Probability reasoning in judicial fact-finding

Ian Hunt[*]   The Hon Mr Justice Mostyn[#]

[*]School of Mathematics, Monash University, Australia

[#]Royal Courts of Justice, UK

Accepted Manuscript: The International Journal of Evidence & Proof (2019)

Abstract

We argue that the laws of probability promote coherent fact-finding and avoid potentially unjust logical contradictions. But we do not argue that a probabilistic Bayesian approach is sufficient or even necessary for good fact-finding. First, we explain the use of probability reasoning in *Re D (a Child)* [2014] EWHC 121 (Fam) and *Re L (A Child)* [2017] EWHC 3707 (Fam). Then we criticise the attack on this probabilistic reasoning found in *Re A (Children)* [2018] EWCA Civ 1718, which is the appeal decision on *Re L*. We conclude that the attack is unjustified and that the probability statements in the two cases were both valid and useful. We also use probabilistic reasoning to enlighten legal principles related to inherent probability, the Binary Method and the blue bus paradox.

Keywords: Fact-finding; Bayes' formula; Laws of Probability; Inference to the Best Explanation; Relative Plausibility; Inherent Probability; Blue Bus Paradox; The Binary Method

Acknowledgements

We wish to acknowledge valuable comments and criticisms on draft forms of this paper from Peter Mirfield, Ho Hock Lai, Edward Cheng, Ronald Allen and an anonymous reviewer.

*Corresponding author: ihunt@bunhill.co.uk;ian.hunt@monash.edu

In this article, we argue that the laws of probability can promote coherent fact-finding and avoid logical contradictions. We assume that the laws of probability hold, that Bayes' formula is valid and that probability is interpreted as subjective degrees of belief. Our argument is essentially that probability reasoning is therapeutic: an "elementary probabilistic model of degrees of belief often contains just the right balance of accuracy and simplicity to enable us to command a clear view of the issues and see where we [are or could be] going wrong" (Horwich, 1993, page 62).

An example of therapeutic probabilistic reasoning in fact-finding is as follows. Assume that there are three mutually exclusive and exhaustive explanations for something that happened and that the judge reckons that the probability of each explanation being true is less than 0.5. Also assume that on pain of incoherence the judge ensures that the sum of her subjective probabilities is one. In this case, no legal facts can be found "on the balance of probabilities"; to find otherwise would imply a contradiction in terms of the laws of probability (in particular that the three probabilities must sum to one) or require a post-hoc fix to the originally reckoned probabilities. Worse than a mere contradiction or a fix-up of the odds, a finding of fact for an event with probability less than 0.5 risks serious injustice.

We analyse two real cases in which judges apply similar reasoning. These cases are *Re D (a Child)* [2014] EWHC 121 (Fam) and *Re L (A Child)* [2017] EWHC 3707 (Fam). We also analyse *Re A (Children)* [2018] EWCA Civ 1718, which is the appeal decision on *Re L*. The ruling in *Re A* admonishes judges to avoid using the laws of probability in findings of fact, going as far to suggest that referring to the probability of a past event is pseudo-mathematics. In this article we respectfully respond to the criticisms launched by *Re A*. We conclude that the probability references in *Re D* and *Re L* were careful, justified and useful. We acknowledge that there is an element of vindicatorship in these conclusions. But we carefully quote from the cases concerned, so a reader can make her own mind up about the bones of contention; and we have made an honest effort to provide reasons for our conclusions.

The rest of this paper is structured as follows. First, we explain the reasoning behind the fact-finding decisions in *Re D* and *Re L*. Secondly, we give a generous interpretation to the objections to this reasoning that were raised in *Re A*. Thirdly, we criticise the best objections and set out our case for using subjective probability arguments in judicial fact-finding. Fourthly, we argue that probabilistic reasoning enlightens, in a therapeutic sense, legal principles related to inherent probability, the Binary Method and the blue bus paradox.

The first three sections are written in the style of an opinionated "case comment". Our responses to the judgments are coloured by our respective experiences as an implicated judge and an independent statistician. The last section is in the context of particular principles related to fact-finding.[1] Here we take the opportunity to further explicate our theoretical position, which is essentially that Bayesian reasoning has an important therapeutic, but not defining, role to play in judicial fact-finding. Implicitly, we also argue that subjective probabilistic reasoning can be well constrained by legal principles; hence, we conclude that there is little risk in allowing judges to use the laws of probability, especially when it helps *explain* their ultimate fact-finding decisions, as it did in *Re D* and *Re L*. Our theoretical position borrows much from the pragmatic philosophical stance of Horwich (1993) and concurs with Friedman (1997), who concludes on page 291 that "[i]t is necessary to keep Bayesian methods in their proper place with respect to juridical proof. For the most part, they are of analytical assistance only, to those who think about and craft evidentiary law — but for that purpose they are of very great assistance indeed."

1 Illustrative examples

The two examples in this section are real cases in which there are three competing explanations for something that happened. In both cases, each scenario has a probability reckoned by the judge and the probabilities aggregate to one, according to the laws of probability.

1.1 Example I

The first example is from *Re D (a Child)* [2014] EWHC 121 (Fam), which Mostyn J tried in January 2014. The factual issue was whether the mother had turned off the oxygen supply to her seriously ill daughter. The parties in the case agreed that there was a closed class of possible scenarios, namely:

i. The oxygen supply was not in fact turned off, and Nurse G was mistaken in believing

[1] We limit our analysis to fact-finding by judges and do not discuss civil jury instructions.

that it was; or

ii. The oxygen supply was accidentally turned off by Student Nurse J; or

iii. The oxygen supply was deliberately turned off by the mother.

Mostyn J analysed the problem as follows at [34] – [39]:

"34. Counsel for the Local Authority asks me to consider scenario (i) first. She invites me to find first on the balance of probabilities that the oxygen supply was indeed turned off and that Nurse G is not mistaken about that. As I will explain, I accept that submission notwithstanding that I have some serious concerns that I may well be wrong. I will find on the barest balance of probability that the supply was turned off. I appreciate that in a different context in *Re B (Care Proceedings: Standard of Proof)* at para 44 Lady Hale stated that 'it is positively unhelpful to have the sort of indication of percentages that the judge was invited to give in this case'. However I do not think that prevents me from indicating, only for the sake of example, that the probability that the supply was turned off was 55% (or as the mathematicians would say P = 0.55 and Q = 0.45). Indeed, were I not to do so I believe that a serious injustice may well arise in this and other cases, for the reasons that follow.

35. If I approach the exercise in the staged way suggested by Counsel for the Local Authority then the 55% probability which I ascribe to scenario (i) is converted by reason of Lord Hoffmann's binary method of judging to a 100% certainty (or P = 1). What is a mere likelihood (in the true sense of the word) is transmuted into a certainty. The 45% probability that the oxygen supply was not turned off simply will not feature in the second stage which inquires into who turned it off.

36. This is a very problematic and an arguably illogical method of proceeding. What it means is that were I to adopt it I would be left with a straight binary choice between J and the mother. If I decide that on the balance of probabilities it was not the mother (i.e. that the probability of her having done it was less than 50%) then it has to follow, so the argument goes, that J did it by accident

as the relevant probabilities of the scenarios (here, in stage 2, only (ii) and (iii)) have to add up to one.

37. But that is a flawed approach. It puts up a false choice. Let us say, for the sake of example (and I am not actually deciding this, for reasons which I will explain), that I conclude that as between the mother and J the probability is 40/60 then the true probabilities of the three scenarios are: i) The oxygen supply was not turned off: P1 = 0.45 (and thus Q1 = 0.55) ii) J turned it off by accident: P2 = (0.6 x Q1) = 0.33 iii) The mother turned it off deliberately: P3 = (0.4 x Q1) = 0.22 It can be seen that the sum of the relevant probabilities (P1 + P2 + P3) is 1, which is what it has to be. The probabilities of the competing scenarios have to add up to 1, no more, no less. There is no scope for some unallocated probability, as the House of Lords in *Re B (Care Proceedings: Standard of Proof)* made abundantly clear. If Counsel for the Local Authority's technique were followed the relevant probabilities of the competing scenarios would add up to more than 1, which is completely impossible.

