Inference to the Best Explanation versus Bayes' Rule in a Social Setting

ARTICLE *in* THE BRITISH JOURNAL FOR THE PHILOSOPHY OF SCIENCE · JULY 2015

Impact Factor: 1.28 · DOI: 10.1093/bjps/axv025

Igor Douven
French National Centre for Scientific Resea…

Inference to the Best Explanation versus Bayes' Rule in a Social Setting

Igor Douven
Sciences, Normes, Décision (CNRS)
Paris-Sorbonne University
igor.douven@paris-sorbonne.fr

Sylvia Wenmackers
Faculty of Philosophy
University of Groningen
s.wenmackers@rug.nl

Abstract

This paper compares inference to the best explanation with Bayes' rule in a social setting, specifically, in the context of a variant of the Hegselmann–Krause model in which agents do not only update their belief states on the basis of evidence they receive directly from the world but also take into account the belief states of their fellow agents. So far, the update rules mentioned have been studied only in an individualistic setting, and it is known that in such a setting both have their strengths as well as their weaknesses. Here, it is shown that in a social setting, inference to the best explanation outperforms Bayes' rule according to every desirable criterion.

It is widely agreed that, at least as a matter of sociological fact, our scientific activities are aimed at more than producing predictively accurate theories. We seek for theories that also *explain* the data, that make us *understand* why the data are as they are. It is more controversial whether prediction and explanation ought to both play a part (whether or not these parts are equal) in the epistemic evaluation of scientific theories. If one theory is predictively as successful as another, but the former is a better explanation of the data than the latter, does that mean that we should put greater confidence in the former?

There is evidence indicating that people do tend to invest more confidence in a hypothesis, the better that hypothesis explains the data.[1] According to some theorists—often called "explanationists"—this tendency has a rational basis. In the opinion of these theorists, the so-called inference to the best explanation (IBE), which in its simplest form licenses an inference to the truth of the hypothesis that best explains the available data (Harman [1965], p. 324), is a key normative principle of both everyday and scientific reasoning.[2]

[1] See, e.g., (Koehler [1991]), (Pennington and Hastie [1992]), (Lombrozo [2006], [2007]), and (Lombrozo and Carey [2006]).

[2] See, among many others, (Boyd [1984]), (Musgrave [1988]), (McMullin [1992]), (Psillos [1999]), and (Lipton [2004]).

The normative status of IBE has long been questioned by advocates of Bayesian confirmation theory. According to these philosophers, IBE faces two challenges, one from dynamic Dutch book considerations, the other from inaccuracy minimization considerations. However, it was shown in (Douven [2013]) that IBE can meet both challenges: even if adherence to IBE were to make one vulnerable to Dutch books, the rule has compensating advantages; and inaccuracy minimization considerations actually favor IBE over Bayes' rule, given various plausible understandings of what it is to minimize inaccuracy. At the same time, the results that have been published so far show that there is still a sense of inaccuracy minimization—expressed in terms of average incurred penalties as measured by some standard scoring rule—in which Bayes' rule may outperform IBE, leaving the question of which of the two rules (if either) is to be preferred still open.

The aforementioned results pertain to comparisons of IBE and Bayes' rule in a strictly individualistic setting, which presupposes agents to update their beliefs on incoming information whilst completely disregarding the beliefs of other agents. Lately, researchers from various quarters have started to draw attention to the fact that this presupposition is often far from realistic: in reality, our beliefs and belief changes may depend as much on our interactions with others—parents, children, teachers, students, colleagues—as on the evidence we receive immediately from the world. This paper considers both IBE and Bayes' rule in a social setting, in which agents can communicate their beliefs to others and thereby influence those others' beliefs. As a formal framework for studying and comparing the rules, we use an extension of the well-known Hegselmann–Krause model for opinion dynamics, which allows an agent to update partly on the basis of the belief states of her epistemic neighbors, that is, agents whose belief states are not too far off (in a certain sense) from the agent's own belief state. The upshot will be that if agents do not only update their degrees of belief on the basis of evidence but also take into account the degrees of belief of their epistemic neighbors, then the noted advantage of Bayesian updating evaporates and IBE does better than Bayes' rule on every reasonable understanding of inaccuracy minimization.

Any comparison of IBE and Bayes' rule requires two issues to be clarified first: (i) given that IBE lacks a canonical formulation, we must know which rule exactly we are to compare with Bayes' rule; and (ii) we must know along which dimension or dimensions the rules are to be compared with each other. The first two sections address these issues, Section 1 presenting a precise probabilistic version of IBE and Section 2 describing the criteria (directly related to the already mentioned Dutch book and inaccuracy minimization considerations) that have been used for comparing IBE and Bayes' rule in a context of individualistic epistemology. Then, in Section 3, we turn to an evaluation of IBE and Bayes' rule in the context of social epistemology.

1. What is inference to the best explanation? In the literature, one finds a number of different principles proposed under the heading of IBE. All these principles share the broad idea—call this "the core idea"—that inference is to be governed by considerations of explanatory goodness. What is further common to most of them is that they are stated in terms of a *categorical* notion of belief or acceptance, as when we are told that "[i]t is reasonable to accept a satisfactory explanation of any fact, which is the best available explanation of that fact, as true" (Musgrave [1988], p. 239; italics omitted), or that "IBE authorises the acceptance of a hypothesis H, on the basis that it is the best explanation of the evidence" (Psillos [2004:83]; italics omitted), or that the guiding idea behind IBE is that "a hypothesis is accepted on the basis of a judgement that it best explains the available evidence" (Psillos [2007], p. 442). These and similar formulations may suggest that IBE belongs to (in Foley's [1992] terminology) the epistemology of belief—which is concerned with categorical beliefs—rather than to the epistemology of *degrees* of belief. If this suggestion were correct, the debate between Bayesians and explanationists might boil down to the more general debate between Bayesians and mainstream epistemologists about whether categorical belief talk has a place in scientific philosophy. But we are not aware of any theorists holding that explanationists are committed to the epistemology of belief. And while explications of the core idea in probabilistic terms are few, they are not altogether absent from the literature.

In fact, a number of authors have recently sought to state IBE in such a way as to make it consistent with Bayes' rule, and thus dissolve the apparent opposition between explanationists and Bayesians. According to these authors, Bayesians should even welcome the insight that explanation has confirmation-theoretic import, as it provides them with a much-needed supplement to their doctrine.[3]

To see why they think this, first recall that to apply Bayes' rule and thereby determine the probability to be assigned to a hypothesis H upon the receipt of evidence E, one must be able to assign unconditional probabilities to H and E (the "prior" of H and the "marginal likelihood" of E) as well as a probability to E conditional on H (the "likelihood" of H on E). Where are these values to come from? For some hypotheses and pieces of evidence, the likelihood can be determined analytically; for instance, the likelihood of "Coin C has a bias for heads of .7" on "The next toss with C will be heads" equals .7. However, the kind of case where semantic analysis suffices to fix priors and likelihoods is special. For the remaining cases, Bayesians have little to offer beyond the contention that, in those cases, any choice of priors and likelihoods is as good as any other, as long as the choice respects the axioms of probability. Most Bayesians are well aware, however, that this subjectivist feature of their doctrine is the single most important reason for others not to endorse it.

The aforementioned authors believe that IBE can help Bayesians precisely in overcoming the subjectivism that adheres to their position. According to these authors, the solution is to adopt as a further principle that priors and likelihoods are to be based on explanatory considerations. For instance, if H is an a priori better explanation than H^* is (e.g., because H is mathematically simpler than H^*), then H ought to receive a higher prior than H^*. Similarly, if H is a better explanation of E than H^* is, then the likelihood of H on E should be higher than the likelihood of H^* on E. Thus understood, IBE is

[3] See, for instance, (Okasha [2000]), (Lipton [2004], Ch. 7), and (Weisberg [2009]).

not in competition with Bayes' rule. It could hardly be, given that on this construal IBE is not an update rule at all, but rather helps with preliminary work that is to be carried out before Bayes' rule can be applied.

Even if the proposal is granted, the constraints it places on priors and likelihoods are still so weak—in general, there will be indefinitely many ways to comply with them—that it will do little to persuade those who are offended by the subjective nature of Bayesian rationality. Moreover, many explanationists will have envisioned a more substantive confirmation-theoretic role for explanatory considerations. They will insist that, while such considerations may help in determining priors and likelihoods, IBE is an update rule in its own right, and as such on a par with Bayes' rule, rather than subservient to it.

Somewhat ironically, the possibly most straightforward explication of IBE as a probabilistic update rule is to be found in the writings of Bas van Fraassen, one of IBE's staunchest critics. In the following, we use the label "IBE" for the probabilistic interpretation of the explanationist's core idea that van Fraassen ([1989], Ch. 6) proposes, and keep using "IBE" for the core idea itself. To state IBE, let $\{H_i\}_{i \leqslant n}$ be a set of self-consistent, mutually exclusive, and jointly exhaustive hypotheses, and let $\Pr(\cdot)$ be one's probability function prior to learning E. Then, for all i, one's new probability for H_i after E (and nothing stronger) has been learnt—designated by "$\Pr[E](H_i)$"—comes from Pr via IBE iff

$$\Pr[E](H_i) \quad = \quad \frac{\Pr(H_i)\Pr(E \mid H_i) + f(H_i, E)}{\sum_{j=1}^{n}\big(\Pr(H_j)\Pr(E \mid H_j) + f(H_j, E)\big)}.$$

Here, f is a function that assigns a bonus $c \geqslant 0$ to the hypothesis that explains the given evidence E best in light of the background knowledge, and that assigns zero to the other hypotheses.

Thus understood, IBE amounts to a generalization of Bayes' rule: the latter is the instance of the above schema with f the constant 0 function. Bayes' rule is known to be equivalent to the conjunction of Certainty and Rigidity. According to Certainty, after updating on E, E has probability 1, and according to Rigidity, probabilities conditional on E remain as they were before the update on E. One readily verifies that IBE violates the latter condition if $c > 0$. For let $\Pr(H_i \mid E) = a/b$ (for some $a \geqslant 0$, $b > 0$). Then, where H_j is the best explanation of E, $\Pr[E](H_i) = (a + 0)/(b + c) < \Pr(H_i \mid E)$ if $i \neq j$, and $\Pr[E](H_i) = (a + c)/(b + c) > \Pr(H_i \mid E)$ if $i = j$.

To be sure, IBE is properly called an update rule only when the function f has been specified. To provide a general definition of this function is a tall order, given that f is to assign bonuses for best explanations, and that there is no unanimity on what is required for a hypothesis to count as the best explanation; worse yet, there is no unanimity on what counts as an explanation. This is not a problem specifically for explanationists. It is a problem for anyone—whatever his or her confirmation-theoretic leanings—who holds that explaining is among the core scientific activities, and that philosophers of science should be able to account for such activities.

Van Fraassen sidesteps this problem by comparing IBE and Bayes' rule in the context of a simple statistical model which affords an arguably uncontentious characterization of best explanation. We follow suit here. The model van Fraassen considers consists of a set $\{H_i\}_{0 \leqslant i \leqslant 10}$ of eleven hypotheses concerning the bias of a given coin C, where for each i, H_i is the hypothesis that C has a bias for heads of $i/10$; it is assumed that one of these hypotheses is true. Then, where E_j indicates that the outcome of the j-th toss in a series of tosses with C is E, van Fraassen defines H_i to best explain heads$_j$ iff $i/10$ is closer to the frequency of heads in the first j tosses with C than $k/10$, for $k \in \{0, \ldots, 10\}$ with $k \neq i$. Note that, given this definition, it can happen that two different bias hypotheses both qualify as best explanations for the same evidence; for instance, if there have been 64 heads in the first 99 tosses, then, in light of the background knowledge, both H_6 and H_7 best explain heads$_{100}$.

