

Inference to the Best Explanation in Uncertain Evidential Situations

Borut Trpin and Max Pellert

Abstract

It has recently been argued that a non-Bayesian probabilistic version of inference to the best explanation (IBE*) has a number of advantages over Bayesian conditionalization (Douven [2013]; Douven and Wenmackers [2017]). We investigate how IBE* could be generalized to uncertain evidential situations and formulate a novel updating rule IBE**. We then inspect how it performs in comparison to its Bayesian counterpart, Jeffrey conditionalization (JC), in a number of simulations where two agents, each updating by IBE** and JC, respectively, try to detect the bias of a coin while they are only partially certain what side the coin landed on. We show that IBE** more often prescribes high probability to the actual bias than JC. We also show that this happens considerably faster, that IBE** passes higher thresholds for high probability, and that it in general leads to more accurate probability distributions than JC.

- 1 *Introduction*
- 2 *Generalizing IBE* to Uncertain Evidential Situations*
- 3 *Detecting the Bias of a Coin*
- 4 *Overall Performance of IBE** vs. JC*
- 5 *Speed of Convergence*
- 6 *The Threshold for High Subjective Probability*

1 Introduction

It has recently been argued in the literature on belief updating that at least in some cases a non-Bayesian probabilistic version of inference to the best explanation (IBE*)¹ outperforms Bayesian conditionalization on a number of pragmatic and epistemic counts (Douven [2013]; Douven and Wenmackers [2017]). As IBE* steers away from Bayesian conditionalization, which arguably leads to rational partial beliefs (see, e.g. Teller [1973]), it has also been criticized from various angles (e.g. vulnerability to Dutch books, Van Fraassen [1989], pp. 160-70, and potential incoherence, Climenhaga [2017]). Nonetheless, its proponents argue that its advantages outweigh its potential disadvantages. One aspect of the probabilistic form of inference to the best explanation that has not yet been addressed, however, is whether it could govern belief updating in uncertain evidential situations (that is, when $\Pr^*(\text{Evidence}) < 1$)². If so, how well would such a rule perform against Jeffrey conditionalization (JC), the almost universally acclaimed generalization of Bayesian conditionalization to uncertain evidential situations (Jeffrey [1983], ch. 11)? Let us first illustrate what is at stake. Consider the following example:

James lives in a safe neighborhood. He recently bought a new car. He wants to see it first thing in the morning, so he parks it below his bedroom window. He wakes up very early when it is still dark, looks outside and has the impression that his new car is missing. He is not fully certain of his observation because of the darkness, but he concludes that his car was probably stolen.

¹We denote a specific probabilistic form of inference to the best explanation defended by Douven and Wenmackers ([2017]) as IBE*, following the authors' notation.

² $\Pr^*(\cdot)$ represents the posterior probability function. $\Pr(\cdot)$ represents the prior probability function.

Putting aside that James might be overreacting, we can ask ourselves how James came to this conclusion. One could argue that he inferred it by induction: If a car is missing (as James' observation suggests) and no one else has the keys, it is usually stolen. However, it seems at least as likely that he came to his conclusion by inference to the best explanation. He never had his car stolen before and thefts are very rare in his neighbourhood. For example, he may have come up with two possible explanations: 1) His car was stolen, 2) His car was towed. To the best of his knowledge cars are not towed during the night, so the theft provides him the best explanation. Be that as it may, what we want to stress is that it intuitively seems that explanatory considerations are also at play when evidence is less than fully certain.

Inference to the best explanation as a principle opens many questions,³ but it is rather uncontroversial that explanatory considerations influence our everyday reasoning (see Lombrozo [2012] for a review of empirical evidence). As the above toy example suggests, it seems that uncertain evidence is not exempt.

The more dividing question is whether explanatory considerations should bear any normative weight in belief updating in uncertain evidential situations. We argue that at least in some cases they should. Our argument is simple: We first present simulations of coin bias detection in low light conditions where the evidence (whether the coin landed heads or tails) is uncertain. We then show that in the context of these simulations the agent who updates by IBE**, our proposed generalization of IBE* (described below), much more often becomes highly confident of the actual bias of the coin after a series of updates than the Bayesian agent who updates by JC. We also show that IBE** leads to higher probabilities ascribed to the true hypotheses and to lesser epistemic inaccuracy. That is, at least in the cases we discuss, the advantages of IBE** do not just outweigh its disadvantages (e.g. vulnerability to Dutch books because IBE** is non-Bayesian; see Armendt [1980]). Even more, we show that in the reported cases IBE** outperforms JC because the agent who updates by JC rarely assigns high

³For instance, what is the nature of explanations (Lipton [2004], ch. 2–3), how to compare the power of competing explanations (Schupbach and Sprenger [2011]), what is the difference between IBE and induction (Weintraub [2013]) and what if the best explanation is false (the bad lot objection; Van Fraassen [1989], pp. 142–3).

probability to the true hypothesis (i.e. the actual bias of the simulated coin).

2 Generalizing IBE* to Uncertain Evidential Situations

How should an explanationist⁴ update partial beliefs in uncertain evidential situations? Why do we need IBE**, the generalization of the probabilistic form of IBE we mentioned earlier (IBE*)? To spell out our proposal we need to first give a more detailed account of IBE*.

Van Fraassen ([1989], ch. 7) was perhaps the first to step towards IBE* (to show that any particular probabilistic form of IBE is inferior to Bayesian updating). Douven ([2013], 431), however, provides a more precise definition. That is, when an agent learns E , she needs to update by:

$$\Pr^*(H_i) = \frac{\Pr(H_i)\Pr(E|H_i) + f(H_i, E)}{\sum_{j=1}^n (\Pr(H_j)\Pr(E|H_j) + f(H_j, E))} \quad (\text{IBE}^*)$$

where $\{H_i\}_{i \leq n}$ is a set of mutually exclusive and jointly exhaustive hypotheses, $\Pr(\cdot)$ the prior and $\Pr^*(\cdot)$ the posterior probability function and f a function that assigns a bonus to the hypothesis (or hypotheses in cases of a tie)⁵ that provides the best explanation of the evidence.

It can be seen from Equation IBE* that standard Bayesian conditionalization is a special case of IBE* when there is no bonus (i.e. when the function f assigns 0 to all hypotheses).⁶ But not all (or even most) of our learning experiences are based on fully certain evidence. Our experiences are often such that we learn something but we cannot express what we learned because we are less than fully certain about the evidence.

Neither SC nor IBE* are defined for cases when evidence is uncertain. The standard Bayesian strategy for belief updating with uncertain evidence is to Jeffrey conditionalize (Jeffrey [1983]). That is, in most situations when a learning experience shifts agent's degrees of belief over a partition of propositions E_1, E_2, \dots, E_n from their prior degrees

⁴The term explanationist is a common shorthand for an agent who favours the hypothesis that best explains the evidence. In our case, the explanationist denotes the agent who updates by IBE* when the evidence is fully certain and by IBE** when it is not.

⁵When more hypotheses best explain the evidence, the bonus is split among them.

⁶Standard conditionalization is defined as $\Pr^*(H_i) = \Pr(H_i)\Pr(E|H_i)/\Pr(E)$.

$\Pr(E_1), \Pr(E_2), \dots, \Pr(E_n)$ to new degrees $\Pr^*(E_1), \Pr^*(E_2), \dots, \Pr^*(E_n)$, she needs to update her degree of belief in any proposition H_i by Jeffrey conditionalization (JC):

$$\Pr^*(H_i) = \sum_{k=1}^n \Pr^*(E_k) \Pr(H_i|E_k) \quad (\text{JC})$$

A probabilistic form of IBE for uncertain evidence has, however, not yet been defined. To formulate such a rule we first note that IBE* is structurally analogous to standard conditionalization, except for the additional function f that awards bonus points to the hypothesis that provides the best explanation. We are thus able to similarly formulate IBE** by adding the explanatory function f to JC, which is a generalization of standard conditionalization to uncertain evidential situations. An explanationist in an uncertain evidential situation should, therefore, update by the following rule, which we will call IBE** (in analogy with IBE*):

$$\Pr^*(H_i) = \sum_{k=1}^m \left(\Pr^*(E_k) \frac{\Pr(H_i) \Pr(E_k|H_i) + f(H_i, E_k)}{\sum_{j=1}^n (\Pr(H_j) \Pr(E_k|H_j) + f(H_j, E_k))} \right) \quad (\text{IBE}^{**})$$

where $\{H_i\}_{i \leq n}$ is, again, a set of mutually exclusive and jointly exhaustive hypotheses, and $\{E_k\}_{k \leq m}$ a partition of the sample space E in m parts.

Note that JC is a special case of IBE** when there are no bonuses (i.e. f assigns 0 to all hypotheses). Further, IBE* is a special case of IBE** when the evidence is fully certain. Standard (Bayesian) conditionalization is a special case of IBE** when the evidence is fully certain and there are no bonuses for the best explanation. In other words, IBE** generalizes standard conditionalization, JC and IBE*.

Let us now revisit the example with James and his car to show how IBE** works in practice. We can represent the effect of James' observation as a shift from his partition of prior probabilities $\{\Pr(\text{parked}), \Pr(\neg \text{parked})\}$, where $\Pr(\text{parked})$ was very high, to a new partition $\{\Pr^*(\text{parked}), \Pr^*(\neg \text{parked})\}$, where $\Pr^*(\text{parked})$ is low. Theft provides the best explanation of the uncertain evidence that prompted this shift, so IBE** instructs James to become more confident that his car was stolen than he would become if he updated in a Bayesian way (by JC).