38. It can also be seen that in neither of scenarios (ii) and (iii) is the probability more than 50%, or anything approaching that, and so on this analysis if I find on the balance of probability that the mother did not turn off the oxygen supply deliberately, then it just does not follow that it is more likely than not that J did so by accident. If the judging exercise is done in parallel, rather than in series, as I believe it must, then it can be seen that the least unlikely explanation is in fact that the oxygen supply was not turned off.

39. Put another way, a way which is less numeric and more linguistic, if there is an alleged primary harmful act and a whodunit between two possible perpetrators then in deciding the whodunit the possibility that the primary act was not in fact harmful has to be taken into account."

We argue in this paper that there is no fault in the reasoning above.[2] And we argue that a serious injustice may have eventuated if the laws of probability had not guided the reasoning of Mostyn J: if an "overall" or "intuitive" or "relatively plausible" approach had

[2]In section 3.3 we discuss the "binary method" to which Mostyn J refers at [35].

been followed, as urged by counsel for the local authority, then the possibility that the tap was not turned off would have been left out of account, with the result potentially being a finding that the mother turned off the tap.

1.2 Laws of probability

We assume that the laws of probability are best defined by Kolmogorov's axioms (Kolmogorov, 1933). These axioms have several practical implications. First, all probabilities must be between zero and one. Secondly, the total probability over a set of mutually exclusive events must sum to one. Thirdly, Bayes' formula can be derived directly from the laws of probability; Bayes' formula simply states that "[t]he probability that two subsequent events will both happen is a ratio compounded of the probability of the 1st, and the probability of the 2d on supposition the 1st happens" (Bayes, 1763, proposition 3).

In this paper, we define a "Bayesian" approach to reasoning as one that simply *uses* Bayes' formula in an epistemic context. This stands in contrast to "Bayesianism" which we regard as a somewhat nebulous doctrine that expresses *all* uncertainty as probability and is often associated directly with decision theory.

1.3 Example 1: comment

The purpose of this section is to highlight the piecemeal approach to probability reasoning in *Re D* and how this relates to Bayes' formula. First, assume the following:

- X = "the oxygen supply [tap] was deliberately turned off by the mother";
- A = "the tap was turned off";
- B = "the mother turned the tap off deliberately"; and
- C = "Nurse J turned the tap off accidentally".

The key scenario to which Mostyn J must assign a probability in this case is whether "the oxygen supply [a tap connected to an infant child] was deliberately turned off by the mother." The probability of X could be estimated directly: a judge could assess the available evidence and background information, and reasonably form a probabilistic belief about whether or not event X happened. Mostyn J does not take this approach. Assume that X is logically equivalent to the conjunction of two events, A and B, where A is “the tap was turned off” and B is “the mother turned the tap off deliberately”. Splitting X in this way is not trivial or without purpose, as the approach of Mostyn J demonstrates.

In the hypothetical example at [37] of the judgment, Mostyn J uses a *piecemeal* approach to find the probability of X. In the example, he calculates the probability that both A and B occurred together, as a way of determining whether X occurred. The formula he uses equates the joint probability of events A and B occurring with the probability of A occurring, multiplied by the probability of B occurring when it is assumed that A is true. This is Bayes' formula. So Mostyn J's approach is consistent with the laws of probability.

In this case all parties were agreed that there were three mutually exclusive scenarios that could explain what happened. For the sake of clarifying the calculations involved, assume that the following sentences are logically equivalent to those three scenarios:

i. the tap was *not* turned off (“not A”); or

ii. the tap was turned off *and* Nurse J turned off the tap by accident (“A and C”); or

iii. the tap was turned off *and* the mother turned off the tap deliberately (“A and B”).

The following three formulas help explain the example calculations in the judgment at [37]. Using the same values as the example, the laws of probability and Bayes' formula, the following relationships hold:

- $\mathbb{P}(not\text{A}) = 1 - \mathbb{P}(\text{A}) = 1 - .55 = .45$;
- $\mathbb{P}(\text{A}and\text{C}) = \mathbb{P}(\text{A})\mathbb{P}(\text{C}given\text{A}) = .55 * .6 = .33$;
- $\mathbb{P}(\text{A}and\text{B}) = \mathbb{P}(\text{A})\mathbb{P}(\text{B}given\text{A}) = .55 * .4 = .22$;

where $\mathbb{P}$ denotes probability, “A*and*B” means both A and B occurring, and “B*given*A” means B occurring conditional upon A having occurred.[3]

Bayes' formula can take a variety of other forms. For any A and B, these forms include:

[3] By the laws of probability, for any A and B we also have that $\mathbb{P}(\text{B}) = \mathbb{P}(\text{A})\mathbb{P}(\text{B}given\text{A}) + \mathbb{P}(not\text{A})\mathbb{P}(\text{B}given\ not\text{A})$; in our case this means that we have $\mathbb{P}(\text{B}) = \mathbb{P}(\text{A}and\text{B})$ because *not*A implies B cannot occur, which in turn means $\mathbb{P}(\text{B}given\ not\text{A}) = 0$. The same follows for $\mathbb{P}(\text{C})$.

- $\mathbb{P}(\text{A}\textit{and}\text{B}) = \mathbb{P}(\text{A})\mathbb{P}(\text{B}\textit{given}\text{A})$

- $\mathbb{P}(\text{A}\textit{and}\text{B}) = \mathbb{P}(\text{B})\mathbb{P}(\text{A}\textit{given}\text{B})$

- $\mathbb{P}(\text{B}\textit{given}\text{A}) = \mathbb{P}(\text{A}\textit{and}\text{B})/\mathbb{P}(\text{A})$

- $\mathbb{P}(\text{A}\textit{given}\text{B}) = \mathbb{P}(\text{A}\textit{and}\text{B})/\mathbb{P}(\text{B})$

Depending on the question at hand, any form of Bayes' formula can be used, supposing that you have the required components i.e. you can calculate the left-hand side by plugging in values on the right-hand side. In this case, Mostyn J finds it appropriate to use the first form of Bayes' formula listed above as an exemplar: he gives values for $\mathbb{P}(\text{A})$ and $\mathbb{P}(\text{B}\textit{given}\text{A})$, and calculates $\mathbb{P}(\text{A}\textit{and}\text{B})$. In other words, the probability of X is calculated on a piecemeal basis. The second version of Bayes' formula above offers an alternative piecemeal approach. There are several steps. First, note that $\mathbb{P}(\text{A}\textit{given}\text{B})$ would equal one (if the mother turned off the tap, then the tap is indeed turned off). Secondly, assume that $\mathbb{P}(\text{B})$ can be estimated directly. In other words, using this form of Bayes' formula is actually the same as estimating the probability of X directly (which, if you analyse what we have called X, A and B, is what you would expect).

Piecemeal probability analysis suits an enquiring mind that is willing to "extend the conversation", as Dennis Lindley, a prominent statistician, used to say (Lindley, 2006). In other words, when considering the probability of an event, it often proves advantageous (and prudent) to consider the truth or falsity of another event that may be related to the first.[4] This is essentially what Mostyn J has done in *Re D* and the concept of "extending the conversation" is consistent with the comment about "whodunit" scenarios at [39].

