Reasonable Doubt and Alternative Hypotheses: A Bayesian Analysis

Ulrike Hahn* Stephan Hartmann†

December 1, 2020

Abstract

A longstanding question is the extent to which 'reasonable doubt' may be expressed simply in terms of a threshold degree of belief. In this context, we examine the extent to which learning about possible alternatives may alter one's beliefs about a target hypothesis, even when no new 'evidence' linking them to the hypothesis is acquired. Imagine the following scenario: a crime has been committed and Alice, the police's main suspect has been brought to trial. There are several pieces of evidence that raise the probability that Alice committed the crime. Her attorney's defense strategy is not to challenge this evidence, but instead to provide personal details about Alice's neighbour, Jane. While Jane is one of many people the police spoke to, they saw no reason to investigate her further. You now learn that Jane, too, had access to the shed where the murder weapon was stored, just like Alice. To what extent should this alter your beliefs about Alice's guilt? In this paper, we provide a formal description of the problem and a solution indicating circumstances under which learning about Jane will more or less impact beliefs about Alice.

1 Introduction

Central to the criminal trial in the anglo-american world is the notion of "reasonable doubt": criminal conviction requires certainty beyond reasonable doubt.[1]

*Department of Psychological Sciences, Birkbeck, University of London, Malet Street, London WC1E 7HX (UK) – u.hahn@bbk.ac.uk

†Munich Center for Mathematical Philosophy, LMU Munich, 80539 Munich (Germany) – http://www.stephanhartmann.org – s.hartmann@lmu.de.

[1]In fact, the U.K.'s Crown Court has given up the phrase "beyond a reasonable doubt" in its guidance to trial judges in directing the jury, in favour of a mention that "the prosecution must

But what makes doubt reasonable and how can this notion be formalised? This question has implications far beyond the legal context: society faces the question of whether scepticism is 'reasonable' for a whole host of fundamental issues ranging from vaccination debates to anthropogenic climate change. On the one hand, there is a long tradition that has held that 'scepticism', and hence doubt, is integral to 'critical thinking' [5, 13]. Taken at face value, such a view might suggest that doubting is always reasonable. However, it is hard to think of a man faced with a hungry tiger ready to pounce as 'reasonable' when embarking on a prolonged deliberation about whether what looks and sounds like a tiger really is a tiger, or whether tigers really are dangerous.

An obvious starting point for a formal treatment of 'reasonable doubt' is to view the notion as expressing a kind of threshold, that is, a degree of belief beyond which continued 'doubt' becomes 'unreasonable'. This seems reminiscent of epistemological notions such as 'full belief' [21], with their attendent problems [12, 23] on the one hand, and of decision-theoretic thresholds for action on the other (see e.g., [8]). In the context of the latter, where the appropriate threshold lies, is determined by the utilities in question: Faced with a question of whether or not to take action, the consequences of uncertainty in our beliefs need to be resolved. Given that the normative framework of utility theory stipulates that we adopt the course of action that maximises expected utility, an implicit belief threshold exists for any set of actions such that beyond this threshold we are convinced enough to act. This threshold corresponds to the degree of belief that needs to be exceeded for the expected utility of an option to exceed the alternative possible actions. Given that expected utility is probability-weighted utility, changing those utilities will necessarily move the threshold.

Consequently, more severe outcomes will have different thresholds than will less severe ones. And for extreme outcomes, such as potentially disastrous consequences of climate change, the severity of the outcome may completely dominate the probability component (see so-called Dismal Theorems in [22]). While it may be useful to think about belief thresholds like this in many contexts, it is unclear that this is exactly the notion that legal systems have in mind [3, 15]. In the recent trial of former Trump campaign manager Paul Manafort, the jury submitted to the presiding judge a formal question asking for a definition of 'reasonable doubt'. In response, U.S. District Judge T.S. Ellis replied that reasonable doubt was "doubt based on reason".[2] This does not sound like a belief threshold; in fact, it does not sound like a belief simpliciter at all. Rather it makes reference to the process of belief formation itself.

Does this extend to our intuitions about reasonableness beyond the courtroom? Arguably, one would consider the doubt (or, conversely, the belief) of an agent 'unreasonable' where that agent failed to fully incorporate her evidence into her beliefs. Faced with two positive and two negative pieces of evidence, an agent

make the jury sure that [the defendant] is guilty" to prove guilt, and that "nothing less will do". [2]

[2]See https://uk.reuters.com/article/uk-usa-trump-russia-manafort-jury/jury-in-manafort-trial-asks-about-reasonable-doubt-as-it-ends-first-day-idUKKBN1L12HW

who underweighted the positive evidence, or ignored it entirely, even though they believed that evidence to be diagnostic in principle, clearly falls short of the demands of a rational, Bayesian agent. As a result, her beliefs will be less accurate than they could (or should) be [14, 16], and it seems appropriate to deem that agent 'unreasonable'.

This case, where evidence has already been obtained, seems clear. But what about the case where the search for evidence was somehow restricted. Imagine, for example, a criminal investigation where police and prosecutor focused only on a single obvious suspect from the outset. Might this limitation be sufficient to create 'reasonable doubt', and could it do so even where the evidence against that suspect merits a high posterior degree of belief in the suspect's guilt?

To sharpen intuitions here, we provide a formal, Bayesian treatment of the belief dynamics in such cases. Specifically, we assume a case such as the following: a crime has been committed and Alice, the police's main suspect has been brought to trial. There are several pieces of evidence that raise the probability that Alice committed the crime. Her attorney's defense strategy is not to challenge this evidence, but instead to provide personal details about Alice's neighbour, Jane. While Jane is one of many people the police spoke to initially, they did not pursue further investigation of her or her possible connections with the crime. You now learn that Jane, too, had access to the shed where the murder weapon was stored, just like Alice. To what extent should this alter your beliefs about Alice's guilt?

