Explanation, updating, and accuracy

Igor Douven
Sciences, Normes, Décision (CNRS), Paris-Sorbonne University, Paris, France

ABSTRACT
There is evidence that people update their credences partly on the basis of explanatory considerations. Philosophers have recently argued that to minimise the inaccuracy of their credences, people’s updates also ought to be partly based on such considerations. However, there are many ways in which explanatory considerations can factor into updating, not all of which minimise inaccuracy. It is an open question whether in their updating, people take explanatory considerations into account in a way that philosophers would deem recommendable. To address this question, we re-analyse data from an experiment reported in Douven and Schupbach, “The role of explanatory considerations in updating” (Cognition, 2015).

1. Introduction

According to Bayesians, our credences (i.e. degrees of belief) ought to be coherent at any given time – that is, they ought to be formally probabilities – and we ought to update them by Bayes’ rule, meaning that, for all propositions A and B, upon the learning of A, our new unconditional credence in B ought to be set equal to our previous conditional credence in B given A (provided A is not deemed false with certainty). Evidence has accumulated that people often comply with these Bayesian norms (e.g. Oaksford and Chater 2007). Often, but not always: there are also reports of clear violations. While most of these reports concern the static norm of obedience to the probability axioms, a number of results indicate violations of Bayes’ rule, the norm of belief dynamics (Baratgin and Politzer 2007; Douven and Schupbach 2015a; Pennington and Hastie 1992; Robinson and Hastie 1985; Zhao, Crupi, Tentori, Fitelson, and Osherson 2012).

Violations of Bayes’ rule might be entirely unsystematic. Alternatively, they could be due to a bias in probabilistic reasoning, perhaps related to the biases that were discovered in work on the static norm (Kahneman, Slovic, and Tversky 1982). But Douven and Schupbach (2015a) found that, at least in some contexts, Bayes’ rule is systematically violated because people’s belief updates are influenced by explanatory considerations. These results did not come as a complete surprise, given that previous research had already shown that explanation plays a variety of roles in belief formation and in cognition more generally (Chi, de Leeuw, Chiu, and LaVancher 1994; Legare and Lombrozo 2014; Lombrozo 2006, 2007, 2012; Pennington and Hastie 1992; for more recent results, see Walker, Lombrozo, Williams, Rafferty, and Gopnik 2016).

In light of recent philosophical work on the normative status of explanatory reasoning, the aforementioned findings might appear encouraging, as they seemingly show that people are doing what, according to philosophers, they ought to do. But that is not necessarily so. What philosophers have argued is that some ways of taking explanatory considerations into account when updating one’s credences tend, in some contexts, to increase the accuracy of those credences as compared to when updating goes by Bayes’ rule. However, there are...
many ways to take explanatory considerations into account which will not lead to having more accurate credences, and can even easily lead to having less accurate credences. Douven and Schupbach’s study only reveals that people’s belief updates are guided, in part, by explanatory considerations, but not that this guidance has any positive effect (or any effect at all) on the accuracy of the resulting credences.

Here, we consider the following question: To the extent that people let their belief updates be influenced by explanatory considerations, what effect does that have on the accuracy of their credences?

2. Theoretical background

Let “descriptive explanationism” designate the claim that people, as a matter of fact, update their credences at least partly on the basis of explanatory considerations, and let “normative explanationism” designate the claim that people ought to update their credences at least partly on that basis. These claims are to be contrasted with descriptive Bayesianism – the claim that people update their credences via Bayes’ rule – and normative Bayesianism, the claim that people ought to (exclusively) update their credences via that rule.2

In philosophy, normative explanationism is closely associated with the so-called inference to the best explanation (IBE), which in its simplest (qualitative) version licenses us to infer the hypothesis that best explains the available data (Harman 1965). With the Bayesian revolution of the 1990s, IBE came under a cloud, if only because all standard versions of the rule were concerned with categorical belief, whereas Bayesians were convinced that the more fundamental notion of belief is one that admits of degrees.3

In psychology, the Bayesian revolution gave rise to what has been termed “New Paradigm Psychology of Reasoning” (Over 2009). It is important to realise, however, that the New Paradigm is not as strongly committed to Bayes’ rule as the frequent use of the labels “Bayesian” and “Bayesianism” by its practitioners suggests. For instance, Oaksford and Chater (2013, p. 374), who are among the New Paradigm’s staunchest advocates, admit that it is still “unclear what are the rational probabilistic constraints on dynamic inference.”