1.4 Example 2

In *Re L* Francis J was concerned with the death by strangulation of a 10-year-old girl. The possible causes considered were suicide, accident, and a perpetrated act by a member of her

[4] For any random outcomes X and Y we have $\mathbb{P}(\text{X}) = \mathbb{P}(\text{Y})\mathbb{P}(\text{X}\textit{given}\text{Y}) + \mathbb{P}(\textit{not}\text{Y})\mathbb{P}(\text{X}\textit{given not}\text{Y})$. To understand this in words, first think about all the events in which X can occur when Y occurs and then all the events in which X can occur when Y does not occur; gather together all these events and you have a set which represents all the ways in which X can occur. The same sort of logic applies to the probability of X occurring, except the ultimate "gathering together" is with a weighted sum, where the weights are $\mathbb{P}(\text{Y})$ and $\mathbb{P}(\textit{not}\text{Y})$.

family. He held at [98]:

> "98. Aggregating, as I must, the probability of suicide together with the probability of accident, I find that the aggregate of these two is more than 50 per cent. Doing the best that I can, I find that the possibility of suicide is about 10 per cent, and the possibility of accident and a perpetrated act are about 45 per cent each. It would be wrong for anyone to regard these figures as in any way accurate, for of course they are not. They persuade me, however, that the local authority has not discharged the burden of proof which is upon it. I am not satisfied, on the balance of probabilities, that this was a perpetrated act, albeit that I recognise that it is one of three possibilities. On the basis, however, that I do not discard the least probable and then allow a competition between the other two options, but that I should look at each of the alternative possibilities and aggregate them together, I am quite satisfied that the burden of proof in this case is not discharged. Accordingly, I do not find that the local authority's case is proved in respect of any of the contested issues. By application of the binary principle, it is the finding of this court that neither the father, the mother, M or N are responsible for a sexual assault on L and nor are they responsible for her death."

We believe that it was legitimate, as well as being mathematically and logically sound, for Francis J to have approached the problem in the way that he did. Probabilities assisted Francis J in the search for truth and his reasoning is straight-forward. While it is tenable to question the *values* of the subjective probabilities that Francis J settled upon, his probabilistic reasoning *process* is clear and valid — the approach is essentially the same as taken by Mostyn J in *Re D*: there are three explanations for something that happened; the judge reckons that the probability of each explanation being true is less than 0.5; and on pain of incoherence, the judge ensures that the sum of the probabilities is one.

1.5 Summary

These two real examples demonstrate how a judge's reasoning can be clearly explained with straight-forward invocations of Bayes' formula. In particular, the piecemeal construction of ultimate probabilities unclothes a judge's subjective reasoning and fact-finding rulings.

2 Arguments against probability

In this section we articulate and criticise the attack on probabilistic reasoning found in *Re A (Children)* [2018] EWCA Civ 1718, which is the appeal decision on *Re L*. There are two attacks. First, that reckoning a probability of a past event is pseudo-mathematics. Secondly, that legal fact-finding should not invoke probability reasoning in any case.

The first attack is fallacious and easily dismissed. The second we interpret as an argument for using "Inference to the Best Explanation" (or "IBE") in judicial fact-finding. We believe that there is great merit in arguing for IBE and we suggest a reconciliation between IBE and probabilistic approaches.

2.1 Source of the two attacks

The following passages contain the two attacks against using probabilistic arguments in fact-finding. In *Nulty & Ors v Milton Keynes Borough Council* [2013] EWCA Civ 15, cited approvingly by King LJ in *Re A* at [56], Toulson LJ says:

> "35. The civil 'balance of probability' test means no less and no more than that the court must be satisfied on rational and objective grounds that the case for believing that the suggested means of causation occurred is stronger than the case for not so believing. In the USA the usual formulation of this standard is a 'preponderance of the evidence'. In the British Commonwealth the generally favoured term is a 'balance of probability'. They mean the same. Sometimes the 'balance of probability' standard is expressed mathematically as '50 + % probability', but this can carry with it a danger of pseudo-mathematics, as the argument in this case demonstrated. When judging whether a case for believing that an event was caused in a particular way is stronger than the case for not so believing, the process is not scientific (although it may obviously include evaluation of scientific evidence) and to express the probability of some event having happened in percentage terms is illusory.

> 36. Mr Rigney submitted that balance of probability means a probability greater than 50%. If there is a closed list of possibilities, and if one possibility is more likely than the other, by definition that has a greater probability than 50%. If there is a closed list of more than two possibilities, the court should ascribe a probability factor to them individually in order to determine whether one had a probability figure greater than 50%.
>
> 37. I would reject that approach. It is not only over-formulaic but it is intrinsically unsound. The chances of something happening in the future may be expressed in terms of percentage. Epidemiological evidence may enable doctors to say that on average smokers increase their risk of lung cancer by X%. But you cannot properly say that there is a 25 per cent chance that something has happened: *Hotson v East Berkshire Health Authority* [1987] AC 750. Either it has or it has not. In deciding a question of past fact the court will, of course, give the answer which it believes is more likely to be (more probably) the right answer than the wrong answer, but it arrives at its conclusion by considering on an overall assessment of the evidence (i.e. on a preponderance of the evidence) whether the case for believing that the suggested event happened is more compelling than the case for not reaching that belief (which is not necessarily the same as believing positively that it did not happen)."

Following that quotation within *Re A*, King LJ concludes as follows.

> "58. In my judgment what one draws from *Popi M* and *Nulty Deceased* is that: i) Judges will decide a case on the burden of proof alone only when driven to it and where no other course is open to him given the unsatisfactory state of the evidence. ii) Consideration of such a case necessarily involves looking at the whole picture, including what gaps there are in the evidence, whether the individual factors relied upon are in themselves properly established, what factors may point away from the suggested explanation and what other explanation might fit the circumstances. iii) The court arrives at its conclusion by considering whether on an overall assessment of the evidence (i.e. on a preponderance of the evidence) the case for believing that the suggested event happened is more compelling than the

> case for not reaching that belief (which is not necessarily the same as believing positively that it did not happen) and not by reference to percentage possibilities or probabilities.
>
> 59. In my judgment the judge fell into error, not only by the use of a 'pseudo-mathematical' approach to the burden of proof, but in any event, he allowed the 'burden of proof to come to [his] rescue' prematurely. "

2.2 Attack 1: A "probability of a past event" is pseudo-mathematics

The first attack is that there is no such thing as the probability of a past event. We argue that dispelling probability reasoning on this basis is unsound: a probability assessment is legitimately and straight-forwardly interpretable as a *subjective* degree of belief. For example, if you toss a coin and it lands heads up, but you cover it up before you look at it, while as a matter of *objective* fact it is certain that it has landed heads up, the probability so far as you are concerned that it has landed heads up, should be 0.5 (at least if one appeals to the symmetry of the coin and fairness in the tossing). And if you roll a fair die and it comes up 5, but you cover it up before you look at it, then so far as you are concerned there should be a one in six chance it is a 5. There is also a one in three chance that it is a 1 or 2, and a one in three chance that it is 3 or 4. The chance that it is a 1, 2, 3 or 4 is, according to the laws of probability, the *aggregation* of the previous two chances, i.e. a two in three chance.

We believe the charge of "pseudo-mathematics" is clearly fallacious: it is untenable to claim that speaking about the probability of a past event is in any sense false, counterfeit, pretended or spurious.

2.3 Attack 2: "Inference to the Best Explanation" is best

The judgment in *Re A* contains a more subtle argument against probability reasoning. This argument is essentially that comparative *explanations* guide beliefs best. A key implication, we take it, is that the laws of probability are redundant. And the practical conclusion at [58] is that, rather than making "reference to percentage possibilities or probabilities", judges should conclude an event happened if the case for it having happened is more "compelling" than the case for it not having happened. In short, we assume the second attack on probability reasoning within *Re A* is that a judge should find a fact on the basis of an "inference to the best explanation" ("IBE").

> "An articulation of Inference to the Best Explanation might proceed in three stages: identification, matching and guiding. First we identify both the inferential and explanatory virtues. We specify what increases the probability of a hypothesis and what makes it a better potential explanation; that is, what makes a hypothesis likelier and what makes it lovelier. Second, we show that these virtues match: that the lovelier explanation is the likelier explanation, and vice versa. Third, we show that loveliness is the inquirer's guide to likeliness, that we judge the probability of a hypothesis on the basis of how good an explanation it would provide." Lipton (2003, page 120)

Lipton (2003) admits that planning to articulate IBE in this fashion is too optimistic, but his text goes a long way to justifying IBE as a general and useful rule of inference.

