Can the Evidence for Explanatory Reasoning Be Explained Away?

Igor Douven

Can the Evidence for Explanatory Reasoning Be Explained Away?

Article *in* IEEE Transactions on Cognitive and Developmental Systems · July 2018
DOI: 10.1109/TCDS.2018.2861832

Igor Douven
French National Centre for Scientific Research

Can the Evidence for Explanatory Reasoning Be Explained Away?

Igor Douven

***Abstract*—Recent evidence appears to show a close connection between explanation and belief revision, specifically, the revision of graded beliefs. Insofar as this is also evidence of violations of Bayesian norms of reasoning, the question arises whether we are facing a new bias here, on a par with previously discovered biases in probabilistic reasoning. We consider an apparently successful attempt by Costello and Watts to explain away a number of known such biases in terms of sampling error, which makes those biases look entirely innocuous and compatible with the descriptive adequacy of Bayesian psychology in any but the most uninteresting way. Specifically, we query whether this attempt can be extended to neutralize the aforementioned evidence allegedly showing that explanatory considerations influence our reasoning in ways inconsistent with Bayesian prescriptions.**

***Index Terms*—Bayes' rule, belief updating, biases, explanation, probability, reasoning.**

I. Introduction

THERE is growing evidence that explanation plays a variety of important roles in human cognition. (For an overview, see [23].) Recent evidence also appears to show a close connection between, on the one hand, explanation and questions of explanatory goodness and, on the other, how people revise (or "update") their degrees of belief or subjective probabilities ([2], [10], [11]). Specifically, people tend to reward hypotheses on the basis of how well they explain the available evidence, where the reward consists of a probability boost over and above what would be warranted by Bayesian norms, in particular Bayes' rule. According to Bayes' rule, we should update our degree of belief in a hypothesis H upon the receipt of evidence E by adopting as our new *unconditional* degree of belief in H whatever our degree of belief in H *conditional on* E was before we received that evidence. People tend to violate this prescription to the extent that they find H a good explanation of E, in which case their new degree of belief is likely to exceed their previous conditional degree of belief, an excess that comes at the expense of one or more of H's rival hypotheses, which is or are believed to a degree less than prescribed by Bayes' rule.

Questions concerning the normative status of explanatory reasoning have been hotly debated for several decades now. According to some (e.g., [8], [13], [7]), explanatory reasoning is fine, and under certain circumstances even recommendable, possible inconsistencies with Bayesian norms notwithstanding. Authors in this camp tend to favor some version of the so-called inference to the best explanation, which licenses inference to, or at least the investment of high confidence in, the hypothesis that explains the available evidence best. But this is a minority position. Most researchers interested in the matter would say that the aforementioned finding betokens a bias (e.g., [22], [31]). Giving extra credits for explanatory goodness may be a deeply entrenched tendency in some (or even all) of us, but from a normative standpoint, that practice is to be condemned.

Normative questions to the side, that explanatory considerations appear to have this influence on people's reasoning puts pressure on the *descriptive adequacy* of Bayesian approaches to rationality. While this pressure may not be felt so much by Bayesian philosophers and economists, who primarily see Bayesian norms as providing an ideal that may be outside our reach but to which we should nevertheless aspire, there is a growing number of psychologists drawn to the view that people often do comply with those norms, at least by and large (e.g., [25], [29]).

This paper looks at recent attempts by Costello and Watts [5], [6] to explain away deviations from Bayesian norms in terms of "noisy" sampling from memory. The deviations addressed by these authors were not directly related to the *updating* of degrees of belief, but there is nothing in their approach per se to suggest that it might not apply equally to registered deviations from Bayes' rule. What would thereby be achieved is *not* that somehow what appear to be deviations fall into place as being in accordance with Bayesian updating after all. Instead, while there would still be evidence of a probabilistic bias, this bias would be of an utterly boring kind, not *really* brought about by some illegitimate influence of explanatory considerations on reasoning, but simply due to the well-recorded fact that human memory is all too fallible. In other words, the bias would be of a kind we should expect to observe even if people did what they could to meet Bayesian standards. If so, then that would show that descriptive Bayesianism can still be correct—to the extent that any theory of rationality can be descriptively correct, given some long-known human imperfections.