Moreover, it would be wrong to think that any version of IBE must be cast in terms of qualitative belief. By now, versions of IBE exist that output new credences. Douven (2013) and Douven and Wenmackers (2016) discuss a version of IBE that is much like Bayes′ rule except that it assigns a bonus to the most explanatory hypothesis. As these authors show, this rule produces coherent credences provided the input credences are coherent. And Douven (2016) presents a number of similar probabilistic versions of IBE that do not simply assign a bonus to the best explaining hypothesis, but that credit each hypothesis under consideration in proportion to its explanatory power.

Indeed, not only is there no good reason left to disregard IBE for those who hold that an understanding of human rationality can only be achieved by taking seriously the notion of graded belief, there are at least two good reasons for actively investigating probabilistic versions of IBE, one reason being normative, the other descriptive.

As for the normative reason, Douven (2016) shows that, in some contexts, the probabilistic versions of IBE proposed in that paper lead people to have more accurate credences than they would have if they updated by means of Bayes′ rule. To be clearer about this, it should first be pointed out that accuracy, as the notion pertains to credences, is commonly understood in terms of some scoring rule. There are a number of such rules, but the most popular one is still the Brier score, which was also the first scoring rule to be proposed (Brier 1950). Let \( \{H_i\}_{i \in \mathbb{N}} \) be a set of mutually exclusive and jointly exhaustive hypotheses, and let \( \llbracket H_i \rrbracket \in \{0, 1\} \) designate the truth value of \( H_i \). Then the Brier rule assigns a score of

\[
\sum_{i=1}^{n} (\llbracket H_i \rrbracket - \Pr(H_i))^2
\]

to a person whose credences are represented by \( \Pr(\cdot) \). For instance, if your cre-

---

2Normative Bayesianism is sometimes further divided into objective (or logical) Bayesianism and subjective Bayesianism, where the latter recognises the probability axioms and Bayes’ rule as the only rationality constraints, while the former holds that rational belief management requires compliance to additional principles. For the purposes of this paper, the objective/subjective distinction is immaterial, however, given that the data we shall draw upon all come from an experimental setting in which credences could be fully based on objective probabilities, and nowadays it is difficult to find any Bayesians who do not adhere to some version of the so-called Principal Principle, according to which credences should be based on objective probabilities, whenever these are available; see Lewis (1980).

3Philosophers in the Bayesian camp also objected to IBE because – they alleged – changing one’s credences by means of any rule other than Bayes′ makes one vulnerable to so-called dynamic Dutch books, which are collections of bets that all seem fair at the time they are offered but together ensure a negative net payoff; see Teller (1973). This dynamic Dutch book argument has been contested, however; see Douven (1999). Meanwhile, Bayesians themselves have come to regard Dutch book arguments as being generally misguided, in that they are concerned with practical rationality – specifically, the avoidance of monetary losses – rather than epistemic rationality; see, for instance, Joyce (1998).
dences in No Precipitation Tomorrow, Mild to Moderate Precipitation Tomorrow, and Heavy Precipitation Tomorrow are 0.2, 0.3, and 0.4, respectively, then, if tomorrow stays dry, your Brier score is \((1 - 0.2)^2 + 0.3^2 + 0.4^2 = 0.89\). Brier scores are penalties, so lower is better, with the minimum being reached when one believes the true hypothesis to the degree 1 (and does not invest any confidence in one of the other hypotheses).4

Douven (2016) compares in these terms Bayes’ rule with instances of the following schema:

\[
Pr(E|H_i) = \frac{Pr(H_i)Pr(E|H_i)}{\sum_{j=1}^n Pr(H_j)Pr(E|H_j) + c Pr(H_i)Pr(E|H_i)M(H_i, E)}.
\]

Here, \(Pr(\cdot)\) is one’s credence function prior to learning evidence \(E\); \(Pr(E|H_i)\) is one’s new credence in hypothesis \(H_i\) immediately after learning \(E\) (and nothing stronger); and \(c \in [0, 1]\) is a constant determining what percentage of \(H_i\)'s credibility is added in proportion to this hypothesis’ power to explain \(E\), where explanatory power is measured by \(M\).

