Quaestio facti. Revista Internacional sobre Razonamiento Probatorio
Quaestio facti. International Journal on Evidential Legal Reasoning
Sección: Conjeturas y refutaciones
2026 | 11 pp. 1-18
Madrid, 2026
DOI: 10.33115/udg_bib/qf.i11.23280
Marcial Pons Ediciones Jurídicas y Sociales

ISSN: 2604-6202
Recibido: 07/05/26 | Aceptado: 16/06/26 | Publicado *online*: __/__/2026

BAYESIAN DECISION THEORY CAN GUIDE LEGAL FACTFINDING

Mario Günther*
Panthéon Sorbonne

ABSTRACT: I argue that Bayesian decision theory can guide legal factfinding. I do so by offering an account of legal proof on which judges should minimize expected justice costs. My account entails a judge's credence threshold for finding guilty and his prior credence of guilt. Hence, it can guide a judge in his decision based on the lawful evidence presented at trial—unlike the Bayesian model Mackor presents.

KEYWORDS: Philosophy of Law, Legal Proof, Bayesianism, Decision Theory, Retributive Justice, Undeserved Punishment.

* Please contact Mario.Gunther@univ-paris1.fr for any inquiries. ORCID: https://orcid.org/0000-0001-6208-448X.

1. INTRODUCTION

Mackor (2026) asks whether 'the' Bayesian model can and should guide the judicial evaluation of evidence in criminal cases as a whole. She focuses on Dutch criminal law, which allows a judge to convict a defendant for an offense only if the judge is convinced that the defendant committed it in light of the lawful evidence presented at trial. However, a judge may not convict even if the presented lawful evidence convinces him that the defendant is guilty. Mackor points out that this freedom of choice leaves the judge "helpless" in his decision (p. 364). For he lacks further guidance on whether or not he should convict the defendant. Dutch criminal law leaves open when the presented lawful evidence is sufficient for a finding of guilt. It leaves open how the judge should decide.

The Bayesian decision criterion is often conceptualized as a threshold for the posterior probability that the defendant committed the alleged offence—the final probability that the defendant is guilty after learning all the lawful evidence presented at trial (Günther & Friedrich, 2026). However, a judge has only a posterior of guilt if he starts out with a prior probability of guilt—the probability of guilt with which a judge begins to evaluate the lawful evidence received at trial. Mackor suggests that the judge's prior of guilt could be based on some reference class. This poses the question what reference class of the many possible ones should be used to estimate the prior (fn.4). If there is no suitable reference class, she suggests relegating the task of determining the prior of guilt to "the court"—without giving any further guidance (pp. 366-7&fn. 11).

Surprisingly, Mackor does not offer a Bayesian standard of proof in terms of a probability threshold. She neither says when a final probability of guilt suffices for finding a defendant guilty, nor what a judge's prior of guilt should be in general. The Bayesian model Mackor presents leaves open how a judge should decide. As it stands, the model cannot guide judges in their decisions.

Here I argue that Bayesian decision theory can guide a judge's factfinding. I do so by offering a decision-theoretic account of legal proof which does not leave open how a judge should decide. The account is based on the heart of Bayesian decision theory—the principle of expected utility maximization (EUM). But it goes beyond the formalism of decision theory by making a substantive assumption about the justice costs of the possible trial outcomes. The normative assumption entails with the principle of EUM a probability threshold for finding guilty from which I derive a prior probability of guilt. The result is a Bayesian account for how a judge should decide in criminal trials. The account shows that Bayesian decision theory can guide legal factfinding in principle.

The plan is as follows. Section 2 explains how the principle of EUM can figure as the decision criterion in legal factfinding. Section 3 argues for an assignment of utilities or justice costs to the possible trial outcomes based on the substantive assumption that a judge's decision-making only aims at establishing retributive justice and avoiding undeserved punishment. The result is a simplified probability threshold for finding guilty in terms of the severity of undeserved punishment. Section 4 outlines what probability assessments a judge should have based on the total lawful evidence presented at trial, how a judge's prior of guilt should be determined, and how strong the total lawful evidence must be for a conviction. This last step allows us to complete the model Mackor presents such that it coincides with ours, and so does not leave open how a judge should decide.

2. THE PRINCIPLE OF EXPECTED UTILITY MAXIMIZATION

How should a judge decide? Bayesian decision theory answers that any agent should maximize her expected utility (Jeffrey, 1983; Ramsey, 1926; Savage, 1972; von Neumann & Morgenstern, 1944). This principle of expected utility maximization (EUM) says that the choices of an agent should solely depend on her value and probability assessments. The value assessments of an agent are her preferences. The probability assessments of an agent are her degrees of belief, or credences for short. If an agent's preferences satisfy certain rationality axioms, her preferences can be represented by a utility function and her credences by a probability function. A rational agent's expected utility of a choice option is the sum of her utilities of the possible outcomes weighted by her credences that the respective outcomes are true. The principle of EUM says that a rational agent chooses an option that has the highest expected utility among the available options.