One of the problems with both IBE* and IBE** is that the rules leave undefined the central function f which awards the bonus points to the best explanation. It should be noted that a number of competing measures of explanatory power have been proposed (for an overview and their own proposal, see Schupbach and Sprenger [2011]). The function f could, therefore, award the bonus to the best explanation after comparing the explanatory power of hypotheses according to any of these measures or in some other way. However, because the measures differ, it is possible that different hypotheses could come out as the best explanations according to different measures.

We believe that this is not a bug but a feature, as the slogan goes. Our standing here is pluralist: different measures of explanatory power may apply in different contexts. Because of this, the function f is not precisely specified. We return to this question later.

We also need to address the worry that IBE** may lead to incoherence and expose an agent to Dutch books because it breaks away from the Bayesian rules for belief updating. As we will show, this is a price that is well worth paying because in our simulations, JC, the main Bayesian update rule for uncertain evidential situations, often leads to very inaccurate partial beliefs after a longer series of updates. That is, neither rule is optimal but we argue that the disadvantages of JC are much bigger than the disadvantages of IBE** because the supposedly rational JC more often prescribes the agent to form a high degree of belief in false rather than true hypotheses.

3 Detecting the Bias of a Coin

Let us now demonstrate that, in at least some cases, IBE** performs much better than JC. Consider the following scenario (adapted from similar examples in Douven [1999], [2013]):

Ibbie, who (as her name suggests) always updates her partial beliefs by IBE**, and Jeff, a die-hard JC updater,⁷ enter a magician's room with no windows. The magician shows them 11 coins that are visually indistinguishable but each is

⁷It is well known that JC is only applicable when the rigidity condition is satisfied, i.e. when $\Pr^*(H|E_i) = \Pr(H|E_i)$ for all E_i in the evidential partition (Jeffrey [1983], p. 174). Jeff only updates by JC if this condition is satisfied.

known to have a different bias to land heads or tails. Specifically, their biases to land heads range from 0 (always tails) to 1 (always heads) in 0.1 increments. The biases for tails are converse (i.e. $\Pr(\text{Tails}|\text{Coin}_i) = 1 - \Pr(\text{Heads}|\text{Coin}_i)$ for all i). The coins can only land heads or tails. The magician tells them that he will select one of the coins and flip it. Their goal is to determine which coin he selected.

In addition to flipping the coin, the magician also tells them that he will adjust the lighting in the room. Ibbie and Jeff are able to clearly see whether a coin landed heads or tails if the lights are at the maximum level of brightness but the more the lights are dimmed, the less certain they become about the outcome of the coin toss. If the lights are off and the room is pitch black, they think it is just as likely that the coin actually landed heads as tails (assuming they hear that the coin landed). That is, their certainty about the outcome is

$\Pr^*(\text{Heads}) = \Pr^*(\text{Tails}) = 0.5$ by the principle of indifference. To simplify, assume that their certainty about the evidence (heads or tails) is in linear relation to the level of lighting in the room: $\Pr^*(\text{Outcome}) = (1 + x)/2$ where x is the level of lighting (from 0 to 1; the LIGHT/CERT assumption). Further, assume that they are both equally certain about the outcome because they are both in the same room, at nearly the same position and because they are idealized.

The bias detection has two variations: it may end when both are very confident that any of the coins were selected (i.e. $\Pr^*(H_i) > T$ for some hypothesis H_i ; T represents the threshold for high subjective probability) or when they are both very confident of which coin was selected. It ends after 1500 tosses if it does not end earlier. How will Ibbie and Jeff perform if they are not allowed to communicate?

First, note that their evidence is binary in the sense that the coin can only land heads or tails. If they are certain that the coin landed heads to some degree x , they are, therefore, also $1 - x$ certain that it may have landed heads. By way of illustration, suppose that the lighting is at 50% and the coin lands tails. By LIGHT/CERT, they are then 0.75 certain that the coin landed tails and 0.25 certain that it landed heads. Second, note that the prior probabilities that any of

	H_0	H_1	H_2	H_3	H_4	H_5	H_6	H_7	H_8	H_9	H_{10}
Bias:	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Pr_{t=0}:	1/11	1/11	1/11	1/11	1/11	1/11	1/11	1/11	1/11	1/11	1/11

Figure 1: Coins that may be used, their bias for heads, the corresponding hypotheses and their prior probability⁸

the 11 coins with different biases was selected are uniformly distributed over the 11 hypotheses for both agents (see Figure 1). This is because they have no way of knowing which coin is used (the principle of indifference). This means that they both start with $\Pr_{t=0}(H_i) = 1/11$ for all i .

Further, the agents cannot simply update by standard conditionalization or IBE* because their evidence is partitioned in the sense that they do not merely learn that the coin landed heads or tails. Instead, they learn that the coin may have landed heads to some degree x and tails to degree $1 - x$. Let us now adapt the updating rules Ibbie and Jeff use for this particular scenario.

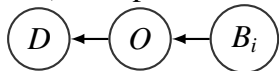
⁸The hypotheses are numbered by their bias toward heads. That is, H_i is $i/10$ biased toward heads. H_0 therefore has no relation to the null hypothesis.

Jeff updates by the following formula:⁹

$$\Pr^*(H_i) = \Pr^*(Heads) \frac{\Pr(H_i)\Pr(Heads|H_i)}{\sum_{j=1}^{11}(\Pr(H_j)\Pr(Heads|H_j))} + \Pr^*(Tails) \frac{\Pr(H_i)\Pr(Tails|H_i)}{\sum_{j=1}^{11}(\Pr(H_j)\Pr(Tails|H_j))} \quad (\text{JC variant})$$

Ibbie, on the other hand, awards bonus points to the hypothesis that best explains her uncertain evidence. She therefore updates in the following way:¹⁰

⁹JC is only appropriate if the rigidity condition is satisfied. That is, when $\Pr^*(H|E_k) = \Pr(H|E_k)$ for all k (Jeffrey [1983], p. 174). It is not straightforward to determine whether the rigidity is satisfied in our scenario. However, Schwan and Stern ([2017]) provide a convincing causal updating norm based on causal Bayes nets, which is helpful in determining when it is appropriate to update by JC. Assume that there is a dummy proposition D that represents the ineffable learning (the observation of the coin landing heads/tails in a given level of lighting). We can then represent Jeff's belief network in terms of the following causal graph where D is a variable representing the values of the dummy proposition (D or $\neg D$), O represents those of the actual outcome and B_i those of the bias of the coin:



That is, the bias of the coin in question (B_i) influences whether it actually lands heads or tails (O), which causes the observation (D). In our case, the causal updating norm is satisfied because the path between D and B_i is d-separated by O for any D .

¹⁰Recall that IBE** generalizes IBE*, JC and standard conditionalization. If we simulate the scenarios where the agents are fully certain about the outcomes (i.e. $\Pr^*(Outcome) = 1$), and set the explanatory bonus for one agent to 0.1 and 0 to the other, then by using IBE** as the general rule we replicate the findings about IBE* vs. standard conditionalization reported in (Douven [2013]).

$$\Pr^*(H_i) = \Pr^*(Heads) \frac{\Pr(H_i)\Pr(Heads|H_i) + f(H_i, Heads)}{\sum_{j=1}^{11} (\Pr(H_j)\Pr(Heads|H_j) + f(H_j, Heads))} + \Pr^*(Tails) \frac{\Pr(H_i)\Pr(Tails|H_i) + f(H_i, Tails)}{\sum_{j=1}^{11} (\Pr(H_j)\Pr(Tails|H_j) + f(H_j, Tails))} \quad (\text{IBE}^{**} \text{ variant})$$

Specifically, the function f assigns a bonus of 0.1 to the hypothesis that best explains the evidence.¹¹

As we mentioned above, it is not agreed upon how to determine which hypothesis provides the best explanation. Ibbie, for example, assigns the explanatory bonus by her own method that works well in this context. Particularly, she uses a modified version of the frequency measure (mentioned and criticized by Van Fraassen [1989]) to accommodate the function f for the less than fully certain evidence. The function f thus compares the relative frequencies of heads or tails as the most likely outcomes ($\text{FREQ}_{\text{Heads}}$, $\text{FREQ}_{\text{Tails}}$) with the relative frequency predicted by each hypothesis and awards the closest prediction(s) by assigning a bonus.¹² If two predictions are equally close, the bonus is split in half.

To illustrate: Suppose Ibbie is 0.8 certain that the coin landed heads on the first toss ($\Pr^*(Heads) = 0.8$). Her $\Pr^*(Tails)$ is therefore 0.2. The most likely outcome is that the coin landed heads because $\Pr^*(Heads) > 0.5$ and $\Pr^*(Tails) < 0.5$. The frequency of heads as the most likely outcome is then $\text{FREQ}_{\text{Heads}} = 1/1$, which is exactly as predicted by H_{10} . The function f , therefore, awards the bonus for providing the best explanation of the most likely outcome (heads) to H_{10} . On the other hand, tails was not the most likely outcome, so $\text{FREQ}_{\text{Tails}} = 0/1$ and H_{10} again receives the bonus as the best explanation of the fact that the coin did not land tails. That is, $f(H_i, Heads)$ and $f(H_i, Tails)$ both assign the bonus to H_{10} .¹³

¹¹We follow Douven ([1999], [2013]) with the bonus of 0.1, although it could also be higher or lower. A systematic investigation is beyond the scope of this paper. However, the data below (e.g. Table 1, Table 3) shows that the arbitrary 0.1 bonus gives an edge to IBE**.