Various probabilistic measures of explanatory power have been proposed in the philosophical literature (see Schupbach 2011). One of the measures used in Douven (2016) is Good’s (1960), which had done particularly well in empirical work (Douven and Schupbach 2015b). According to Good’s measure, the explanatory power of a hypothesis \(H\) in light of evidence \(E\) equals \(\ln(Pr(E|H)/Pr(E))\).

In Douven (2016), the instance of schema (1) with Good’s measure and \(c = 0.1\) is among the rules that are compared with Bayes’ rule by means of computer simulations. The simulations concerned a binomial process, like the repeated tossing of a coin with unknown bias or the turning of a roulette wheel to determine whether the table is well balanced or rather tends to favour the landing of the ball on one of red and black. In the simulations, it was antecedently given that there are 11 hypotheses, \(\{H_i\}_{0 \leq i \leq 10}\), concerning the long-run relative frequency of a specific one of the two outcomes (e.g. the coin landing heads, or the ball landing on red), with \(H_i\) being the hypothesis that, in the long run, the relative frequency of the designated outcome will converge to \(i/10\). The 11 hypotheses were assumed to be mutually exclusive and jointly exhaustive and to be initially equally likely. For each \(H_i\), 1000 simulations of 500 repeated trials were conducted, where the chance for the designated outcome (e.g. the coin coming up heads, or the ball landing on red) to occur was \(i/10\).

The question these simulations were meant to answer is what difference it would make to the development, and especially the accuracy, of a person’s credences whether he or she uses one update rule or the other for adapting his or her credences in light of the incoming results from the binomial process. The general answer was that, under rather general conditions, updating one’s credences by means of some version of IBE can have significant epistemic advantages – including achieving greater accuracy – as compared to updating one’s credences by means of Bayes’ rules. To illustrate this finding, Figure 1 shows the Brier penalties incurred by Bayes’ rule and the aforementioned instance of Equation (1), after each of the 500 trails, and averaged over 1000 simulations, where the outcomes were generated on the assumption that \(H_3\) was the true hypothesis (left panel) and that \(H_5\) was the true hypothesis (right panel, see Douven 2016 for the full results). Figure 1 also compares the so-called Bayes factor, after every trial, for the differences in scores between the rules.5 Thus, after every trial, the 1000 scores for one rule at that point were compared with the 1000 scores for the other, and then the Bayes factor was calculated. If the Bayes factor is below 0, that indicates support (to some extent) for the null hypothesis that there is no difference in means of the scores, and the further below 0 the Bayes factor is, the stronger is the support for the null hypothesis. If the Bayes factor is greater than 0, this indicates support for the alternative hypothesis that there is a difference, and the greater the Bayes factor, the stronger the support is. Of course, the difference can then be in favour of one rule or of the other, where the difference being in favour of a given rule here means that rule has a reliably lower Brier score. (Recall that Brier scores are penalties and so that lower is better.) To show which of these

---

4 There is a second popular scoring rule, the so-called log score. Where \(H_i\) is the true hypothesis, this rule assigns a penalty of \(-\ln(Pr(H_i))\) to a person whose degrees of belief are given by \(Pr(\cdot)\). In Douven (2016), both the Brier and the log score are used for measuring accuracy. To save space, we only consider the Brier score in this paper. It is worth noting that scoring rules were originally proposed for the purpose of eliciting credences. However, apart from this ex ante use (as Winkler et al. 1996 calls it), scoring rules can also be used ex post, for evaluating credences. In Douven (2016), the Brier and log rules were used ex post.

5 See Lee and Wagenmakers (2013, 104 ff) for details on Bayes factors.
is the case, Figure 1 uses black dots to indicate support for the hypothesis that the difference is in favour of Bayes’ rule and dark grey dots to indicate support for the hypothesis that the difference is in favour of the version of IBE we are considering. For the \( p = 0.4 \) case (so the case where \( H_4 \) is true), we see that after a short initial segment in which the difference is in favour of Bayes’ rule, there is a long stretch in which the difference is in favour of IBE; then for about the last 200 trails, there is no difference between the rules. For the \( p = 0.5 \) case, the results are even clearer in favour of IBE: again after a brief initial segment in which Bayes’ rule is more accurate, IBE has a reliably lower Brier score for the rest of the trials.