Let us consider the decision situation a rational judge faces in a criminal trial as a whole. As Mackor (2026, p. 368) says, the judge needs to decide between the ultimate hypothesis *G* that the defendant is guilty and the ultimate hypothesis *I* that the defendant is not guilty or 'innocent' based on the lawful evidence presented at trial. The two ultimate hypotheses are negations of each other, and so they mutually exclude each other and jointly exhaust the space of logical possibilities. The judge has the options of finding guilty (FG) or not (FI). Hence, the possible outcomes of the trial are a true finding of guilt (*TG*), a false finding of guilt (*FG*), a true finding of not guilty (*TI*), and a false finding of not guilty (*FI*). The judge's decision can be summed up in the following decision table:

	guilty (*G*)	not guilty (*I*)
finding guilty (FG)	*TG*	*FG*
finding not guilty (FI)	*FI*	*TI*

The principle of expected utility maximization provides a decision criterion: the judge should find guilty if and only if (iff) doing so has higher expected utility than finding not guilty:

$$EU(\mathrm{FG}) > EU(\mathrm{FI}) \qquad (1)$$ [1]

The expected utility of finding guilty is:

$$EU(\mathrm{FG}) = P(G) \cdot U(TG) + P(I) \cdot U(FG)$$

The expected utility of not finding guilty is:

$$EU(\mathrm{FI}) = P(G) \cdot U(FI) + P(I) \cdot U(TI).$$

A strength of decision theory is that it allows to determine credence thresholds for finding guilty in terms of the utilities. Proposition 1 of the Appendix shows that the inequality (1) is equivalent to:

$$P(G) > \frac{U(TI) - U(FG)}{U(TG) - U(FG) + U(TI) - U(FI)} = \theta \qquad (2)$$

The principle of expected utility maximization recommends a judge to find guilty just in case his credence $P(G)$ of guilt meets the threshold θ obtained from the utility values he assigns to the possible outcomes. I make a proposal for what utility values a judge should assign in the next section.

3. JUSTICE COSTS

A judge should make a just decision. Bayesian decision theory on its own does not say what decisions are just. But it offers a framework for spelling out one's notion of justice in terms of a judge's utility assignments to the possible trial outcomes (Kaplan, 1968; Kaye, 1980, 1982, 1999). I will do so by making the substantive assumption that a judge's decision has only two aims: to establish retributive justice and to avoid undeserved punishment. My assumption implies that *all* of the consequences unrelated to the two aims that a judge's decision may have should not matter for his utility assignments. It is substantive by excluding consequences, such as the social benefit of deterring and incapacitating dangerous or recidivismapt criminals. Indeed, these consequences are unrelated to establishing retributive justice for the case under consideration and avoiding undeserved punishment for the defendant. My substantive assumption leads to specific costs to justice, as I will explain in what follows.

Retributive justice makes two core claims. First, wrongdoers should be punished only in virtue of having done wrong. Second, wrongdoers deserve punishment in proportion to the harm they caused—in proportion to the severity of the offense they committed. So the severity of the punishment should be proportional to the

[1] Strictly speaking, the principle of expected utility maximization allows a judge both to find either guilty or not guilty if both findings have the same expected utility. We stipulate that in such rare cases of ties, a judge should find not guilty.

severity of the wrong done. In sum, the fundamental idea of retributive justice is this: a person deserves punishment iff she committed a wrongful act; and she deserves a more severe punishment for more severe wrongful acts. To punish an innocent person and not to punish a guilty person are therefore deviations from establishing retributive justice.

In the context of a criminal trial, the defendant deserves punishment iff she is guilty of committing a crime; and she deserves punishment in proportion to the severity of the crime. Unless otherwise stated, I assume that the punishment is proportional to the severity of the crime and so appropriate. On this assumption, a judge's true findings establish retributive justice, whereas his false findings do not. Any defendant deserves a true finding of not guilty and the appropriate punishment of a true finding of guilt. By contrast, no defendant deserves a false finding of not guilty and the punishment incurred by a false finding of guilt. The latter trial outcome incurs undeserved punishment on the defendant and so violates the judge's second aim to avoid such.

By my substantive assumption, there is only one way for a judge's decision to be unjust: the decision does not establish retributive justice. And only an unjust finding of guilt incurs undeserved punishment to the defendant. Let the *justice cost* of a possible trial outcome measure the extent to which it is unjust. We can think of costs to justice as deviations from establishing retributive justice weighted by the severity of undeserved punishment. In particular, I propose a twotier account of how unjust a possible trial outcome is:

(i) A possible trial outcome bears a cost of justice iff it is unjust.

(ii) The justice cost of an unjust trial outcome is greater in proportion to the severity of undeserved punishment it incurs on the defendant.

(i) says that a possible trial outcome bears a cost of justice iff it does not establish retributive justice for the case under consideration. Defendants deserve the trial outcomes of true findings but not those of false findings. A false finding of not guilty incurs a justice cost, but it does not incur an undeserved punishment to the defendant for the simple reason that there is no punishment in findings of not guilty. A false finding of guilt incurs a justice cost and an undeserved punishment for the defendant. By (ii), false findings of guilt bear a greater cost to justice than false findings of not guilty because the severity of some undeserved punishment is always higher than that of no punishment.

I have made a rough proposal for how to measure the justice costs of possible trial outcomes. This allows us to think of a judge's decision as follows: he should minimize the expected justice costs. I will refine and integrate the proposal into my decisiontheoretic account in the next two subsections.

3.1. A Simpler Credence Threshold

What justice costs should a judge assign to true findings? On the assumption that the severity of the punishment is proportional to that of the crime, a judge's true findings establish retributive justice perfectly: a true finding of not guilty establishes retributive justice perfectly and exactly as much as a true finding of guilt. Hence, true findings bear no cost to retributive justice and incur no undeserved punishment.