As for the definition of f, let $f(H_i, E_j) = c$ if H_i is the unique best explanation of E_j; let $f(H_i, E_j) = f(H_k, E_j) = 0.5c$ if, for $i \neq k$, both H_i and H_k best explain E_j; and let $f(H_i, E_j) = 0$ otherwise. As stated, IBE can be said to explicate IBE only if $c > 0$; later on, we will be more specific about the value of c.

It is worth noting that no claim is being made here to the effect that IBE is the correct or best explication of IBE.[4] Even if there are better explications, we believe that IBE merits attention in the context of the present discussion given that (i) it offers an update rule at variance with Bayes' rule which (ii) is at least inspired by the explanationist's core idea that explanation has confirmation-theoretic import, (iii) has been targeted and claimed to be untenable by one of the leading opponents of that idea, and (iv) will be seen to be not only tenable but to have a number of important virtues in comparison with IBE and, when combined with the idea of social updating, to do better than Bayesian updating (also conceived as a social procedure) in every desirable way.

2. Judging the rules—by which lights? This brings us to our second question, the question of which criteria we are to assume in evaluating the rules at issue. Presumably, we will want an update rule to be formally adequate, at least in that it outputs a probability function when given a probability function as input. Bayes' rule is known to pass this criterion. As for IBE, first note that it is safe to assume that, for all E, $f(H, E) = 0$ whenever H is a tautology or a contradiction. If we make the further formal assumption that, for all E, $f(H \vee H^{\star}, E) = f(H, E) + f(H^{\star}, E)$ whenever H and $H^{\star}$ are mutually exclusive,[5] then it is easy to prove that updating a probability function via IBE leads again to a probability function.

However, according to what for a long time has been the main Bayesian argument against IBE, or indeed against any update rule at variance with Bayes' rule, updating via any version of IBE or any other non-Bayesian rule is bound to lead to something almost as bad as formal inadequacy in the above sense, namely, to *synchronic incoherence*. The notion of synchronic incoherence is standardly operationalized by means of so-called

[4] For instance, one could propose an alternative to the function f that assigns partial bonuses to multiple hypotheses, or a bonus that depends on the amount of information—suitably measured—that is available.

[5] We call this a "formal assumption" because it may not enjoy much intuitive support. On the other hand, in practice the assumption may also play no role, given that we do not tend to consider the explanatory power of disjunctions of hypotheses.

dynamic Dutch books. A Dutch book is a collection of bets that jointly guarantee a net loss, the distinctive feature of a *dynamic* Dutch book being that the bets are placed at different points in time, typically before and after a belief change on the basis of newly acquired evidence. David Lewis ([1999]) argued that one is vulnerable to a dynamic Dutch book unless one updates via Bayes' rule. Van Fraassen ([1989], Ch. 6) makes the argument vivid by means of IBE and the above statistical model: he presents an agent who updates on the outcomes of tosses with C via IBE and as a result falls victim to a Dutch bookie who, at different points in time, offers the agent bets on the outcomes of specific tosses with C. As van Fraassen shows, the bets are such that they jointly cannot but have a negative net payoff, and yet each of them seems fair to the agent at the moment at which it is offered. For bettors, losing is all in the game, but accepting bets that *guarantee* a net loss, where one could have seen the loss coming, is surely irrational. Whence the conclusion that Bayes' rule is the only rational update rule.

This argument is problematic for a variety of reasons. For one, suppose updating via some non-Bayesian rule exposes one to Dutch bookies, and thus may lead to monetary losses.[6] Then it is still to be noted that nothing anyone has said rules out that updating via some non-Bayesian rule has compensating advantages or even advantages that vastly outweigh the exposure to Dutch bookies.[7] In fact, in recent work using the above coin-tossing model, it was shown that if one updates via IBE, one will on average be faster—typically *much* faster—in assigning a high probability (say, a probability of .9) to the true bias hypotheses than if one updates via Bayes' rule (Douven [2013]). If those are right who take high probability to be necessary for assertion or for action more generally, this means that, for instance, a scientist who updates via IBE may typically be sooner in a position to announce the results of her research than her Bayesian competitors who are trying to determine the truth of the same hypothesis or hypotheses—which can greatly improve her career prospects. In the same work, it is also shown more directly how the difference between Bayes' rule and IBE as regards speed of convergence to the truth can place persons updating via the latter rule in a financially more advantageous position. As a result, any losses an IBE explanationist may incur at the hands of a Dutch bookie may be more than made up for by gains she would not have had were she to update via Bayes' rule.[8]

[6]"May," because in actuality it may never happen that a bookie offers one a set of bets that together form a dynamic Dutch book.

[7]As an anonymous referee noted, Bayesians have traditionally taken vulnerability to dynamic Dutch books to expose some kind of inconsistency (see, e.g., Skyrms [1987]). The question that still needs answering, however, is why anyone would want to avoid being inconsistent in *that* sense (whichever precise sense it is).

[8]Roche and Sober ([2013]) seek to buttress the conclusion of van Fraassen's version of the dynamic Dutch book argument by arguing that explanatory considerations cannot provide the kind of probabilistic boost that they are supposed to provide according to explanationists wedded to IBE or similar rules. Specifically, they argue that the evidence E for a hypothesis H screens off from H any fact F about explanatory connections between E and H, meaning that $\Pr(H \mid E \wedge F) = \Pr(H \mid E)$. It is to be noted, however, that their argument for this claim proceeds from start to end on Bayesian assumptions. If giving bonuses to best explanations makes no sense from a Bayesian perspective, why should that bother the explanationist? It is also to be noted that there may be no reading of $\Pr(H \mid E \wedge F)$ and $\Pr(H \mid E)$ that is

This is all supposing that updating via IBE or some other non-Bayesian rule *does* expose one to dynamic Dutch books. It need not do so. Igor Douven ([1999]) argues that update rules like Bayes' rule and IBE are not to be assessed in isolation but as parts of *packages* of principles which, next to one or more update rules, also include decision-theoretic principles. Building on work by Patrick Maher ([1992]), Douven presents a package of principles that includes IBE but that nonetheless guards one against Dutch bookies; the package allows one to infer to the best explanation—as formalized by IBE or possibly also some other probabilistic update rule—free of charge, *pace* Lewis and van Fraassen.[9]

In short, it is far from clear that we should adopt as a criterion for evaluating update rules that they keep us out of the hands of Dutch bookies. But even if we do adopt that criterion, explanationists can make sure—by taking on board certain additional, decision-theoretic principles—that their rule does just as well on this score as Bayes' rule.

The appeal to decision-theoretic principles in the present context already hints at a reason why, over the past ten years or so, many in the Bayesian community have become increasingly dissatisfied with the dynamic Dutch book approach to vindicating Bayes' rule. Update rules are rules for changing our *beliefs*, or degrees of belief, which makes it a question of *epistemic* rationality whether or not we are justified in relying on them (if we do). Whether reliance on any such rule makes us liable to financial losses rather seems to be a question of *practical* or *prudential* rationality. It is not a priori that acting as an epistemically responsible agent is prudentially always the best thing to do. Trying to see the world as it is may be epistemically mandatory, but may be less than prudent if a realistic picture of the world makes one more prone to suffering from depression than fostering a rosier—albeit distorted—picture would do. In other words, the dynamic Dutch book argument seems to be addressing exactly the wrong issue.

This has led a number of philosophers to develop a different approach to defending Bayes' rule, one inspired by the thought that update rules, like other epistemic principles, are to be assessed in light of their conduciveness to our epistemic goal.[10] All else being equal, if one epistemic principle is more likely to help us achieve our epistemic goal than another, then where we have to choose between the two, it is epistemically rational to adopt the former rather than the latter. And what these philosophers have tried to show is exactly that Bayes' rule is more conducive to our epistemic goal than any other update rule. If that is right, then implicitly this shows those who rely on IBE, or on some other explication of IBE at odds with Bayes' rule, to be epistemically irrational.

However, the new strategy appears to be hardly less problematic than the dynamic Dutch book strategy. In mainstream epistemology, there is widespread agreement that

neutral in the present debate. For instance, in Douven's ([1999]) proposal, $\Pr(H \mid E)$ is defined so as to equal $\Pr[E](H)$—which, you may recall, is the result of updating H on E in accordance with IBE—whereas for Bayesians it equals $\Pr_E(H)$, which is the result of updating H on E in accordance with Bayes' rule; and as mentioned in the previous section, $\Pr[E](H)$ and $\Pr_E(H)$ are, in general, not equal.

[9] See also (Tregear [2004]).

[10] See, e.g., (Rosenkrantz [1992]) and (Leitgeb and Pettigrew [2010]).

our epistemic goal is to believe all that is true and nothing that is false (e.g., Lehrer [1974], p. 202, Foley [1993], p. 19).[11] This is conceived as an ideal that every epistemic agent ought to aim at, even if they will always fall short of realizing it. But Bayes' rule and IBE have their homes not in mainstream epistemology—which is the previously mentioned epistemology of belief simpliciter—but rather in the epistemology of *degrees* of belief. And it is at least not immediately obvious how to translate the epistemic goal, as traditionally conceived, into the vocabulary of the epistemology of degrees of belief.

The same doubts that have arisen in relation to the dynamic Dutch book defense of Bayes' rule have also arisen in relation to the older (non-dynamic) Dutch book defense of the static part of Bayesianism, that is, the part according to which our degrees of belief at any one moment should obey the probability axioms. In the context of mounting a new, distinctively nonpragmatic and epistemic defense of that static part, Jim Joyce ([1999]) proposed to translate the epistemic goal as traditionally understood in terms of minimizing the inaccuracy of our degrees of belief. In his proposal, we ought to aim at having degrees of belief that are as accurate as possible. The notion of accuracy at stake is operationalized by reference to a scoring rule, specifically the so-called Brier rule. Given a set $\{H_i\}_{i \leqslant n}$ of mutually exclusive and jointly exhaustive hypotheses, and $[\![H_i]\!] \in \{0, 1\}$ being the truth value of H_i, the Brier rule assigns a penalty of $\sum_{i=1}^{n}\big([\![H_i]\!] - \Pr(H_i)\big)^2$ to an agent whose degrees of belief are given by $\Pr(\cdot)$. Intuitively and roughly, what this amounts to is that, all else being equal, one is more accurate the higher one's degree of belief in the true hypothesis, and again ceteris paribus, one is more accurate the lower one's degree of belief in any false hypothesis.

In a recent attempt to defend the dynamic part of Bayesianism—Bayes' rule—in a similar nonpragmatic fashion, Hannes Leitgeb and Richard Pettigrew ([2010]) adopt Joyce's formulation of our epistemic goal. The claim they then argue for is that Bayes' rule does better than any other update rule in achieving our epistemic goal, understood in Joyce's sense. To be more precise, they argue that by updating according to Bayes' rule, an agent minimizes her expected inaccuracy, meaning that, as judged by her current degrees of belief, the inaccuracy of her new degrees of belief updated on evidence she receives is going to be lower if she updates via Bayes' rule than if she updates via some non-Bayesian rule. However, whilst this may be enough for achieving the epistemic goal on one reading of inaccuracy minimization, there are other readings that seem at least as legitimate and therefore are at least as relevant to the present discussion.

Note that, when understood in a dynamic context, Joyce's formulation of our epistemic goal is ambiguous in multiple ways. That one ought to minimize the inaccuracy of one's degrees of belief could for instance be interpreted as meaning any of the following:

1. that every update ought to minimize *expected* inaccuracy;
2. that every update ought to minimize *actual* inaccuracy;
3. that every update ought to contribute to the long-term project of coming to have a minimally inaccurate representation of the world.

[11] This view is widely but not universally shared. For instance, Sartwell ([1992]) and Kelp ([2014]) argue for an epistemic goal in terms of knowledge rather than true belief.

