Without such a modularized sequence of arguments, legal decision-making will not be readily intelligible to all.

It is, of course, hard to say when a possibility is relevant or reasonable and when not.[23] However, recall that Staffel *assumes* that the flip of the fair coin is irrelevant for whether or not a person gatecrashed. So a substantively rational fact-finder would abstract away from the coin flip, and this particular instance of the problem of fine-graining vanishes for the stability theory. More generally, the stability theory in conjunction with a more substantial rationality norm about considering only relevant and reasonable possibilities would explain a great deal of arguments in a court of law. After all, disputes about which pieces of evidence and which possibilities are relevant and reasonable often take center stage in legal proceedings.

In sum, we have defused Staffel's challenges for our account of rational belief. Her "purely statistical" evidence represented by uneven statistics may well be individual evidence on our account of evidence, and so allow for rational belief. Her complaint that merely considering irrelevant propositions may result in a loss of previously rational belief is correct. In response to this perhaps more serious worry, we have complemented the stability theory by a substantive rationality norm about relevance, roughly speaking: abstract away from irrelevant and unreasonable propositions and keep the deliberations simple. This solves the problem of fine-graining due to irrelevant and unreasonable propositions. The underlying idea is simply that a substantively rational fact-finder would neither consider irrelevant nor unreasonable propositions in the first place. The complemented stability theory explains why legal arguments about which possibilities are taken to be relevant and reasonable are commonplace. The stability theory of belief seems to be in a rather strong position when amended with the substantive rationality norm and a contextually appropriate threshold: it can explain why we have our intuitions about statistical and individual evidence, why these intuitions are rational, and it gives rise to standards of legal proof in terms of rational belief.

6. Comparison to Moss's Knowledge of Guilt

We have argued that legal proof is tantamount to *rational belief* of guilt. A defendant should be found guilty just in case it is rational for the fact-finder to believe that the defendant is guilty. We have spelled out how we understand rational belief. It is belief according to Leitgeb's stability theory, plus a contextually determined threshold and a more substantive rationality norm of relevance and reasonableness. In brief, a rational agent believes a proposition $A \subseteq \Pi$ iff $B_W \subseteq A$ and B_W is P_Π-stable, where B_W is the agent's strongest believed proposition and P_Π her credence function is defined over a partition Π of reasonable and relevant possibilities.

[23] Günther (2024) develops a general theory of what possibilities should and should not be considered as relevant in light of the available evidence. Our account of rational belief can be amended by this theory of evidence.

Moss (2021) argues for a similar account. She defends the thesis that legal proof is tantamount to *knowledge* of guilt. "Conviction [in a criminal trial]," she says, "requires proving beyond a reasonable doubt that the defendant is guilty, and this conclusion is proved if and only if the judge or jury knows it." (p. 2) On her view, a defendant is to be found guilty just in case the fact-finder *knows* that the defendant is guilty.[24]

According to Moss, it "is widely agreed that the merely statistical evidence in *Prisoners* cannot sustain a verdict of guilt." (p. 1) But why should the randomly picked prisoner not be found guilty? We would say because it is not rational for the fact-finder to believe that he is guilty. In addition, the belief is not rational because of the symmetry induced by the uniform probability distribution. On Moss's account, by contrast, the fact-finder does not know that the randomly picked prisoner is guilty, and therefore he should not be found guilty. Moreover, the finder does not know because she cannot rule out the possibility that the defendant is the innocent prisoner.

What is knowledge according to Moss (2021)? She adheres to the knowledge-first epistemology championed by Williamson (2000). Knowledge-firsters think that knowledge is unanalyzable, more fundamental than belief, and more important than belief. It is, therefore, unsurprising that she does not analyze knowledge in terms of belief plus other ingredients. A bit more surprising is that she does not explain how we come to know a proposition based on our evidence. This being said, Moss draws on Lewis's (1996) account of elusive knowledge. An agent knows a proposition A iff her truthful evidence eliminates any possibility in which $\neg A$. *Any* possibility? Even the most far-fetched $\neg A$-possibilities that arise only from considering conspiracy theories? There are virtually always uneliminated possibilities of error lurking. If those error possibilities were relevant, we would hardly ever know anything. However, Lewis and Moss want to say that we know a lot. After all, Moss wants to say that every legitimate legal conviction in a criminal trial is based on knowing the accused's guilt. Hence, we need to ignore a fair deal of the many uneliminated error possibilities. Lewis says that these ignored error possibilities "are outside of the domain" of *any*, "they are irrelevant to the truth of A" (p. 553). In an attempt to whisper that which must remain unmentioned, Lewis explicitly restricts the domain of "any" in his definition. An agent knows A iff her truthful evidence eliminates any possibility in which $\neg A$—Psst!—except for those possibilities we are ignoring.