¹²The assignment of the bonus thus combines the strengths of frequentism with those of Bayesianism.

¹³The function f which awards the explanatory bonus in the described way always awards the same hypothesis for both parts of agent's uncertain evidence because

Now suppose that the most likely sequence of the first four tosses was $HTHH$ where H stands for heads and T for tails. $FREQ_{\text{Heads}}$ is then $3/4$ (0.75) and H_7 and H_8 according to which the coin has the bias 0.7 and 0.8 for heads both get half the bonus on the fourth update for providing the best explanation of the observed outcomes.

If Ibbie used some other likelihood-based measure of explanatory power, the function f would assign the bonus to hypotheses at either side of the extremes, that is, the hypotheses that the coin is fully or almost fully biased toward heads or tails because these hypotheses are the most likely given the outcome of each single coin toss.¹⁴ However, to determine the bias of the coin, a longer series of outcomes needs to be taken into account, and this is where Ibbie's measure (FREQ) shines.

Further, if Ibbie wanted to provide the best explanation of her actual observations (and not the most likely outcomes), then the function f would lead her astray. To illustrate the point, suppose the coin lands heads on every toss in the first 10 rounds and she is always 0.8 certain that it landed heads. The best explanation of her actual observations is then that the coin has the 0.8 bias for heads and H_8 gets the explanatory bonus in each round. This is, however, obviously wrong because she is not interested in explaining her certainty. What she is interested in is, instead, how biased the coin actually is and she, therefore, needs to look for the best explanation of the most likely outcomes.

4 Overall Performance of IBE** vs. JC

With the particularities of Ibbie's and Jeff's updating in place, we can now move on to their actual performance. We developed a computer simulation to test how well they perform in detecting the bias of a coin in the scenario described above. We ran simulations with a number of different parameters:

$$FREQ_{\text{Tails}} = 1 - FREQ_{\text{Heads}}.$$

¹⁴For example, H_{10} (the coin is fully biased toward heads) provides the best explanation according to the measure proposed by Good ([1960]) in all but the second round of the sequence $HTHH$. H_0 provides the best explanation in the second round when tails is the most likely outcome.

- The certainty of the outcome (the level of lighting) may be constant throughout the bias detection or randomized on every coin toss.
- Any of the 11 coins with different biases may be used.
- The belief updating may stop when the agents' probability for any of the coins peaks above the threshold for high probability; or when it peaks above the threshold for the actually selected coin; or after a preset number of rounds.
- The threshold probability when they settle on any (or the true) hypothesis may vary. The default threshold probabilities we used were 0.99 (following Douven [2013]) and 0.95 (more on the thresholds in section 6).

We are most interested in knowing how often the agents first assign high probability (over 0.95 or 0.99) to the true hypothesis (true positive), how often they assign high probability to one of the false hypotheses and how often they remain neutral in the sense that no hypothesis is assigned high probability after a large number of rounds¹⁵ (we call this neutrality). Our assumption is that were Jeff and Ibbie to actually participate in a similar real-world scenario, they would become confident that one of the coins was selected when their subjective probability passes the threshold (0.95 or 0.99).

We ran 55,000 simulations with maximally 1500 rounds of updates in a number of different settings. Particularly, we ran 1000 simulations for each of the 11 possible coins in 5 different levels of lighting, so that the agents were either constantly 0.6, 0.7, 0.8 or 0.9 certain about the outcomes (heads or tails), or that the level of certainty was randomized in every round ($\text{Pr}^*(\text{Outcome})$ was pseudorandomly generated from a uniform distribution in the range of (0.5, 1)).

At first glance, our overall findings do not look particularly bad for the Jeffreyan agent (see Table 1). Ibbie became confident of the true hypothesis more often than Jeff, that is in approximately 50% of the cases, while Jeff only did in 16% or 11% of the cases (depending

¹⁵We simulated 1500 rounds because we found that the probabilities assigned to hypotheses receive very minor changes after so many updates. Further, we also believe that 1500 rounds of updates is already a very long sequence.

	JC		IBE**	
$\Pr^*(H_{true}) > 0.95$	8789	16%	27,034	49%
$\Pr^*(H_{true}) > 0.99$	5951	11%	28,385	52%
$\Pr^*(H_{false}) > 0.95$	13,001	24%	16,707	30%
$\Pr^*(H_{false}) > 0.99$	9140	17%	11,978	22%
$\Pr^*(H_i) < 0.95$ for all i	33,210	60%	11,259	20%
$\Pr^*(H_i) < 0.99$ for all i	39,909	73%	14,637	27%

Table 1: Number and percentage of simulations where agents assigned probability above threshold to the true or false hypothesis and when they did not reach the threshold

on the threshold). However, Ibbie’s updating strategy was bolder, so she more often became very confident of the false hypothesis (in 30% or 22% of the cases depending on the threshold; compare to Jeff’s 24% or 17% cases). Jeff, on the other hand, appears to be more cautious: if the evidence is not reliable in the sense that he is not fully certain about it, JC often prescribes him to stay neutral (i.e. to not assign very high probability to any hypothesis).

However, we believe that this is not the best way to look at our findings. Our epistemic goals may differ, but there are two generally agreed upon desiderata: (1) to assign high probability to true hypotheses, and (2) to not assign high probability to false hypotheses. Now suppose the magician agrees with this principle and awards 1 point to each agent who becomes confident of the true hypothesis and takes 1 point back if they become confident of any of the false hypotheses. The magician neither awards nor takes any points for neutrality (i.e. for not passing the threshold). Ibbie wins with flying colours because Jeff, unlike her, more often becomes highly confident of the false hypothesis than of the true hypothesis.

A critical reader could raise the objection that Jeff’s neutrality should also be taken into account. We can accept this objection. But how should neutrality be scored? We believe that the following two principles are hard to disagree with: (1) it is worse to be neutral than confident of the true hypothesis, and (2) it is better to be neutral than confident of the false hypothesis. The points for neutrality should therefore be somewhere in between the points awarded or taken for high confidence of the true and the false hypotheses.

Depending on our epistemic values, we may take neutrality to be either a virtue or a vice. If we follow the former, we believe that points should be awarded for neutrality, otherwise we think the points should be taken. We are pluralists on this topic in the sense that neutrality may

Scoring rule	T – F		2 × T – 2 × F + N		2 × T – 2 × F – N	
	JC	IBE**	JC	IBE**	JC	IBE**
Threshold						
0.95	-4212	10,327	24,786	31,913	-41,634	9395
0.99	-3189	16,407	33,531	47,451	-46,287	18,177

Table 2: Different scoring rules for neutrality and true/false hypotheses above the threshold

be desirable in some and undesirable in other contexts. In the present context, however, Ibbie still wins even if neutrality is seen as a virtue. Suppose the magician awards two points when an agent is confident of the true and takes two points when she is confident of the false hypothesis. If the magician awards or takes one point when they remain neutral, Jeff still loses on both counts (Table 2).¹⁶

5 Speed of Convergence

As we can see in Table 1, Ibbie assigns high probability to false hypotheses slightly more often. We believe that this is offset by the fact that she much more often assigns high probability to the true hypothesis than her Bayesian companion Jeff (see the scoring rules above; Table 2). However, a very critical reader could still object that assigning high probability to a false hypothesis is the worst sin a rational agent could commit. In this case, JC would be vindicated simply because it more often cautiously prescribes lower probabilities to any hypothesis and, hence, less often leads to high probability for false hypotheses. While we do not share these views, we can provide further arguments in favour of IBE** over JC.

One such advantage is pragmatic in nature. IBE** converges toward high probability much faster than JC. Particularly, in our simulations Jeff never passed the threshold for high probability faster than Ibbie. The difference in the speed was vast. There were only 30 simulations where Ibbie was 'only' up to 100 updates faster in passing the threshold. In all other simulations where at least one of them passed the threshold, Ibbie passed the threshold over 100 updates faster than her companion. Summarising, Jeff passed the threshold

¹⁶It is easy to verify that in our simulations Jeff loses unless neutrality is interpreted as a very valuable virtue, that is, unless two points are awarded/taken for true/false hypotheses above the threshold and more than 1.33 or 1.55 points are awarded for neutrality (depending on the threshold).

	JC		IBE**	
$\Pr^*(H_{true}) > 0.95$	9126 (+337)	17% (+0.6%)	41,001 (+13,967)	75% (+25%)
$\Pr^*(H_{true}) > 0.99$	5971 (+22)	11% (+0.04%)	38,258 (+11,894)	70% (+22%)
$\Pr^*(H_{true}) < 0.95$	45,874	83%	13,999	25%
$\Pr^*(H_{true}) < 0.99$	49,029	89%	16,742	30%

Table 3: Number and percentage of simulations where agents assigned probability above threshold to the true hypothesis and when they failed to do so (improvements over the first variation – see Table 1 – are marked in parentheses)

probability in fewer cases and when he did, he passed it much slower than Ibbie.