Moreover, if our epistemic goal is understood in the sense of the third interpretation, there are further questions to be asked about how to balance precision and speed of convergence. Most notably, should we aim to have minimally inaccurate degrees of belief in the long run, disregarding entirely how long the run may be, or should we aim to minimize the inaccuracy of our degrees of beliefs in a more limited timeframe, even if that comes at the expense of precision (i.e., even if we could achieve greater accuracy if time constraints played no role)?[12]

Minimizing expected inaccuracy of the post-update belief state and minimizing actual inaccuracy of that belief state will not in general amount to the same thing, and neither amounts to realizing the epistemic goal in the sense of interpretation 3 (minimizing actual long-term inaccuracy).[13] So, granted that Bayes' rule is most conducive toward realizing our epistemic goal in the first sense—as Leitgeb and Pettigrew show—IBE or some other formalization of IBE may still be more conducive toward realizing our epistemic goal in one of the other senses. This means that, absent a reason to deem realizing our epistemic goal in the former sense more important than realizing it in the latter sense, the inaccuracy minimization defense of Bayes' rule is ineffective.

In fact, in light of Robbie Williams' ([2012], p. 835) observation that it is unclear what is so desirable about having maximized one's chances of being accurate as judged from a perspective one has already abandoned, it would seem that the second and third interpretations above provide more reasonable conceptions of our epistemic goal than the first one. Even apart from the fact that that perspective has been abandoned, that a perspective is involved at all makes inaccuracy minimization on this interpretation quite disanalogous to our epistemic goal as traditionally conceived. The latter, after all, posits something objective that we should aspire to, not something that is intrinsically dependent on our beliefs or degrees of belief, however misguided these might be from the point of view of an objective bystander.

Moreover, if the epistemic goal as proposed by Joyce is indeed to mirror what traditional epistemologists have taken to be our epistemic goal, then, it seems, we should go with interpretation 3 (regardless of how that interpretation is further specified). For in traditional epistemology, our epistemic goal is commonly conceived as an ideal to be attained, or rather approached, in the long run (see, e.g., Latus [2000] and Goldman [2010]).

It might be thought that while it has been shown that Bayes' rule is most conducive to our epistemic goal in the sense of interpretation 1, it is so far mere speculation that IBE might outperform Bayes' rule when our epistemic goal is understood in one of the other two senses. That is not quite true, however. Douven ([2013]) reports the outcomes

[12] See on the trade-off between precision and speed of convergence, (Zollman [2007], [2010]) and (Douven [2010]).

[13] Given some scoring rules, the long-term project of inaccuracy minimization is at least closely related to the goal of convergence to the truth as understood earlier in this section. For instance, the so-called log score rule looks *only* at the probability an agent assigns to the true hypothesis: according to this rule, an agent's penalty equals $\ln(\Pr(H_i))$, with H_i the true hypothesis. This is not so for the Brier rule, however, which looks at the probability the agent assigns to the true hypothesis, but also looks at how the remaining probability is distributed over the false hypotheses.

of computer simulations in which IBE does much better on average than Bayes' rule in realizing the long-term goal of having maximally accurate degrees of belief. These simulations assume the statistical model described in Section 1, and further assume a Bayesian agent and an IBE explanationist to update their degrees of belief after each toss of a coin C which in total is tossed 1000 times. For each bias that the coin can have according to the said model, 1000 such simulated sequences of 1000 coin tosses were run. In each simulation, Brier penalties incurred after the 100th, 250th, 500th, 750th, and 1000th updates were calculated for both the Bayesian and the explanationist. For all biases that were considered, the explanationist at *each* of those reference points got assigned a lower penalty than the Bayesian in the *vast* majority of simulations. For anyone concerned with realizing the long-term project of arriving at a maximally accurate degrees-of-belief function, that would seem an excellent reason to prefer IBE over Bayes' rule.

There is a complicating factor, however. For the same paper shows that when, for each of the possible biases, the *average* penalty—averaged over all 1000 simulated sequences—incurred at the reference points is calculated, there is mostly no difference, but *when* there is a difference between the explanationist's and the Bayesian's scores, it is always a small difference in favor of the Bayesian. At first, this may seem surprising. But the explanation is fairly straightforward.

The three columns of Figure 1 show the results of three simulations with a coin bias of .1, .5, and .9, respectively.[14] The upper row shows the developments of the degrees of belief assigned to the true hypothesis by two agents who receive exactly the same evidence (if one sees heads coming up at a given toss, then so does the other, and analogously for tails), where one agent is a Bayesian (represented by the black circles) and the other is an explanationist (represented by the colored squares). The lower row shows the corresponding Brier penalties that they incur at every step. In the bias .1 and bias .9 simulations, it is quite clear that the explanationist does better than the Bayesian at least insofar as the explanationist's confidence in the true bias hypothesis converges faster to 1 than the Bayesian's confidence in that hypothesis. It is also clear that in those cases, the explanationist incurs lower Brier penalties than the Bayesian after every or almost every update.

However, the bias .5 simulation nicely illustrates a characteristic feature of the explanationist, to wit, that in a clear sense, she is an *enthusiastic* learner, or at least a more enthusiastic learner than the Bayesian. In that simulation, we clearly see the same pattern in how the explanationist's and the Bayesian's degree of belief in the true hypothesis change over time. It is just that the pattern is more pronounced in the case of the explanationist. Whereas this enthusiasm is precisely what in general leads the explanationist to assign a high probability to the true hypothesis faster than the Bayesian (in the long or medium-length run), it is also what leads her further *astray* than the Bayesian in the event that a sequence of tosses contains a longer subsequence

[14]Given the relation between the bias for heads (p) and the bias for tails ($1 - p$), biases .1 and .9 lead to similar results. Showing results for both gives us an opportunity to inspect the variation between different runs of the simulation.

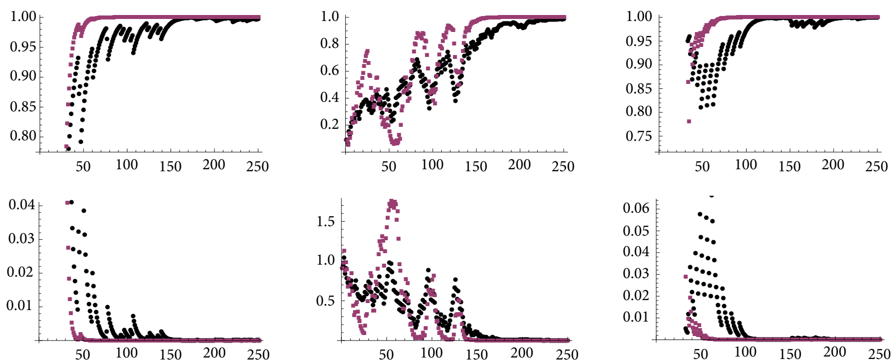

Figure 1: Simulations with a randomly chosen sequence of 250 tosses with a coin with bias $p = .1$ (left column), bias $p = .5$ (middle column), and bias $p = .9$ (right column). Upper row: probability of true bias hypothesis; lower row: Brier penalties. Colored squares: explanationist; black circles: Bayesian.

of consecutive tosses whose relative frequency of heads deviates strongly from the coin's bias for heads (in the short run). In other words, where we have said that the explanationist is a more enthusiastic learner than the Bayesian, the Bayesian may rightly point out that the explanationist is at the same time a less *cautious* learner than the Bayesian.

The second panel in the bottom row of Figure 1 makes it plain how this may lead to relatively large differences in the Brier penalties incurred by the agents. When in the beginning of the simulation the degrees of belief for the true bias hypothesis of both the explanationist and the Bayesian start to drop, the drop is much greater for the explanationist, and around the fiftieth toss, the explanationist even comes close to believing the truth to a degree of 0 when her Bayesian colleague's degree of belief for the same hypothesis is still above .2. In the graph, we see this visualized as an eruption in the explanationist's Brier score, with the explanationist coming close to incurring the maximal Brier penalty around the fiftieth toss.

To put the point in general terms, due to her enthusiasm, the explanationist mostly does better than the Bayesian, where doing better also means that she incurs a lower Brier penalty. But *if* things go wrong—in the form of the occurrence of subsequences of the type mentioned above—then they tend to go wrong much more badly for the explanationist than for the Bayesian, and in fact so much so that, averaged over all simulations, the explanationist still incurs a slightly greater Brier penalty than the Bayesian.

The aim of the work summarized here was not to establish the superiority of IBE. It was the more modest goal of exhibiting some flaws in the extant arguments for Bayes' rule. Indeed, nothing said in that work gives reason to believe that inaccuracy minimization in the sense of incurring minimal average Brier penalties is somehow outweighed by inaccuracy minimization in one or more other senses. Nor will we provide such a reason here. Rather, our goal in the following is to show that if Bayes' rule and IBE are compared in a setting that takes into account *social* aspects of our belief management, the comparison may well turn out in favor of IBE in *all* important respects.

3. From an individualistic to a social perspective. Following pioneering work by Alvin Goldman (see his [1999] for an overview of the early work), from the 1990s onwards epistemologists have started paying explicit attention to the role or roles social practices play in our epistemic lives. By now, it is broadly accepted that an epistemology that considers epistemic agents as isolated from their social environment—as traditional epistemology has long done—cannot but miss important aspects of the belief- and knowledge-forming processes: our epistemic states are not just shaped by evidence we receive directly from the world, but also by our interactions with other epistemic agents. This insight (as one can surely call it) has led to the study of such topics as the conditions under which we are warranted in accepting a person's testimony, different ways of aggregating opinions and judgments, the role of experts in society, and the resolution of disagreements amongst epistemic peers.

It is important to note that no social epistemologist—as those following Goldman's lead are now commonly called—has ever suggested that individualistic epistemology has not led to any true understanding. Some epistemically significant episodes of a person's life are well captured by the traditional model of the thinker facing, and trying to make sense of, the world on her own. Indeed, it could hardly be said that, because they leave out of the picture possible epistemic contributions from the updater's social surroundings, the various arguments pro and con Bayes' rule, IBE, and other update mechanisms are of no value. After all, sometimes we do take into account, over a certain period of time, one piece of evidence after the other, without in the interim consulting or otherwise interacting with others.

But often enough, we do not. Scientists conduct a study or experiment relevant to a hypothesis they are testing, register the data they get, discuss their work over lunch with colleagues investigating the same hypothesis or (probably less frequently) discuss it at a conference with researchers from others labs also focussing on the hypothesis, go back to their lab to gather further data, and so on. During the process, the confidence they invest in their hypothesis may be affected as much by the incoming data as by the exchanges they have with their colleagues. The question of whether the social dimension might make a difference to the standing of this or that update rule, and specifically whether it might make a difference to the standings of Bayes' rule and IBE or other explications of IBE, has never been asked. It is the topic of the remainder of this paper.

More specifically, we compare Bayesian and explanationist updating again along the dimensions of inaccuracy minimization and speed of convergence, but this time the comparison will be made within a social setting. Our main tool of investigation will be computer simulations, where the simulations will all pertain to the earlier statistical model of eleven bias hypotheses. To model the distinctively social aspect, we use an extended version of the Hegselmann–Krause (HK) model for studying opinion dynamics in societies of agents who interact doxastically by basing their opinions wholly or partly on the opinions of others. We begin by describing this model as originally presented by Hegselmann and Krause.[15]

3.1. The Hegselmann–Krause model. The HK model has been studied extensively by philosophers, social scientists, mathematicians, and physicists. Questions that have been investigated by means of this model include the question under which conditions belief polarization tends to occur in societies of doxastically interacting agents, the question under which conditions the opinions of such agents tend to converge, questions concerning the speed with which opinions diverge or converge, the question of whether it makes a difference to such processes if data are noisy, the question of the influence of what exactly the truth is, and many others.[16]

The versions of the model developed and studied by Hegselmann and Krause assume communities of agents who are trying to determine the value τ of some unspecified parameter, where it is antecedently given that $\tau \in (0,1]$.[17] In the simplest model, all agents in the community update repeatedly and simultaneously by averaging the opinions of the agents that are in their Bounded Confidence Interval (BCI), which is the case precisely when an agent's opinion is not too far off the agent's own opinion. In most studies with the model (both this simple model and more complicated variants), being "not too far off" is simply taken to mean that the absolute difference between the opinions is below some threshold value ϵ.