Which uneliminated $\neg A$-possibilities may not be ignored? Which ones are the relevant alternatives? Lewis (1996, pp. 554–567) attempts to give a general account of relevant possibilities. Among other criteria, there is the Rule of Belief. A possibility is relevant if a rational agent assigns it a sufficiently high credence, and not just because the possibility is unspecific.

How high is "sufficiently high"? It depends on how much is at stake. As Lewis (1996, p. 556) puts it: "When error would be especially disastrous, few possibilities may be properly ignored. Then even quite a low degree of belief may be "sufficiently high."" The stakes, and more generally the epistemic context, determine in part which alternatives are considered relevant. If you attend to an uneliminated $\neg A$-possibility, however far-fetched, you cannot know that A. I know, for example, that I have two hands. However, consider the possibility that I am a *brain-in-a-vat*: my experience is just as it is but I do not have hands. My knowledge dissolves in face of such a skeptical possibility. For knowledge is infallible: I know a proposition only if I am not aware of any error possibility. In

[24]Moss (2021, p. 23) modifies her account for civil lawsuits, where the evidential standard is preponderance of the evidence. This burden of proof does not require for conviction that the fact-finder knows that the defendant is liable. She rather claims that a civil penalty requires that the fact-finder knows that the defendant is *probably* liable (Moss, 2013, 2018). However, her account delivers only the desired results because of a substantive *rule of consideration*: "in many situations where you are forming beliefs about a person, you morally should keep in mind the possibility that they might be an exception to statistical generalizations" (p. 221). So, you do not even know that one of the prisoners is *probably liable* because you cannot rule out the possibility that this prisoner standing trial is "an" exception to the statistical generalization that 99 out of 100 prisoners killed the guard—you cannot rule out the possibility that he is 100% innocent. Our account has no need for such a rule. For further discussion of the rule of consideration, see Smartt (2022).

https://doi.org/10.1017/can.2025.6 Published online by Cambridge University Press

general, consider previously ignored uneliminated possibilities of error and your (Lewisian) knowledge vanishes. It is elusive.

We have just seen that bringing far-fetched brain-in-a-vat possibilities into a deliberation risks a loss of Lewisian knowledge. Unlike Lewis, Moss uses a notion of knowledge that is protected against such unreasonable doubts. For her, knowledge is tantamount to a proof beyond reasonable doubt. In particular, proving a defendant's guilt beyond reasonable doubt is what it means to know it. Thus, her knowledge is only elusive in the face of reasonable error possibilities, not in the face of unreasonable ones. The unreasonable ones must be properly ignored, even if we are aware of them.

The standard of proof beyond reasonable doubt is usually understood as a means to protect the accused against hasty conviction. Any reasonable doubt needs to be dispelled. Less salient but equally important is that the standard also deters the fact-finder from considering unreasonable possibilities. There will virtually always be far-fetched possibilities that a fact-finder should set aside, for instance, brain-in-a-vat possibilities. The standard is meant to exclude such unreasonable doubts. Its function is to constrain the deliberation context of a fact-finder to those possibilities that are reasonable to consider.

In a court of law, advocates pursue sometimes the strategy to cast doubt on the defendant's guilt by calling the fact-finder's attention to far-fetched possibilities of error. The fact-finder then has a decision to make: is the error possibility reasonable? If not, the error possibility is disregarded; otherwise, it is considered to be relevant and may cast reasonable doubt on the defendant's guilt. As there will virtually always be unreasonable error possibilities, the fact-finder will hardly ever know that the defendant is guilty. And yet, depending on the fact-finder's decision, the beyond reasonable doubt standard either guards her deliberation against unreasonable error possibilities, or else protects the defendant from being too hastily convicted. However, no matter how she decides, her Lewisian knowledge of guilt vanishes in light of the presented error possibilities. So, as long as an advocate pursues the strategy of pointing out such possibilities, Lewis must agree that a fact-finder cannot know whether or not a defendant is guilty. Like all Lewisian knowledge, knowledge of guilt is elusive and thus inappropriate as a criterion for conviction. By contrast, Mossian knowledge is only elusive with respect to reasonable possibilities. As a consequence, her notion of knowledge is fallible: there may be uneliminated error possibilities, which have been classified as unreasonable but may be true, and yet she speaks of knowledge.[25]

Our substantively rational belief is fallible like Mossian knowledge. We have said that it is substantively rational to abstract away from irrelevant and unreasonable possibilities. A substantively rational belief may, of course, be false if an ignored unreasonable possibility happens to be true. However, rational belief is supposed to be fallible, unlike factive notions of knowledge.