An argument could be made that the speed of convergence toward high probability is practically important but only so for the true hypothesis. Assigning high probability to a false hypothesis cannot be praised, be it fast or slow. We agree with this. However, our simulations suggest that IBE** is also able to answer to this objection.

Suppose Ibbie and Jeff participate in a slightly different scenario of bias detection. Everything remains unchanged except that the magician only cares about correct guesses. To illustrate, suppose the coin is 0.7 biased toward heads. After 400 rounds, Ibbie becomes highly confident that the coin is 0.6 biased toward heads, writes down her answer and passes it to the magician. He takes note, but the coin tossing continues. After 100 more rounds, Ibbie changes her mind and becomes highly confident that the coin is actually 0.7 biased toward heads. The magician takes note. The tossing continues for another 1000 rounds and Jeff still does not assign high probability to any hypothesis. Because the magician only considers correct answers, he disregards Ibbie’s first answer and records that she called the correct coin. He also records that Jeff was not confident about the actually selected coin – he was, after all, not confident of any coin.

As we can see in Table 3, Ibbie often corrects her false guess and comes to assign high probability to the true hypothesis later (compare with the data for the first guesses in Table 1). Her opponent Jeff, on the other hand, barely improves even if his wrong calls are disregarded. Specifically, Ibbie now makes approximately 20 – 30% more correct guesses (depending on the threshold), while Jeff’s correct guesses only increase by less than 1%.

The comparative speed of passing the threshold remains similar to that in the original example. Ibbie is, again, much faster, although Jeff now overtakes her in a small number of

Updates (IBE** – JC)	less than -100	-100 to 0	0	0 to 100	more than 100
$\Pr(H_{true}) \geq 0.95$	8	15	752	71	40,155
$\Pr(H_{true}) \geq 0.99$	28	41	133	146	37,910

Table 4: Difference between IBE** and JC in the number of updates needed to pass the threshold for the true hypothesis

simulations. However, when Ibbie is the first to pass the threshold for the true hypothesis (as in the large majority of the cases), she typically passes the threshold much faster than Jeff: she is typically over 100 steps faster. Jeff, on the other hand, is typically only up to 100 steps faster (Table 4).

The speed of convergence toward the threshold can thus be seen as not a merely pragmatic matter because Ibbie’s bolder updating allows her to often recover from the initially wrong call. Further, even if both agents become confident of the true hypothesis, the speed advantage still works in Ibbie’s favour. In a more realistic competitive scenario, Ibbie would be able to announce her scientific findings faster than Jeff. She would still be faster even if she allowed herself to keep updating beliefs for a bit longer, because, for example, she meta-epistemologically observed that IBE** sometimes initially prescribes her to endorse a false hypothesis.

Nonetheless, an objection could still be raised that in realistic scenarios (e.g. in scientific research) an agent who assigns very high probability to some hypothesis after a number of observations will typically stop further investigations (for a similar point about categorical beliefs, see Harman [1986], pp. 46-50). If we add to this that there are typically no magicians who already know which hypothesis is true so that they can ignore false calls, we then return to our initial findings: Ibbie is faster, but also more often endorses the false hypothesis.

We accept this objection. One of the fundamental goals in belief updating is to avoid assigning high probability to false propositions. However, when all things are considered, we still come to the conclusion that (1) adopting IBE** more often leads to endorsing the true hypothesis, and that (2) this happens much faster than with JC.

6 The Threshold for High Subjective Probability

We assume that Jeff and Ibbie call one of the coins if their subjective probability for that coin is sufficiently high (see Lewis [1976], p. 303; Jackson [1979], p. 565, but also Douven [2006], pp. 457–9). We set the default thresholds at 0.95 and 0.99, which seem reasonably very high (see Douven [2013], p. 433 in support of 0.99). However, high subjective probability is a vague term, so we also compared how the agents perform when the threshold is lower.

The talk of thresholds for sufficiently high subjective probability is familiar from the Lockean thesis (Foley [1992]; Hawthorne [2009]).¹⁷ We take it that the threshold for high subjective probability needs to be higher than 0.5.¹⁸ 0.5 is a rather low threshold but we found that if the threshold is only slightly lowered (from 0.99 to 0.95), IBE** already leads to many more cases where the probability for the true hypothesis passes it. JC, on the other hand, in general requires the threshold to be much lower and even then often fails to pass it for the true hypothesis. For instance, if the evidence is constantly 0.8 certain and the coin is fully biased, JC does not even pass the minimum threshold of 0.5 (Figure 3). IBE** is therefore more robust with respect to different threshold values.

This also shows an interesting change in performance related to thresholds when the evidence is fully certain (compare Douven [2013]). In this case, the two agents nearly always pass even the very high threshold of 0.99 regardless of the bias of the coin used in the simulation.

On the other hand, if the evidence is less than fully certain (like in our simulations) the thresholds that may be reached get lower. Additionally, the thresholds now also hinge on the actual bias of the coin used in the simulation.

¹⁷We remain largely silent about the Lockean thesis here because we are only addressing the relation between assertion and high probability.

¹⁸If we instead settled on a relational threshold, then an agent could call the coin that is merely 9.1% probable as long as the other coins are less probable. But this would at the same time mean that a (rational) agent is 90.9% confident about the negation, that is, that the most probable coin was not selected (100% – 9.1%). It, therefore, does not make much sense to set the threshold below 0.5.

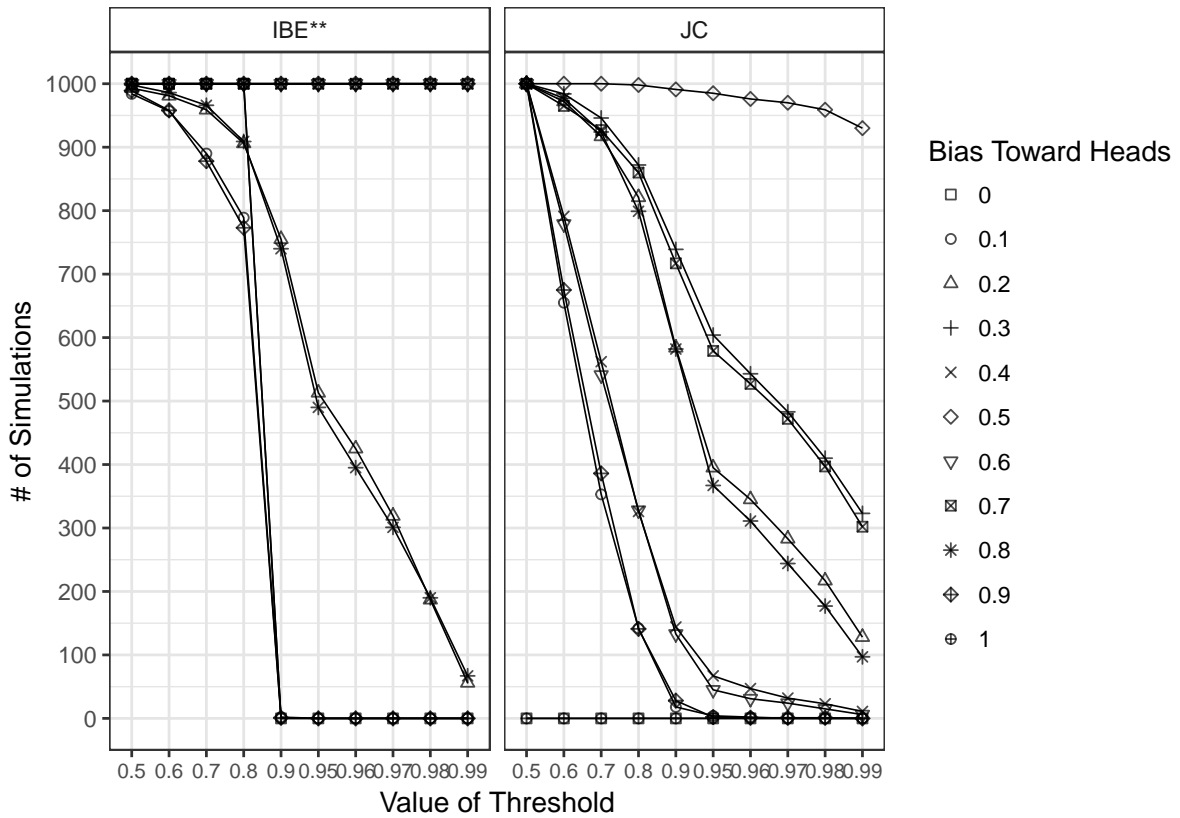


Figure 2: Number of simulations with probability for any hypothesis above threshold (certainty of outcomes constant at 0.8)

For illustration, consider the simulations where the agent’s certainty about the outcomes was fixed at 0.8 in every round of updating. As we can see in Figures 2 and 3, Ibbie has problems passing the higher thresholds (Threshold > 0.9) when the coin is fully or almost fully biased. Jeff, on the other hand, has problems passing even the minimum threshold of 0.5.