In a more interesting version of the model, the agents again update their opinions repeatedly and simultaneously by taking into account the opinions of agents within their BCI, but now they take a *weighted* average instead of the straight average of those opinions and the value of τ.[18] More exactly, the opinion of agent x_i after the $(u+1)$-st update is given by

$$x_i(u+1) \;=\; \alpha \frac{1}{|X_i(u)|} \sum_{j \in X_i(u)} x_j(u) \;+\; (1-\alpha)\tau, \qquad \text{(HK)}$$

where $x_i(u)$ is the opinion of agent x_i after the u-th update, $\alpha \in [0,1]$, and $X_i(u) \coloneqq \{j : x_j(u) \in \text{BCI}\}$.[19]

[15] See (Hegselmann and Krause [2002], [2005], [2006], [2009]).

[16] See for example, next to the papers by Hegselmann and Krause cited in the previous note, (Dittmer [2001]), (Fortunato [2004]), (Lorenz [2007]), (Douven and Riegler [2010]), and (Kurz and Rambau [2011]).

[17] The exclusion of 0 from the interval makes computations easier but otherwise has no significance.

[18] The idea is not, obviously, that the agents *know* the value of τ. Rather, it is that the agents receive information that somehow points in the direction of that value. See Hegselmann and Krause [2006, Sect. 1] for a detailed account of how to interpret this part of their model.

[19] We are slightly simplifying here in that, in some computer experiments, Hegselmann and Krause allow both the confidence interval and the weighting factor α to vary for different agents.

The belief states of agents in the versions of the HK model studied by Hegselmann and Krause always consist, at any given point in time, of one value. But their model is very flexible and it has been extended to one in which agents are equipped with richer belief states in that they hold opinions on various related or unrelated matters (Riegler and Douven [2009] and Wenmackers, Vanpoucke, and Douven [2012], [2014]). Also, while Hegselmann and Krause in their writings concentrate on purely descriptive questions—like the ones mentioned earlier, for instance, concerning the circumstances under which communities of initially disagreeing agents tend to converge—in the meantime, the model has also been used to address various normative questions, such as what the right response is to the discovery that one disagrees about an issue with one or more epistemic peers.[20] We here recruit a new extension of the HK model in the service of addressing another normative issue, to wit, the issue of which update procedure someone attempting to be rational ought to deploy.

3.2. A probabilistic extension of the HK model. In the new extension of the model, the belief states of agents at any given time do not consist of a single real number—as they do in the original HK model—or of a vector of 0's and 1's, encoding whether the agents do or do not believe particular propositions, as in Riegler and Douven's ([2009]) extension of the HK model. Rather, in the new model, belief states are represented by probability functions. The language these functions are defined on consists of the eleven bias hypotheses introduced toward the end of Section 1 and the truth-functional combinations of these hypotheses, although now it is a bit more realistic (though not strictly necessary—see note 25) to suppose that the hypotheses concern the bias, not of a single coin, but of a set of coins, one for each agent, which perhaps all come from the same factory and at any rate are known to have the same (at least initially unknown) bias.

Just as in the model with the update procedure (HK), the agents in the extended model take into account both evidence from the world—in this case, evidence about the bias of their respective coin, which comes to them in the form of tosses they observe—and the opinions of those agents that are in their BCI. Obviously, the BCI now must be defined differently from before. There are many options here, with which one may want to experiment. In our simulations, we went for the very straightforward definition which lets two agents be in each other's BCI iff the sum of the absolute differences in the probabilities they assign to the eleven bias hypotheses is below a given threshold value. Put more formally, agent i with belief state $\Pr_i(\cdot)$ is within the BCI of agent j with belief state $\Pr_j(\cdot)$ iff

$$\Delta\left(\Pr_i(\cdot), \Pr_j(\cdot)\right) \;\coloneqq\; \sum_{k=0}^{10} \left|\Pr_i(H_k) - \Pr_j(H_k)\right| \;\leqslant\; \epsilon.$$

Note that, as in the original HK model, an agent is always in her own BCI.

[20] See (Douven [2010]), (Douven and Riegler [2010]), (Douven and Kelp [2011]), and (Kelp and Douven [2012]).

As will be described in more detail below, in our simulations we study communities of agents who update repeatedly and simultaneously. The communities are characterized by the fact that their members all update via the same procedure, which consists of two steps, a first step—the "evidential" step—in which the agent updates on incoming evidence—the outcome of a toss with the agent's coin—and a second, "averaging" step in which the agent averages over the belief states of the agents in her BCI. More precisely, let $\Pr_i^u(\cdot)$ be the belief state of agent i at stage of inquiry u, and let $X\big(\Pr_i^u(\cdot)\big)$ be the set of agents j such that $\Delta\big(\Pr_i^u(\cdot), \Pr_j^u(\cdot)\big) \leqslant \epsilon$, for some given value of ϵ. Then where agent j at stage u learns evidence E_j^u, the update rule we are proposing is given by this algorithm:

Algorithm: update rule (in pseudocode)

---

for each $i \in \{0, \dots, 10\}$ **do**

$$\Pr_j^{u_{\text{evid}}}[E_j^u](H_i) \;\leftarrow\; \frac{\Pr_j^u(H_i)\Pr_j^u(E_j^u \mid H_i) + f(H_i, E_j^u)}{\sum_{k=0}^{10}\big(\Pr_j^u(H_k)\Pr_j^u(E_j^u \mid H_k) + f(H_k, E_j^u)\big)}$$

end for

for each $i \in \{0, \dots, 10\}$ **do**

$$\Pr_j^{u_{\text{av}}}[E_j^u](H_i) \;\leftarrow\; \frac{\sum_{m \in X(\Pr_j^{u_{\text{evid}}})} \Pr_m^{u_{\text{evid}}}[E_m^u](H_i)}{\Big|X(\Pr_j^{u_{\text{av}}})\Big|}$$

end for

for each $i \in \{0, \dots, 10\}$ **do**

$$\Pr_j^{u+1}(H_i) \;\leftarrow\; \Pr_j^{u_{\text{av}}}[E_j^u](H_i)$$

end for

It will be clear that this is actually just a schema, yielding a different update rule for each possible combination of ϵ and explanation bonus c assigned by f. The communities to be studied will differ from each other in the values their members assign to c—which is relevant in the evidential step—and ϵ, which is relevant in the averaging step. Note that if both ϵ and c are 0, we obtain Bayes' rule. With $\epsilon = 0$ and $c > 0$, we obtain an instance—the instance for the specific value of c—of the IBE schema defined earlier. A community of agents for which $c = 0$ but $\epsilon > 0$ can be considered to be a community of "averaging Bayesians," that is, agents who update by Bayes' rule on the evidence they receive from the world but then immediately also let their belief states be further shaped by the belief states of other agents in their community.[21] If instead both $c > 0$ and $\epsilon > 0$, we have a community of "averaging explanationists," that is, explanationists who are also willing to take into account what beliefs certain other agents in their community hold.

We have two further comments on this. First, we know from Section 2 that, given minimal assumptions on f, if $\Pr_j^u(\cdot)$ is a probability function, then so is $\Pr_j^{u_{\text{evid}}}(\cdot)$. And given that (i) any convex combination of probability functions is itself a probability function,[22] and (ii) the straight average of a number of probability functions is a convex combination of these functions, it follows that $\Pr_j^{u_{\text{av}}}(\cdot)$, and hence also $\Pr_j^{u+1}(\cdot)$, is a probability function whenever $\Pr_m^{u_{\text{evid}}}(\cdot)$ is a probability function for each $m \in X(\Pr_j^{u_{\text{evid}}}(\cdot))$.

And second, taking a convex combination of the probability functions of the individual agents in a group is the best studied method of forming social probability functions.[23] Authors concerned with social probability functions have mostly considered assigning different weights to the probability functions of the various agents, typically in order to reflect agents' opinions about other agents' expertise or past performance. The averaging part of our update rule is in some regards simpler and in others less simple than those procedures. It is simpler in that we form probability functions from individual probability functions by taking only *straight* averages of individual probability functions, and it is less simple in that we do not take a straight average of the probability functions of *all* given agents, but only of those whose probability function is close enough to that of the agent whose probability is being updated.

3.3. Simulations. We have conducted simulations with communities of agents who repeatedly and simultaneously update on coin tosses.[24] Each agent has her own coin,

[21] The update rule that we consider for "social Bayesians" or "averaging Bayesians" consists of a social averaging component and a Bayesian component; it is not Bayesian as a whole. We do not rule out the possibility of an entirely Bayesian way for an agent to aggregate the indirect evidence, as obtained via social interactions. For instance, one may think of a Bayesian who considers various ways in which the other agents may be updating, thereby applying ideas from model averaging—ideas that are native to Bayesian statistics. However, to implement this suggestion in a practical situation requires making many special assumptions about the situation, whereas we aim for a learning rule that is very broadly applicable (without much modification).

[22] Given a set $\{\lambda_i\}_{1\leqslant i\leqslant n}$ such that $\lambda_i \in [0,1]$ for all i and $\sum_{i=1}^n \lambda_i = 1$, and given a set $\{\Pr_i(\cdot)\}_{1\leqslant i\leqslant n}$ of probability functions, the function $\Pr(\cdot) = \sum_{i=1}^n \lambda_i \Pr_i(\cdot)$ is said to be a *convex combination* of the $\Pr_i(\cdot)$.

[23] It now often goes by the name of "linear opinion pooling," which was introduced in (Stone [1961]). Discussion of this method goes back at least to Laplace's *Essai Philosophique sur la Probabilité* from 1814 (Bacharach [1979]). See (Genest and Zidek [1986]) and (Cooke [1991]) for useful overviews of methods of aggregating probability judgments.

[24] In this paper, we are interested in finite groups of agents (e.g., fifty peers) who consider a relatively small number of hypotheses (e.g., eleven bias hypotheses) and who interact intermittently. Since there exist relatively few analytical solutions for problems involving discrete variables, this type of questions is usually addressed by numerical simulations. A formulation in terms of continuous variables can be obtained by considering a limit situation, such as an infinite population, a continuum of bias hypotheses, and/or continual interaction. Although analytical solutions may be more readily obtainable in the continuous case, they do not apply to the original problem directly. Hence, we have opted here for a purely computational approach. (This, of course, does not preclude that further insights may be gained by means of analytical

but the coins are antecedently known to all have the same bias, where a bias of x is interpreted as indicating that the objective probability for heads on any given coin toss equals x.[25] All agents in all communities start with a flat distribution over the eleven bias hypotheses, which, as stated, are now interpreted as being about the bias of the *coins* instead of about the bias of a single coin. All communities consist of agents that update by the *same* specific instance of the algorithm described above, that is, the algorithm with fixed values for ϵ and c. One could also consider mixed communities, whose agents have different values for ϵ and c, but we highlight this possibility here only as a possible avenue for future research.