Similar to Mossian knowledge, substantively rational belief is elusive. It may vanish by considering further possibilities as relevant and reasonable, which amounts to a fine-graining of the underlying partition of possibilities (see Sections 4 and 5). However, a substantively rational agent only considers a very fine-grained partition of possibilities when she considers a great number of distinctions to be reasonable and relevant. The greater the number of distinctions is, the more fine-grained the partition, and the higher the credence necessary for rational belief. To be precise, suppose the strongest believed proposition B_W is a union of many very fine-grained "partition cells," or better possibilities, w. For any $w \in B_W$, the proposition $\{w\} \cup \neg B_W$ is then epistemically possible. Hence, a belief in some proposition $A \supseteq B_W$ would require to have a stably high probability in A conditional on any $\{w\} \cup \neg B_W$. In such a context, where destabilizing error possibilities are lurking everywhere, stably high credence requires (near) maximal credence.

Furthermore, we agree with Moss that high credence of guilt is not sufficient for conviction. From there, however, she jumps to the conclusion that the "criminal standard of proof cannot be

[25]This argument applies mutatis mutandis to Moss's modified account for civil lawsuits, where legal proof requires knowledge of probable liability.

https://doi.org/10.1017/can.2025.6 Published online by Cambridge University Press

defined in terms of any threshold notion of confidence"[26] (p. 11). Pace Moss, we have proposed such a schema for standards of proof in Section 5: a proposition A meets a standard of proof iff A is P-stable and $P(A) > s$, where s is a threshold appropriate for the standard at hand. At the end of Section 3, we have already given a preliminary notion of believing beyond reasonable doubt: a rational agent believes now a proposition A beyond reasonable doubt when she is confident in it now and she anticipates no *relevant* possibility that would lower her confidence below a certain threshold. The final notion requires in addition the substantive norm on reasonableness and relevance, and to fix an appropriate threshold. To believe beyond reasonable doubt that the defendant is guilty is thus identified with having a rational belief in his guilt in a criminal trial. By adjusting the threshold, we obtain another evidential standard: to believe by preponderance of evidence that the defendant is liable is tantamount to rational belief in his liability in a tort case.

There is a problem for Moss's account to which ours is not susceptible. Mossian knowledge is supposed to be factive, even though it is fallible. Only true beliefs may count as knowledge. If a defendant is in fact innocent, you can never know that she is guilty.[27] Hence, an innocent defendant can never be legitimately convicted on Moss's account. Her knowledge account does not allow for wrongful convictions that are legitimate. The impossibility of wrongful but legitimate convictions is hard to swallow. To see this, consider a legal case where the defendant is innocent and, yet, there is compelling but misleading evidence that the defendant is guilty. As unfortunate as it is, it is rational to wrongfully convict the defendant (Blome-Tillmann, 2015). And indeed, rational but wrongful convictions based on misleading evidence exist. So, it is a problem for Moss that she cannot account for wrongful but legitimate convictions.[28]

The criterion of rational belief allows, of course, for legitimate wrongful convictions. If the evidence clearly points to the guilt of a defendant, it is rational for a fact-finder to (falsely) believe that the defendant is guilty, and so to find the defendant guilty, independently of whether the defendant is in fact guilty or innocent. In a way, the whole problem a fact-finder faces is that she *cannot know* whether or not the defendant under consideration is guilty. This is why Lewis (1996, p. 560) points out that

> what matters most to us jurors is not whether we can truly be said to know; what really matters is what we should believe to what degree, and whether or not we should vote to convict.

We have proposed an account that explains "what we should believe" based on "what we should believe to what degree," and "whether or not we should vote to convict." As compared to knowledge, our account requires merely rational belief of guilt. Even if no unreasonable possibilities come up in a legal trial, knowledge of guilt requires ruling out *all* relevant possibilities in which the defendant is innocent. By contrast, our rational belief of guilt only requires a partition of relevant possibilities such that the defendant's guilt is P-stable and above a contextually determined threshold. It does not require ruling out all relevant possibilities, such as the possibility that the prisoner is innocent in the eyewitness example. Our rational belief of guilt merely requires that this possibility is, to a certain extent, less likely than the one of the prisoner's guilt.