It is interesting to note that the more the coins are biased either toward heads or tails, the more problematic it is for both agents to assign high probability to any hypothesis. This is particularly problematic for JC, although IBE** also somewhat demonstrates this trend. This comes as a bit of a surprise because we intuitively expect the agents to perform better at least when the coins are fully biased. The sequence of uncertain observations is, after all, simpler (always somewhat certainly heads/tails). For example, if we limit ourselves to the case of a coin that always lands heads when the outcomes are 0.8 certain (Figure 4), we can see that Jeff becomes more confident of the hypotheses that the coin is more biased toward heads than tails. However, after approximately 15 rounds his expectations of heads and tails stabilize very

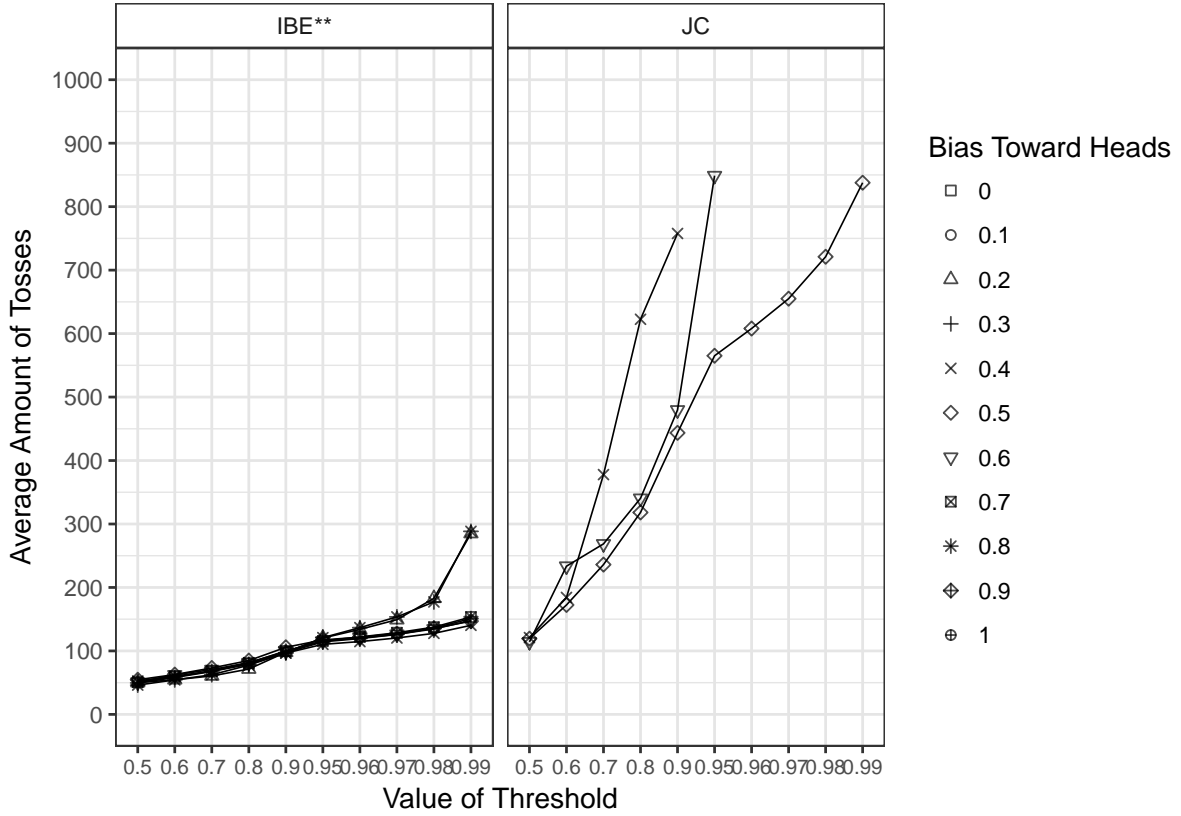


Figure 3: Average number of tosses needed to pass different thresholds (certainty of outcomes constant at 0.8)

close to 0.8 and 0.2, respectively. That is, $\Pr(\text{Heads}) = \sum_{i=1}^{11} (\Pr(H_i)\Pr(\text{Heads}|H_i)) \approx 0.8$; and $\Pr(\text{Tails}) \approx 0.2$.

JC updating comes to a halt when the newly learned evidence is in line with prior expectations. If the expectation is very close to the learned evidence, it similarly only prescribes minor updates. This is easy to verify. Suppose the expectation of the evidence matches the certainty of evidence: $\Pr(\text{Heads}) = \Pr^*(\text{Heads}) = 0.8$ and $\Pr(\text{Tails}) = \Pr^*(\text{Tails}) = 0.2$. Jeff then follows the formula:

$$\begin{aligned} \Pr^*(\text{Heads}) \frac{\Pr(\text{Heads}|H_i)\Pr(H_i)}{\Pr(\text{Heads})} + \Pr^*(\text{Tails}) \frac{\Pr(\text{Tails}|H_i)\Pr(H_i)}{\Pr(\text{Tails})} \\ = \Pr(\text{Heads}|H_i)\Pr(H_i) + \Pr(\text{Tails}|H_i)\Pr(H_i) = \Pr(H_i) \end{aligned}$$

for all H_i . Jeff, therefore, stops updating (Figure 4). Ibbie, on the other hand, assigns a small

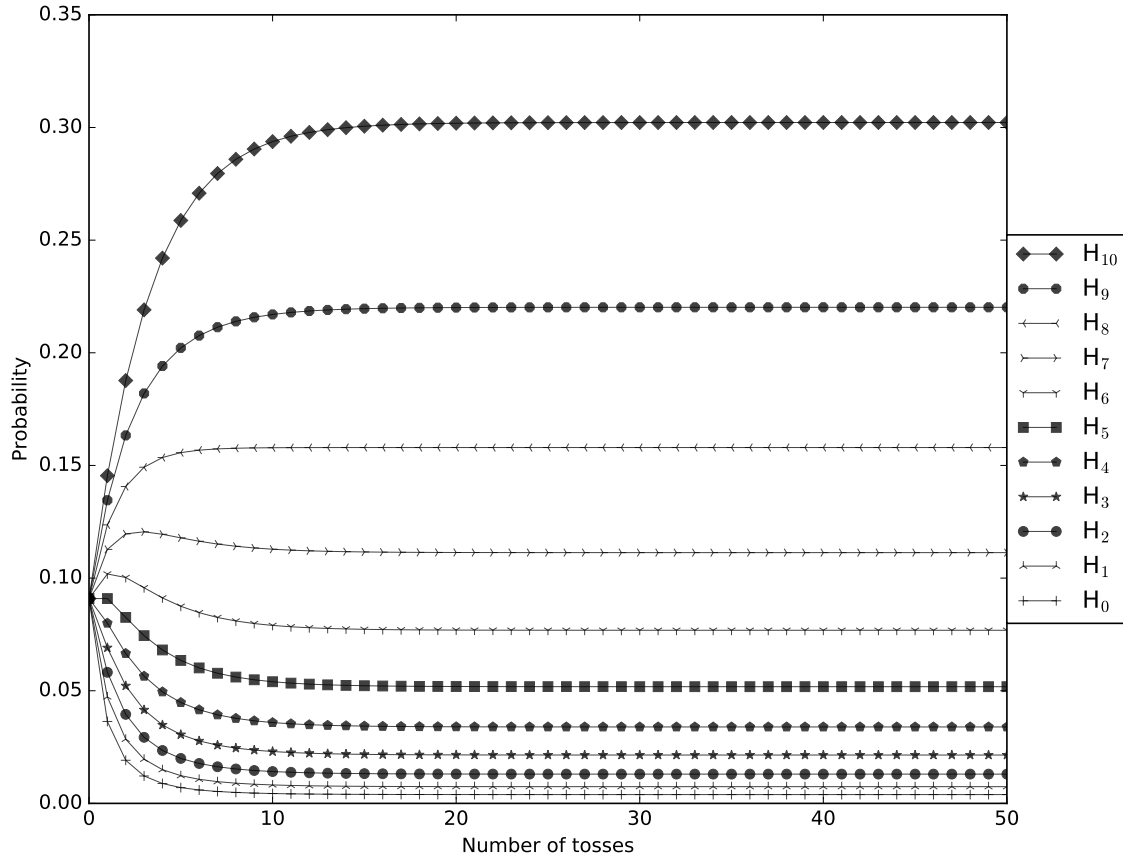


Figure 4: Subjective probabilities prescribed by JC when the coin is fully biased toward heads and the certainty of evidence is constantly at 0.8

bonus to the best explanation and thus manages to reach the threshold of 0.9 in contrast to Jeff who stays around 0.3.

IBE** thus helps counteract the problem of uncertainty of evidence. Imagine a case where Jeff and Ibbie are trying to diagnose a patient instead of a biased coin. They both inspect a number of symptoms. Because they are only partially certain about their observations of the symptoms (e.g. due to technical limitations of their diagnostic tools), they update their beliefs according to JC and IBE**. However, Jeff's findings are inconclusive, while Ibbie on the other hand identifies the correct disease with high probability. An objection could be raised that Ibbie might be ascribing high probability to the false hypothesis – but, as we have seen, Jeff also almost as often ascribes high probability to a false hypothesis in addition to often not passing the threshold.