The justice costs for each possible trial outcome are the respective utility assignments. There are no justice costs of true findings:

$$U(TG) = U(TI) = \mathbf{0}$$ [2]

As a result, the formula to determine the credence threshold for finding guilty in equation (2) simplifies to:

$$P(G) > \frac{U(FG)}{-\,U(FG) - U(FI)} = \theta \qquad (3)$$ [3]

On my account, the credence threshold is always above 1/2. For this to be seen, note that a false finding of guilt incurs some punishment for the defendant and a false finding of not guilty does not. Some undeserved punishment is always more severe than none. So the justice cost of a false finding of guilt is greater than that of a false finding of not guilty. Hence, my account excludes the case where the credence threshold for finding guilty is below 1/2. This case means trouble for other decision-theoretic accounts: it says that a judge should find guilty if his credence meets the low credence threshold—even if the judge thinks it less likely that the defendant is guilty than not (Günther, 2024b, p. 139). On my account, a rational judge never finds guilty when he thinks that the defendant is less likely to be guilty than not.

3.2. The Severity of Undeserved Punishment

What justice costs should the judge assign to false findings? To recap, any false finding does not establish retributive justice and so bears a cost of justice by (i) of my proposal. A false finding of not guilty does not incur any punishment on the defendant, and so no undeserved punishment. Its justice cost is by (ii) not greater than

---

[2] We can set the justice costs of true findings to zero without loss of generality because utility functions are only unique up to positive affine transformation:

$$U(O) = aU'(O) + b, \text{ where } a, b, \in \mathrm{R} \text{ and } a > 0$$

[3] Kaplan (1968) also considers only the justice costs or 'disutilities' of false findings but without the argument from retributive justice. He has been criticized for doing so, among others, by Tribe (1971), Lillquist (2002), Laudan and Saunders (2009), and Nance (2016). However, their counterarguments assume that the utilities of true convictions may differ from those of true acquittals—contrary to our notion of retributive justice.

the baseline justice cost for false findings required by (i). As utility functions are only unique up to positive affine transformations, we can set the baseline cost to justice to

$$\boldsymbol{U(FI) = -1}$$

By doing so and setting the cost of true findings to zero, we only determine that a rational judge prefers any true finding over a false finding of not guilty.

A false finding of guilt incurs the baseline justice cost by (i) and is greater than that cost in proportion to the severity of undeserved punishment it incurs on the defendant by (ii). Let us represent the severity of undeserved punishment by a severity factor s. The severity factor s equals one if there is no undeserved punishment and is strictly greater than one if there is undeserved punishment. As the baseline cost to justice is represented by the cost of a false finding of not guilty, we obtain:

$$\boldsymbol{U(FG) = U(FI) \cdot s, \text{ where } s \geq 1.} \qquad (4)$$

The severity of undeserved punishment is nothing but the ratio of the justice cost of a false finding of guilt to that of a false finding of not guilty.

The second tier of my proposal is motivated by a principle of minimizing undeserved harm: a judge should minimize the undeserved punishment to the defendant. Hence, the justice cost of a false finding of guilt grows

with the amount of undeserved punishment it incurs on the defendant. On my account, the severity of an undeserved punishment can explain why falsely convicting a defendant of murder bears a greater justice cost than falsely convicting a defendant of shoplifting expensive headphones. Murder is a more severe crime than shoplifting, and so murderers deserve a harsher punishment than shoplifters by retributive justice.

As a matter of fact, law codes determine the severity of punishments against defendants for different crimes. Dutch law, for example, codifies that murder is punishable by life imprisonment or a fixed-term prison sentence of up to 30 years, while shoplifting is only punishable up to 4 years in prison and a fine up to €27.500. An imprisonment for 30 years is a more severe punishment than an imprisonment for 4 years plus the fine. By (ii), the justice cost to falsely imprison a defendant for 30 years is greater than to falsely imprison a defendant for 4 years plus the fine. A false finding in a murder case therefore has a greater justice cost than a false finding in a shoplifting case—also in the Netherlands.

Note the asymmetry to false findings of not guilty. There is no punishment in false acquittals, and so its justice cost cannot scale with the severity of undeserved punishment to the defendant. My account has this general upshot: a judge's justice cost of falsely convicting a defendant increases with the severity of the undeserved punishment borne by the defendant; however, his cost of a false acquittal does not.

On my account, the credence threshold for a finding of guilt is raised by a higher justice cost of false findings of guilt, but it is never lowered by the justice cost of false findings of not guilty because the latter cost remains constant for a given crime[4]. It follows that my account allows us to express the credence threshold in terms of the severity factor alone. By substituting in (3) the justice cost $U(FG)$ of a false finding of guilt with that $U(FI)$ of a false finding of not guilty weighted by the severity s of the undeserved punishment, we obtain:

$$P(G) > \frac{-U(FI) \cdot s}{-U(FI) \cdot s - U(FI)} = \frac{-s}{-s-1} = \frac{s}{s+1} \qquad (5)$$

The more severe the undeserved punishment is, the higher the credence threshold for finding guilty. It is only the severity of the possible undeserved punishment borne by the defendant that shifts the credence threshold for finding guilty.