In sum, our account shares many features and merits of Moss's. However, ours is not susceptible to the problem of legitimate wrongful convictions. In addition, while Moss is silent on how evidence that may come in probabilistic form relates to knowledge, we have worked out how probabilistic evidence relates to rational belief. Finally, a fact-finder may, on our account, rationally believe that a

[26]And Moss is not alone. Recall, for example, what Buchak (2014, p. 291) says in fn. 7.

[27]Moss (2021, p. 28) points out that knowledge of probable liability is also "factive": "The civil standard of proof by a preponderance of the evidence is also factive, in the sense that a defendant cannot be proved probably liable unless the defendant is probably liable."

[28]The problem of wrongful convictions does not apply to Blome-Tillmann's (2017) account. A fact-finder's evidential probability that she knows that a defendant is guilty or liable may be sufficiently high for conviction even if the defendant is innocent.

https://doi.org/10.1017/can.2025.6 Published online by Cambridge University Press

defendant is liable in a tort case, and yet not know it. So when Moss (2021, p. 21) says that "legal proof seems to require *something that looks an awful lot like knowledge*," we are inclined to answer: "yes, it requires rational belief."[29]

7. Conclusion

We have defended the thesis that a defendant should be found guilty just in case it is rational for the fact-finder to believe that the defendant is guilty. The fact-finder's belief is understood to be doubly rational: she should have rational credences, because evidence comes in probabilistic form, and she should have rational all-or-nothing beliefs, because her verdict is binary. Such a notion of doubly rational belief has been put forth by Leitgeb (2014b): rational belief is stably high credence. Where the probability distribution P represents the fact-finders credences, she rationally believes a proposition A, iff A is entailed by the strongest and P_{Π}-stable proposition B_W.

On this picture, a high credence in the defendant's guilt is necessary for rational belief of guilt, but not sufficient. A high probability of guilt may not be stable. One reason for instability is that the probability distribution is uniform over the relevant possibilities, which is the case in paradigmatic cases of statistical evidence. Furthermore, belief may be unstable—despite high credence—when the agent considers many distinctions to be relevant. A fine-graining of the underlying partition Π of possibilities may thus induce a loss of formerly rational belief. Rational belief is elusive.

Staffel (2016) takes the elusiveness of rational belief to challenge Leitgeb's stability theory. In response, we amended the notion of rational belief by a substantive norm: roughly, abstract away from irrelevant and unreasonable possibilities and keep the deliberations simple. Together with contextually determined probabilistic thresholds, the notion of substantively rational belief gives rise to an analysis of different standards of proof. Belief beyond reasonable doubt amounts to rational belief in a criminal trial. Belief by preponderance of evidence is rational belief in a civil trial.

Our account of legal proof justifies the following intuition: even if the probability of guilt is very high, a defendant should not be found guilty based on statistical evidence alone, while a defendant may well be found guilty based on individual evidence alone. Blome-Tillmann (2017) thinks this intuition corresponds to extant legal practice. If so, our account can be read as a justification of the current standards of legal proof. Ross (2021), by contrast, argues that current legal practice *sometimes* allows for conviction based on purely statistical evidence. If he is right and shares our understanding of statistical evidence, our account can be read as a proposal to revise those standards.

In sum, we have examined the idea that legal proof is tantamount to rational belief of guilt. Along the way, we proposed a simple distinction between statistical and individual evidence, and analyzed evidential standards in terms of rational belief. While our account shares the merits of Moss's (2021) account of legal proof, it avoids the pitfalls of hers. Rational belief of guilt is a less demanding requirement than Mossian knowledge of guilt, and so allows for wrongful but legitimate convictions. In addition, unlike Mossian knowledge, our criterion of rational belief of guilt has a straightforward connection to probabilistic evidence. Therefore, we hope that we are rationally permitted to believe that rational belief is a better criterion for legal proof than knowledge. However, we will never know it.