Now imagine Jeff telling the patient that she most likely has the disease X, but another

disease Y is also almost as likely, and there is a number of less likely diseases – however, it is actually more likely that she has neither of these diseases. This is not very useful because Jeff is (in this example) not the kind of doctor who entertains a number of what-if counterfactual scenarios for his own speculative joy. Rather, he needs to be the kind of doctor who comes up with a conclusive diagnosis and later adapts it if it turns out to be incorrect.

It is also of interest why the agents perform better when the coin is less biased (preferably fair; see Figure 5). The explanation of this phenomenon goes beyond the scope of this paper but it is again related to the mechanism of JC, which is structurally so similar to IBE** that some of its shortcomings are reflected in both rules (though IBE** still performs better). In short: if prior expectation of heads and tails keeps increasing and decreasing because of the observations and subsequent updates, then the hypothesis with the closest prediction receives smaller increases and decreases (corrections) than the others and prevails after a number of rounds.

On the other hand, if the coin has a stronger bias, there are fewer corrections and the hypothesis that is the closest to the expectations of evidence has less chance of passing the threshold. IBE** overcomes these problems to some extent because it favours the hypotheses that are in line with the sequence of the most likely outcomes. Nonetheless, the more biased coins still have problems passing the higher thresholds (> 0.9). However, when following JC means not even passing the 0.5 threshold, IBE is still favoured.

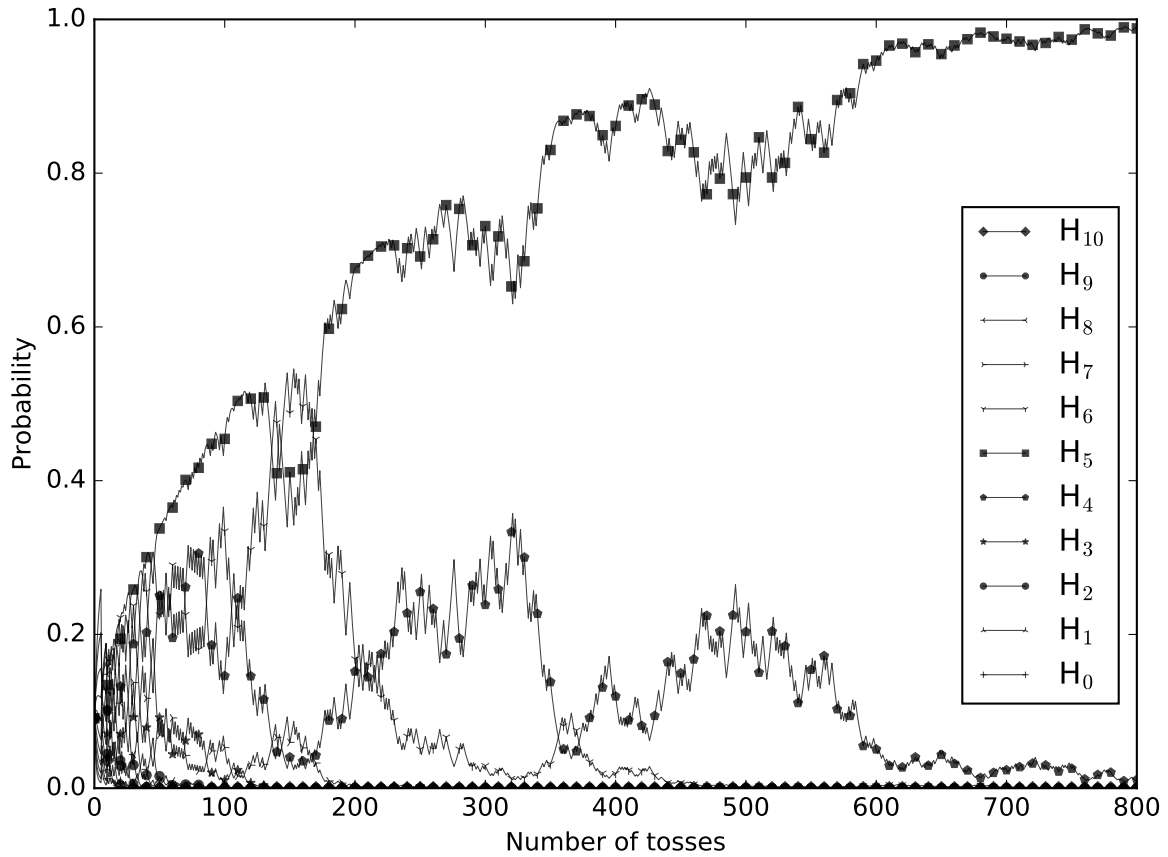


Figure 5: Subjective probabilities prescribed by JC when the coin is fair (H_5) and the certainty of evidence is constantly at 0.8 in a randomly selected simulation

7 Epistemic Inaccuracy

To counter the last line of defense in favour of JC, we need to also address epistemic inaccuracy (in the sense of Joyce [1998]; Leitgeb and Pettigrew [2010a]). The defense of JC has often been based on pragmatic grounds, particularly because it is invulnerable to Dutch books (Armendt [1980]; Skyrms [1987]). We have demonstrated that this advantage is offset by a number of problems, for instance, the problem of threshold probabilities for true hypotheses, slower convergence to the truth and, particularly, frequent assigning of high probability to false hypotheses (even if the agent’s false calls are disregarded; Table 3).

The latter brings us to the problem of epistemic inaccuracy. In short, epistemic inaccuracy measures how (in)accurate an agent’s subjective probabilities are in relation to the truth. For example, if it rains and you are fully certain that it rains ($\Pr(\text{rain})=1$), then your subjective probability for rain is completely accurate. If you are certain that it does not rain, then it is

completely inaccurate. But this is just the actual (synchronic) epistemic inaccuracy.

Leitgeb and Pettigrew ([2010b]) prove that standard Bayesian conditionalization minimizes expected inaccuracy. The authors also show that JC, on the other hand, does not necessarily minimize the expected inaccuracy. They then propose an alternative update rule instead; let us call it the Leitgeb-Pettigrew conditionalization (LPC). Levinstein ([2012]), however, shows that LPC may lead to absurd updates and instead endorses JC because there supposedly is no better alternative; Levinstein goes on to instead question the inaccuracy scoring measure (the Brier score). It seems that the question of epistemic inaccuracy of JC is, therefore, at least still open.

Nonetheless, we can still compute both Brier scores and logarithmic scores and simply compare how the two updating rules (JC and IBE**) fare in this aspect. Brier scores are a simple way to measure the accuracy of probability predictions. If the prediction is correct, the score is 0. If it is false, it increases the further off the prediction is; this is also why it is otherwise known as the Brier penalty. It originated from meteorology where it was used to assess the precision of weather predictions.

We use a version of the Brier score for binary events because the coins in our simulations may only land heads or tails.¹⁹ We compute Brier penalties after every belief update with the following formula:

$$BS = (1 - \Pr(H_t))^2 + \sum_{i=1}^{10} (0 - \Pr(H_f))^2 \quad (\text{BS})$$

where H_t represents the true hypothesis (the coin that is actually used in the simulation) and H_f represents the other (false) hypotheses.

It can be seen from the Equation BS that the agent who ascribes higher probability to the true hypothesis and lower probability to false hypotheses incurs a lower penalty and is, therefore, more accurate.²⁰

¹⁹In the original formulation of Brier ([1950], p. 1), a formula is used that allows for more than two classes of events happening.

²⁰There are also other scoring rules that may be used to assess the epistemic accuracy but they are less useful for the present simulations. We mentioned the log score which calculates

Update nr.	100	250	500	750	1000
IBE**	0.35	0.12	0.04	0.03	0.02
JC	0.77	0.86	0.97	1.03	1.06

Table 5: Average Brier scores in 55,000 simulations of different certainties of evidence and biases

Brier scores reveal an interesting insight in favour of IBE** in comparison to JC. If we take into account all 55,000 simulations in different settings, we see that, as the simulations progress, the Brier penalties for Ibbie on average decrease while Jeff’s increase (Table 5). This is a clear sign that JC does not minimize epistemic inaccuracy; on the contrary: it leads to ever higher inaccuracy.

To see exactly how this happens, consider the Brier scores in simulations where the certainty of evidence changes every time the coin is tossed (Figure 6; suppose the level of lighting in the room randomly changes all the time). Regardless of the coin that is used in the simulation, Ibbie sooner or later assigns high probability to the true coin even if she doesn’t reach the very high levels and the Brier penalties reflect this. Jeff, on the other hand, detects the true coin if the coin is fair and also moves toward higher probabilities for the true coin if it is almost fair (bias 0.4 or 0.6). Because the penalties keep increasing for all the other coins, his average Brier penalties, however, turn out to be much larger than Ibbie’s. This suggests that JC in our simulations on average prescribes ever more inaccurate subjective probabilities, while IBE** moves toward higher accuracy.

Note that our argument for lesser epistemic inaccuracy of IBE** in comparison to JC is based on the initial example introduced in section 3 in which all of the coins had an equal prior $\frac{1}{2}$ for the negative logarithm for the probability of the true hypothesis, $LS = -\log \Pr(H_t)$. The range of LS is in $[0, \infty)$. However, the agents sometimes assign near-zero probability to the true hypothesis after a number of rounds and incur a very large penalty. This, in turn, distorts the average log-penalties incurred over a number of simulations. For example, if an agent assigns near-zero probability to the true hypothesis in only one of the 1000 simulations, the average log score of all simulations turns out to be misleadingly high. LS is, therefore, not very representative in this context. Nonetheless, IBE** also outperforms JC when log scores are compared.