3.3. The Severity of the Crime and Appropriate Punishments

My account offers the means to derive the magnitude of the justice costs from the appropriate severity of punishments for crimes. By retributive justice, the severity of punishment should depend on how severe the crime is. Stealing a bottle of water is a less severe crime than stealing a large amount of money or committing criminal tax fraud. In many law codes, the former is punishable by a small fine, while the latter is punishable by jail. If we assume that these law codes are appropriate, the severity of the punishment is directly proportional to the severity of the crime. If so, the severity of the punishment, or equivalently of the crime, can be used to determine the appropriate justice costs, and so to determine the appropriate credence threshold for finding guilty.

On my account, the credence threshold for finding guilty varies with the appropriate severity of the punishment. If Dutch law, for example, encodes the appropriate severity of a punishment for a crime, we can derive how much greater the justice cost of a false finding of guilt for the crime is compared to that of a false finding of not guilty. For illustration, suppose we appropriately measure the severity of punishment in terms of years in jail as follows: any 3 years in jail add a unit of severity, and one year in prison corresponds to a fine of €500.000. A 30-years imprisonment for murder gives us then a severity factor of 10. This result is reminiscent of the famous thought by Blackstone (1753, p. 358) that “the law holds that it is better that ten guilty persons escape than that one innocent suffer.” If this is an appropriate severity assessment, we obtain that the justice cost $U(FG)$ of a false finding of guilt is ten times greater than the baseline justice cost $U(FI)$. And so the credence threshold for a murder finding is 10/11, or approximately 0.91.

[4] Ross (2023, pp. 1090-1, 1097) comes to a similar conclusion even though he rejects “popular” decision-theoretic accounts of the standard of proof.

A 4-years imprisonment plus a fine of €27.500 gives us a severity factor of $4/3 + (1/3 \cdot 27.500/500.000) \approx 1.352$. If this severity assessment is appropriate, we obtain that the justice cost $U(FG)$ of a false finding of guilt is roughly 1.352 times greater than the baseline cost $U(FI)$. And so the credence threshold for a theft finding is roughly 0.57. This is, of course, only a proof of concept. It is tricky to find a good unit of severity and translations between years in jail and monetary fines. The main issues are how the severity of the punishment should be measured and what the exact functional form of severity growth should be—linear, logarithmic, quadratic, exponential, some combination, or something else altogether. I leave an appropriate assessment of the unit of severity and its functional form of growth for future research.

We have seen how appropriate amounts of punishment can be translated into justice costs of false findings of guilt on my account. My account thereby offers a method to estimate the appropriate ratio of the justice costs for a given crime from its severity. This being said, my account can also be used in the reverse direction. If we know the ratio of justice costs for a given crime, we can determine its appropriate severity. Hence, my account is compatible with a method of reflective equilibrium between a crime's justice costs, its severity, and the credence threshold for finding guilty (Jellema, 2025).

4. BAYESIAN EVIDENCE EVALUATION

We have seen that taking the principle of expected utility maximization as a decision criterion for a judge implies that his standard of proof is equivalent to a credence threshold. Given the appropriate justice costs, a judge should find the defendant guilty iff his credence in the defendant's guilt meets the threshold in light of the lawful evidence presented at trial.

The open question is what credences a judge should have based on the lawful evidence presented at trial? Decision theory only stipulates that a rational agent's credences conform to the probability axioms. Bayesianism adds that she updates her credences by conditionalization or a generalization thereof. There are no further epistemic constraints on a rational agent imposed by Bayesian decision theory.

A rational judge determines his credence of guilt by updating on the total lawful evidence presented at trial. A rational judge starts out with a prior credence function P_0, which represents his first credences before learning any evidence presented at trial[5]. He conditionalizes his prior credence $P_0(G)$ that the defendant is guilty on the total lawful evidence E_T to obtain the final credence $P(G)$ of guilt:

$$\mathbf{P(G) = P_0(G \mid E_T)} \qquad (6)$$

[5] Strictly speaking, the prior credence function $P_0(\cdot) = P_{00}(\cdot \mid B)$ encodes some background information B about the world but no evidence that is relevant to the hypothesis G that the defendant is guilty in the absence of the evidence presented at trial.

The conditional probability is given by the ratio definition of conditional probability:

$$P_0(G \mid E_T) = \frac{P_0(G \cap E_T)^{6}}{P_0(E_T)} \qquad (7)$$

The total lawful evidence E_T is equivalent to the conjunction of all individual pieces of lawful evidence $E_1 \cap E_2 \cap \ldots \cap E_n$ successively presented at trial. It is easy to see that a rational judge's final credence of guilt is the same whether he conditionalizes on the total lawful evidence or on all individual pieces successively.

4.1. How to Determine the Prior Credence of Guilt?

A rational judge's final credence of guilt depends on his prior. For this to be seen, observe that $P_0(G \cap E_T) = P_0(E_T \mid G) \cdot P_0(G)$, and so by (7)

$$P_0(G \mid E_T) = \frac{P_0(E_T \mid G) \cdot P_0(G)}{P_0(E_T)} \qquad (8)$$

This formula is known as Bayes's Theorem. It is often used to calculate the posterior credence after learning some evidence because the likelihood $P_0(E_T \mid G)$ of the evidence given that the defendant is guilty is often easier to assess in practice than the credence $P_0(G \cap E_T)$ of the conjunction of guilt and evidence. Bayes's Theorem makes explicit that a rational judge's posterior and final credence of guilt depends on his prior.