Given that (normative) Bayesians have come to put great emphasis on the importance of having accurate credences, such results should give them pause. At a minimum, and to repeat, these results should be reason to actively study probabilistic versions of IBE, such as the instances of Equation (1).

As for the descriptive reason, Schupbach (2011) reports a small-scale experiment in which ten balls were sequentially drawn, without replacement, from an urn with black and white balls, where the urn had been randomly selected from two urns – named “urn A” and “urn B” in the experiment – each containing 40 balls, but with different ratios of black and white balls (30/10 for urn A, 15/25 for urn B). The contents, but not the identities, of the urns were known to the participants, and the participants were asked, after each draw, to judge in light of the outcomes so far: (i) the explanatory goodness of the hypothesis that urn A had been selected; (ii) the explanatory goodness of the hypothesis that urn B had been selected; and (iii) the likelihood that urn A had been selected. Participants were requested to give judgements (i) and (ii) by making a mark on a continuous scale from \(-1\) to 1, which had, at equal distances, five labelled points, the leftmost label reading that the hypothesis at issue was an extremely poor explanation of the evidence so far, the right-most reading that the hypothesis was an extremely good explanation, and the labels in between reading that the hypothesis was a poor/neither poor nor good/good explanation, in the obvious order.\(^6\) The goal of the study was to determine the extent to which various probabilistic measures of explanatory goodness capture people’s subjective verdicts of explanatory goodness.

In Douven and Schupbach (2015a), the same data were used to see whether judgements of explanatory goodness influenced people’s belief updates. It turned out that they did, which motivated a follow-up experiment basically repeating the experiment from Schupbach (2011) but now with a much larger cohort of participants. This experiment yielded further strong evidence for the influence of judgements of explanatory goodness on how people update their credences. A regression analysis showed that a model that had, in addition to

\[\text{Figure } 1. \text{ Brier scores after each trial, averaged over 1000 simulations, on the assumption that the designated outcome has an objective probability of 0.4 (left) or 0.5 (right) of occurring. Black lines indicate scores for Bayes’ rule, grey lines scores for the version of IBE with Good’s measure. Shown on the alternative } y\text{-axis are the results (the logarithms of Bayes factors) of running a Bayesian } t\text{-test on the scores after each trial. Black dots indicate that the difference between the means is in favour of Bayes’ rule, dark grey dots that the difference is in favour of IBE, and light grey dots that there is no significant difference between the means.}\]
objective probabilities, judgements of explanatory goodness among the independent variables was significantly more accurate in predicting participants’ credences after each draw than a model with objective probabilities as the only independent variable. The best model had both objective probabilities and the difference in explanatory goodness between the hypotheses at issue as independent variables. According to descriptive Bayesianism, the model with only objective probabilities as independent variable should have performed best, and adding as a variable judgements of explanatory goodness, or the difference in explanatory goodness between the two hypotheses, should not have led to any noteworthy improvement. Instead, model fit was greatly improved by adding those variables, and that supports descriptive explanationism.

It might thus seem that people do what they ought to do, namely attribute weight to explanatory considerations in updating their credences. As noted in the introduction, however, this conclusion would be rash. For as Douven (2016) also shows taking explanatory considerations into account is not guaranteed to make one more accurate and can even make one less accurate. It all depends on how explanatory considerations factor into belief updates. For instance, the rule termed EXPL in Douven (2016), which is a probabilistic version of IBE, leads in computer simulations to less accurate credences than Bayes’ rule or various other versions of IBE do. Consequently, from the fact that Douven and Schupbach (2015a) found support for descriptive explanationism, we cannot simply infer that people update in a way that tends to make them more accurate than they would be were they Bayesian updaters. Perhaps people factor explanation into their updates by following something like EXPL, in which case they may end up with less accurate credences than they would have by following Bayes’ rule. Section 3 has another look at Douven and Schupbach’s (2015a) data to address the question of whether people’s taking explanatory considerations into account in updating helps them arrive at more accurate credences.