We provide a formal description of this problem and then detail a solution. This formalization distinguishes those circumstances under which learning about Jane will impact beliefs about Alice, and those, in which it will not. We next present the model and formal treatment, before we return to its implications for the wider debate on rational scepticism and the reasonableness of doubt.

The remainder of this article is organized as follows. Section 2 presents our baseline model and discusses what follows from it. Section 3 then refines the baseline model and shows under which conditions something we call the 'focal fallacy' obtains. Section 4 discusses our findings and makes some more general observations. Finally, Section 5 summarizes our main results and suggests some future work.

2 The Baseline Model

Abstracting from the concrete case mentioned above, we consider the following situation. There is a hypothesis H ("Alice committed the crime") which is supported by a piece of evidence E ("Alice had access to the shed where the murder weapon was stored"). There may, however, also be alternative hypotheses. For example, it could be known that n people where in the area at the time the crime took place. The agent therefore assumes that there are n alternative hypotheses $\mathrm{H_1}, \ldots, \mathrm{H_n}$. Formally, these alternative hypotheses are contained in the "catch all" hypothesis, i.e.

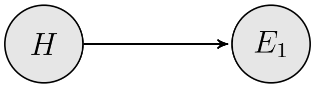

Figure 1: The Bayesian network representation of the relation between H and E_1.

$$\neg \mathrm{H} \equiv \mathrm{H}_1 \vee \cdots \vee \mathrm{H}_\mathrm{n}. \tag{1}$$

Here we assume that exactly one of the hypotheses $\mathrm{H}, \mathrm{H}_1, \ldots, \mathrm{H}_\mathrm{n}$ is true.[3] These are the values of the $n+1$-valued propositional variable H.[4]

Next we introduce the binary propositional variable E_1 which has the values E_1: "There is evidence that suspect no. 1 committed the crime" and $\neg\mathrm{E}_1$: "There is no evidence that suspect no. 1 committed the crime". The relation between the propositional variables H and E_1 is represented in the Bayesian network in Figure 1.

The prior probability distribution P (after having learned E, that is the evidence implicating the focal suspect Alice, which the present analysis considers to be part of the background knowledge) over H and E_1 is given by

$$P(\mathrm{H}) = h \quad , \quad P(\mathrm{H}_i) = \overline{h}/n \tag{2}$$

and

$$\begin{aligned} P(\mathrm{E}_1|\mathrm{H}_1) &= 1 \\ P(\mathrm{E}_1|\mathrm{H}_\mathrm{i}) &= \alpha \quad \text{for } i = 0, 2, \ldots, n, \end{aligned} \tag{3}$$

with $n \geq 2$ and $\alpha \in (0, 1)$. We have also used the convenient definitions $\overline{h} := 1 - h$ and $\mathrm{H}_0 := \mathrm{H}$.[5]

Here α is the probability of a false positive [6] that is, the probability that the seeming evidence against suspect no. 1 obtains even though another suspect, i.e. if no. 0 or no. $2, 3, \ldots$ or n committed the crime. Note that this likelihood is the same for suspect no. 0 and the other suspects: The agent who assigns these likelihoods does not distinguish between suspect no. 0 (who has already been investigated by the police) and the other suspects who are, so far, unknown to the police.

Next, the agent learns E_1. In other words, evidence against alternative suspect no. 1 (say, Jane) is put forward. How shall the agent update her beliefs? The following proposition specifies the new probability distribution P' which follows from Conditionalization on E_1, i.e. (in this case) $P'(\cdot) = P(\cdot|\mathrm{E}_1)$.

[3] Note that the model can be modified if the agent assumes that two hypotheses are true, that is, for example, if the agent assumes that two people jointly committed the crime.

[4] Here and in the remainder we denote propositional variables in italics and their values in roman script.

[5] It is plausible to assume that $h > 1/(n+1)$ as the probability distribution under consideration resulted from an update on other evidence supporting H. We will see that the qualitative results stated in the propositions below do not depend on this assumptions. Our quantitative results, however, depend on the value of h, as we will discuss below.

Proposition 1 *An agent considers the propositional variables H and E_1 with a prior probability distribution P defined in eqs. (2) and (3). The agent then learns E_1 and updates P to P' using Conditionalization. Then $P'(\mathrm{H}) < P(\mathrm{H})$, $P'(\mathrm{H_1}) > P(\mathrm{H_1})$ and $P'(\mathrm{H_j}) < P(\mathrm{H_j})$ for $j = 2, \ldots, n$.*

We see that evidence for the claim that suspect no. 1 (i.e., in our case, Jane) committed the crime will always decrease the probability that one of the other suspects, including no. 0 (i.e., in our case, Alice herself), committed the crime.[6] At the same time, and not surprisingly, the probability that suspect no. 1 committed the crime increases. Hence, if Alice's defence attorney succeeds in providing evidence for the claim that Jane committed the crime, then Alice's chances increase that she will not be convicted. Note, however, that this result presupposes that the probability that the seeming evidence against Jane obtains is the same given that any one of the other suspects actually committed the crime (i.e. Alice or any one of the nameless $n-1$ further suspects). Finally, we note that it would be even better for Alice if the probability that one of the remaining suspects (i.e. nos. $2, \ldots$) committed the crime also increased. We will see that this happens in some of the scenarios considered in the next section. Before we consider such scenarios, however, we generalize Proposition 1.