There is a lively and current line of academic literature that argues that IBE should supplant probabilistic reasoning in legal fact-finding. This literature provides a nuanced and rich set of arguments that compliment the related conclusions in *Re A*. An informal introduction to the debate can be found in Park et al. (2010). Formally developed arguments for IBE in fact-finding are found in Allen and Pardo (2019b) and Allen and Stein (2013). For counter arguments and support for probabilistic reasoning see Nance (2016), Cheng (2013), Kaye (2016), Schwartz and Sober (2017), Clermont (2015), Friedman (1997) and Sullivan (2019).

The key difficulty with IBE, in any context, is understanding what "explains better than" or "is more compelling than" really means. Lipton's shift in terms to "loveliness" does not help in this respect; nor does the term "relative plausibility", used by Allen and Pardo (2019b), help. And the judgment in *Re A* gives no guidance either. In other words, IBE is left vague.

The main difficulty with the purely probabilistic approaches to fact-finding is that there are so many different accounts; most of these accounts can be construed as a species of "Bayesianism" or its cousin "likelihoodism" but there are few principles or results with which a judge could select a particular version to follow.

Another problem with the academic debate, we believe, is that there is neither a clear winner in terms of *describing* how judges really do things, nor is there a winner in terms of compelling *normative* reasons for doing things in any particular way. Ultimately, IBE comes up short on description (many judgments refer explicitly to probability and we argue below that a widely held interpretation of the standard of proof contains an absolute level-of-likeliness component) and is under-developed in terms of normative force (IBE is vague). Probabilistic approaches also fail to describe behaviour very well (many judges avoid probabilistic reasoning or get simple examples of it wrong) and the idea of basing judicial fact-finding purely on probability is in a sense over-developed as a normative prescription, due to the plethora of paradigms. Whether or not we are right about these conclusions, we believe that consideration of the middle-ground is called for.

2.4 Our deflationary position

We argue that IBE and general probabilistic approaches to fact-finding can work together. A reconciliation that treats IBE and probabilistic reasoning as rough-equals will leave Bayesianism's fundamentalists incredulous (van Fraassen, 1989; Farmakis and Hartmann, 2005), but there are well grounded arguments linking Bayesian epistemology with IBE (Sprenger and Hartmann, 2019; Hartmann et al., 2017; Henderson, 2013).

In fact, we believe that probability and explanation are joined at the hip — a position that seems to be shared by Lipton:

> "Bayes' theorem [formula] provides a constraint on the rational distribution of degrees of belief, but this is compatible with the view that explanatory considerations play a crucial role in the evolution of those beliefs, and indeed a crucial role in the mechanism by which we attempt, with considerable but not complete success, to meet the constraint. That is why the Bayesian and the explanationist should be friends." Lipton (2003, page 120)

Furthermore, in the realm of forensic science and criminal justice, Jackson et al. (2015) demonstrate a compatibility between Bayesian and explanatory reasoning. We see no reason why an analogous coalition cannot work well in civil fact-finding.

In *Re L*, a *naive* application of "inference to the best explanation" is that the judge would, at the least, infer the disjunction that "either that it was an accident or a family member did it". But by overlaying an explicit probability analysis, one understands that neither of these explanations has, in the judge's mind, a probability exceeding 0.50. This is because the judge's rational degrees of belief were constrained by an alternative, albeit less likely, explanation that the death was caused by suicide. Similarly, in *Re D* the judge explains how *naively* inferring the best explanation may mean concluding that the nurse turned the tap off by accident. We argue that in both cases, a straight-forward save for common sense and justice, is the coherent logic provided by Bayes' formula. In other words, these cases involve therapeutic applications of Bayes' formula (*à la* Horwich, 1993).

But *sophisticated*, rather than naive, applications of IBE may have resulted in similar conclusions in these cases. We give two general examples of how this might work. First, a judge could initially assess what the best explanation is and then test it to ensure a particular *standard* is met. If the applicable standard was "on the balance of probability", then Bayes' formula could naturally assist in the second stage. Tuzet (2019) suggests two-step IBE processes deserve attention and further scrutiny[5] — otherwise the suggestion of using IBE is too easily dismissed on the basis that the best of a bad bunch of explanations need not be very good at all. Secondly, a judge could use IBE to compare the plaintiff's explanation versus a compound explanation containing all other possible explanations.[6]

A contentious issue is how the standard "on the preponderance of evidence", which does not mention the word probability and is typically favoured by IBE proponents, is different from the standard "on the balance of probability". Leading proponents of IBE advocate finding facts with a preponderance-style standard of proof and a naive or one-step IBE method. For example, Pardo (2013) and Allen and Pardo (2019b) offer the following example.

> "[S]uppose a plaintiff offers a story that a reasonable jury concludes is 0.4 likely and the defendant offers a story that the jury concludes is 0.2 likely. The 'greater than 0.5' standard implies that the plaintiff should lose — even though the plain-

[5]An absolute assessment stage for IBE, in vague terms of 'satisfactory' and 'good enough', has been suggested by Musgrave (1988) and Lipton (2003, 1993).

[6]Allowing parties to argue "disjoint" cases is explicitly allowed for by Allen and Pardo (2019b). Our point is that probabilistic reasoning may be of particular assistance in these cases — and nothing in Allen and Pardo (2019b) indicates that they would disagree.

> tiff's account is twice as likely to be true as the defendant's alternative account. This frustrates the goal of equalizing the risk of error; plaintiffs should not bear the risk of error for all of the unknown probability space. The mistake is to assume that any unknown possibilities favor the defendant (or the party without the burden of proof). This is inconsistent with equalizing the risk of error." Pardo (2013, pages 592-593)

This argument is interesting for three reasons. First, the argument invokes probability explicitly, but sticks to finding a fact with a naive IBE — which at least in theory suggests a role for numerical probabilities in assessing what is "best". Secondly, a purely probabilistic approach to fact-finding could also follow a standard that incorporates a "risk of error" principle — so IBE is not the only available rule of inference in this circumstance.

Thirdly, we do not believe the actual standard of proof works in this fashion. We argue that, ultimately, there is an *absolute* assessment of likeliness that guides finding of facts. The judgment in *Rhesa Shipping Co SA v Edmond and Another: The Popi M* [1985] 1 WLR 948, HL is enlightening. The "best explanation" for the fact in issue appears to be the main one offered by the plaintiff — that their ship was wrecked in an accident with an unidentified submarine; but since this event was deemed *improbable*, the plaintiff had not, according to the House of Lords, discharged their burden of proof. And in the words of Lord Brandon at [956]:

> "[T]he legal concept of proof of a case on a balance of probabilities must be applied with common sense. It requires a judge of first instance, before he finds that a particular event occurred, to be satisfied on the evidence that it is more likely to have occurred than not."

We believe this common sense approach accurately represents the current standard of proof, however it is worded.[7]

Finally, the judgment in *Re A* at [58] emphasises that judges should take into consideration the "whole picture, including what gaps there are in the evidence"; which we take

[7]Ho (2019) cites *The Popi M* as a counter-example to IBE (or relative plausibility) being a good explanation of judicial fact-finding. In their rejoinder to Ho, (Allen and Pardo, 2019a, page 211) treat the *The Popi M* as an outlier and claim "the case is absurd". Our argument is unrelated to how unusual the case is or whether IBE is a good explanatory theory of fact-finding: we believe that the judgment in *The Popi M* helps to clarify a *general principle* about the standard of proof.

to imply that a judge should only find a fact after comparing an event with all the events implied by its logical *negation*.[8] And this, in combination with our reading of IBE, renders little practical difference between standards based on preponderance of evidence or balance of probability. For example, in *Re D* the judge compared all possible scenarios, within reason, to the scenario that the mother turned off the tap. The snag in general is that we may have probability gaps or pure uncertainty related to some events within a "negation"; but in these situations we suspect that there will be significant problems with explanations, or "explanatory gaps", as appears to be the case in *The Popi M*. In practice, residual uncertainty can be dealt with by an agreement amongst parties about what constitutes the "whole picture"; such agreements are demonstrated in *Re D* and *Re L*.