For illustration, suppose the rational judge's likelihood of the total lawful evidence given that the defendant is guilty is 0.7, his prior credence of this evidence is 0.1, and his credence threshold for finding guilty is 0.9. Then his final credence of guilt is 0.91 for a prior of 0.13 and 0.84 for a prior of 0.12. The small difference in the prior credence of guilt determines whether or not his final credence meets the threshold. A rational judge's credence of guilt is sensitive to his prior. As long as there is not a unique prior, it often remains unclear whether or not a judge should find guilty—on the Bayesian model Mackor presents and any other. But legal decisions should not be arbitrary in this way (Dahlman, 2018, p. 19). This is the *problem of the prior* in the legal context.

[6] It follows that

$$P(E_T) = P_0(E_T \mid E_T) = \frac{P_0(E_T \cap E_T)}{\mathrm{P}_0(\mathrm{E}_\mathrm{T})} = 1$$

For cases, where the rational judge does not become certain of the evidence after learning it, many Bayesians use Jeffrey conditionalization—a generalization of conditionalization due to Jeffrey (1983).

Opponents of Bayesian accounts of legal factfinding justifiedly take the problem of the prior as an argument against them. The prior credence of guilt cannot be determined by the lawful evidence presented at trial. The prior must already be in place before any evidence presented at trial is received by a rational judge at trial. Otherwise the Bayesian machinery does not get off the ground. Hence, the problem of the prior is perhaps the most serious problem for Bayesian accounts of legal factfinding.

What prior credence of guilt should a judge have? I think a fundamental principle of criminal procedure can help. The presumption of innocence says that a defendant should be considered innocent until proven guilty by the lawful evidence presented at trial. It implies that a defendant is to be considered not guilty until the evidence proves otherwise. The presumption of innocence thereby constrains the appropriate prior of guilt. On a credence threshold view, it says at least this: a rational judge presumes a defendant 'innnocent' or not guilty at the beginning of the trial iff his prior credence $P_0(G)$ of guilt is below the credence threshold θ for finding guilty. If $P_0(G) < \theta$ and no evidence were presented, the rational judge would consider the defendant innocent in the sense of finding not guilty. The problem with this notion of the presumption of innocence is that it does not single out a prior credence of guilt. Indeed, it leaves way too many candidates for the prior credence of guilt on the table—infinitely many.

I propose to understand the presumption of innocence as follows: a rational judge's prior credence of guilt should be $P_0(G) = 1 - \theta$, where θ is the credence threshold for finding guilty. It follows that the higher the credence threshold for a finding of guilt, the lower a rational judge's prior credence of guilt, and so the more demanding it is to find the defendant guilty. As θ is always strictly above 1/2 on my account, a rational judge's prior of guilt is always strictly below 1/2. Indeed, a rational judge should at least assume that the defendant is less likely to be guilty than not at the outset of a trial. This ensures that a rational judge never finds a defendant guilty without considering at least some lawful evidence.

My method of setting the prior credence of guilt is again motivated by the severity of (the appropriate punishment for) the crime under dispute. The more severe the crime, the higher the credence threshold for finding guilty and the lower the prior credence of guilt. My account entails that a judge should presume a defendant not guilty to the degree that the crime is severe. My notion of the presumption of innocence figures as a safeguard for not incurring undeserved punishment on the defendant, just like the credence threshold for finding guilty does. And so my Bayesian account for the evaluation of evidence got off the ground.

My method for setting a normative prior credence of guilt determines a *presumed prior* in the sense of Stein (1996), Posner (1999), Dahlman (2018), and Dahlman and Kolflaath (2021). My presumed prior follows Friedman's (2000) criticism in rejecting Posner's reading of the presumption of innocence as a prior credence of guilt of 1/2. I also part way with Dahlman and Kolflaath who assume without argument that the prior credence of guilt must be much smaller than 1/2 (p. 296). My presumed prior is determined by how high the credence threshold for a finding of guilt should be in terms of justice costs. Hence, it is not arbitrary[7]. My standard of proof is determined by both, the prior credence of guilt and the credence threshold for finding guilty.

4.2. The Strength of the Evidence Required for Conviction

My account sets a rational judge's credence threshold θ for a finding of guilt and his prior credence $P_0(G) = 1 - \theta$ of guilt. So we can think of the standard of proof in terms of how strong the total lawful evidence presented at trial must be to increase the prior credence of guilt such that it meets the credence threshold. As Dahlman (2018, p. 26) shows, the two numeric values allow us to compute the strength of the evidence required for a conviction if the strength is measured by the likelihood ratio of the total lawful evidence. This ratio compares the likelihood of the total lawful evidence if the defendant is guilty to that if the defendant is not guilty:

$$\frac{P_0(E_T \mid G)}{P_0(E_T \mid \neg G)} \qquad (9)$$

In the Appendix, I prove a theorem that refines Dahlman's result in the confines of my account:

Theorem 1 A rational judge's final credence of guilt meets the threshold $P_0(G \mid E_T) > \theta$ for a finding of guilt iff

$$\frac{P_0(E_T \mid G)}{P_0(E_T \mid \neg G)} > \frac{\theta^2}{(1 - \theta)^2}$$

On my account of legal proof, a rational judge finds a defendant guilty iff his likelihood ratio of the total lawful evidence is greater than his ratio of the squared