Let us assume that Alice's defense attorney does not only provide evidence for the claim that Jane committed the crime, but also that several other suspects such as Bob, Clarence and Dave committed the crime in order to raise doubts about Alice having committed the crime. For example, the defense attorney could point out that all five suspects (and not only Alice and Jane) had access to the shed where the murder weapon was stored. We ask: what is the posterior probability that Alice committed the crime in this case? Do we get qualitatively the same results as in the case of Proposition 1, or does one find some new behavior?

To address these questions, we consider the situation where the agent learns $\mathrm{E_1}, \ldots, \mathrm{E_k}$ with $1 \leq k \leq n$ and assume that the corresponding propositional variables E_i are conditionally independent of each other given the value of the hypothesis (see Figure 2). Generalizing eqs. (3), we furthermore assume

$$\begin{aligned} P(\mathrm{E_j}|\mathrm{H_j}) &= 1 \quad \text{for } j = 1, \ldots, k \\ P(\mathrm{E_j}|\mathrm{H_i}) &= \alpha \quad \text{for } i = 0, \ldots, n \text{ and } j = 1, \ldots, k \text{ and } i \neq j. \end{aligned} \tag{4}$$

The following proposition specifies the new probability distribution P'.

Proposition 2 *An agent considers the propositional variables H and $E_1, \ldots, E_k$ with $1 \leq k \leq n$ with a prior probability distribution P defined in eqs. (2) and (4) and $E_i \perp\!\!\!\perp E_j | H$ for $i \neq j = 1, \ldots, k$. The agent then learns $\mathrm{E_1}, \ldots, \mathrm{E_k}$ and*

[6] It is interesting to see (from the proof of Proposition 1) that the posterior probability of H can be smaller than the original prior probability, i.e. $1/(n+1)$, if $\alpha < \overline{h}/(\overline{h} + n^2\, h - n)$. This is the case, e.g., if $\alpha < 1$ for $h = 1/n$ or if $\alpha < (n-1)/(2n-2)$ for $h = 1/(n-1)$ or if $\alpha < 1/(n-1)^2$ for $h = 1/2$. Plugging in numbers for n, one sees that the posterior probability of H can be smaller than the original prior probability of H for relatively large values of α.

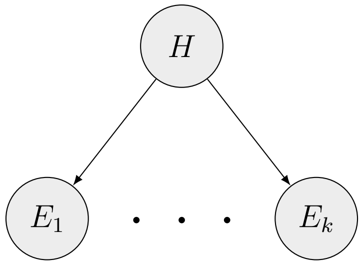

Figure 2: The Bayesian network representation of the relation between H and $E_1, \ldots, E_k$.

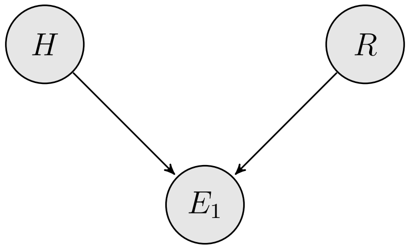

Figure 3: The Bayesian network representation of the relation between H, E_1 and R.

updates P to P' using Conditionalization. Then $P'(\mathrm{H}) < P(\mathrm{H})$, $P'(\mathrm{H_j}) > P(\mathrm{H_j})$ for $j = 1, \ldots, k$ and $P'(\mathrm{H_j}) < P(\mathrm{H_j})$ for $j = k + 1, \ldots, n$.

We conclude that bringing up evidence against other subjects does not change the qualitative results of Proposition 1. That is, the probability that one of the specific suspects now highlighted by the defense attorney (i.e. no. $1, \ldots, k$) committed the crime increases and the probability that any one of the remaining suspects (including Alice) committed the crime decreases. However, the amount of change depends on the number k of these newly prominent suspects. We see, for example, from the proof of proposition 2 that the new probability of H is a decreasing function of k/n (see Appendix). Similarly, the new probability of H_j (for $j = k+1, \ldots, n$) is proportional to its prior probability (i.e. $\overline{h}/n$) and to k/n.

In closing this section, we would like to generalize Proposition 1 in another respect and allow the evidence to be the testimony of a witness. That is, we assume in our example that the witness claims that Jane committed the crime. (This witness could be Alice' defense attorney who states that Jane committed the crime.) It is then up to the judge to take this witness report properly into account. Here we assume that the witness is a partially reliable information source. This is a strong assumption which may not hold. Our model therefore represents a baseline indicating what is possible in the ideal case for Alice.

Following the witness model discussed in [1, ch. 3], we introduce the additional binary propositional variable R (which stands for the reliability of the witness and has the values R–the witness is fully reliable–and ¬R–the witness is fully unreliable) and assume that the relation between the three variables is represented by the Bayesian network in Figure 3. To proceed, we specify the prior probabilities of all root notes, i.e.

$$P(\mathrm{H}) = h \quad , \quad P(\mathrm{H}_i) = \overline{h}/n \tag{5}$$

and

$$P(\mathrm{R}) = r \tag{6}$$

and the conditional probabilities of the child node E_1, given the values of its parents:

$$\begin{aligned} P(\mathrm{E_1}|\mathrm{H_1},\mathrm{R}) &= 1 \\ P(\mathrm{E_1}|\mathrm{H_k},\mathrm{R}) &= 0 \quad \text{for } k = 0, 2, \ldots, n, \\ P(\mathrm{E_1}|\mathrm{H_i},\neg\mathrm{R}) &= a \quad \text{for } i = 0, 1, \ldots, n, \end{aligned} \tag{7}$$

Here we have assumed that the witness is a truth-teller if she is (fully) reliable, and that she randomizes with a probability a (the *randomization parameter*) if she is (fully) unreliable. Note that this model does not allow for a systematic bias against H_1. With this one can show that the same qualitative results obtain as the ones stated in Proposition 1.