3. Explanation and accuracy

Schupbach’s (2011) experiment had only 26 participants, which was not enough for the analysis envisaged in Douven and Schupbach’s work on explanationism. Also not ideal for the purposes of that analysis was that in Schupbach’s experiment each participant received a unique series of draws. For these reasons, the larger follow-up experiment mentioned above was conducted. In this experiment, there were 259 participants, who were divided into four groups, with each group receiving a different one of four series of draws that had been randomly chosen from the 26 series that had occurred in Schupbach’s original experiment. Just as in that experiment, participants were informed that the balls would be drawn without replacement from an urn that had been chosen from two urns – again called “urn A” and “urn B” – on the basis of the flip of a fair coin. They were also fully informed about the contents of each urn. They were then shown, one at a time, the 10 draws and asked, after each draw, to make the judgements labelled (i) to (iii) above.7

Of Douven and Schupbach’s 259 participants, 206 remained after exclusion of participants on the basis of a number of standard criteria (being nonnative speakers of English, etc.), and 167 remained after further exclusion for failing to answer some comprehension questions correctly. Regressing these participants’ credences that urn A had been selected, first, onto the objective probabilities that urn A had been selected, and second, onto those objective probabilities as well as the difference between (a) the judged explanatory goodness of the hypothesis that urn A had been selected and (b) the judged explanatory goodness of the hypothesis that urn B had been selected, led to the results already mentioned in the previous section: the larger model vastly outperformed the “Bayesian” model with only objective probabilities as independent variable, and it did so across all conventional measures of model fit. Most notably, the Bayesian model had an AIC value of −1627.55 while the larger, “explanationist” model had an AIC value of −2422.07.

Douven and Schupbach were interested in the empirical adequacy of descriptive explanationism; they were not concerned with the issue of accuracy. However, it is possible to bring their data to bear on that issue, as follows: For each of the four series of draws used in the experiment, it is known from which urn they had come in Schupbach’s smaller experiment. (Groups 1, 2, and 4 had received series of draws that had come from urn B; group 3 had received a series of draws from urn A.)

7The data of this experiment are available at http://www.sciencedirect.com/science/article/pii/S0010027715000955#MMCvFirst.
Consequently, for each of the participants, we can calculate the Brier score incurred over the 10 draws. We can then run a separate regression analysis for each participant, again with credences as dependent variable and objective probabilities and difference in judged explanatory goodness between the hypotheses as independent variables. The β weights for the independent variables we thereby obtain can be interpreted as the relative weights a participant assigns to probability and explanation in determining his or her new credence after seeing the outcome of a draw. And finally, it can be checked how these weights relate to how accurate – in terms of Brier penalties – the participant was. This should illuminate our research question: we might find that giving more weight to explanation tends to increase accuracy, or the opposite, or we might find no relation at all.

Nine participants had zero variance in their responses to the questions of how likely it was that urn A had been selected (they consistently gave .5 as an answer after each draw), which made it impossible to reliably estimate β coefficients for the probability and explanation variables for them. For that reason, these participants were excluded from the analysis, leaving 158 participants.

Note that, for any coherent credence function, the maximum Brier score is 2 and the minimum is 0. Hence, the maximum total Brier score incurable after all 10 updates that were recorded in the experiment is 20 (the minimum is, of course, again 0). Table 1 gives summary statistics for these scores, both split up for the groups separately and for all participants together. It does the same for the β coefficients for objective probabilities and difference in judged explanatory goodness that were obtained from running the individual regression analyses mentioned in the previous paragraph.

A one-way ANOVA with total Brier scores as dependent variable and group as independent variable showed that the differences between the group means for the total Brier scores were significant: MSE = 30.25, MSE = 30.25, p < 0.0001; the effect size was $\eta^2_p = 0.209$, which counts as very large. Post hoc comparisons using Tukey’s HSD indicated that in particular the differences in mean total Brier score between groups 1 and 4, groups 2 and 3, and groups 3 and 4 were significant at an α level of 0.001, while the difference in means between groups 1 and 2 was significant at an α level of .05.8 Two further one-way ANOVAs with group as independent variable and probability and explanation, respectively, as the dependent variable were nonsignificant: $F(3, 154) = 1.94, MSE = 0.26, p = .13$, for probability; $F(3, 154) = 1.18, MSE = 0.21, p = .32$, for explanation.