Specifically, we conducted, for each bias for heads $p \in \{0, .1, .2, .3, .4, .5\}$,[26] for each value of $\epsilon \in \{0, 0.1, 0.2, \ldots, 1\}$, and for each value of $c \in \{0, 0.1, 0.2, \ldots, 1\}$, 50 simulations with communities of 50 agents, all having their own coin with bias p. The agents tossed their coins simultaneously and immediately updated on the outcome, using the instance of the algorithm schema for the given values of ϵ and c; they repeated this 500 times. Thus, altogether we conducted $6 \times 11 \times 11 \times 50 = 36{,}300$ simulations.[27]

It is to be noted that for each possible bias, we generated 50 sets of 50 sequences (one for each agent) of 500 tosses. For that bias, these same 50 sets were used for all combinations of ϵ and c that we investigated. To give a concrete example of what this entails: in the 14th simulation for $p = .4$, $\epsilon = 0.7$ and $c = 0.1$, the 50 agents were fed the same 50 sequences (one per agent) as the 50 agents in the 14th simulation for $p = .4$, $\epsilon = 0.2$ and $c = 0.9$. An intuitive way to think about this is that, given a bias p and two particular combinations of values for ϵ and c, the agents considered in the n-th simulation for the one combination are in a sense the epistemic counterparts of the agents considered in the n-th simulation for the other combination: for each agent in one simulation there is an agent in the other simulation such that both agents receive exactly the same evidence during the whole simulation; it is just that both update on that evidence by means of different instances of the algorithm schema.

What the simulations ultimately do is compare 121 ($= 11 \times 11$) update rules, one for each combination of ϵ and c. As stated above, these rules can be thought of as different ways to socialize Bayes' rule (as long as we keep $c = 0$) as well as the explanationist idea (for $c > 0$). All rules were compared with respect to speed of convergence and cumulative Brier penalties, as was done for Bayes' rule (the combination of $\epsilon = 0$ and $c = 0$) and IBE (the combination of $\epsilon = 0$ and $c = 0.1$) in Douven ([2013]).

methods.)

[25] Alternatively, we could assume that one coin has been used to generate a very long sequence of tosses, which is then cut into as many equal segments as there are agents, giving one segment to each agent.

[26] It was unnecessary to run simulations for all eleven possible bias hypotheses. Simulation outcomes for the assumption that the objective probability for heads is p can also be interpreted as simulation outcomes for the assumption that that bias is $1 - p$, simply by switching the interpretation of which code (in our simulations 0 and 1) stands for heads and which for tails.

[27] The simulations were programmed in *Mathematica* 9 and were run on the Millipede Cluster at the University of Groningen. We thank the cluster team and in particular Bob Dröge for technical assistance.

4. Results and discussion. We first look at Brier penalties. After each update, we calculated the Brier score for each agent and then averaged over the 50 agents. We summed the averages for all 500 tosses of the given simulation and finally calculated the average over the 50 simulations. (Note that the maximum that could be reached in this way is 50,000: the maximum Brier penalty for an agent after an update is 2, so for all 50 agents it is 100, so summed over 500 updates it is 50,000.) Figure 2 represents the results.

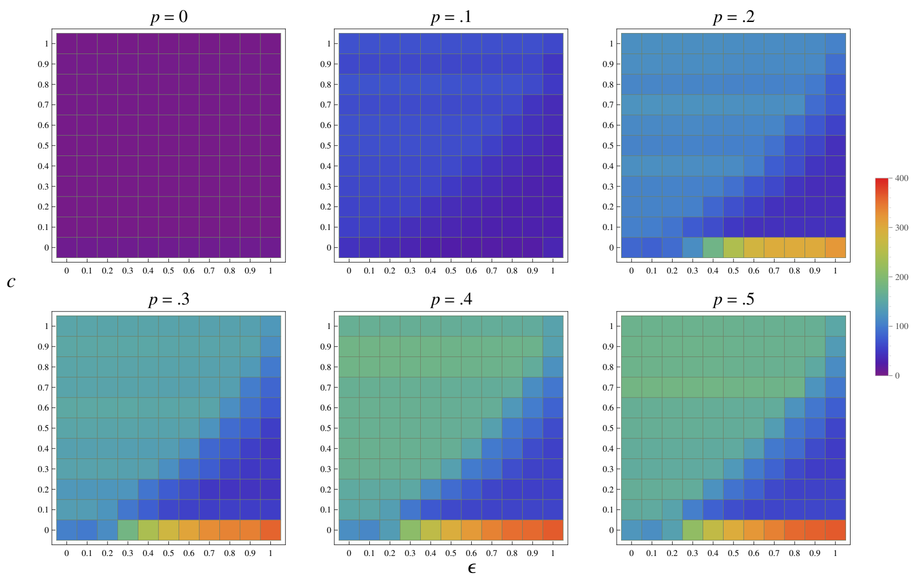

Figure 2: Average Brier scores per agent, for 6 different possible biases ($p = x$ indicates that the objective probability for each of the coins to come up heads is x).

Inspection of the leftmost columns of the various panels confirms the findings summarized in Section 2: the square in the left lower corner of the panels represents Bayes' rule, and the square just above that represents IBE as it was defined for the simulations described in Section 2. It is clear that if there is a difference in color between the squares, the lower one is closer to purple, indicating that the average total Brier penalty incurred over all 500 tosses is never higher for Bayes' rule and sometimes lower. In fact, given that in each of the leftmost columns, the square at the bottom is at least as close to purple as any other square in the same column, we get an immediate strengthening of the previous results: assigning a higher explanation bonus than 0.1 will not lead to better results than can be obtained with Bayes' rule.

The most significant finding is that by listening to the agents within their BCI, explanationists can do better in terms of average total Brier penalty minimization than Bayesians. Indeed, for many combinations of values for ϵ and c, explanationists

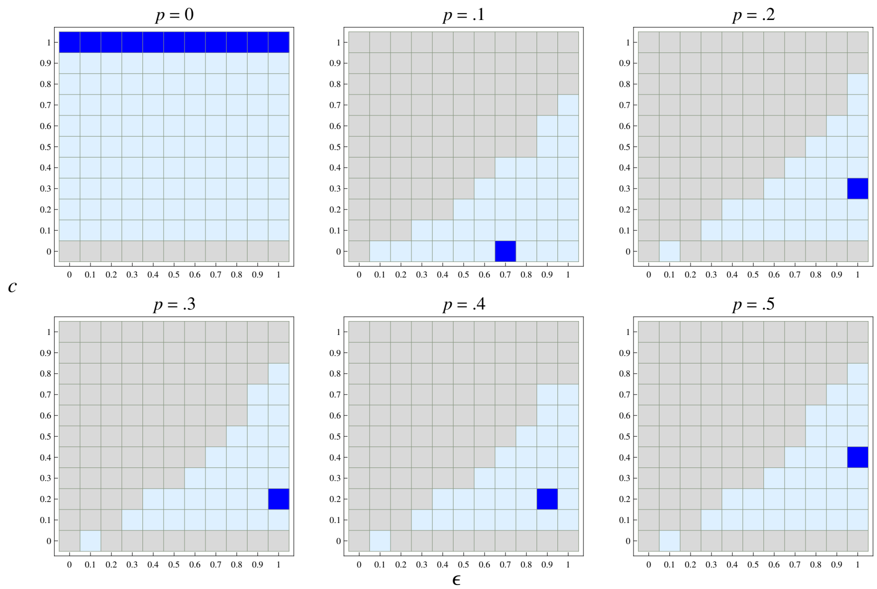

Figure 3: Combinations of ϵ and c that do better than Bayes' rule ($\epsilon = 0$, $c = 0$) are represented as light blue squares; combinations that do best for a given bias are represented as dark blue squares.

will typically do better than Bayesians. While for those combinations, explanationists typically do *somewhat* better than non-averaging Bayesians, they typically do a *whole lot* better than averaging Bayesians. Indeed, the difference in average total Brier penalty over the 500 tosses between the averaging explanationist rule with $\epsilon = 1$ and $c = 0.1$ and the averaging Bayesian rule with $\epsilon = 1$ and $c = 0$ is over 300 (with a theoretical maximum of 1000).

The plots in Figure 3 show both for which combinations of c and ϵ the explanationist does better in terms of average total Brier penalty minimization than the pure (i.e., non-averaging) Bayesian—these combinations are marked as light blue squares—and also which combination does best in absolute terms: this is marked as a dark blue square.[28]

[28]For the bias values $.1 \leqslant p \leqslant .5$, the position of the dark blue square varies in a nonsystematic way. This shows that 50 simulations are insufficient to identify the exact combination of variables that leads to the smallest Brier penalty. This lack of numerical stability is due to the fact that lower Brier penalties are more sensitive to small variations. The other results presented in this work do not depend crucially on such small fluctuations, so we have kept the number of simulations equal to 50 throughout to keep the required computation time within limits.

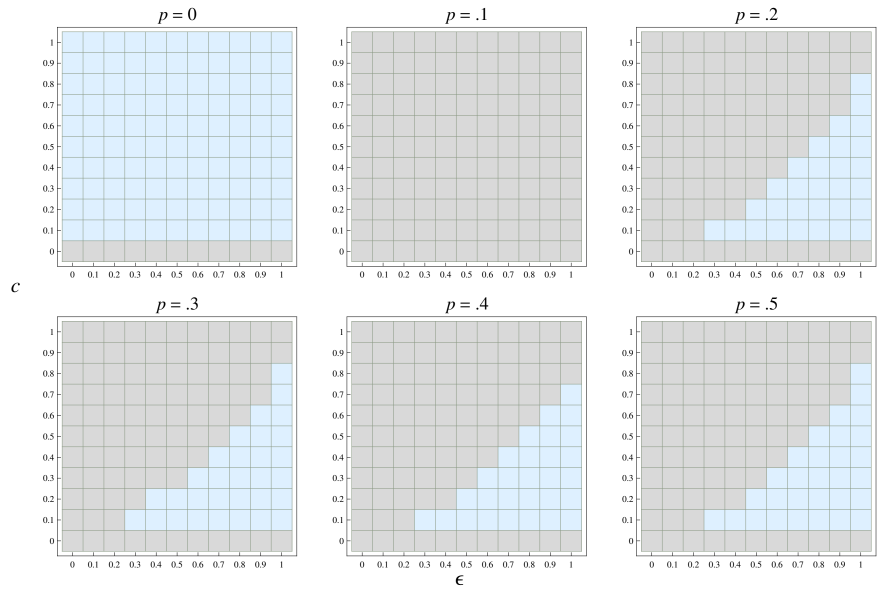

Figure 4: Combinations of ϵ and c that do better than any Bayesian rule ($c = 0$ with any value for ϵ) are represented as light blue squares.

It is also interesting to know which combinations of ϵ and c do better than *any* version of Bayes' rule, whether averaging or non-averaging (thus better than any value of ϵ combined with $c = 0$). This question is answered by the plots in Figure 4.

It is evident from the plots in Figures 3 and 4 that, with the exception of the $p = .1$ case, some explanationist rule *always* does better in terms of minimizing average total Brier penalties than *all* (averaging or non-averaging) Bayesian rules. Naturally, it could still be that in the $p = .1$ case, the difference in favor of the averaging version of Bayes' rule with $\epsilon = 0.7$ (which gives the best result, as Figure 3 shows) is so large that, averaging over all possible biases the coins can have, it would still be better to go with that version of Bayes' rule than with any explanationist rule. That is not so, however. In our simulations, the difference between Bayes' rule with $\epsilon = 0.7$ and any explanationist rule with $c = 0.1$ and $\epsilon \in \{0.7, 0.8, 0.9, 1\}$ or with $c = 0.2$ and $\epsilon \in \{0.8, 0.9, 1\}$ is minimal: the former rule incurs an average total penalty of approximately 22 over the 500 tosses, each of the latter rules incurs an average total penalty of approximately 28 over the 500 tosses.