[7] I largely follow Dahlman and Kolflaath's (2021) arguments against determining priors of guilt on empirical grounds, such as the number of possible perpetrators (Fenton *et al.*, 2017; Lindley, 1977). The main problem is that empirical facts make the priors somewhat arbitrary: one judge might assign an empirical prior of guilt based on one reference class for a case, where another would assign a higher or lower such prior based on a different reference class. Reichenbach's (1949) reference class problem is looming large here (Colyvan *et al.*, 2001; Hájek, 2007). Building on Reichenbach, Franklin (2011, pp. 559-61) proposes that the ideal reference class for finding guilty or not should be defined by all features correlated with the defendant's guilt as long as the probability estimate of guilt remains reliable. This seems to counter the charge that reference classes are arbitrary in principle. But they are still arbitrary in practice because what features are correlated with the guilt hypothesis depends on the available data. More importantly, there are hypothetical cases, where even an ideal reference class might give us a prior of guilt which surpasses the threshold for finding guilty. No evidence is then needed to find a defendant guilty. And so the empirical prior based on the ideal reference class conflicts with the presumption of innocence.

credence threshold for finding guilty to the squared prior credence of guilt. For illustration, suppose that the credence threshold for finding guilty is 0.9. Then a rational judge finds the defendant guilty iff he thinks the total lawful evidence presented at trial is 81 times more likely if the defendant is guilty than if the defendant is not guilty.

We are now in a position to complete the Bayesian model Mackor (2026) presents. The model is the odds formulation of Bayes's Theorem of equation (8):

$$\frac{P_0(G \mid E_{\mathrm{T}})}{P_0(I \mid E_{\mathrm{T}})} = \frac{P_0(G)}{P_0(I)} \cdot \frac{P_0(E_{\mathrm{T}} \mid G)}{P_0(E_{\mathrm{T}} \mid I)} \qquad (10)$$

The model allows us to calculate the posterior odds in terms of the likelihood ratio of the total lawful evidence and the prior odds. If the likelihood ratio is greater (lower) than one, the posterior odds are greater (lower) than the prior odds. The likelihood ratio measures the upward or downward impact of the total lawful evidence on the posterior odds. As Mackor and Prakken (2026) know, the model allows us to deduce the final probability of guilt but only because the two ultimate hypotheses exclude each other and jointly exhaust the logical space—as Proposition 2 in the Appendix shows.

One can use my account to establish a normative credence threshold for a finding of guilt, and so a normative prior credence of guilt. If one uses the odds formulation of Bayes's Theorem together with the threshold θ and the prior 1 – θ of guilt, the package coincides with my account of when a judge should find a defendant guilty. As $P_0(G \mid E_{\mathrm{T}}) > \theta$ iff $P_0(I \mid E_{\mathrm{T}}) < 1 - \theta$, we obtain $P_0(G \mid E_{\mathrm{T}}) > \theta$ iff

$$\frac{P_0(G \mid E_{\mathrm{T}})}{P_0(I \mid E_{\mathrm{T}})} > \frac{1-\theta}{\theta} \cdot \frac{\theta^2}{(1-\theta)^2} = \frac{\theta}{1-\theta}$$

Hence, my account has the resources to turn the model Mackor presents into an account of legal proof, which does not leave open how a judge should decide.

5. CONCLUSION

I offered a Bayesian account of legal proof on which a judge's decision should minimize expected justice costs. The principle of expected utility maximization and my notion of justice costs entail a credence threshold for finding guilty in terms of the severity of undeserved punishment. The prior credence of guilt is derived from the normative credence threshold and implements a presumption of innocence. The normative prior and credence threshold determine how strong the lawful evidence presented at trial must be for a finding of guilt. They also make the model Mackor presents equivalent to mine, and so capable of guiding a judge through a whole criminal trial.

I have left several issues for future work. Among them are to explicitly show that my account covers civil cases as well; to defend the normative foundations of Bayesian decision theory for legal factfinding (Fenton & Lagnado, 2021); to defend my specific notion of justice costs against others; to examine the extent to which actual judges can be guided in their decisions by the regulative ideal of a rational judge; and to tackle open problems for decision-theoretic accounts, such as the proof paradoxes (Günther, 2024a, 2024c; Redmayne, 2008). But my account does show that Bayesian decision theory can guide a rational judge in his decision and so can serve as a guide to legal factfinding in principle.

APPENDIX

Proposition 1

The inequality $EU(\mathrm{FG}) > EU(\mathrm{FI})$ is equivalent to (2) in the main text.

Proof. Suppose $EU(\mathrm{FG}) > EU(\mathrm{FI})$. By the definition of expected utility, we obtain

$$P(G) \cdot U(TG) + P(I) \cdot U(FG) > P(G) \cdot U(FI) + P(I) \cdot U(TI),$$

which is equivalent to

$$P(G) \cdot U(TG) - P(G) \cdot U(FI) > P(I) \cdot U(TI) - P(I) \cdot U(FG).$$

By the Law of Total Probability, $P(I) = 1 - P(G)$. So

$$P(G) \cdot U(TG) - P(G) \cdot U(FI) > (1 - P(G)) \cdot U(TI) - (1 - P(7)) \cdot U(FG)$$

By multiplying out the right-hand side, we obtain

$$P(G) \cdot U(TG) - P(G) \cdot U(FI) > U(TI) - P(G) \cdot U(TI) - U(FG) + P(G) \cdot U(FG)$$

Rearranging all probabilty weighted utilities to the left-hand side gives us

$$P(G) \cdot U(TG) - P(G) \cdot U(FI) + P(G) \cdot U(TI) - P(G) \cdot U(FG) > U(TI) - U(FG),$$

and so

$$P(G) \cdot (U(TG) - U(FI) + U(TI) - U(FG)) > U(TI) - U(FG)$$

Dividing both sides by $U(TG) - U(FI) + U(TI) - U(FG)$ results in

$$P(G) > \frac{U(TI) - U(FG)}{U(TG) - U(FI) + U(TI) - U(FG)}$$

Proposition 2

The final probability of guilt can be calculated from the prior odds and the likelihood ratio of the total lawful evidence only because the two ultimate hypotheses of guilt G and not guilty $\neg G$ are mutually exclusive and jointly exhaustive.