Proposition 3 *An agent considers the propositional variables H, E_1 and R with a prior probability distribution P defined in eqs. (5), (6) and (7). The agent then learns E_1 and updates P to P' using Conditionalization. Then $P'(\mathrm{H}) < P(\mathrm{H})$, $P'(\mathrm{H_1}) > P(\mathrm{H_1})$ and $P'(\mathrm{H_j}) < P(\mathrm{H_j})$ for $j = 2, \ldots, n$.*

Proposition 3 generalizes Proposition 1 as the prior probability distribution used in Proposition 1 obtains if one sets $a = 1$.[7]

3 Biases and the Focal Fallacy

The Baseline Model makes two important assumptions. First, it assumes that $P(\mathrm{E_1}|\mathrm{H_i}) = \alpha$ for $i = 2, \ldots, n$. This is a reasonable assumption as there is no reason to differentiate between the various suspects. They are all basically unknown to the judge and it is therefore reasonable to assign both the same prior probability to them and to also assume that the likelihoods $P(\mathrm{E_1}|\mathrm{H_i})$ are the same for all of them. The evidence against Jane is equally likely to obtain for each of them.

Second, the Baseline Model assumes that $P(\mathrm{E_1}|\mathrm{H}) = \alpha$, i.e. that the likelihood that the evidence against neighbour Jane would obtain given that she did not actually commit the crime would be the same regardless of whether it was, in fact, Alice or any one of the other suspects that committed the crime. This is clearly a controversial assumption and a judge may not want to make it. There are, in

[7] To see this, we calculate $P(\mathrm{E_1}|\mathrm{H_1}) = r + a\,\overline{r}$ and $P(\mathrm{E_1}|\mathrm{H_i}) = a\,\overline{r}$ for $i = 0, 2, \ldots, n$ and note that this coincides with the prior probabilities specified in eqs. (3) if $a = 1$ and $\alpha = \overline{r}$.

fact, reasons for setting $P(\mathrm{E}_1|\mathrm{H})$ to a higher or lower value than $P(\mathrm{E}_1|\mathrm{H_i})$. For example, the judge may recognize that Alice is a special suspect. The police found evidence against her and knows much more about her. Alice is different from the other suspects, which justifies assigning her a different prior and a different likelihood on the new evidence (against Jane). More specifically, one would expect the prior to be greater than $1/(n+1)$ as this prior is the posterior conditional on the evidence E that obtained against Alice.

Concerning the likelihood $P(\mathrm{E}_1|\mathrm{H})$, there are reasons to set it to a higher or to a lower value than $P(\mathrm{E}_1|\mathrm{H_i})$ (for $i = 2, \ldots, n$). For example, one may set $P(\mathrm{E}_1|\mathrm{H})$ to a rather low value as the judge sees no reason why E_1 should obtain if Alice is guilty. $P(\mathrm{E}_1|\mathrm{H_i})$, on the other hand, may be set to a higher value as the judge does not know much about the other suspects so it may well be possible that E_1 obtains if one of them committed the crime. But there is also a different consideration: the judge may not trust Alice's defense attorney and consider it to be more likely that the evidence against Jane obtains if Alice committed the crime than if one of the other suspects committed the crime. In this case the judge would set $P(\mathrm{E}_1|\mathrm{H}) > P(\mathrm{E}_1|\mathrm{H_i})$.

In short, there are any number of reasons to assume that these likelihoods could plausibly be different. Consequently, we next study in detail the consequences of assuming that $P(\mathrm{E}_1|\mathrm{H})$ and $P(\mathrm{E}_1|\mathrm{H_i})$ differ. To do so, we set

$$P(\mathrm{E}_1|\mathrm{H}) =: \alpha \quad , \quad P(\mathrm{E}_1|\mathrm{H_i}) =: \beta \tag{8}$$

for $i = 2, \ldots, n$ and $\alpha, \beta \in (0, 1)$. The rest remains as in Section 2. Obviously, the Baseline Model is a special case of the present model if $\alpha = \beta$.

We can then show the following proposition:

Proposition 4 *An agent considers the propositional variables H and E_1 with a prior probability distribution P defined in eqs. (2) and (8). The agent then learns the proposition* E_1 *and updates P to P′ using Conditionalization. Then* $P'(\mathrm{H}) < P(\mathrm{H})$ *iff* $\alpha < \beta + x_n\,\overline{\beta}$, $P'(\mathrm{H}_1) > P(\mathrm{H}_1)$ *and* $P'(\mathrm{H_i}) < P(\mathrm{H_i})$ *iff* $\alpha > \beta - y_n(h)\,\overline{\beta}$ *for* $i = 2, \ldots, n$ *with* $x_n = 1/n$ *and* $y_n(h) = \overline{h}/(n\,h)$.

Figure 3 summarize the main insights from Proposition 4. In this phase plot, the blue line represents the curve $\alpha = \beta + x_n\,\overline{\beta}$ and the orange line represents the curve $\alpha = \beta - y_n(h)\,\overline{\beta}$. The phase plot exhibits three regions: In region I, $P'(\mathrm{H}) > P(\mathrm{H})$ and $P'(\mathrm{H_i}) < P(\mathrm{H_i})$. In region II, $P'(\mathrm{H}) < P(\mathrm{H})$ and $P'(\mathrm{H_i}) < P(\mathrm{H_i})$. Note that region II is the region around the diagonal $\alpha = \beta$ (not plotted) familiar from Proposition 1. In region III, $P'(\mathrm{H}) < P(\mathrm{H})$ and $P'(\mathrm{H_i}) > P(\mathrm{H_i})$.

It is interesting to note that the blue line is independent of the prior of H (i.e. h). The orange line, however, does depend on the prior. To illustrate the dynamics here, Figure 4 shows the point β_0 where the orange line meets the x-axis as a function of h for $n = 5, 10, 30$ and 100.[8] One sees that β_0 moves further to

[8]It is easy to see that $\beta_0 = \overline{h}/\left[1 + (n-1)\,h\right]$.