In Douven ([2013]), it was shown that, in an individualistic setting ($\epsilon = 0$), IBE does better than Bayes' rule in some respects, but not in all. In particular, we saw that, in the simulations reported in that paper, Bayes' rule *on average* incurred a slightly lower total Brier penalty than IBE. On the other hand, we saw that, in those simulations, Bayes' rule *typically* led to a (slightly) higher Brier score than IBE. Moreover, IBE was also seen to do better in terms of speed of convergence to the truth than Bayes' rule. We have seen by now that, as far as average total Brier penalties are concerned, the explanationist can easily do better than the Bayesian when updating proceeds in a social setting. The question still to be answered is whether, in such a setting, the explanationist can have it all, and so can choose values for ϵ and c such that she not only does best with regard to average total Brier penalties but also with regard to the chance that she will incur a lower Brier penalty than her Bayesian counterpart as well as with regard to the speed with which probabilities converge to the truth.

As for Brier scores, we picked as an explanationist rule the algorithm with $\epsilon = 1$ and $c = 0.1$—which comes out as one of the best combinations with regard to average total Brier scores, as was seen, and thus the one which the explanationist might want to pick—and compared this with all social versions of Bayes' rule, that is, the algorithm with any combination of ϵ and c such that $\epsilon > 0$ and $c = 0$. The comparison proceeded as follows: in each simulation and after each round of updates, we summed the agents in the explanationist community who had incurred a lower Brier penalty than their counterparts[29] in the Bayesian community, summed the agents in the Bayesian community who had incurred a lower Brier penalty than their explanationist counterparts, subtracted the latter from the former, and then averaged over all 50 simulations. This yields for each round of updates a number between 50 and -50, where the maximum is reached for a given update u if all 50 explanationist agents have a lower Brier score than their 50 Bayesian counterparts after the u-th update in all 50 simulations, and where the minimum is reached if the converse situation obtains, that is, if all 50 Bayesian agents in all 50 simulations have a lower Brier score after the u-th update than their 50 explanationist counterparts in those simulations after the same update.[30]

From a (social) Bayesian perspective, the comparison turned out most favorably for $\epsilon = 0.1$ and least favorably for $\epsilon = 1$. Figure 5 plots the results of the above simulations for these (for the Bayesian) best and worst settings in blue and purple, respectively, for each of the biases we considered; for each of these biases, the graphs for the other values of ϵ lie almost in their entirety between the two plotted in the panels of the figure. It appears that for almost the whole range of updates, all agents in the explanationist community do better than their counterparts in the $\epsilon = 1$ Bayesian community for all biases with the exception of $p = .1$ (in which case they still do better in the long run). The comparison with the $\epsilon = 0.1$ Bayesian community is slightly more nuanced. Still, although for biases .2 to .5 there is a brief episode during which the Bayesian community does better overall than the explanationist community, in the long run, virtually all explanationists do better than their $\epsilon = 0.1$ Bayesian counterparts.

[29] Recall that agents are each other's counterparts iff they receive the same sequence of 500 coin tosses. Recall further that, given how the simulations are set up, for each possible bias and each simulation carried out for that possibility, every agent in a community has a counterpart in every other community.

[30] We indicated that we are interested in the chance that an explanationist incurs a lower Brier score than her Bayesian counterpart. However, the quantity plotted in Figure 5 is not a frequency, so it cannot be interpreted as (an estimation of) the chance directly. However, it is a linear function of the relevant frequency. In particular, to transform the difference plotted in the figure into a frequency, we have to add the number of agents (here 50) to it and divide the total by twice the number of agents (here 100).

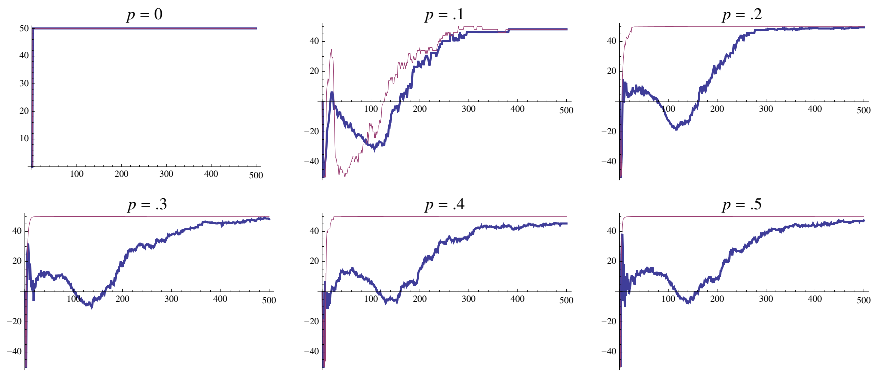

Figure 5: Average differences over 50 simulations between explanationists ($\epsilon = 1$, $c = 0.1$) who are faster than their Bayesian counterparts, where purple indicates the results under the assumption that for all Bayesians $\epsilon = 1$ and blue the results under the assumption that for all Bayesians $\epsilon = 0.1$ ($c = 0$ for all Bayesians). (For $p = 0$, there is no difference between the two graphs.)

Finally, we turn to the dimension of speed of convergence. In Douven [2013], speed of convergence was operationalized in terms of how fast agents come to assign a probability above .9 to the true bias hypothesis. We went with this operationalization and compared in this respect each of the averaging versions of Bayes' rule with (again) the explanationist rule that sets $\epsilon = 1$ and $c = 0.1$.

The results found in the earlier paper—to wit, that IBE updaters are in general faster than Bayesians in assigning a high probability to the true hypothesis—were found to generalize: the averaging explanationist rule in general has one assign a high probability to the truth faster than any of the averaging Bayesian rules. However, the difference between these averaging rules is even more pronounced than the one found for the non-averaging rules. We again show the results for the Bayesian rules only for $\epsilon = 0.1$ and $\epsilon = 1$, which are again the best and worst case, respectively, for the Bayesian. From top to bottom, the three rows of graphs in Figure 6 show for, respectively, $\epsilon = 1$ in combination with $c = 0$, $\epsilon = 0.1$ in combination with $c = 0$, and $\epsilon = 1$ in combination with $c = 0.1$, the percentage of agents in a population who assign, after the various updates, a probability above .9 to the truth, averaged over the 50 simulations. It is seen that when the coins have a strong bias (i.e., close to 0 or 1), both the Bayesian community (for either value of ϵ) and the explanationist community come rather quickly to assign in their entirety a high probability to the truth.

By contrast, when the coins have a more moderate bias, then whereas the explanationists all assign a high probability to the truth after 250 to 350 tosses, the whole Bayesian community—in the $\epsilon = 1$ case—or at least a large part of that community—in the $\epsilon = 0.1$ case—assigns a probability beneath the .9 threshold to the truth for the

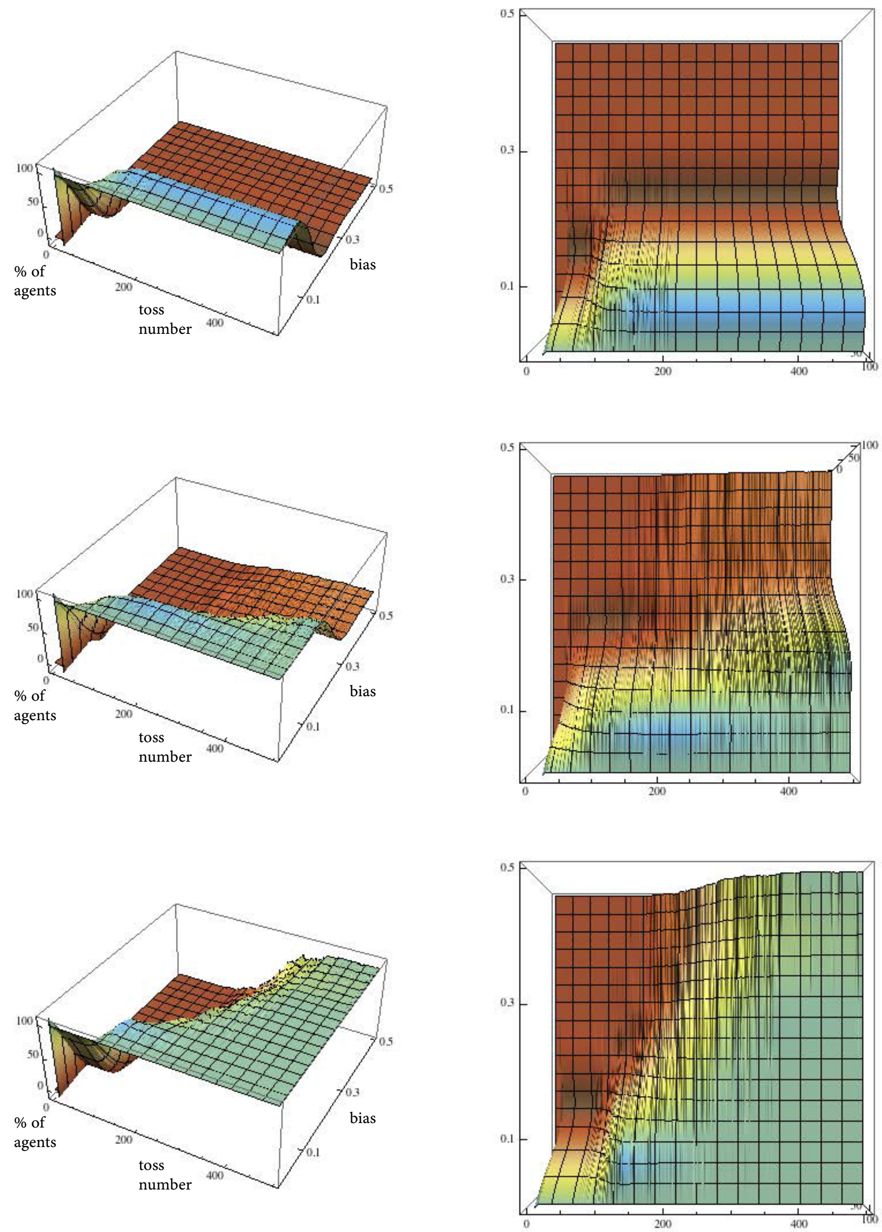

Figure 6: Percentage of population that assigns a probability above .9 to the true bias hypothesis, for possible biases $p = 0$ to $p = .5$. Top row: $\epsilon = 1$, $c = 0$; middle row: $\epsilon = 0.1$, $c = 0$; bottom row: $\epsilon = 1$, $c = 0.1$. (Each row shows the same graph twice, from two different angles.)

whole series of 500 tosses. To make clearer that the explanationists do much better also than the $\epsilon = 0.1$ Bayesians (which may not be immediately clear from the graphs), we mention that, on average over all biases and all updates, at the end of the series of updates 64 per cent of the explanationists assign a high probability to the truth as compared to 41 per cent of those Bayesians.