To determine the relationship between, on the one hand, the weights the participants gave to probability and to explanation and, on the other, the total Brier score they incurred, linear models were fitted with, as dependent variable, the total Brier scores of all 158 participants, and, as independent variables, one or both of these participants’ β weights for probability and the β weights for explanation that resulted from the individual regression analyses described above. To control for the significant differences in group means for Brier scores, we included group as a factor in the models.

In a model that had as independent variables, next to the β weights for probability and explanation, also their interaction, this interaction was not significant. Therefore, we consider as the full model the one with β weights for probability and for explanation, but not their interaction, as independent variables. Besides this model, we also fitted a linear model with only the β weights for probability as independent variable as well as a linear model with only the β weights for explanation as independent variable. The full model clearly topped the others, with an AIC value of 529.56 as opposed to 582.66 for the probability-only model and 556.53 for the explanation-only model, and a

---

Table 1. Summary statistics for total Brier scores and β coefficients for probability and explanation.

|       | N | Brier score M | Probability (β) M | Explanation (β) M | SD | SD |
|-------|---|--------------|-------------------|-------------------|----|----|
| All   | 158 | 3.18         | −0.51             | 0.37              | 0.28 | 0.42 |
| Gr1   | 46  | 2.82         | −0.43             | 0.40              | 0.28 | 0.50 |
| Gr2   | 30  | 3.79         | −0.57             | 0.32              | 0.16 | 0.35 |
| Gr3   | 44  | 2.28         | −0.48             | 0.39              | 0.35 | 0.44 |
| Gr4   | 38  | 4.18         | −0.61             | 0.32              | 0.30 | 0.34 |

---

For those wondering what might explain these differences, note that Brier scores can vary greatly just depending on the actual sequence of outcomes. To illustrate, consider the sequences wbwbwbwbwb and bbbbbwwww (with b for a draw of a black ball and w for a draw of a white ball), which have a roughly equal chance of occurring on the supposition that urn B (with 15 black balls and 25 white ones) was selected. If someone’s credence for the hypothesis that urn B has been selected would, after any given draw, be exactly equal to the objective probability for that hypothesis, after the given draw, then she would incur a Brier penalty of 1.07 if she received the first sequence of draws and a Brier penalty of 7.46 if she received the second.
I. DOUVEN

−

some otherwise ordinary people are much more concerning geopolitical questions, it was shown that tory considerations tended to be also more accurate. that participants who gave more weight to explana-

tion in updating their credences. But in general less accurate the more weight they give

nations (Lombrozo2007; Sloman1994, 1997). In the inclined to overestimate the likelihood of best expla-

ability had shown that people are sometimes the relationship between explanation and prob-

score by over two points. The effect of attending measured in standard deviations) lowers the Brier

Every extra unit of weight given to explanation (as every extra unit of weight given to explanation (as measured in standard deviations) lowers the Brier score by over two points. The effect of attending to objective probabilities in the model was, although similarly directed, somewhat smaller. So, supposing objective probabilities are available, it is certainly a good idea to attend to them, in line with what normative Bayesians would recommend. However, contrary to what those Bayesians would recommend it is also a good idea to attend at the same time to explanatory considerations as doing so is likely to further increase the accuracy of one’s credences.

4. General discussion

Nothing about accuracy followed from Douven and Schupbach’s (2015a) finding that explanation was a significant predictor for their participants’ repeatedly updated credences. In fact, previous work on the relationship between explanation and probability had shown that people are sometimes inclined to overestimate the likelihood of best explanations (Lombrozo 2007; Sloman 1994, 1997). In the light of those results, Douven and Schupbach’s finding might lead one to suspect that people are in general less accurate the more weight they give to explanation in updating their credences. But this suspicion was not confirmed by our new analysis. To the contrary, a regression analysis showed that participants who gave more weight to explanatory considerations tended to be also more accurate.

In recent publications on prediction tournaments concerning geopolitical questions, it was shown that some otherwise ordinary people are much more accurate forecasters than even professional intelligence analysts (Mellers et al. 2015; Tetlock and Gardner 2015).9 A key objective of the research was to determine what distinguishes the most accurate forecasters – superforecasters, as they have been called – from the rest of the population. It turned out that while superforecasters stood out, rather predictably, on IQ, in particular fluid intelligence, they also had characteristic cognitive styles and skills that can be acquired or practised. Among those, it was noted that superforecasters updated their credences more frequently than others and were also better at translating information into probability judgements (e.g. relating qualitative causal judgements to probabilities).