2.5 Invoking probability is legitimate and natural

We argue that invoking probability within the process of fact-finding is a legitimate and natural process for a judge to consider. In part, this is because

> "[p]robability is virtually ubiquitous. It plays a role in almost all the sciences. It underpins much of the social sciences — witness the prevalent use of statistical testing, confidence intervals, regression methods, and so on. It finds its way, moreover, into much of philosophy. In epistemology, the philosophy of mind, and cognitive science, we see states of opinion being modelled by subjective probability functions, and learning being modelled by the updating of such functions. Since probability theory is central to decision theory and game theory, it has ramifications for ethics and political philosophy. It figures prominently in such staples of metaphysics as causation and laws of nature. It appears again in the philosophy of science in the analysis of confirmation of theories, scientific explanation, and in the philosophy of specific scientific theories, such as quantum mechanics, statistical mechanics, and genetics. It can even take center stage

[8] In several passages, it appears that Allen (2018) believes these sorts of negations play little or no role in fact-finding. For example, "the alternative in American litigation virtually never is that the plaintiff's explanation is false. Rather, the alternative is that the defendant's explanation is true, and thus the plaintiff's explanation is false" (Allen, 2018, page 999); and '[i]f there is no competition, the one explanation standing, whether in civil or criminal cases, wins" (Allen, 2018, page 1000).

in the philosophy of logic, the philosophy of language, and the philosophy of religion." Hájek (2018)

Probability interpretation is no problem: we can straight-forwardly interpret probability statements in fact-finding as subjective degrees of belief. And laws governing fact-finding in civil cases typically refer explicitly to probability. Furthermore, the expression of subjective probabilities is justified within the realm of expert forensic evidence (Aitken and Nordgaard, 2018; Biedermann et al., 2017).[9] Hence, we conclude that using probabilistic reasoning in civil fact-finding is *prima-facie* uncontroversial.

2.6 Bayes' formula works

We insist that subjective probabilities should be constrained by the laws of probability. This implies Bayes' formula should be adhered to. But this constraint is no inferential shackle: the usefulness of Bayes' formula is undeniable (McGrayne, 2012). For example, Simpson (2010) eloquently explains how Bayes' formula was used at Bletchley Park to decode Japanese and German communications during World War II. While Alan Turing and I. J. Good were at the forefront of developing Bayesian applications during and after the war, Simpson (2010) explains that many others at Bletchley Park were also using the laws of probability.[10]

Just because Bayes' formula works somewhere doesn't entail that it works in judicial fact-finding. But we wish to emphasise that it is an important and practical part of our conceptual heritage.

2.7 Slippery slope?

Finally, there is in *Re A*, perhaps, a hint of a "slippery slope" argument that says it is too easy to be "rescued" by "the burden of proof" when a judge refers to the laws of probability. We respectfully reject any such argument on the basis that we assume judges *are* capable of appropriately framing the issues and facts in their cases; and we do not believe that

[9]For attacks on subjectivity see Pardo and Allen (2008), Stein (1997) and the references within Aitken and Nordgaard (2018) and Biedermann et al. (2017).

[10]Francis J, the judge in *Re L*, informs us that his mother "was one of the Enigma Codebreakers, having spent two years there in the latter part of the war, and was finally decorated for it about 20 years ago when the secrets were eventually revealed".

referencing probability calculus increases the risk of unjustly sliding towards dead ends, where no facts are found. Specifically, we see no evidence in *Re D* and *Re L* that the judges failed to set appropriate contexts, within which they employed their probabilistic reasoning. This does not entail that the judges' subjective probabilities were "correct" in these cases; but we believe that the respective fact-finding processes were well *explained* themselves by the probability reasoning within the judgments.

Furthermore, we believe that probabilistic reasoning can be explanatory in itself. For example, Lord Brandon in *The Popi M* states that "[t]here are cases, however, in which, owing to the unsatisfactory state of the evidence or otherwise, deciding on the burden of proof is the only just course for him to take". The hypothetical example in *Re D* explains *how* the evidence of a case might be "unsatisfactory" in this sense.

2.8 Summary

We do not believe probability defines fact-finding. But we find no convincing argument in *Re A* that a judge should *not* invoke subjective probability, especially when explaining their fact-finding processes. Explanatory analysis and probabilistic reasoning are joined at the hip: both are important to fact-finding.

3 Legal principles, paradox and partial beliefs

In this section we seek to further demonstrate the value of probabilistic reasoning in terms of interpreting, and working with, legal principles. The principles we analyse relate to: inherent probability, which is linked to the reference class problem; the Binary Method; and the proverbial blue bus paradox. Our argument in this section, as above, is essentially that an "elementary probabilistic model of degrees of belief often contains just the right balance of accuracy and simplicity to enable us to command a clear view of the issues and see where we [are or could be] going wrong" (Horwich, 1993, page 62).

3.1 Background

Before we analyse specific principles, we need to give a disclaimer that delimits our discussion and we also assert that subjective probabilities can be constrained by principles above-and-beyond the laws of probability.

Describing how principles constrain beliefs is difficult. "Speakers and scientists employ diverse principles, but they are not conscious of them. The situation is similar in the case of inductive inference generally. Although we may partially articulate some of our inferences if, for example, we are called upon to defend them, we are not conscious of the diverse principles of inductive inference we constantly use" (Lipton, 2003, page 12). We do our best in the context of legal principles and probabilistic beliefs — from the perspectives of a judge and a statistician; and we try to articulate our position using common sense language and simple examples. Finally, we are *not* trying to outline or argue for any particular interpretation of subjective probability or credence;[11] and we sidestep the debate about what the "weight of evidence" may mean.[12]

We assume that it is not illogical, or contrary to the laws of probability, to adjust your probabilistic beliefs according to principles. For example, assume that you are subjectively reckoning a set of probabilities and that you are told that you must, according to some meaningful rule, fix a particular probability in this set equal to one (or zero or some other value between zero and one).[13] You can still always adhere to the laws of probability if you appropriately massage, change or revise some of the other probabilities within the set. Another example would be if you were told to ignore a particular piece of evidence, or at least not alter your probabilistic beliefs in case you acknowledge the existence of the evidence. The only conceptual difference these adjustments and constraints make is that your probabilities would be in a sense *conditional on following the rule or principle involved.*

[11] Though, various "logical" and "imprecise" probability interpretations are arguably related to the views we express when considering the examples of inherent probability, the reference class problem, the blue bus paradox and standpoints — for example, see Keynes (1921) and Kyburg and Teng (2001).

[12] Keynes introduces weight as follows. "As the relevant evidence at our disposal increases, the magnitude of probability may either decrease or increase, according as the new knowledge strengthens the unfavourable or favourable evidence; but something seems to have increased in either case — we have a more substantial basis on which to rest our conclusion ... New evidence will sometimes decrease the probability of [the hypothesis] but will always increase its 'weight"' (Keynes, 1921, page 77). Weight is analysed from a philosophical perspective in Joyce (2005) and with respect to judicial fact-finding it is discussed in Nance (2016).

[13] This rule is similar to the Binary Method which we discuss in section 3.3.

Finally, we acknowledge the criticism that *pure* subjective Bayesianism is "in danger of putting the cart before the horse" (Cox, 2000, page 323); a quip in which the cart is the apparatus of belief change and the horse is evidence. We are not arguing that subjective Bayesianism defines fact-finding and we see little value in either unchecked beliefs or unbridled evidence. The principles we discuss relate to how fact-finders *ought* to adjust or constrain their partial beliefs with respect to evidence, over and above the coherence constraint implicit in Bayes' formula.