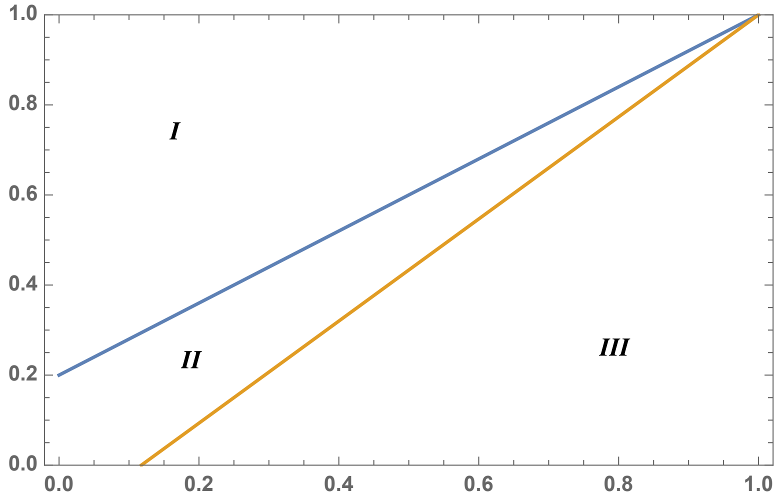

Figure 4: The α vs. β-phase plot for $n = 5$ and $h = .6$. (Note that β is plotted on the x-axis and α on the y-axis.)

the left the larger h is. That is, the area of region III increases with increasing h and therefore more assignments of α and β lead to a decrease of the probability of H and an increase of the probability of the other H_i's.

Note also that, as n increases, the blue line and the orange line in Figure 3 move closer and closer to the diagonal $\alpha = \beta$ (as x_n and $y_n(h)$ go to zero in the limit). In this case, region II disappears from Figure 3 and $P'(\mathrm{H}) < P(\mathrm{H})$ and $P'(\mathrm{H_i}) > P(\mathrm{H_i})$ iff $\alpha < \beta$.

In short, once the possibility of differential likelihoods is taken into account, judging the overall impact of evidence for a new suspect becomes considerably more complicated, and it seems all too easy to get the consequences wrong. The best strategy for Alice' attorney is to suggest that α is small and in any case smaller than β. That is, Alice' attorney should talk about the other suspects and not leave it to the imagination of the judge how to assess the probability of the observation that the evidence against Jane obtained if in fact one of the other suspects committed the crime. This is an important quantity and it has to be taken into account. Neglecting β may lead to an undesired result.

At the same time, it is difficult to assess β, especially for the judge who may not know much more about the other people than their total number. After all, the police regarded these people as not suspect and decided to not investigate them further. So what is a rational assignment of β? Should the judge simply assume that β is low? This is problematic as β is *independent* of the prior probability of H_i: The prior probability that one of the other people committed the crime may be low (which is the conclusion the police arrived at), and yet the probability that evidence against Jane obtains given that, say, Dave committed the crime may be large. Neglecting to properly assess this likelihood amounts to committing the *focal fallacy*. In this case, the judge and the police are too concerned with Alice

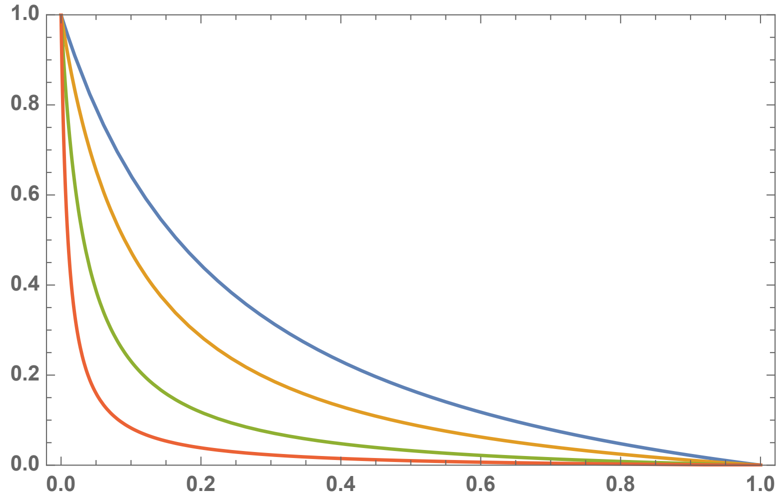

Figure 5: β_0 as a function of h for $n = 5$ (blue), $n = 10$ (orange), $n = 30$ (green), and $n = 100$ (red).

and do not properly assess the likelihood of E_1 given H_j with $j \geq 2$ after having established that the prior probability that one of these people committed the crime is low.

4 Discussion

In this paper, we examined formally the impact of considering a new piece of evidence against a suspect previously outside the main focus of the investigation. This formal analysis reveals multiple interrelated ways in which an undue focus on a single prime suspect can make conclusions about guilt unreliable. Learning that an alternative suspect who was previously considered only as part of a "catch all" alternative hypothesis may be linked to the crime can substantially alter the probability that the initial, focal suspect committed it. It is thus not enough to simply consider how strong the evidence is regarding that focal suspect, while not considering other possibilities. We refer to failures to reckon properly with alternative suspects as committing the focal fallacy.

One may commit the focal fallacy by focusing too much on one key suspect (or hypothesis) in a number of ways:

(i) By setting the number of serious alternatives at too low a value. The reason for this might be that one settles too quickly on the first serious contender.

(ii) By setting the value of β at too low a value (or even to 0). The reason for this might be a confusion of the prior probability of the hypothesis with the likelihood. However, the prior and the likelihood are *independent*.