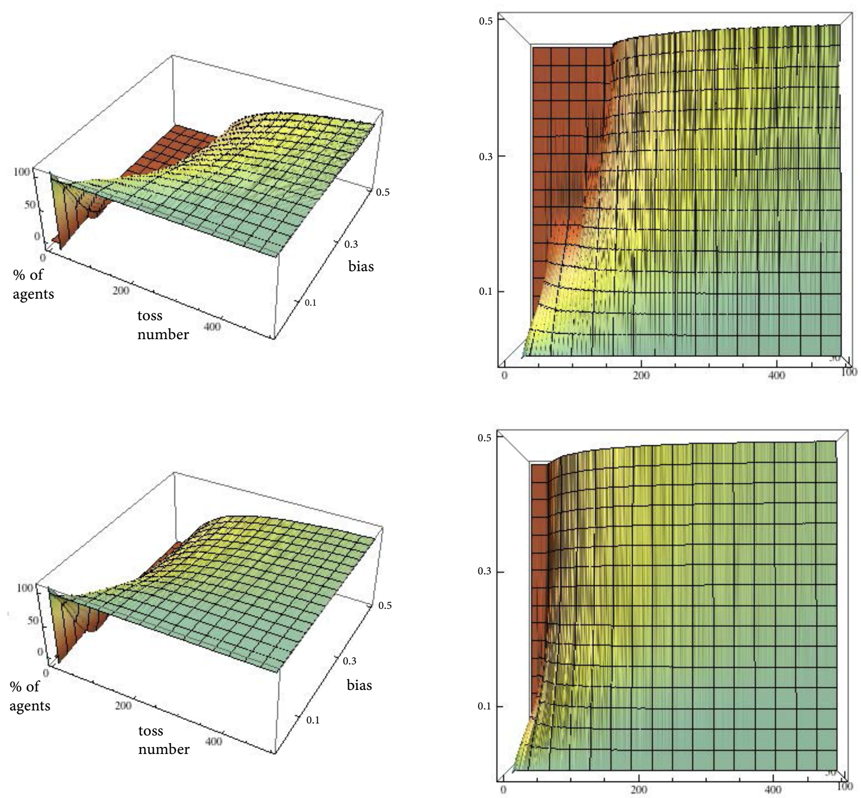

Figure 7: Percentage of population that assigns a probability above .9 to the true bias hypothesis, for possible biases $p = 0$ to $p = .5$. Upper row: $\epsilon = 0$, $c = 0$; lower row: $\epsilon = 0$, $c = 0.1$

Supposing the results obtained so far to have some generality that goes beyond the statistical model of Section 1, they strongly suggest that, in a social setting, one ought to be an explanationist. This is not only so if one's goal is to maximize the speed with which one comes to assign a high probability to the true hypothesis, but also if one's goal is to minimize one's inaccuracy, whether this is understood in terms of maximizing the chance that one incurs lower Brier penalties than one's Bayesian counterparts (or colleagues, in real life) or in terms of incurring on average a lower Brier penalty than one's Bayesian counterparts. This is a stronger conclusion than the one which was reached for updating in an individualistic setting, where simulations resulted in a more fragmented picture.

The results may also raise the hope that an independent argument in favor of social updating is in the offing. As stated, social epistemologists have drawn attention to the epistemic importance of various social practices: we often (have to) rely on others to extend our knowledge. It is no tenet of social epistemology that we *should* rely on others whenever possible. But the results visualized by Figures 2–4 might seem to suggest that, whenever we have a choice—which will not always be the case—we should not only go by the evidence we receive directly from the world but also let our beliefs be influenced by colleagues and others whose beliefs are close enough to our own.

However, this is *not* generally supported by our data. Figure 7 represents the results from our simulations concerning speed of convergence of non-averaging Bayesian communities and non-averaging explanationist communities. Comparison of the top row in this figure with the second and third row in Figure 6 shows that, in terms of speed of convergence, the non-averaging Bayesian communities do better than the (in the present respect) best averaging Bayesian communities as well as one of the (in the present respect) best averaging explanationist communities. Of course, as the lower row in Figure 7 shows, the non-averaging explanationist communities do better still. (That non-averaging explanationist communities do better in the current respect than non-averaging Bayesian communities was already known from the simulations discussed in Section 2.)

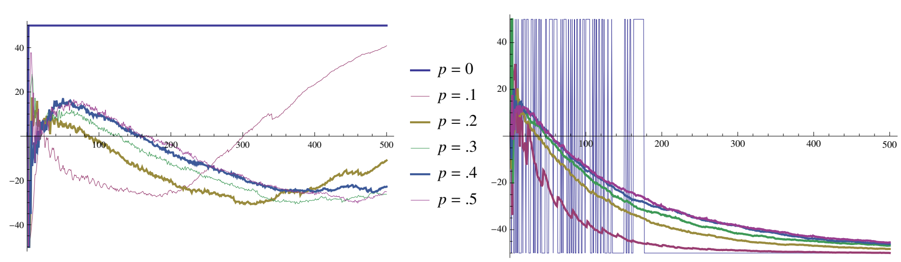

Figure 8: Average differences over 50 simulations between averaging explanationists ($\epsilon = 1$, $c = 0.1$) and their non-averaging Bayesian counterparts (left panel), respectively their non-averaging explanationist counterparts (right panel).

We can also compare, after any given update, the number of explanationists in an averaging community that have a Brier score that is lower than their Bayesian counterpart's score after the given update, as we did in Figure 5 for averaging explanationists and averaging Bayesians. It is clear from the left panel of Figure 8 that, overall, non-averaging Bayesians do better than the averaging explanationists with $\epsilon = 1$ and $c = 0.1$ (which are among the best averaging explanationists in this respect). We knew from the simulations described in Douven [2013] that non-averaging explanationists do better in this respect than non-averaging Bayesians, so the right panel of Figure 8 shows, not surprisingly, that non-averaging explanationists do *much* better in the respect of Brier score minimization than their averaging explanationist counterparts.

But note that even if we do not have a general argument in favor of social updating, we still find something interesting, to wit, that if you want to understand the epistemic goal of truth approximation in terms of either speed of convergence or increasing the chance of incurring a low Brier score, then, if you can, you should ignore the beliefs of others in your community. As was shown in Douven [2013], however, in those cases you have a reason to be an explanationist. If, on the other hand, you want to understand the epistemic goal in terms of minimizing average total Brier score, then, as far as the above results go, you have reason to also take into account the beliefs of at least certain others in your community (if possible). But then, as we saw, you *also* have reason to be an explanationist. So, if you have the option of taking into account the beliefs of others in your community, it may not always be best to take those beliefs into account, but it *is* always best to be an explanationist.

5. Interpretation. In the previous section, we offered some discussion along with the simulation results. We observed that—in a social setting and given various important conceptions of our epistemic goal—IBE outperforms Bayes' rule. Here, we wish to offer a more in-depth discussion of the causes underlying this observation.

To this end, we address two specific questions, both related to Figures 2 and 3, namely: (1) Why is the average Brier score per agent for most biases (other than $p = .1$ and $p = .9$) lower for an averaging explanationist than for a Bayesian? (2) In particular, why is there such a large difference in the average Brier score between an averaging Bayesian with a certain ϵ-value (bottom row in Figure 2) and an explanationist with the same ϵ-value who assigns a small explanation bonus ($c = 0.1$; i.e., second row from the bottom in Figure 2)?

To answer these questions, we first explain how it is possible at all to obtain a lower expected average Brier score than an averaging Bayesian. We do this by introducing a third option besides those of the averaging Bayesian and the averaging explanationist. This reference strategy is Bayesian, but it is only available to an agent who has full knowledge of the outcomes of the tosses observed by all the agents in the population. We show how the strategy of the averaging explanationist leads to an average Brier score that is intermediate between that of the averaging Bayesian and that of the reference strategy.

Assume that there is an agent who is "omniscient" in the sense that she has access to information concerning the toss results of all the agents in the population.[31] If the only epistemic goal is to minimize the expected Brier score, Bayesian updating is the

[31] Observe that in the simple model we discuss, any agent can get access to this information: based on another agent's probability function, the relative frequency and number of tosses can be deduced. However, this will no longer work in more realistic cases: an agent will typically be unable to deduce the information on which another agent's probability function is based, since the other agent's probability assignments are only known partially and approximately and the details of each agent's update process will typically vary among agents and be unknown to others.

optimal strategy to achieve it in a non-social setting (Douven [2013]). Hence, also for the omniscient agent (for whom the social context is of little relevance), Bayesian updating based on the totality of information about toss results is the optimal strategy to achieve a minimal expected Brier score.

Compared to this reference strategy, the averaging Bayesian is much more conservative, which leads to a Brier score that is higher on average. This should be easy to see. As the omniscient Bayesian updates on the basis of results of an increasing number of peers, the resulting peak in the probability function becomes *more pronounced* (larger difference between maximal and minimal value of the probability function) than before the update. This corresponds to the fact that the probability of the evidence's being misleading drops as the number of tosses increases.[32] Hence it is clear that on average this strategy will lead to the lowest Brier penalty. In contrast, by averaging over the probability functions of peers (who have received the same results as in the reference case), the peak position in an averaging Bayesian's probability function will be identical to that in the reference case, but the peak will become *less pronounced* than before the update.

In situations where the agents do not have access to the toss results of others (and cannot infer them either, as they could in our toy example), the reference strategy is not available to them. Averaging over other agents' probability functions may then seem an attractive alternative strategy, given that by averaging, the agent still takes into account more independent evidence (albeit indirectly, by looking at other agents' probability functions). This ensures that the position of the peak is based on more information and thus less likely to be based on misleading evidence. And yet averaging does not reflect this increase in information, in that the probability function obtained by averaging does not get more pronounced. In the example on which the numerical simulations were based, the agents observe different series of tosses all produced by coins with the same bias, so the agents receive statistically independent information. In such a context, it is clear that averaging over probability functions, which results in smoothing out rather than boosting the peak at the average position, is too conservative. If the reference strategy is not available, this leaves open the question of a better strategy (in the sense of lowering the expected Brier score). Averaging over probability functions may be a more natural idea in situations where the agents' opinions are based on information that is not (fully) independent and in which their way to arrive at probability assignments is not necessarily based on Bayesian considerations. But in such situations, too, there may be an update strategy that outperforms averaging Bayesians.

Informally, we are looking for a way to boost the signal of the maximum in the probability function that results from the social updating rule based on averaging. One way to achieve this is by introducing an explanation bonus. It increases the probability

[32] In the current context, we may define "misleading evidence" as an initial sequence of toss results with an observed relative frequency that deviates by more than .1 from the actual bias. For a fixed number of toss results, the probability of producing misleading evidence is zero for extreme biases ($p = 0$ and $p = 1$) and rises toward the middle bias ($p = .5$). For nonextreme biases, the more tosses the relative frequency is based on, the lower the probability for this sample to constitute misleading evidence.

of the hypothesis (or pair of hypotheses, in case of a tie) that receives the maximal probability value after updating in a social way, which is closest to the overall relative frequency.

In the toy model, it is not hard to see why this approach (usually) leads to a lower Brier score. The reference strategy gives rise to probability functions that get more sharply peaked as data from more agents is taken into account. The bonus can be regarded as a crude approximation to this: it adds a single peak to the relatively flat averaged function at the position where it reaches its maximal value. This also helps to understand in which cases the bonus approach may fail: the bonus introduces a constant peak, whereas the peak of the reference strategy is smaller at the beginning and more pronounced for a larger total number of tosses (from all the agents), and also depends on the position of the relative frequency (accounting for the variable probability of misleading evidence).

In cases where the bias has an extreme value (i.e., $p = 0$ or $p = 1$), the observed frequency is always exactly equal to the actual bias of the coin. Hence, for these bias values, explanationists converge faster on the true hypothesis as compared to the (more conservative) Bayesians. The closer the bias is to the middle value (i.e., $p = .5$), the more likely it is for agents to receive misleading evidence. In this case, the conservative Bayesians may accumulate fewer penalty points as compared to explanationists, especially at the initial stages of tossing.

Applied to our toy model, this analysis may suggest a modification of IBE, which would allow for using a variable explanation bonus, dependent on the number of tosses and the size of the peer group (total number of agents and ϵ). In general, a variable explanation bonus may codepend on factors that are harder to quantify in a numerical model, such as the reliability of other agents, the total amount of information, and the independence of their evidence.

We can now answer the questions posed at the beginning of this section. As for question (1), the average Brier score per agent for most biases is lower for an averaging explanationist than for an averaging Bayesian, because the averaging explanationist behaves more like a reference Bayesian with direct access to all the information than the averaging Bayesian does: the latter is too conservative and does not reflect the amount of independent evidence the position of the maximum is based on. As for question (2), the large difference in the average Brier score between an averaging Bayesian with a certain ϵ-value and an explanationist with the same ϵ-value who assigns a small explanation bonus arises because the probability function of the reference strategy has a small peak at the beginning which gets more pronounced over time (as more evidence is taken into account): this behavior can be approximated by adding a small bonus at each iteration. In fact it may be the case that the optimal bonus has a non-zero value smaller than 0.1. In particular, for smaller groups and smaller values of ϵ, the slower growth of the peak in the reference solution could be approximated by assigning a smaller value to c.