3.2 "Inherent probability"

There is a legal principle in fact-finding that the "inherent probability" of an event in question must be taken in consideration — in the words of Peter Jackson J in *Re BR (Proof of Facts)* [2015] EWFC 41 at [7]:

> "The court takes account of any inherent probability or improbability of an event having occurred as part of a natural process of reasoning."

But what is an "inherent probability"? Arguably, a reasonable response is that an inherent probability of an event is the frequency with which it would be *expected* to be found within a collective, via random sampling (we assume the sampling can be hypothetical, which may incorporate qualitative features such as symmetry and proportional membership of sets); and that it is natural to constrain our beliefs by such a frequency. This idea can be traced to Frank Ramsey, among others, who as an example insisted that "it will in general be best for his degree of belief that a yellow toadstool is unwholesome to be equal to the proportion of yellow toadstools which are in fact unwholesome" (Ramsey, 1926, page 50).

We concur with the principle that probability beliefs should be constrained by expected frequencies, whether the frequencies are sustained by statistical observation, hypothetical argument or some combination of the two.[14] But this principle is vague and must be sharp-

[14] While we claim this is reasonable, we acknowledge that there are many thorny issues involved — most of which are out of the scope of our article. The discussions in Hájek (2007) and Hájek (2009) are good entry points into the philosophical literature on probability interpretations and frequency constraints; also see Gillies (2000) for a general discussion. There are also psychological challenges related to blending frequencies with single case probabilities — see in particular Gigerenzer (1994). Finally, there is a related literature within the statistical community — classic examples are Dawid (1982) and Efron (1986).

ened by a judge if it is to be useful. The main practical problem is finding an appropriate "reference class" with which an expected frequency may be derived:

> "It is obvious that every individual thing or event has an indefinite number of properties or attributes observable in it, and might therefore be considered as belonging to an indefinite number of different classes of things ... This variety of classes to which the individual may be referred owing to his possession of a multiplicity of attributes, has an important bearing on the process of inference." (Venn, 1888, pages 225-226)

We will analyse the reference class problem with an example in which different *contexts* justify different reference class selections: consider the case of J. M. Keynes who explains that "[t]o a stranger the probability that I shall send a letter to the post unstamped may be derived from the statistics of the Post Office; for me those figures would have but the slightest bearing on the question" (Keynes, 1921, page 322). A subjective interpretation of probability is suitably pluralistic here, because there is no need to assume that there is a *singular* objective or "inherent" probability that applies to Keynes.

First, take the *standpoint* of the post-master, who has no information about individual customers: the observed frequency of unstamped letters across all customers is of primary concern because Keynes belongs to the class "customer". But from the perspective of Keynes, the relevant context may, for argument's sake, focus less on anonymous statistical data and more on explanatory links between his own habits, how he deals with correspondence and a model of the proverbial absent-minded professor.

We would characterise a judge's fact-finding standpoint not as one of a stranger, like that of the post-master; rather, a judge will typically make an effort to incorporate detailed explanatory evidence, analogous to what Keynes himself would consider in the example, into her probability assessments. In concrete cases, this means that crude statistical frequencies, such as the general incidence of incestuous rape or the population-wide levels of infanticide by mothers, may *not* be important to judicial fact-finding despite superficial links to the cases concerned.

In summary, we believe that the term "inherent probability" is unnecessary essentialism; but the principle of calibrating beliefs to expected frequencies is justifiable if suitable reference classes are identified. But in some cases there may be events that have no relevant statistical or hypothetical frequencies and no natural symmetries or set memberships. In these cases we see four options for a judge to explore. First, adjust the context of the fact-finding exercise — broadening or narrowing reference classes as required. Secondly, settle for vague probability beliefs based on probability ranges or comparative assessments. Thirdly, settle for purely subjective probabilities. Fourthly, use a pure explanatory analysis of the facts in issue; which requires uncertainty to be expressed in non-probabilistic language.

3.3 The "Binary Method"

An important and operationally clear principle in judicial fact-finding is called the "Binary Method". Following this principle entails treating a fact found on the balance of probability as something with probability equal to one. The principle was affirmed by Lord Hoffmann in *Re B (Children)* [2009] 1 AC 11 at [2]:

> "If a legal rule requires a fact to be proved (a 'fact in issue'), a judge or jury must decide whether or not it happened. There is no room for a finding that it might have happened. The law operates a binary system in which the only values are 0 and 1. The fact either happened or it did not. If the tribunal is left in doubt, the doubt is resolved by a rule that one party or the other carries the burden of proof. If the party who bears the burden of proof fails to discharge it, a value of 0 is returned and the fact is treated as not having happened. If he does discharge it, a value of 1 is returned and the fact is treated as having happened."

This is consistent with, for example, the reasoning of Lord Diplock in *Mallett v McMonagle* [1970] AC 166 at [176]:

> "In determining what did happen in the past the court decides on the balance of probabilities. Anything that is more probable than not it treats as certain."

The Binary Method, in any given instance, can be understood as a straight-forward constraint on probabilistic beliefs: if a judge initially believes a relevant event has a probability exceeding 0.50 then her probability becomes one, otherwise the probability becomes zero.

Partial beliefs, in terms of subjective probabilities, are not ruled out in a reckoning phase; but *conditional on following the rule*, subsequent probability values are binary.

There is no need for conflict with this principle and reasoning with subjective probability in general. The key to avoiding problems is to follow the reasoning of Mostyn J in *Re D* at [38], in which he recommends that "the judging exercise is done in parallel, rather than in series." This entails only applying the Binary Method to the *final* probabilities of the events under consideration. To do otherwise has two serious downsides. First, the judge may violate the laws of probability when conjunctions of probabilities are involved.[15] Secondly, there is a risk of finding facts that are absurd or unjust, as explained in *Re D* at [37] and in *Re L* at [98].

A practical option for a judge facing potentially paradoxical conjunctions is to appropriately restrict which events must be decided upon. For example, we believe that in *Re D* there was no need to apply the Binary Method to the event that is *solely* concerned with the whether the tap was really turned off or not. This restriction clarifies what really matters for the outcome of this case, under both a probabilistic and explanatory analysis; but does not discard the uncertainty about whether the tap was turned off in the first place.[16] Another option, that may suit a pure IBE approach the most, is to take the event that "the mother turned the tap off" and compare it to the disjunction that "*either* the nurse turned it off *or* in fact the tap was never turned off in the first place".

3.4 The blue bus paradox

In this final section we address the so-called "blue bus paradox". This is a set-up that has a statistical frequency on which most people would base a probability. The paradox is that the supposed probability, despite being beyond the requisite 0.50 threshold, does not appear

[15] Applying the Binary Method to each element of a conjunction is essentially the "conjunction paradox", which Allen and Pardo (2019a) admit is a problem for both probabilistic and IBE accounts of judicial fact-finding. They say that the "'solution' to the problem within relative plausibility [IBE] is to notice that the legal system does not 'solve' the conjunction problem in some mysterious way ... [r]ather, the legal system reduces and domesticates the problem by distributing it over both parties' cases by essentially requiring the articulation of alternative explanations, and by relying on the parties to reduce the scope of litigated ambiguity" (Allen and Pardo, 2019a, page 213).

[16] We are unsure if this is consistent with how Allen and Pardo (2019b,a) understand the conjunction paradox can be dealt with in practice (or even if *Re D* is an example in which they would find a troubling conjunction).

to be adequate for the finding of a fact. An indicative version of the blue bus paradox is as follows.

> "Mrs. Brown is run down by a bus on Orange Street; 60 percent of the buses that travel along this street are owned by the blue bus company, and 40 percent by the red bus company. The only witness is Mrs. Brown, who is color-blind. Mrs. Brown appears to be able to establish a 0.6 probability that she was run down by a blue bus. Yet the overwhelming intuition is that the 60 percent statistic is not sufficient for Mrs. Brown to prove her case in a civil trial. Thus, the argument goes, proof involves something more than just probability." Redmayne (2008, pages 281-282).