While there is no explicit mention of superforecasters’ being Bayesian updaters in Mellers et al. (2015), Tetlock and Gardner (2015, 170 ff) do note that many superforecasters are familiar with Bayes’ theorem and are also able to use it in updating their credences. They add, however, that “[w]hat matters far more to the superforecasters than Bayes’ theorem is Bayes’ core insight of gradually getting closer to the truth by constantly updating in proportion to the weight of the evidence” (p. 171). As was seen in Section 2, that core insight is also implemented by probabilistic versions of IBE, and some of those will even allow us to approach the truth more accurately than Bayes’ rule does. It is impossible to tell from the results on superforecasting that have come out so far whether superforecasters might also differ from others in the weight they assign to explanatory factors in their updates. But the analysis from the previous section gives some reason to take this possibility seriously and to look into it in future research.

In closing, I would like to mention two further avenues for future research. The first concerns the question of why explanatory considerations should guide us closer towards the truth, as at least in some contexts they appear to do. At the most general level, to explain is to provide understanding, to help make sense of things (de Regt and Dieks 2005). Why should the truth be such that it provides understanding to creatures with our cognitive endowments? According to a once popular conception of truth – the so-called pragmatic conception of truth – the link between truth and understanding holds by definition; we call

9Here, too, accuracy was measured by means of the Brier scoring rule.
something “true” only if it “[helps] us to get into satisfactory relations with other parts of our experience” (James 1907/1975, p. 34). However, this conception of truth has lost much, if not all, of its erstwhile appeal. Douven (2002) argues that, if there is a link between truth and explanation indeed, it must hold as an empirical matter of fact, not analytically, because of the meaning of the truth predicate. If this is correct, then empirical research rather than philosophical speculation seems needed to properly explore the designated link.

Second, we have considered the effects of explanatory considerations on accuracy in a specific type of context. It was also mentioned (in note 6) that it is probably wrong to conceive of explanation as a unitary concept. Moreover, we just said that there is evidence that sometimes explanatory considerations affect the accuracy of people’s beliefs negatively. So, how far can the results reported in this paper be generalised? Which types of explanatory considerations foster accuracy in which types of contexts? Here, too, progress is most likely to come from further empirical investigation.

Acknowledgments

I am greatly indebted to Linden Ball, Shira Elqayam, Henrik Singmann, Christopher von Bülow, and three anonymous referees for valuable comments on earlier versions of this paper.

Disclosure statement

No potential conflict of interest was reported by the author.