Since we considered a reference strategy that leads to even lower Brier scores than those of the averaging explanationist, at this point the reader may wonder why we do not consider the reference strategy as our preferred learning rule. Although we agree that the reference strategy offers an optimal solution for this very particular case, this suggestion goes against the spirit of what we are aiming to achieve here. After all, we are not looking for an optimal strategy for cases that are analytically tractable, in which all the evidence is available to all the agents, and in which the only epistemic goal is to minimize expected Brier penalty. Rather, we are looking for a learning rule that applies in a wide range of situations, in which the model is known to be an idealization, in which information regarding other agents' evidence is incomplete and mediated by probability assignments, in which we have to estimate how much independent data the assignment of each agent is based on, in which the data may be noisy, and so on. We are using the analytical toy model as a means to set up a simulation, not as the ultimate case of interest.

6. Conclusion. Earlier work on IBE suggested that this rule has two clear advantages as compared to Bayes' rule: updating by IBE will in general take one faster to the truth, in that one will faster assign a high probability to the true hypothesis; and it will in general minimize one's Brier penalties. But in that work IBE did not come out as a winner on all counts. In particular, it was found that, while explanationists will in general incur slightly lower Brier scores than their Bayesian counterparts, in the infrequent cases in which explanationists incur a *higher* Brier score than their Bayesian counterparts, the difference tends to be large enough to make a comparison in terms of *average* Brier scores turn out in favor of Bayes' rule.

But these results all pertained to a comparison of the two rules in an individualistic setting. This paper set out to compare them in a social setting, which takes into account the important role that others play in how we shape and reshape our belief states. It was seen that, in such a setting, assuming the same statistical model that was used in the individualistic setting, explanationist updating outperforms Bayesian updating on *all* desirable counts: it takes one, in general, faster to the truth; it minimizes, in general, one's Brier scores; and it leads, in general, to a lower *average* Brier score.

Admittedly, much still needs to be done. For one thing, we need a definition of "best explanation" that is more general than the one assumed in the simulations reported in this paper. For another, once a more general definition of "best explanation" is available, it remains to be seen how far our results generalize beyond the simple coin model that was used here. Meanwhile, the present results should give pause to all those who take it as almost a platitude that Bayesianism provides the one true confirmation theory.

Acknowledgments. We are greatly indebted to Christopher von Bülow and two anonymous referees for very valuable comments. Earlier versions of this paper were presented at the workshop Epistemic and Scientific Groups: Interdisciplinary Perspectives, which was held in Nancy (France), at the École Normale Supérieure, Paris, and at the VU University, Amsterdam. We thank the audiences on those occasions for stimulating questions and discussions. Sylvia Wenmackers's work was supported financially by a

Veni grant from the Netherlands Organization for Scientific Research (NWO project 639.031.244 "Inexactness in the exact sciences").

References

Bacharach, M. [1979]: 'Normal Bayesian Dialogues', *Journal of the American Statistical Association*, **74**, pp. 837–46.

Boyd, R. [1984]: 'The Current Status of Scientific Realism', in J. Leplin (*ed.*) *Scientific Realism*, Berkeley, CA: University of California Press, pp. 41–82.

Cooke, R. [1991]: *Experts in Uncertainty*, Oxford: Oxford University Press.

Dittmer, J. C. [2001]: 'Consensus Formation under Bounded Confidence', *Nonlinear Analysis*, **7**, pp. 4615–21.

Douven, I. [1999]: 'Inference to the Best Explanation Made Coherent', *Philosophy of Science*, **66**, pp. S424–35.

Douven, I. [2010]: 'Simulating Peer Disagreements', *Studies in History and Philosophy of Science*, **41**, pp. 148–57.

Douven, I. [2013]: 'Inference to the Best Explanation, Dutch Books, and Inaccuracy Minimisation', *Philosophical Quarterly*, **69**, pp. 428–44.

Douven, I. and Kelp, C. [2011]: 'Truth Approximation, Social Epistemology, and Opinion Dynamics', *Erkenntnis*, **75**, pp. 271–83.

Douven, I. and Riegler, A. [2010]: 'Extending the Hegselmann–Krause Model I', *Logic Journal of the IGPL*, **18**, pp. 323–35.

Foley, R. [1992]: 'The Epistemology of Belief and the Epistemology of Degrees of Belief', *American Philosophical Quarterly*, **29**, pp. 111–24.

Foley, R. [1993] *Working Without a Net*, Oxford: Oxford University Press.

Fortunato, S. [2004]: 'The Krause–Hegselmann Consensus Model with Discrete Opinions', *International Journal of Modern Physics* C, **15**, pp. 1021–9.

Genest, C. and Zidek, J. V. [1986]: 'Combining Probability Distributions: A Critique and an Annotated Bibliography', *Statistical Science*, **1**, pp. 114–35.

Goldman, A. [1999]: *Knowledge in a Social World*, Oxford: Oxford University Press.

Goldman, A. [2010]: 'Epistemic Relativism and Reasonable Disagreement', in R. Feldman and T. Warfield (*eds*), *Disagreement*, Oxford: Oxford University Press, pp. 187–215.

Harman, G. [1965]: 'The Inference to the Best Explanation', *Philosophical Review*, **74**, pp. 88–95.

Hegselmann, R. and Krause, U. [2002]: 'Opinion Dynamics and Bounded Confidence: Models, Analysis, and Simulations', *Journal of Artificial Societies and Social Simulation*, **5** (available at `http://jasss.soc.surrey.ac.uk/5/3/2.html`).

Hegselmann, R. and Krause, U. [2005]: 'Opinion Dynamics Driven by Various Ways of Averaging', *Computational Economics*, **25**, pp. 381–405.

Hegselmann, R. and Krause, U. [2006]: 'Truth and Cognitive Division of Labor: First Steps towards a Computer Aided Social Epistemology', *Journal of Artificial Soci-*

eties and Social Simulation, **9** (available at `http://jasss.soc.surrey.ac.uk/9/3/10.html`).

Hegselmann, R. and Krause, U. [2009]: 'Deliberative Exchange, Truth, and Cognitive Division of Labour: A Low-resolution Modeling Approach', *Episteme*, **6**, pp. 130–44.

Joyce, J. [1998]: 'A Nonpragmatic Vindication of Probabilism', *Philosophy of Science*, **65**, pp. 575–603.

Kelp, C. [2014]: 'No Justification for Lottery Losers', *Pacific Philosophicalo Quarterly*, **95**, pp. 205–17.

Kelp, C. and Douven, I. [2012]: 'Sustaining a Rational Disagreement', in H. de Regt, S. Hartmann, and S. Okasha (*eds*), *EPSA Philosophy of Science: Amsterdam 2009*, Berlin: Springer, pp. 101–110.

Koehler, D. J. [1991]: 'Explanation, Imagination, and Confidence in Judgment', *Psychological Bulletin*, **110**, pp. 499–519.

Kurz, S. and Rambau, J. [2011]: 'On the Hegselmann–Krause Conjecture in Opinion Dynamics', *Journal of Difference Equations and Applications*, **17**, pp. 859–76.

Latus, A. [2000]: 'Our Epistemic Goal', *Australasian Journal of Philosophy*, **78**, pp. 28–39.

Lehrer, K. [1974]: *Knowledge*, Oxford: Clarendon Press.

Leitgeb, H. and Pettigrew, R. [2010]: 'An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy', *Philosophy of Science*, **77**, pp. 236–72.

Lewis, D. [1999]: 'Why Conditionalize?', in his *Papers on Metaphysics and Epistemology*, Cambridge: Cambridge University Press, pp. 403–407.

Lipton, P. [2004]: *Inference to the Best Explanation* (2nd ed.), London: Routledge.

Lombrozo, T. [2006]: 'The Structure and Function of Explanations', *Trends in Cognitive Sciences*, **10**, pp. 464–70.

Lombrozo, T. [2007]: 'Simplicity and Probability in Causal Explanation', *Cognitive Psychology*, **55**, pp. 232–57.

Lombrozo, T. and Carey, S. [2006]: 'Functional Explanation and the Function of Explanation', *Cognition*, **99**, pp. 167–204.

Lorenz, J. [2007]: 'Continuous Opinion Dynamics under Bounded Confidence: A Survey', *International Journal of Modern Physics* C, **18**, pp. 1819–38.

Maher, P. [1992]: 'Diachronic Rationality', *Philosophy of Science*, **59**, pp. 120–41.

McMullin, E. [1992]: *The Inference that Makes Science*, Milwaukee, WI: Marquette University Press.

Okasha, S. [2000]: 'Van Fraassen's Critique of Inference to the Best Explanation', *Studies in History and Philosophy of Science*, **31**, pp. 691–710.

Pennington, N. and Hastie, R. [1992]: 'Explaining the Evidence: Tests of the Story-model for Juror Decision Making', *Journal of Personality and Social Psychology*, **62**, pp. 189–206.

Psillos, S. [1999]: *Scientific Realism: How Science Tracks Truth*, London: Routledge.

Psillos, S. [2004]: 'Inference to the Best Explanation and Bayesianism', in F. Stadler (*ed.*), *Induction and Deduction in the Sciences*, Dordrecht: Kluwer, pp. 83–91.

Psillos, S. [2007]: 'The Fine Structure of Inference to the Best Explanation', *Philosophy and Phenomenological Research*, **74**, pp. 441–8.

Riegler, A. and Douven, I. [2009]: 'Extending the Hegselmann–Krause Model III: From Single Beliefs to Complex Belief States', *Episteme*, **6**, pp. 145–63.

Roche, W. and Sober, E. [2013]: 'Explanatoriness is Evidentially Irrelevant, or Inference to the Best Explanation Meets Bayesian Confirmation Theory', *Analysis*, **73**, pp. 659–68.

Rosenkrantz, R. D. [1992]: 'The Justification of Induction', *Philosophy of Science*, **59**, pp. 527–39.

Sartwell, C. [1992]: 'Why Knowledge Is Merely True Belief', *Journal of Philosophy*, **89**, pp. 167–80.

Skyrms, B. [1987]: 'Coherence', in N. Rescher (*ed.*), *Scientific Inquiry in Philosophical Perspective*, Lanham, MD: University of America Press, pp. 225–41.

Stone, M. [1961]: 'The Opinion Pool', *Annals of Mathematical Statistics*, **32**, pp. 1339–42.

Tregear, M. [2004]: 'Utilising Explanatory Factors in Induction?', *British Journal for the Philosophy of Science*, **55**, pp. 505–19.

van Fraassen, B. C. [1989]: *Laws and Symmetry*, Oxford: Clarendon Press.

Weisberg, J. [2009]: 'Locating IBE in the Bayesian Framework', *Synthese*, **167**, pp. 125–43.

Wenmackers, S., Vanpoucke D. E. P. and Douven, I. [2012]: 'Probability of Inconsistencies in Theory Revision: A Multi-agent Model for Updating Logically Interconnected Beliefs under Bounded Confidence', *European Physical Journal* B, **85**, 44, doi: 10.1140/epjb/e2011-20617-8.

Wenmackers, S., Vanpoucke D. E. P. and Douven, I. [2014]: 'Rationality: A Social-epistemology Perspective', *Frontiers in Psychology*, **5**, 581, doi: 10.3389/fpsyg.2014.00581.

Williams, J. R. G. [2012]: 'Generalized Probabilism: Dutch Books and Accuracy Domination', *Journal of Philosophical Logic*, **41**, pp. 811–40.

Zollman, K. [2007]: 'The Communication Structure of Epistemic Communities', *Philosophy of Science*, **74**, pp. 574–87.

Zollman, K. [2010]: 'The Epistemic Benefit of Transient Diversity', *Erkenntnis*, **72**, pp. 17–35.