Proof. By Bayes's Theorem

$$P_0(G \mid E_T) = \frac{P_0(G) \cdot P_0(E_T \mid G)}{P_0(E_T)}$$

Only because G and $\neg G$ are mutually exclusive and jointly exhaustive, we obtain by the Law of Total Probability

$$P_0(E_T) = P_0(G) \cdot P_0(E_T \mid G) + (1 - P_0(G)) \cdot P_0(E_T \mid \neg G)$$

Substituting $P_0(E_T)$ in Bayes's Theorem yields

$$P_0(G \mid E_T) = \frac{P_0(G) \cdot P_0(E_T \mid G)}{P_0(G) \cdot P_0(E_T \mid G) + (1 - P_0(G)) \cdot P_0(E_T \mid \neg G)}$$

Divide the numerator and denominator by $P_0(E_T \mid \neg G)$:

$$P_0(G \mid E_T) = \frac{P_0(G) \cdot \dfrac{P_0(E_T \mid G)}{P_0(E_T \mid \neg G)}}{P_0(G) \cdot \dfrac{P_0(ET \mid G)}{P_0(E_T \mid \neg G)} + (1 - P_0(G))}$$

This shows that the final probability of guilt can be calculated from the prior probability of guilt and the likelihood ratio of the total lawful evidence if the two ultimate hypotheses are G and $\neg G$. However, observe that the derivation only works because the two ultimate hypotheses are mutually exclusive and jointly exhaustive. Otherwise, one cannot substitute $P_0(E_T)$ in Bayes's Theorem using the Law of Total Probability. Similarly, one cannot derive the prior probability of guilt from the odds ratio alone unless $P_0(\neg G) = 1 - P_0(G)$.

In general, some posterior odds

$$\frac{P_0(H_1 \mid E_T)}{P_0(H_2 \mid E_T)}$$

do not give us the posterior probability $P_0(H_1 \mid E_T)$. To obtain the posterior probability, one would need to know the value x of the posterior odds and that $P_0(H_2 \mid E_T) = 1 - P_0(H_1 \mid E_T)$. Only if so:

$$
\begin{aligned}
P_0(H_1 \mid E_T) &= x \cdot (1 - P_0(H_1 \mid E_T)) \\
P_0(H_1 \mid E_T) &= x - x \cdot P_0(H_1 \mid E_T) \\
P_0(H_1 \mid E_T) + x \cdot P_0(H_1 \mid E_T) &= x \\
P_0(H_1 \mid E_T) \cdot (1 + x) &= x \\
P_0(H_1 \mid E_T) &= \frac{x}{1+x}
\end{aligned}
$$

Theorem 1. A rational judge's final credence of guilt meets the threshold $P_0(G \mid E_T) > \theta$ for a finding of guilt iff

$$\frac{P_0(E_T \mid G)}{P_0(E_T \mid \neg G)} > \frac{\theta^2}{(1-\theta)^2}$$

Proof. Suppose $P_0(G \mid E_T) > \theta$. By Bayes's Theorem and the Law of Total Probability, we obtain

$$\frac{P_0(G) \cdot P_0(E_T \mid G)}{P_0(G) \cdot P_0(E_T \mid G) + (1 - P_0(G)) \cdot P_0(E_T \mid \neg G)} > \theta$$

By multiplying both sides with the denominator, we get

$$P_0(G) \cdot P_0(E_T \mid G) > \theta \cdot P_0(G) \cdot P_0(E_T \mid G) + \theta \cdot (1 - P_0(G)) \cdot P_0(E_T \mid \neg G)$$

By subtracting the term $\theta \cdot P_0(G) \cdot P_0(E_T \mid G)$ on both sides, we obtain

$$P_0(G) \cdot P_0(E_T \mid G) - \theta \cdot P_0(G) \cdot P_0(E_T \mid G) > \theta \cdot (1 - P_0(G)) \cdot P_0(E_T \mid \neg G),$$

which we can rewrite as

$$(1 - \theta) \cdot P_0(G) \cdot P_0(E_T \mid G) > \theta \cdot (1 - P_0(G)) \cdot P_0(E_T \mid \neg G).$$

Dividing both sides by $(1 - \theta) \cdot P_0(G)$ and $P_0(E_T \mid \neg G)$ gives us

$$\frac{P_0(E_T \mid G)}{P_0(E_T \mid \neg G)} > \frac{\theta \cdot (1 - P_0(G))}{(1 - \theta) \cdot P_0(G)}$$

On my account, $P_0(G)$ is defined as $1 - \theta$. Note that we can go through the entire proof backwards. Hence, we obtain the desired result.

REFERENCES

Blackstone, W. (1753). *Commentaries on the laws of england: In four books* (Vol. 2). J.B. Lippincott.