References

Baratgin, J., & Politzer, G. (2007). The psychology of dynamic probability judgment: Order effect, normative theories, and experimental methodology. Mind & Society, 6, 53–66.
Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. Monthly Weather Review, 78, 1–3.
Chi, M. T. H., de Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. Cognitive Science, 18, 439–477.
Colombo, M. (2016). Experimental philosophy of explanation rising: The case for a plurality of concepts of explanation. Cognitive Science. doi: 10.1111/cogs.12340
de Regt, H. W., & Dieks, D. (2005). A contextual approach to scientific understanding. Synthese, 144, 137–170.
Douven, I. (1999). Inference to the best explanation made coherent. Philosophy of Science, 66, S424–S435.
Douven, I. (2002). Testing inference to the best explanation. Synthese, 130, 355–377.
Douven, I. (2013). Inference to the best explanation, Dutch books, and inaccuracy minimisation. The Philosophical Quarterly, 63, 428–444.
Douven, I. (2016). Inference to the best explanation: What is it? And why should we care? In T. Poston & K. McCain (Eds.), Best explanations: New essays on inference to the best explanation. Oxford: Oxford University Press.
Douven, I., & Schupbach, J. N. (2015a). The role of explanatory considerations in updating. Cognition, 142, 299–311.
Douven, I., & Schupbach, J. N. (2015b). Probabilistic alternatives to Bayesianism: The case of explanationism. Frontiers in Psychology, 6, 459. doi:10.3389/fpsyg.2015.00459
Douven, I., & Wenmackers, S. (2016). Inference to the best explanation versus Bayes’ rule in a social setting. The British Journal for the Philosophy of Science. doi: 10.1093/bjps/axv025
Good, I. J. (1960). Weight of evidence, corroboration, explanatory power, information and the utility of experiment. Journal of the Royal Statistical Society, B22, 319–331.
Harman, G. H. (1965). The inference to the best explanation. The Philosophical Review, 74, 88–95.
James, W. (1907/1975). Pragmatism: A new name for some old ways of thinking. Cambridge, MA: Harvard University Press.
Joyce, J. M. (1998). A nonpragmatic vindication of probabilism. Philosophy of Science, 65, 575–603.
Kahneman, D., Slovic, P., & Tversky, A. (1982). Judgment under uncertainty: Heuristics and biases. Cambridge: Cambridge University Press.
Kern-Isberner, G. (2001). Revising and updating probabilistic beliefs. In M. A. Williams & H. Rott (Eds.), Frontiers in belief revision (pp. 393–408). Dordrecht: Kluwer.
Lee, M., & Wagenmakers, E.-J. (2013). Bayesian cognitive modeling. Cambridge: Cambridge University Press.
Legare, C. H., & Lombrozo, T. (2014). Selective effects of explanation on learning during early childhood. Journal of Experimental Child Psychology, 126, 198–212.
Lewis, D. K. (1980). A subjectivist’s guide to objective chance. In R. C. Jeffrey (Ed.), Studies in inductive logic and probability (pp. 83–132). Berkeley: University of California Press.
Lombrozo, T. (2006). The structure and function of explanations. Trends in Cognitive Sciences, 10, 464–470.
Lombrozo, T. (2007). Simplicity and probability in causal explanation. Cognitive Psychology, 55, 232–257.
Lombrozo, T. (2012). Explanation and abductive inference. In K. J. Holyoak & R. G. Morrison (Eds.), Oxford handbook of thinking and reasoning (pp. 260–276). Oxford: Oxford University Press.
Mellers, B., Stone, E., Murray, T., Minster, A., Rohrbaugh, N., Bishop, M., … Tetlock, P. (2015). Identifying and cultivating superforecasters as a method of improving probabilistic predictions. Perspectives on Psychological Science, 10, 267–281.
Oaksford, M., & Chater, N. (2007). Bayesian rationality. Oxford: Oxford University Press.
Oaksford, M., & Chater, N. (2013). Dynamic inference and everyday conditional reasoning in the new paradigm. *Thinking & Reasoning, 19*, 346–379.
Over, D. E. (2009). New paradigm psychology of reasoning. *Thinking & Reasoning, 15*, 431–438.
Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology, 62*, 189–206.
Robinson, L. B., & Hastie, R. (1985). Revision of beliefs when a hypothesis is eliminated from consideration. *Journal of Experimental Psychology: Human Perception and Performance, 11*(4), 443–456.
Schupbach, J. N. (2011). Comparing probabilistic measures of explanatory power. *Philosophy of Science, 78*, 813–829.
Sloman, S. A. (1997). Explanatory coherence and the induction of properties. *Thinking & Reasoning, 3*, 81–110.
Teller, P. (1973). Conditionalization and observation. *Synthese, 26*, 218–258.
Tetlock, P., & Gardner, D. (2015). *Superforecasting: The art and science of prediction*. London: Penguin Random House.
Walker, C., Lombrozo, T., Williams, J. J., Rafferty, A., & Gopnik, A. (2016). Explaining constrains causal learning in childhood. *Child Development*. doi: 10.1111/cdev.12590
Walliser, B., & Zwirn, D. (2002). Can Bayes’ rule be justified by cognitive rationality principles? *Theory and Decision, 53*, 95–135.
Winkler, R. L., Muñoz, J., Cervera, J. L., Bernardo, J. M., Blattenberger, G., Kadane, J. B., … Rios-Insua, D. (1996). Scoring rules and the evaluation of probabilities. *Test, 5*, 1–60.
Zhao, J., Crupi, V., Tentori, K., Fitelson, B., & Osherson, D. (2012). Updating: Learning versus supposing. *Cognition, 124*, 373–378.