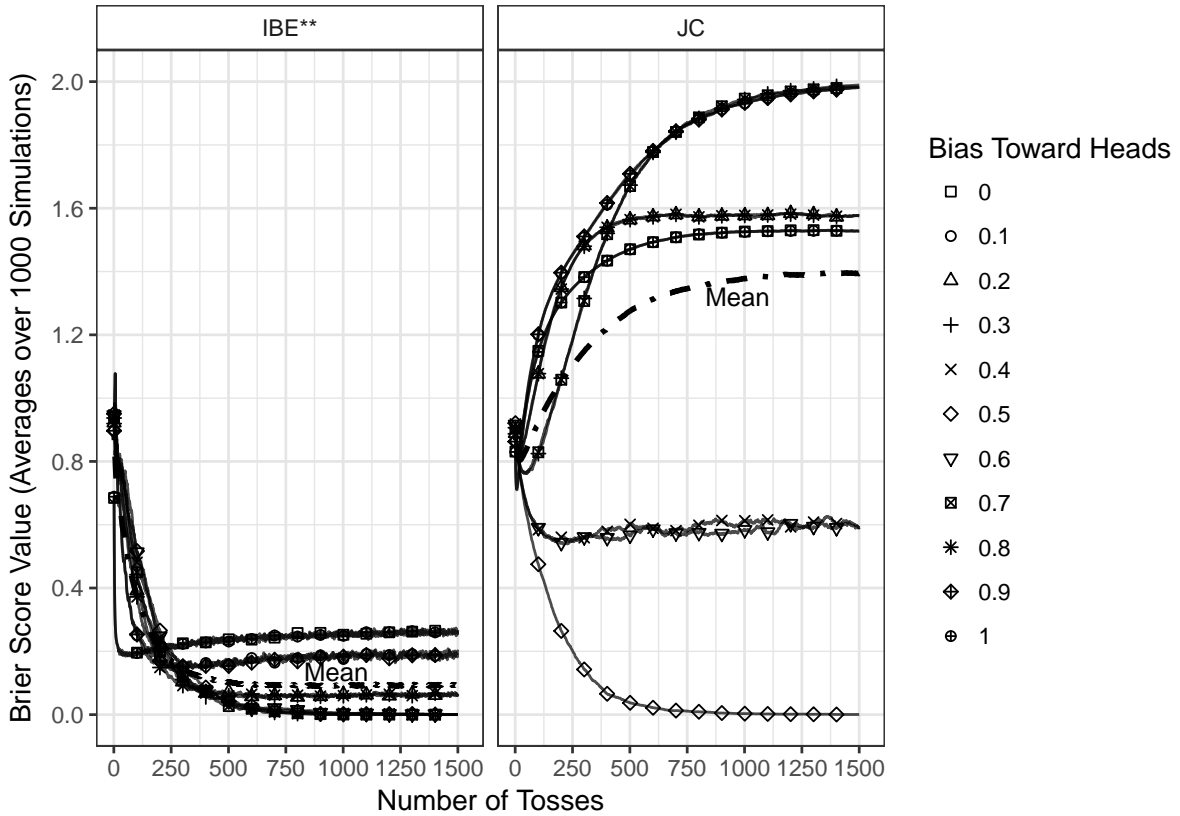


Figure 6: Average Brier scores (over 1000 simulations) when the certainty of evidence is randomized in every round

probability of being flipped and the selected coin was flipped many times (up to 1500 times). The priors were, therefore, uninformative and the evidence was abundant. An argument could be made that IBE** fared much better than JC because it appears to be an update rule that (a) gives extra weight to evidence, and (b) less weight to prior probabilities. If this is the case, then IBE** may only lead to lesser epistemic inaccuracy when both (a) and (b) are met, as was the case in our simulations. A defender of JC could then argue that JC would turn out to be a better updating rule in situations when (a) and (b) are not met, and further, that in the same situations IBE** might lead an agent astray and thus to greater overall inaccuracy.

We can address this worry by considering a variation of our original example. Suppose, for example, that the coins still have the same biases (i.e. biases toward heads from 0 to 1 in 0.1 increments) except that there are now 100 coins: 50 coins that are 0.5 biased toward heads, 21 coins that are 0.6 biased toward heads, 21 coins that are 0.4 biased toward heads and 1 of each of the other coins. Further, suppose that both agents, Jeff and Ibbie, are aware of the

	H_0	H_1	H_2	H_3	H_4	H_5	H_6	H_7	H_8	H_9	H_{10}
Bias:	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Quantity:	1	1	1	1	21	50	21	1	1	1	1
$\Pr_{t=\theta}$:	0.01	0.01	0.01	0.01	0.21	0.5	0.21	0.01	0.01	0.01	0.01

Figure 7: Coins that may be selected, their bias for heads, their quantity, the corresponding hypotheses and their prior probability

Update nr.	100	250	500	750	1000
IBE**	0.54	0.20	0.04	0.01	0.005
JC	0.48	0.39	0.36	0.3550	0.3560

Table 6: Average Brier scores in 500,000 simulations of the scenario where prior probabilities are informative

distribution of coins and their prior probabilities reflect that (see Figure 7). We can now simulate 100 trials (one with each coin) and inspect how JC and IBE** fare when priors are informative.²¹

To do so, we ran 500,000 simulations with 1500 rounds of updates in a number of different settings. Particularly, we again simulated 5 different levels of lighting, in the sense that the agents were either constantly 0.6, 0.7, 0.8 or 0.9 certain about their evidence (whether a coin landed heads or tails), or that the level of their certainty was randomized in every round. We then ran 50,000 simulations with a fair coin (H_5), 21,000 simulations with each of the coins that are 0.6 and 0.4 biased toward heads (H_6 and H_4) and 1000 simulations of each of the other coins in each of the 5 described evidential setups. Our findings show that when priors are informative JC performs better than when they are not. It leads to lesser overall epistemic inaccuracy than it did in the original example – its average Brier scores (penalties) for all five different levels of evidential uncertainty initially decrease and then stabilize around 0.4 (compared to increasing penalties that stabilized around 1 in the original example; Table 5). IBE**, on the other hand, initially leads to greater epistemic inaccuracy than JC (especially in the first approximately 50 updates, see Figure 8). This insight is also in line with findings reported by Douven ([2013], p. 441), where the author reports that an explanationist updating

²¹We are thankful to an anonymous referee for expressing the worry described in the previous paragraph and for suggesting this variation of our original example.

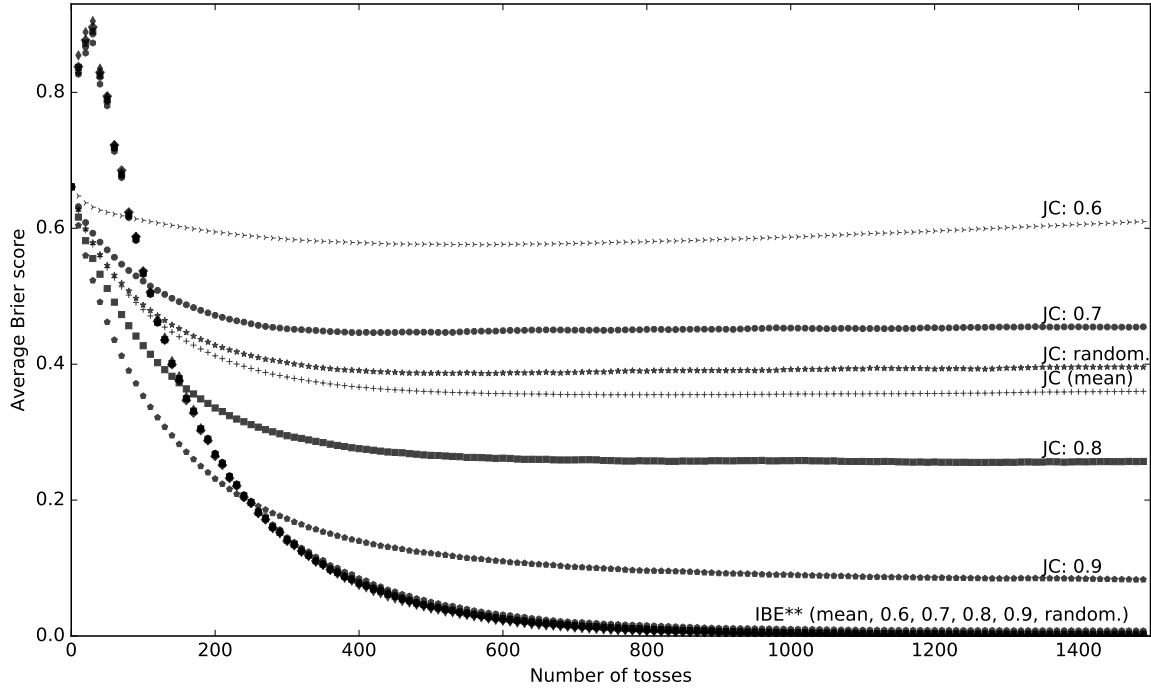


Figure 8: Average Brier scores (over 500,000 simulations) incurred by JC and IBE** given different levels of evidential uncertainty

	JC		IBE**	
$\Pr^*(H_{false}) > 0.95$	15,401	3%	233,882	47%
$\Pr^*(H_{false}) > 0.99$	5779	1%	184,437	37%

Table 7: Number and percentage of simulations where agents initially assigned very high probability to the false hypothesis

by IBE* (a special case of IBE**) reacts to evidence in bolder fashion and is thus often initially lead astray. However, when the updating continues, IBE** leads to an almost fully accurate probability distribution (i.e. its Brier penalties converge toward 0). JC, however, continues to prescribe such belief updates that the agent remains at more or less the same level of epistemic inaccuracy (see Table 6 and Figure 8).