At the same time, the belief dynamics of including new evidence about alternative suspects are complex and themselves easy to get wrong. In particular, without formal analysis it seems impossible to discern that, as the prior (based on other evidence) for the focal suspect increases, learning evidence that implicates a specific alternative suspect is less and less likely to uniquely raise the probability of that alternative suspect alone (see Figure 3). Instead it becomes more and more likely that *either* the probability of the initial focal suspect committing the crime decreases but that of other potential suspects in the "catch all" actually increases *or* the probability for the focal suspect actually goes up, while it decreases for those in the "catch all".

In other words, while the intuitive strategy of a defense attorney which points out that "others could have done it" does have a normative basis on our analysis, it is also too simplistic to assume that merely pointing out that some other individual has a potential evidential link with the crime is enough to cast doubt on the guilt of the focal suspect. Pointing out such new evidence may, as Proposition 4 above shows, actually increase the probability that the focal suspect committed the crime, depending on the underlying likelihoods assigned. Merely identifying evidence that implicates an alternative suspect may 'backfire' and raise the probability that the focal suspect is guilty instead, if only the likelihood that the evidence could arise also if a member of the remaining "catch all" is deemed high enough.

All of this raises multiple issues for both threshold theories of 'reasonable doubt' and for attempts to normatively reconstruct the judgment practice of the criminal trial in broadly Bayesian terms more generally. The proper relationship between Bayesianism and legal reasoning, in particular evidential reasoning has a long and chequered history which we cannot do justice here (see e.g., [10, 19, 20]). However, given the links between Bayesianism and the accuracy of our beliefs (see [14]) it is hard to see how consideration of belief dynamics in Bayesian terms would not, at the very least, fall under the scope of "reasonable" as introduced by trial judge Ellis in our Introduction above. That, in and of itself, would seem to make the various different kinds of responses to alternative evidence outlined by the belief dynamics of our Bayesian formalisation 'fair game'. But this, in turn, arguably challenges the appropriateness of a simple threshold view of belief. Bayesian inference simply cannot proceed without consideration of alternative ways in which the evidence may have come about, and the specific application to 'unpacking' of a 'catch all' hypothesis examined here, simply serves to underscore that general fact. In that sense, Bayesianism seems much more attuned to considerations of due process that legally seek to preclude a criminal procedure that too readily "makes up its mind" than legal commentators typically assume. However, a Bayesian analysis, as conducted here, guards also against undue scepticism by indicating how and why not every alternative raised will meaningfully undermine belief in a focal hypothesis. Rational belief formation, from a Bayesian perspective, requires that all relevant likelihoods are appropriately reckoned with.

Here, the analyses of the present paper contain a broader, cautionary note. The outcome of Bayesian inference often, though not always (see e.g., [11]), hinges critically on what those likelihoods are, and estimating the likelihoods associated with false positives generated by an unexplicated “catch all” hypothesis, or alternatively, a set of generically defined alternatives, seems even harder than estimating the likelihood of the evidence if, in fact, the hypothesis is true (on the general issue of estimating likelihoods see [7]).

The problems associated with the “catch all” hypothesis, including its potential practical intractability, have been discussed before. In particular, Salmon ([17]) has argued that, in a science context, estimating the catch all probability amounts to guessing the future of science an endeavour with little to no realistic constraint. Moreover, the mere possibility of future, perfectly diagnostic, evidence for as yet unconceived, alternative, hypotheses renders the calculations based on extent evidence meaningless wherever that evidence is less than perfectly diagnostic. One solution, here, is to abandon posteriors in favour of posterior odds between presently considered, known alternatives. This solution, which is the one advocated by Salmon for science, may or may not be considered a practically adequate solution in its own right. We do not take a stand on this question here. Rather, we point out that a theoretical focus on posterior odds obscures important new questions about belief dynamics questions that themselves raise other normatively relevant issues. As we argue here, a new focus on normative consequences of unpacking contributes to the understanding of ongoing questions about the nature of reasonable doubt, undue scepticism, and burdens of proof. These questions have both theoretical and practical relevance in a wide range of domains.

5 Conclusion

In closing, we would like to mention that the implications of our discussion may be even more acute for science than for law, given that here there is an arguably even stronger case for the relevance of a Bayesian normative perspective (see e.g., [9, 18]). Applied to the scientific context, our results suggest that in order to properly assess a scientific theory, scientists should not only have justified beliefs about the number of alternative theories (see the discussion of the no alternatives argument in [4]), but also about the likelihoods of the alternatives as the posterior probability of the theory in question crucially depends on them. Given that one does not know much about these alternatives and has at best some mildly justified beliefs about their number, this is a difficult task. In any case, our discussion reminds us that it is unreasonable to settle on the first working theory that is found and focus on it come what may if one’s goal is to find the true theory. It will be interesting to explore the consequences of our discussion for the scientific realism debate. We leave this for another occasion.[9]

[9]For various reasons, the scientific case is much harder to analyze than the legal case. For example, alternative scientific theories typically do not form a partition (as they may overlap) which is one of the assumptions we made in the present analysis.