Acknowledgments. The author would like to thank Katie Steele, Philip Pettit, Alan Hájek, Pamela Robinson, Will Bateman, Hannes Leitgeb, Christian List, Richard Bradley, Lewis Ross, Anna Mahtani, Atoosa Kasirzadeh, Michael Blome-Tillmann, Jennifer Lackey, Anne Ruth Mackor, Christian Dahlmann, Adam Bjorndahl, Kevin Kelly, Teddy Seidenfeld, and the two anonymous referees for their helpful comments improving this work. The author is grateful for having had many opportunities to present parts of this work, namely at the Pittsburgh Formal Epistemology Workshop at Carnegie Mellon University, the North Sea Group Online Seminar hosted by the Universities of Groningen and Lund, the Workshop on Theories of Legal Proof

[29] See also Moss (2018, pp. 206–8).

https://doi.org/10.1017/can.2025.6 Published online by Cambridge University Press

at the Catholic Academy in Bavaria, the Center for Advanced Studies and the Research Seminar of Action and Decision Theory of LMU Munich, the Cologne Center for Contemporary Epistemology and the Kantian Tradition of the University of Cologne, the Central Division Meeting of the American Philosophical Association in 2023, the 11th International Congress of the Society of Analytic Philosophy at the Humboldt University Berlin, the 96th Joint Session of the Aristotelian Society and the Mind Association at the University of St. Andrews, the Philosophy of Science Conference at the Inter-University Center Dubrovnik, the Center for Logic, Language, and Cognition of the University of Turin, and the Humanising Machine Intelligence seminar at the Australian National University. The author would like to thank all of the wonderful audiences for their excellent and stimulating feedback.

Funding statement. The publication was supported by a Junior Residency in the Center for Advanced Studies and LMUexcellent, funded by the Federal Ministry of Education and Research (BMBF) and the Free State of Bavaria under the Excellence Strategy of the Federal Government and the Länder.

Mario Günther is an assistant professor at the Munich Center for Mathematical Philosophy of LMU Munich and a visiting assistant professor at the Philosophy Department of Carnegie Mellon University. His research focuses on Epistemology, Metaphysics, and the Philosophies of Science, Language, Law, and Artificial Intelligence. He received the 2022 Wolfgang Stegmüller Award for outstanding contributions to Analytic Philosophy.

References

Blome-Tillmann, M. (2015). Sensitivity, causality, and statistical evidence in courts of law. *Thought: A Journal of Philosophy*, 4(2), 102–112.

Blome-Tillmann, M. (2017). 'More likely than not' knowledge first and the role of bare statistical evidence in courts of law. In J. Adam Carter, E. C. Gordon, & B. Jarvis (Eds.), *Knowledge first: Approaches in epistemology and mind*. Oxford University Press.

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

Cheng, E. (2013). Reconceptualizing the burden of proof. *Yale Law Journal*, 122(5), 1254–1279.

Christensen, D. (2004). *Putting logic in its place: Formal constraints on rational belief*. Oxford University Press.

Di Bello, M. (2019). Trial by statistics: Is a high probability of guilt enough to convict? *Mind*, 128(512), 1045–1084.

Douven, I. (2002). A new solution to the paradoxes of rational acceptability. *British Journal for the Philosophy of Science*, 53(3), 391–410.

Enoch, D., & Fisher, T. (2015). Sense and 'sensitivity': Epistemic and instrumental approaches to statistical evidence. *Stanford Law Review*, 67, 557–611.

Enoch, D., & Spectre, L. (2019). Sensitivity, safety, and the law: A reply to Pardo. *Legal Theory*, 25(3), 178–199.

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

Gardiner, G. (2018). Legal Burdens of proof and statistical evidence. In J. Chase, & D. Coady (Eds.), *The Routledge handbook of applied epistemology*. Routledge.

Günther, M. (2017). Learning conditional and causal information by Jeffrey imaging on Stalnaker conditionals. *Organon F*, 24(4), 456–486.

Günther, M. (2018). Learning conditional information by Jeffrey imaging on Stalnaker conditionals. *Journal of Philosophical Logic*, 47(5), 851–876.

Günther, M. (2022). Causal and evidential conditionals. *Minds and Machines*, 32(4), 613–626.

Günther, M. (2024). Epistemic sensitivity and evidence. *Inquiry*, 67(6), 1348–1366. https://doi.org/10.1080/0020174X.2021.1936158.

Günther, M., & Sisti, C. (2022). Ramsey's conditionals. *Synthese*, 200(2), 165.

Günther, M., & Trpin, B. (2023). Bayesians still don't learn from conditionals. *Acta Analytica*, 38(3), 439–451.