It is also worth noting that the epistemic accuracy of JC depends on the level of evidential uncertainty. That is to say that in our simulations of coin bias detection, JC prescribed epistemically more accurate degrees of belief when the agent was more certain about the evidence (and vice versa) as we can see in Figure 8. This reveals an interesting insight in comparison to IBE** whose epistemic performance was not affected by different levels of evidential uncertainty (note the overlapping scores for IBE** in Figure 8).

8 Conclusions

We have shown that in cases of uncertain evidence IBE** performs much better than JC for a number of reasons: it is faster, it is more accurate (at least in the long run), it passes the thresholds for high probabilities more often and, most importantly, it much more often prescribes the agent to form high probability in the true hypothesis.

Our crucial finding is that assigning bonus points for the best explanation seems to counteract the problem of uncertain evidence which is apparently often problematic for JC. That is to say that even if the agents are able to update their beliefs many times the JC updater more often assigns very high probability to the false hypothesis than to the true hypothesis. The IBE** updater, on the other hand, performs much better on both epistemic and pragmatic counts because she much more often assigns high probability to the true hypothesis (at least over a large number of updates) and because she is able to assign high probability to the true hypothesis faster. Additionally, our simulations also suggest that the epistemic performance of JC (but not that of IBE**) deteriorates with increasing uncertainty of evidence (Figure 8).

This suggests that explanatory considerations also need to be considered in uncertain evidential situations. IBE** updating offers advantages not only from a practical viewpoint (e.g. concerning speed) but also on epistemic grounds: Its adoption leads to much more often assigning high subjective probability to true than false hypotheses; it does, however, more often initially lead to high probability for the false hypothesis than JC (see Table 1 and Table 7) because IBE** gives extra weight to evidence.

This indicates that updating by IBE** is particularly successful when evidence is abundant and when an agent is able to update her degrees of belief many times. IBE**, however, is not particularly apt as an updating rule if an agent has a limited access to the evidence and, hence, only updates her degrees of belief on a limited number of occasions. The evidence may be misleading (in the shorter run), so giving it extra weight may initially lead the agent astray. On the other hand, IBE** (unlike JC) typically prescribes the agent to assign very high subjective probability to one of the hypotheses in her hypothesis space relatively fast (see, e.g. Figure 3). As we have seen (e.g. Table 3), this means that even if the agent may initially be led astray, she typically ends up assigning very high subjective probability to the true hypothesis. We can

thus expect IBE** to provide a particularly useful update rule for uncertain evidential situations in which it is important that an agent clearly identifies one of the hypotheses as the strongest contender. If, in addition, an agent has access to abundant evidence, we can also expect IBE** to identify the true hypothesis.

An interesting venue for future research can thus be established with the analysis of how JC and IBE** perform when they are forced to make decisions on the basis of their inferences. The idea is that the agents make a choice after reaching some threshold. How often will they choose to do the right or the wrong thing or withhold from making any decisions? The performance evaluation of each updating rule would depend on the positive or negative utilities assigned to right or wrong decisions (or no decisions). These utilities, in turn, depend upon the context of the scenario being considered – is it a diagnosis that has to be ascribed (caution is required), or is it a scenario in which any decision for action is better than being passive (action is required)? Our illustrative scoring (see Table 2) suggests that at least in the longer run IBE** will generally perform better than JC (but see also Footnote 16). Further investigations could, however, identify the specific settings which are particularly apt for IBE** or JC when it comes to decision making. Be that for example in relation to the risk-benefit of decisions, the thresholds for full-blown decision-making, the role of the size of explanatory bonuses assigned by function f (for IBE**) or the number of belief updates that the agents are able to undergo before their decision (or lack thereof).

In summary, our findings indicate that, at least in certain well-specified situations, IBE** is superior to JC and should be seriously considered.

Acknowledgements

We are grateful to Anton Donchev, Igor Douven, Branden Fitelson, Olga Markič and two anonymous referees for their helpful feedback on earlier versions of this paper, and Robbie Hopper for her proofreading. We also want to thank the audience at the Inferentialism, Bayesianism and Scientific Explanation Workshop in Munich, Germany, 2017 for their stimulating questions and suggestions.

Borut Trpin
University of Ljubljana
Ljubljana, Slovenia
borut.trpin@ff.uni-lj.si

Max Pellert
Complexity Science Hub Vienna
Vienna, Austria
Medical University of Vienna
Vienna, Austria
pellert@csh.ac.at

References

- Armendt, B. [1980]: ‘Is There a Dutch Book Argument for Probability Kinematics?’, *Philosophy of Science*, **47**(4), pp. 583–8.
- Brier, G. W. [1950]: ‘Verification of Forecasts Expressed in Terms of Probability’, *Monthly Weather Review*, **78**(1), pp. 1–3.
- Climenhaga, N. [2017]: ‘Inference to the Best Explanation Made Incoherent’, *Journal of Philosophy*, **114**(5), pp. 251–73.
- Douven, I. [1999]: ‘Inference to the Best Explanation Made Coherent’, *Philosophy of Science*, **66**(Supplement), pp. 424–35.
- Douven, I. [2006]: ‘Assertion, Knowledge, and Rational Credibility’, *The Philosophical Review*, **115**(4), pp. 449–85.
- Douven, I. [2013]: ‘Inference to the Best Explanation, Dutch Books, and Inaccuracy Minimisation’, *Philosophical Quarterly*, **63**(252), pp. 428–44.
- Douven, I. and Wenmackers, S. [2017]: ‘Inference to the Best Explanation versus Bayes’s Rule in a Social Setting’, *The British Journal for the Philosophy of Science*, **68**(2), pp. 535–70.

- Foley, R. [1992]: ‘The Epistemology of Belief and the Epistemology of Degrees of Belief’, *American Philosophical Quarterly*, **29**(2), pp. 111–21.
- Good, I. J. [1960]: ‘Weight of Evidence, Corroboration, Explanatory Power, Information and the Utility of Experiments’, *Journal of the Royal Statistical Society. Series B (Methodological)*, **22**(2), pp. 319–31.
- Harman, G. [1986]: *Change in View*, Cambridge: MIT Press.
- Hawthorne, J. [2009]: ‘The Lockean Thesis and the Logic of Belief’, in F. Huber and C. Schmidt-Petri (eds), *Degrees of Belief*, Synthese Library, Springer, pp. 49–74.
- Jackson, F. [1979]: ‘On Assertion and Indicative Conditionals’, *The Philosophical Review*, **88**(4), pp. 565–89.
- Jeffrey, R. [1983]: *The Logic of Decision*, Chicago: The University of Chicago Press.
- Joyce, J. M. [1998]: ‘A Nonpragmatic Vindication of Probabilism’, *Philosophy of Science*, **65**(4), pp. 575–603.
- Leitgeb, H. and Pettigrew, R. [2010a]: ‘An Objective Justification of Bayesianism I: Measuring Inaccuracy’, *Philosophy of Science*, **77**(2), pp. 201–35.
- Leitgeb, H. and Pettigrew, R. [2010b]: ‘An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy’, *Philosophy of Science*, **77**(2), pp. 236–72.
- Levinstein, B. A. [2012]: ‘Leitgeb and Pettigrew on Accuracy and Updating’, *Philosophy of Science*, **79**(3), pp. 413–24.
- Lewis, D. [1976]: ‘Probabilities of Conditionals and Conditional Probabilities’, *The Philosophical Review*, **97**(4), pp. 497–520.
- Lipton, P. [2004]: *Inference to the Best Explanation*, London: Routledge, second edition.
- Lombrozo, T. [2012]: ‘Explanation and Abductive Inference.’, in K. Holyoak and R. Morrison (eds), *Oxford Handbook of Thinking and Reasoning*, Oxford: Oxford University Press, pp. 260–76.

- Schupbach, J. N. and Sprenger, J. [2011]: ‘The Logic of Explanatory Power’, *Philosophy of Science*, **78**(1), pp. 105–27.
- Schwan, B. and Stern, R. [2017]: ‘A Causal Understanding of When and When Not to Jeffrey Conditionalize’, *Philosophers’ Imprint*, **17**(8), pp. 1–21.
- Skyrms, B. [1987]: ‘Dynamic Coherence and Probability Kinematics’, *Philosophy of Science*, **54**(1), pp. 1–20.
- Teller, P. [1973]: ‘Conditionalization and Observation’, *Synthese*, **26**(2), pp. 218–58.
- Van Fraassen, B. C. [1989]: *Laws and Symmetry*, Oxford: Oxford University Press.
- Weintraub, R. [2013]: ‘Induction and Inference to the Best Explanation’, *Philosophical Studies*, **166**(1), pp. 203–16.