A Appendix

A.1 Proof of Proposition 1

Setting again $\mathrm{H_0} := \mathrm{H}$, we first calculate

$$
\begin{aligned}
P(\mathrm{E_1}) &= \sum_{l=0}^{n} \sum_{H_l} P(\mathrm{E_1}|\mathrm{H_l})\, P(\mathrm{H_l}) \\
&= h\,\alpha + \frac{\overline{h}}{n}\,[1 + (n-1)\,\alpha] \\
&= \frac{1}{n}\,(n\,\alpha + \overline{h}\,\overline{\alpha}).
\end{aligned}
$$

With $P'(\mathrm{H_i}) = P(\mathrm{H_i}|\mathrm{E_1}) = \mathrm{P(E_1|H_i)\,P(H_i)/P(E_1)}$ (for $i = 0, \ldots, n$) we can now calculate

$$
\begin{aligned}
P'(\mathrm{H}) &= \frac{n\,h\,\alpha}{n\,\alpha + \overline{h}\,\overline{\alpha}} \\
P'(\mathrm{H_1}) &= \frac{\overline{h}}{n\,\alpha + \overline{h}\,\overline{\alpha}} \\
P'(\mathrm{H}_j) &= \frac{\alpha\,\overline{h}}{n\,\alpha + \overline{h}\,\overline{\alpha}}\,,
\end{aligned}
$$

for $j = 2, \ldots, n$. Next, we calculate

$$
P'(\mathrm{H}) - P(\mathrm{H}) = -\frac{h\,\overline{h}\,\overline{\alpha}}{n\,\alpha + \overline{h}\,\overline{\alpha}} < 0.
$$

Similarly,

$$
P'(\mathrm{H_1}) - P(\mathrm{H_1}) = \frac{(n - \overline{h})\,\overline{h}\,\overline{\alpha}}{n\,(n\,\alpha + \overline{h}\,\overline{\alpha})} > 0.
$$

Finally, for $j = 2, \ldots, n$,

$$
P'(\mathrm{H_j}) - P(\mathrm{H_j}) = -\frac{\overline{h}^2\,\overline{\alpha}}{n\,(n\,\alpha + \overline{h}\,\overline{\alpha})} < 0.
$$

This completes the proof of Proposition 1. ∎

A.2 Proof of Proposition 2

As before, we first calculate

$$
P(\mathrm{E_1}, \ldots, \mathrm{E_k}) = \frac{1}{n}\,(n\,\alpha + k\,\overline{h}\,\overline{\alpha}).
$$

With this, we calculate

$$\begin{aligned}
P'(\mathrm{H}) &= \frac{n\,h\,\alpha}{n\,\alpha + k\,\overline{h}\,\overline{\alpha}}\\
P'(\mathrm{H}_i) &= \frac{\overline{h}}{n\,\alpha + k\,\overline{h}\,\overline{\alpha}}\\
P'(\mathrm{H}_j) &= \frac{\overline{h}\,\alpha}{n\,\alpha + k\,\overline{h}\,\overline{\alpha}}\,,
\end{aligned}$$

for $i = 1, \ldots, k$ and $j = k+1, \ldots, n$. Hence,

$$\begin{aligned}
P'(\mathrm{H}) - P(\mathrm{H}) &= -\frac{h\,\overline{h}\,\overline{\alpha}}{n\,\alpha + k\,\overline{h}\,\overline{\alpha}} < 0\\
P'(\mathrm{H_i}) - P(\mathrm{H_i}) &= \frac{\overline{h}\,(n - \overline{h}\,k)\,\overline{\alpha}}{n\,(n\,\alpha + k\,\overline{h}\,\overline{\alpha})} > 0\\
P'(\mathrm{H_j}) - P(\mathrm{H_j}) &= -\frac{k\,\overline{h}^2\,\overline{\alpha}}{n\,(n\,\alpha + k\,\overline{h}\,\overline{\alpha})} < 0,
\end{aligned}$$

for $i = 1, \ldots, k$ and $j = k+1, \ldots, n$. This completes the proof of Proposition 2. ∎

A.3 Proof of Proposition 3

Setting again $\mathrm{H}_0 := \mathrm{H}$, we first calculate

$$\begin{aligned}
P(\mathrm{E}_1) &= \sum_{l=0}^{n} \sum_{H_l} \sum_{R} P(\mathrm{E}_1|H_l, R)\,P(R)\,P(H_l)\\
&= a\,\overline{r}\,h + (r + a\,\overline{r})\,\frac{\overline{h}}{n} + (n-1)\,a\,\overline{r}\,\frac{\overline{h}}{n}\\
&= \frac{1}{n}\,(n\,a\,\overline{r} + r\,\overline{h}).
\end{aligned}$$

With $P'(\mathrm{H_i}) = P(\mathrm{H_i}|\mathrm{E}_1) = \sum_R P(\mathrm{E}_1|\mathrm{H_i}, R)\,P(R)\,P(\mathrm{H_i})/P(\mathrm{E}_1)$ (for $i = 0, \ldots, n$) we can now calculate

$$\begin{aligned}
P'(\mathrm{H}) &= \frac{a\,\overline{r}\,n\,h}{n\,a\,\overline{r} + r\,\overline{h}}\\
P'(\mathrm{H}_1) &= \frac{(r + a\,\overline{r})\,\overline{h}}{n\,a\,\overline{r} + r\,\overline{h}}\\
P'(\mathrm{H}_j) &= \frac{a\,\overline{r}\,\overline{h}}{n\,a\,\overline{r} + r\,\overline{h}}\,,
\end{aligned}$$

for $j = 2, \ldots, n$. Next, we calculate

$$P'(\mathrm{H}) - P(\mathrm{H}) = -\frac{h\,\overline{h}\,r}{n\,\alpha + \overline{h}\,\overline{\alpha}} < 0.$$

Similarly,

$$P'(\mathrm{H}_1) - P(\mathrm{H}_1) = \frac{(n - \overline{h})\,\overline{h}\,r}{n\,(n\,a\,\overline{r} + r\,\overline{h})} > 0.$$

Finally, for $j = 2, \ldots, n$,

$$P'(\mathrm{H_j}) - P(\mathrm{H_j}) = -\frac{\overline{h}^2\,r}{n\,(n\,a\,\overline{r} + r\,\overline{h})} < 0.$$