Günther, M. (2024). Legal proof should be justified belief of guilt. *Legal Theory*, 30, 129–141. https://doi.org/10.1017/S1352325224000089.

Haack, S. (2014). *Evidence matters: Science, proof, and truth in the law*, Law in Context. Cambridge University Press.

Hedden, B., & Colyvan, M. (2019). Legal probabilism: A qualified defence. *Journal of Political Philosophy*, 27(4), 448–468.

Hempel, C. G. (1962). Deductive-nomological vs statistical explanation. In H. Feigl, & G. Maxwell (Eds.), *Minnesota studies in the philosophy of science* (Vol. 3, pp. 98–169). University of Minnesota Press.

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

Kyburg, H. E. J. (1961). *Probability and the logic of rational belief*. Wesleyan University Press.

Leitgeb, H. (2013). Reducing belief simpliciter to degrees of belief. *Annals of Pure and Applied Logic*, 164(12), 1338–1389.

Leitgeb, H. (2014a). The review paradox: On the diachronic costs of not closing rational belief under conjunction. *Noûs*, 48(4), 781–793.

Leitgeb, H. (2014b). The stability theory of belief. *The Philosophical Review*, 123(2), 131–171.

Leitgeb, H. (2015). I The Humean thesis on belief. *Aristotelian Society Supplementary*, 89(1), 143–185.

Leitgeb, H. (2017). *The stability of belief*. Oxford University Press.
Levi, I. (1967). *Gambling with truth*. MIT Press.
Lewis, D. (1996). Elusive knowledge. *Australasian Journal of Philosophy*, 74(4), 549–567.
Moss, S. (2013). Epistemology formalized. *Philosophical Review*, 122(1), 1–43.
Moss, S. (2018). *Probabilistic knowledge*. Oxford University Press.
Moss, S. (2021). Knowledge and legal proof. In *Oxford studies in epistemology*. Oxford University Press.
Nesson, C. R. (1979). Reasonable doubt and permissive inferences: The value of complexity. *Harvard Law Review*, 92(6), 1187–1225.
Pardo, M. S. (2018). Safety vs. sensitivity: Possible worlds and the law of evidence. *Legal Theory*, 24(1), 50–75.
Pritchard, D. (2015). Risk. *Metaphilosophy*, 46(3), 436–461.
Pritchard, D. (2018). Legal risk, legal evidence and the arithmetic of criminal justice. *Jurisprudence*, 9(1), 108–119.
Redmayne, M. (2008). Exploring the proof paradoxes. *Legal Theory*, 14(4), 281–309.
Ross, L. (2021). Rehabilitating statistical evidence. *Philosophy and Phenomenological Research*, 102(1), 3–23.
Smartt, T. (2022). Reconsidering the rule of consideration: Probabilistic knowledge and legal proof. *Episteme*, 19, 303–318. https://doi.org/10.1017/epi.2020.28.
Smith, M. (2010). What else justification could be. *Noûs*, 44(1), 10–31.
Smith, M. (2018). When does evidence suffice for conviction? *Mind*, 127(508), 1193–1218.
Smith, M. (2020). Against legal probabilism. In J. Robson, & Z. Hoskins (Eds.), *Truth and trial*. Routledge.
Smith, M. (2022). The hardest paradox for closure. *Erkenntnis*, 87(4), 2003–2028. https://doi.org/10.1007/s10670-020-00287-4.
Staffel, J. (2016). Beliefs, buses and lotteries: Why rational belief can't be stably high credence. *Philosophical Studies*, 173(7), 1721–1734.
Steele, K., & Colyvan, M. (2023). Meta-uncertainty and the proof paradoxes. *Philosophical studies*, 180(7), 1927–1950.
Thomson, J. J. (1986). Liability and individualized evidence. *Law and Contemporary Problems*, 49(3), 199–219.
Titelbaum, M. G. (2020). The stability of belief: How rational belief coheres with probability, by Hannes Leitgeb. *Mind*, 130, 1006–1017. https://doi.org/10.1093/mind/fzaa017.
Wheeler, G. (2007). A review of the lottery paradox. In W. Harper, & G. Wheeler (Eds.), *Probability and inference: Essays in honour of Henry E. Kyburg, Jr* (pp. 1–31). King's College Publications.
Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.

Cite this article: Günther, M. 2024. Probability of Guilt. *Canadian Journal of Philosophy* 54: 189–206, doi:10.1017/can.2025.6