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. *https://doi.org/10.1111/1467-9760.00123*

Dahlman, C. (2018). Determining the base rate for guilt. *Law, Probability and Risk*, *17*(1), 15–28. *https://doi.org/10.1093/lpr/mgx009*

Dahlman, C., & Kolflaath, E. (2021). The problem of the prior in criminal trials. In *Philosophical foundations of evidence law* (pp. 287–300). Oxford: Oxford University Press.

Fenton, N., & Lagnado, D. (2021, September). Bayesianism: Objections and rebuttals. In *Philosophical foundations of evidence law*. Oxford University Press. *https://doi.org/10.1093/oso/9780198859307.003.0019*

Fenton, N., Lagnado, D., Dahlman, C., & Neil, M. (2017). The opportunity prior: A simple and practical solution to the prior probability problem for legal cases. *Proceedings of the 16th Edition of the International Conference on Articial Intelligence and Law*, 69–76. *https://doi.org/*10.1145/3086512.3086519

Franklin, J. (2011). The objective Bayesian conceptualisation of proof and reference class problems. *Sydney Law Review*, *33*(3), 545–561.

Friedman, R. D. (2000). A presumption of innocence, not of even odds.

Stanford Law Review, *52*(4), 873–887.

Günther, M. (2024a). Epistemic sensitivity and evidence. *Inquiry*, *67*(6), 1348–66. *https://doi.org/10.1080/0020174X.2021.1936158*

Günther, M. (2024b). Legal proof should be justified belief of guilt. *Legal Theory*, *30*(3), 129–141. *https://doi.org/10.1017/S1352325224000089*

Günther, M. (2024c). Probability of guilt. *Canadian Journal of Philosophy*, *54*(3), 189–206. *https://doi.org/10.1017/can.2025.6*

Günther, M., & Friedrich, C. (2026). A Bayesian justification for the scenario approach to legal proof. *Proceedings of the Workshop on AI for Evidential Reasoning (AI4EVIR 2025)*, *4157*.

Hájek, A. (2007). The reference class problem is your problem too. *Synthese*,

156(3), 563–585. *https://doi.org/10.1007/s11229-006-9138-5*

Jeffrey, R. C. (1983). *The logic of decision* (Second). University of Chicago Press.

Jellema, H. (2025). Standards of proof for future crimes and decision theory. *Quaestio facti*, (10). *https://doi.org/10.33115/udg* bib/qf.i10. 23144

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

Kaye, D. H. (1980). Naked statistical evidence. *The Yale Law Journal*, *89*(3), 601–611.

Kaye, D. H. (1982). The limits of the preponderance of the evidence standard: Justifiably naked statistical evidence and multiple causation. *American Bar Foundation Research Journal*, *7*(2), 487–516.

Kaye, D. H. (1999). Clarifying the burden of persuasion: What Bayesian decision rules do and do not do. *The International Journal of Evidence & Proof*, *3*(1), 1–28. *https://doi.org/10.1177/136571279900300101*

Laudan, L., & Saunders, H. D. (2009). Rethinking the criminal standard of proof: Seeking consensus about the utilities of trial outcomes. *International Commentary on Evidence*, *7* (2). *https://doi.org/doi:10.2202/1554-4567.1099*

Lillquist, E. (2002). Recasting reasonable doubt: Decision theory and the virtues of variability. *UC Davis Law Review*, *36*, 85–197.

Lindley, D. V. (1977). Probability and the law. *Journal of the Royal Statistical Society. Series D (The Statistician)*, *26*(3), 203–220.

Mackor, A. R. (2026). Bayesian modelling of criminal cases as a whole: A philosophical reflection on Dutch case law. *Quaestio facti. International Journal on Evidential Reasoning*, (10). *https://doi.org/10.33115/udg bib/qf.i10.23229*

Mackor, A. R., & Prakken, H. (2026). On reporting likelihood ratio's of exhaustive and non-exhaustive hypotheses about rare events in criminal cases. *Proceedings of the Workshop on AI for Evidential Reasoning (AI4EVIR 2025)*, *4157*.

Nance, D. A. (2016). *The burdens of proof: Discriminatory power, weight of evidence, and tenacity of belief*. Cambridge University Press.

Posner, R. A. (1999). An economic approach to the law of evidence. *Stanford Law Review*, *51*, 1477–1546.
Ramsey, F. P. (1926). Truth and probability. In R. Braithwaite (Ed.), *The foundations of mathematics and other logical essays* (1999 electronic edition, pp. 156–198). Harcourt.
Redmayne, M. (2008). Exploring the proof paradoxes. *Legal Theory*, *14*(4), 281–309.
Reichenbach, H. (1949). *The theory of probability*. University of California Press.
Ross, L. (2023). Criminal proof: Fixed or flexible? *The Philosophical Quarterly*, *73*(4), 1077–1099. *https://doi.org/10.1093/pq/pqad001*
Savage, L. J. (1972). *The foundations of statistics* (Second Revised). Wiley Publications in Statistics.
Stein, A. (1996). Judicial factfinding and the bayesian method: The case for deeper scepticism about their combination. *The International Journal of Evidence & Proof*, *1*(1), 25–47. *https://doi.org/10.1177/136571279600100103*
Tribe, L. H. (1971). Trial by mathematics: Precision and ritual in the legal process. *Harvard Law Review*, *84*(6), 1329–1393.
von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.