We accept that "proof involves something more than just probability". But this conclusion does not rule out judges referring to the laws of probability, nor does it entail that judges should never appeal to partial beliefs in terms of subjective probabilities. Similarly, we have no trouble with Judea Pearl's conclusion, in the context of causality and Simpson's paradox (Malinas, 2001),[17] that "causality is governed by its own logic and that this logic requires a major extension of probability calculus" (Pearl, 2000, page 180) — but probability arguments and Bayes' formula still have an important role to play in casual claims, as Pearl himself demonstrates in his book.

So what then, are important roles that probabilistic reasoning can play in the context of legal proof and bare statistical frequencies? We see several. First, combining a variety of related statistical observations into particular probabilities and distributions.[18] Secondly, coherently explaining how subjective beliefs change in light of auxiliary probabilistic arguments, conditional on principles. For example, in real judicial processes there can be multiple statistical frequencies, some measured with more error than others, related to a fact in issue; and a wide variety of expert evidence is presented in probabilistic terms.

A pragmatic judge can grapple with statistical frequencies by referring to principles. For example, Glanville Williams said in a discussion related to the blue bus paradox and its

[17]This paradox was named after the same Simpson who was using Bayes' formula to help defeat the Japanese forces in World War II (Simpson, 2010).

[18]The simplicity of the blue bus paradox, in terms of being limited to one piece of statistical evidence, clarifies the demonstration that proof is something more than probability. But the same simplicity renders the set-up silent about the merits of probabilistic reasoning, in absolute terms or relative to an "explanatory approach", in cases with more than one piece of statistical evidence.

cousin the gatecrasher paradox:

> "Evidently, statistics cannot make good a deficiency of evidence involving the particular defendant. The true reason why the proof fails in the gatecrasher case and the Blue Bus case is that it does not sufficiently mark out the defendant from others ... This requirement that evidence should focus on the defendant must be taken to be a rule of law relating to proof, distinct from the general rule governing the quantum of proof." Williams (1979, page 305)

The "focus" principle to which Williams refers is vague and we are not suggesting that he dissolved the blue bus paradox forty years ago.[19] Unfortunately, this principle and other vague or incomplete ones from the nascent literature on "statistical evidence" may be the best we have at this stage.[20]

But vague principles are not vacuous and, given supporting explanations, they can still constrain partial beliefs in a reasonable fashion. For example, Enoch and Fisher (2015) promote a principle called "sensitivity", which we interpret here as meaning that a statistical frequency focuses on a fact in issue to the extent that the statistic involved is "counterfactually sensitive to the truth [of the fact in issue]" (Enoch and Fisher, 2015, page 557). If a judge decides that this principle applies to a particular case, he can explain why in terms of epistemic or instrumental concerns, making reference to the details of the case and legal precedent as appropriate. The blue bus case is illustrative. It could be argued by a judge in this case that the market-share statistic of 60 percent would be true whether or not the blue bus company was responsible for running down Mrs. Brown. Accordingly, a judge might conclude that, conditional on the principle of "sensitivity", his partial belief that the

[19] Hamer (1994) warns that applications of "focus rules" appear to be arbitrary. Though in many cases we would characterise judicial reference class selections, standpoints and context framing as being *ad hoc* (which is not a four letter word) and pragmatic, rather than arbitrary.

[20] We do not denigrate the academic debate on how to connect individuals to particular statistical frequencies in a judicial context, a debate in which the blue bus paradox plays an important role. This debate is decades old but is currently lively in law journals (Enoch et al., 2012; Enoch and Fisher, 2015; Stein, 2015; Brennan-Marquez, 2017; Pardo, 2019; Picinali, 2016; Pundik, 2008, 2011, 2016; Levanon, 2019) and has cross-overs with recent philosophical literature (Ho, 2008; Nance, 2016; Blome-Tillmann, 2015, 2017; Littlejohn, 2017; Smith, 2018; Pritchard, 2018; Gardiner, 2018; Moss, 2018; Jackson, 2018; Bolinger, 2018; Di Bello, 2018). We believe an excellent framing of the issues, and reasons why they remain contentious, can be found in the psychology literature from the 1990s (Wells, 1992; Gigerenzer, 1994; Koehler, 1996; Koehler and Shaviro, 1990). Given the obvious links between basing probabilities on statistical frequencies and the reference class problem, we also see value in analysing historical thought, such as Ramsey (1926) and the case of his yellow toadstools, along with Keynes (1921) and his unstamped letters (see section 3.2, above).

blue bus company is responsible for the accident does not meet the conventional standard of proof. Further *explanation* for this conclusion might be demanded if the case was real. The contentious issues would include how the sensitivity principle shapes the evidence, what alternative principles are available and which precedents are being followed. But the use of probability reasoning *per se* is straight-forward. In other words, we believe that in this case, and in general, the real bones of contention are about the horse (evidence), rather than the cart (partial beliefs and the standard of proof).

In summary, we acknowledge that there is a lack of precise and uncontentious principles with respect to using statistical frequencies in fact-finding.[21] But that does not render probabilistic *reasoning* useless or invalid, nor does it entail that subjective degrees of belief are left unconstrained.

4 Summary

We do not argue that a probabilistic Bayesian approach is sufficient or even necessary for fact-finding — "there is a difference between using probability theory as an illustrative or analytical tool and using probability theory as a sufficient basis for determining what the fact-finder should believe" (Schwartz and Sober, 2017, page 692). We have argued that probabilistic reasoning has therapeutic merit with respect to judicial fact-finding; indeed, in some cases the best explanation of a judgment might include explicit probabilistic reasoning.[22] But because of the ruling in *Re A*, whatever private inner thought processes the judge may entertain, he or she is not allowed to articulate in the judgment any reliance on the laws of probability when reaching his or her decision. What cannot be gainsaid is that there has been as a result a significant curtailment of judicial freedom in the fact-finding sphere, and that freedom is, as Winston Smith famously said in Nineteen Eighty-Four, the freedom to say that two plus two makes four.

[21] The problem of connecting statistics to probabilities is general and not new. For example, Hacking (2006, pages 12–13) argues that probability has a Janus-faced heritage; one face is aleatory and concerned with statistical frequencies and physical contingencies; the other face is epistemic, being concerned with what we reasonably believe to be the case and intertwined with what we use probability for.

[22] Professor Allen (of Allen and Pardo, 2019b), having kindly reviewed our paper, informs us that he doesn't see much to disagree about. He re-iterated that relative plausibility theorists directly embrace probability, just not as the only tool of rational thought. And he agreed that when a Bayesian approach actually works, it should be employed.

References

Aitken, C. and A. Nordgaard (2018). The roles of participants' differing background information in the evaluation of evidence. *Journal of Forensic Sciences 63*, 648–649.

Allen, R. (2018). The declining utility of analyzing burdens of persuasion. *Seton Hall Law Review 48*, 995–1015.

Allen, R. and M. Pardo (2019a). Clarifying relative plausibility: A rejoinder. *The International Journal of Evidence & Proof 23*, 205–217.

Allen, R. and M. Pardo (2019b). Relative plausibility and its critics. *The International Journal of Evidence & Proof 23*, 5–59.

Allen, R. and A. Stein (2013). Evidence, probability, and burden of proof. *Arizona Law Review 55*, 557–602.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London 53*, 370–418.

Biedermann, A., S. Bozza, F. Taroni, and C. Aitken (2017). The meaning of justified subjectivism and its role in the reconciliation of recent disagreements over forensic probabilism. *Science & Justice 57*, 477–483.

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

Blome-Tillmann, M. (2017). More likely than not: knowledge first and the role of bare statistical evidence in courts of law. *Preprint, University of Cambridge / McGill University*.

Bolinger, R. J. (2018). The rational impermissibility of accepting (some) racial generalizations. *Synthese* (May 2018), 1–17.

Brennan-Marquez, K. (2017). Plausible cause: Explanatory standards in the age of powerful machines. *Vanderbilt Law Review 70*, 1249.