This completes the proof of Proposition 3. ∎

A.4 Proof of Proposition 4

As before, we first calculate

$$P(\mathrm{E}_1) = \frac{1}{n}\,\left(n\,(h\,\alpha + \overline{h}\,\beta) + \overline{h}\,\overline{\beta}\right).$$

With this, we can now calculate

$$\begin{aligned} P'(\mathrm{H}) &= \frac{n\,h\,\alpha}{n\,(h\,\alpha + \overline{h}\,\beta) + \overline{h}\,\overline{\beta}} \\ P'(\mathrm{H}_1) &= \frac{\overline{h}}{n\,(h\,\alpha + \overline{h}\,\beta) + \overline{h}\,\overline{\beta}} \\ P'(\mathrm{H}_i) &= \frac{\overline{h}\,\beta}{n\,(h\,\alpha + \overline{h}\,\beta) + \overline{h}\,\overline{\beta}}\,, \end{aligned}$$

for $i = 2, \ldots, n$.

Next, we calculate

$$P'(\mathrm{H}) - P(\mathrm{H}) = \frac{h\,\overline{h}\cdot\left[n\,(\alpha - \beta) - \overline{\beta}\right]}{n\,(h\,\alpha + \overline{h}\,\beta) + \overline{h}\,\overline{\beta}}.$$

Similarly,

$$P'(\mathrm{H}_1) - P(\mathrm{H}_1) = \frac{\overline{h}\cdot\left[n\,h\,\overline{\alpha} + (n-1)\,\overline{h}\,\overline{\beta}\right]}{n\,\left(n\,(h\,\alpha + \overline{h}\,\beta) + \overline{h}\,\overline{\beta}\right)} > 0.$$

Finally, for $i = 2, \ldots, n$,

$$P'(\mathrm{H_i}) - P(\mathrm{H_i}) = -\frac{\overline{h}\cdot\left[n\,h\,(\alpha - \beta) + \overline{h}\,\overline{\beta}\right]}{n\,\left(n\,(h\,\alpha + \overline{h}\,\beta) + \overline{h}\,\overline{\beta}\right)}.$$

From this, Proposition 4 follows immediately. ∎

References

[1] Luc Bovens and Stephan Hartmann. *Bayeian Epistemology*. Oxford University Press, Oxford, 2003.

[2] Judicial College. The crown court compendium – part i: Jury and trial management and summing up. *Online at www.judiciary.gov.uk/publications/crown-court-bench-book-directing-the-jury-2/*, pages 5–8, February 2017.

[3] Barbara Davidson and Robert Pargetter. Guilt beyond reasonable doubt. *Australasian Journal of Philosophy*, 65(2):182–187, 1987.

[4] Richard Dawid, Stephan Hartmann, and Jan Sprenger. The no alternatives argument. *The British Journal for the Philosophy of Science*, 66(1):213–234, 2015.

[5] Robert H Ennis. Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher*, 18(3):4–10, 1989.

[6] David M Green and John A Swets. *Signal detection theory and psychophysics*. Wiley, New York, 1966.

[7] Ulrike Hahn, Christoph Merdes, and Momme von Sydow. How good is your evidence and how would you know? *Topics in Cognitive Science*, 10(4):660–678, 2018.

[8] Ulrike Hahn and Mike Oaksford. The burden of proof and its role in argumentation. *Argumentation*, 21(1):39–61, 2007.

[9] Colin Howson and Peter Urbach. *Scientific Reasoning: The Bayesian Approach*. Open Court, Chicago, 2005.

[10] David A Lagnado, Norman Fenton, and Martin Neil. Legal idioms: a framework for evidential reasoning. *Argument & Computation*, 4(1):46–63, 2013.

[11] Lucien LeCam. On some asymptotic properties of maximum likelihood estimates and related bayes estimates. *Univ. California Pub. Statist.*, 1:277–330, 1953.

[12] Hannes Leitgeb. The stability theory of belief. *Philosophical Review*, 123(2):131–171, 2014.

[13] John E McPeck. *Critical Thinking and Education*. Routledge, London, 2016.

[14] Richard Pettigrew. *Accuracy and the Laws of Credence*. Oxford University Press, Oxford, 2016.

[15] Federico Picinali. Two meanings of reasonableness: Dispelling the floatingreasonable doubt. *The Modern Law Review*, 76(5):845–875, 2013.

[16] Roger D Rosenkrantz. The justification of induction. *Philosophy of Science*, 59(4):527–539, 1992.

[17] Wesley Salmon. Rationality and objectivity in science or Tom Kuhn meets Tom Bayes. In W. Savage, editor, *Scientific Theories*, pages 175–204. University of Minnesota Press, Minneapolis, MN, 1990.

[18] Jan Sprenger and Stephan Hartmann. *Bayeian Philosophy of Science*. Oxford University Press, Oxford, 2019.

[19] Franco Taroni, Colin GG Aitken, Paolo Garbolino, and Alex Biedermann. *Bayesian Networks and Probabilistic Inference in Forensic Science*. Wiley Chichester, 2006.

[20] Peter Tillers and Eric D Green, editors. *Probability and Inference in the Law of Evidence. The Uses and Limits of Bayesianism*. Springer, Berlin, 1988.

[21] Bas C. Van Fraassen. Fine-grained opinion, probability, and the logic of full belief. *Journal of Philosophical Logic*, 24(4):349–377, 1995.

[22] Martin L Weitzman. On modeling and interpreting the economics of catastrophic climate change. *The Review of Economics and Statistics*, 91(1):1–19, 2009.

[23] Gregory Wheeler. A review of the lottery paradox. *Probability and Inference: Essays in Honour of Henry E. Kyburg, Jr*, pages 1–31, 2007.