Cheng, E. (2013). Reconceptualizing the burden of proof. *The Yale Law Review 122*, 1256–1279.

Clermont, K. (2015). Trial by traditional probability, relative plausibility, or belief function? *Case Western Reserve Law Review 66*, 353–391.

Cox, D. R. (2000). Comment on: The philosophy of statistics by Dennis Lindley. *Journal of the Royal Statistical Society. Series D 49*, 321–324.

Dawid, A. P. (1982). The well-calibrated Bayesian. *Journal of the American Statistical Association 77*, 605–610.

Di Bello, M. (2018). Trial by statistics: Is a high probability of guilt enough to convict? *Mind*.

Efron, B. (1986). Why isn't everyone a bayesian? *The American Statistician 40*, 1–5.

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

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

Farmakis, L. and S. Hartmann (2005). Book review: Inference to the Best Explanation by P. Lipton. *Notre Dame Philosophical Reviews, 01 June 2005*.

Friedman, R. D. (1997). Answering the bayesioskeptical challenge. *The International Journal of Evidence & Proof 1*, 276–291.

Gardiner, G. (2018). Legal burdens of proof and statistical evidence. *The Routledge Handbook of Applied Epistemology. London: Routledge*.

Gigerenzer, G. (1994). Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In G. Wright and P. Ayton (Eds.), *Subjective probability*, pp. 129–161. John Wiley & Sons.

Gillies, D. (2000). *Philosophical theories of probability.* Routledge.

Hacking, I. (2006). *The emergence of probability: A philosophical study of early ideas about probability, induction and statistical inference.* Cambridge University Press.

Hájek, A. (2007). The reference class problem is your problem too. *Synthese 156*, 563–585.

Hájek, A. (2009). Fifteen arguments against hypothetical frequentism. *Erkenntnis 70*, 211–235.

Hájek, A. (2018). Interpretations of probability. *The Stanford Encyclopedia of Philosophy*.

Hamer, D. (1994). The civil standard of proof uncertainty: Probability, belief and justice. *Sydney Law Review 16*, 506–536.

Hartmann, S., M. Tesic, and B. Eva (2017). Confirmation by explanation: A bayesian justification of IBE. *Preprint — http://philsci-archive.pitt.edu/13328/*.

Henderson, L. (2013). Bayesianism and inference to the best explanation. *The British Journal for the Philosophy of Science 65*, 687–715.

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

Ho, H. L. (2019). How plausible is the relative plausibility theory of proof? *The International Journal of Evidence & Proof 23*, 191–197.

Horwich, P. (1993). Wittgensteinian bayesianism. *Midwest Studies in Philosophy 18*, 62–77.

Jackson, E. (2018). Belief, credence, and evidence. *Synthese*, 1–20.

Jackson, G., A. Aitken, and P. Roberts (2015). *Communicating and Interpreting Statistical Evidence in the Administration of Criminal Justice: 4. Case Assessment and Interpretation of Expert Evidence*. The Royal Statistical Society.

Joyce, J. M. (2005). How probabilities reflect evidence. *Philosophical Perspectives 19*, 153–178.

Kaye, D. (2016). Digging into the foundations of evidence law. *Michigan Law Review 115*, 915–934.

Keynes, J. M. (1921). *A Treatise on Probability*. Macmillan (1948), London.

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences 19*(1), 1–53.

Koehler, J. J. and D. N. Shaviro (1990). Veridical verdicts: Increasing verdict accuracy through the use of overtly probabilistic evidence and methods. *Cornell Law Review 75*, 247–279.

Kolmogorov, A. (1933). *Foundations of the theory of probability.* New York: Chelsea.

Kyburg, H. E. and C. M. Teng (2001). *Uncertain inference.* Cambridge University Press.

Levanon, L. (2019). Statistical evidence, assertions and responsibility. *The Modern Law Review 82*(2), 269–292.

Lindley, D. (2006). *Understanding uncertainty.* Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken.

Lipton, P. (1993). Is the best good enough? In *Proceedings of the Aristotelian Society*, Volume 93, pp. 89–104.

Lipton, P. (2003). *Inference to the best explanation* (2 ed.). Routledge.

Littlejohn, C. (2017). Truth, knowledge, and the standard of proof in criminal law. *Synthese*, 1–34.

Malinas, G. (2001). Simpson's paradox: A logically benign, empirically treacherous hydra. *Monist 84*, 265–283.

McGrayne, S. B. (2012). *The Theory That Would Not Die: How Bayes' Rule Cracked the Enigma Code, Hunted Down Russian Submarines, and Emerged Triumphant from Two Centuries of Controversy.* Yale University Press.

Moss, S. (2018). *Probabilistic knowledge.* Oxford University Press.

Musgrave, A. (1988). The ultimate argument for scientific realism. In *Relativism and realism in science*, pp. 229–252. Springer.

Nance, D. (2016). *The Burdens of Proof: Discriminatory Power, Weight of Evidence, and Tenacity of Belief.* Cambridge University Press.

Pardo, M. (2013). The nature and purpose of evidence theory. *Vanderbilt Law Review 66*, 547–613.

Pardo, M. (2019). The paradoxes of legal proof: A critical guide. *Boston University Law Review 99*, 233–291.

Pardo, M. and R. Allen (2008). Juridical proof and the best explanation. *Law and Philosophy 27*, 223–268.

Park, R. C., P. Tillers, F. C. Moss, D. M. Risinger, D. H. Kaye, R. J. Allen, S. R. Gross, B. L. Hay, M. S. Pardo, and P. F. Kirgis (2010). Bayes wars redivivus — an exchange. *International Commentary on Evidence 8*(1).

Pearl, J. (2000). *Causality: Models, Reasoning and Inference* (1 ed.). Cambridge University Press.

Picinali, F. (2016). Base-rates of negative traits: Instructions for use in criminal trials. *Journal of Applied Philosophy 33*(1), 69–87.

Pritchard, D. (2018). Legal risk, legal evidence and the arithmetic of criminal justice. *Jurisprudence 9*, 108–119.

Pundik, A. (2008). Statistical evidence and individual litigants: a reconsideration of Wasserman's argument from autonomy. *The International Journal of Evidence & Proof 12*(4), 303–324.

Pundik, A. (2011). The epistemology of statistical evidence. *The International Journal of Evidence & Proof 15*(2), 117–143.

Pundik, A. (2016). Freedom and generalisation. *Oxford Journal of Legal Studies 37*(1), 189–216.

Ramsey, F. (1926). Truth and probability. In Kyburg and Smokler (Eds.), *Studies in Subjective Probability*. Kreiger (1980).

Redmayne, M. (2008). Exploring the proof paradoxes. *Legal Theory 14*, 281–309.

Schwartz, D. and E. Sober (2017). The conjunctions problem and the logic of jury findings. *William & Mary Law Review 59*, 619–692.

Simpson, E. (2010). Edward Simpson: Bayes at Bletchley Park. *Significance (the Royal Statistical Society and the ASA) 7*, 76–80.

Smith, M. (2018). When does evidence suffice for conviction? *Mind 127*(508), 1193–1218.

Sprenger, J. and S. Hartmann (2019). *Bayesian philosophy of science (forthcoming)*. Oxford University Press.

Stein, A. (1997). Bayesioskepticism justified. *The International Journal of Evidence & Proof 1*, 339–342.

Stein, A. (2015). The new doctrinalism: Implications for evidence theory. *University of Pennsylvania Law Review*, 208512107.

Sullivan, S. P. (2019). A likelihood story: The theory of legal fact-finding. *University of Colorado Law Review 90*, 1–66.

Tuzet, G. (2019). Abduction, IBE and standards of proof. *The International Journal of Evidence & Proof 23*, 114–120.

van Fraassen, B. C. (1989). *Laws and Symmetry.* Oxford University Press.

Venn, J. (1888). *The logic of chance.* London: Macmillan and Co. and New York.

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

Williams, G. (1979). Mathematics of proof I. *Criminal Law Review* (May), 297–308.