II. Explanation and updating

Evidence that people update their degrees of belief in ways inconsistent with Bayes' rule dates back at least to the 1960s. Phillips and Edwards [27] used an at the

Igor Douven is with Sorbonne University, CNRS, Sciences, Normes, Décision, 75005 Paris, France (e-mail: igor.douven@sorbonne-universite.fr).

time new experimental paradigm to show that people's probability estimates after the receipt of new evidence differed systematically from what they should have been according to Bayes' rule. In this paradigm, participants are informed that one container (e.g., an urn or a bookbag) holds two different types of objects (e.g., red chips / blue chips, or black balls / white balls) in one specific ratio, and another container holds those same types of objects in a different ratio. The participants are then shown a collection of objects sampled randomly from one of the containers, without being told from which, and are asked to estimate the probability that the sample comes from the first container rather than from the second. That the estimates deviate reliably from Bayesian updates has been replicated in numerous experiments since (see, e.g., [14], [18], [28]).

None of the early work on updating was concerned with explanatory reasoning. The first researchers to look into the connection between updating and explanation were Pennington and Hastie, in their influential work on juror decision making [26]. This work showed the importance of the order in which evidence is presented to jurors, specifically, that participants are much more inclined to judge a defendant guilty if the prosecutors present their evidence in an order that facilitates the mental construction of an explanatory story of how the crime unfolded. Pennington and Hastie also found evidence that how the different pieces of evidence impacted their participants' degrees of belief differed consistently from what descriptive Bayesianism would predict.

In an even more direct attempt to compare Bayesian and explanatory reasoning, Bes et al. [2] gave participants information both about causal relations among three random variables and about statistical correlations among those variables. Their results were strong evidence that the participants had tended to ignore the statistical information and to base their probability judgments strictly on the causal information. Although they were not thereby violating Bayes' rule (which is not meant to cover the provision of statistical information; see [16]), they *were* violating a closely related principle, almost equally widely endorsed by Bayesians, to wit, the so-called principal principle [21], which dictates that one's *degree of belief* in H on the supposition that H's *statistical probability* equals x, ought to equal x as well.

According to Bes et al., the effect of the causal information on their participants' probability judgments is likely related to the extent to which the participants could process that information into an explanatory story. It was not part of these authors' design to ask participants to judge the explanation quality of the statements whose probabilities they were asked to estimate. Had they done so, the data might have revealed a strong correlation between those judgments and the probability estimates that their participants did provide.

That, at least, is suggested by the experimental results from Douven and Schupbach [10]. These authors fell back upon the "bookbags-and-pokerchips" paradigm [14] used in the earliest studies on probabilistic updating, though they added some twists. In this paradigm, Douven and Schupbach designed a sequential probabilistic updating task that would allow them also to measure the degree to which explanatory factors influenced the updating. More specifically, they interviewed participants individually, showing them, at the start, two urns, each of which contained 40 balls. Participants were shown that one urn ("urn A") contained 30 black balls and 10 white ones, and that the other ("urn B") contained 15 black balls and 25 white ones. This information stayed available, in the form of a picture of those contents, during the whole interview. Then the experimenter flipped a coin and, based on the outcome, chose one or the other urn, outside of the participants' view, after which 10 balls were drawn, one by one, from the selected urn and were lined up before the participants. The participants were asked to answer three questions after each draw. In order, these were (i) how well the hypothesis that urn A had been selected explained the results from the drawings so far; (ii) how well the hypothesis that urn B has been selected explained those results; and (iii) how probable it was, in their opinion, that urn A had been selected, in view of the results so far. Participants had to answer the questions about explanatory goodness by marking a point on a continuous scale with anchors "extremely poor explanation" and "extremely good explanation."

In the main part of their analysis, Douven and Schupbach fitted three mixed-effects models, each with the collected participants' responses to the third question as fixed effect, at least the objective conditional probabilities as predictor variable, and random slopes and intercepts for participants. In one model (model MMO, in [10]), objective conditional probabilities were the only predictor. A second model further included both the collected responses to the first question and the collected responses to the second question as predictors (MMOAB). A third model had next to objective conditional probabilities the computed *difference* between the participants' responses to the first question and their responses to the second question as predictor (MMOD). Because including more predictors will in general lead to better model fit, these models were compared using criteria that penalize for extra predictors, in particular the so-called Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), which weigh (in slightly different ways) model fit against model complexity. The results from the model comparisons are summarized in Table I.

AIC values and BIC values are to be interpreted as penalties, meaning that lower is better. Also, they only make sense comparatively, and then only for models that are fit to the same data. The difference in AIC value between MMO and either of the other models was greater than 10, which according to Burnham and Anderson [3] is to be interpreted as indicating that the former model enjoys *no* support from the data, given the availability of the other models. The pattern is the same for the BIC values.

These model comparisons would seem alarming for

Table I
MODEL COMPARISON RESULTS CONCERNING THE MAIN MODELS FROM DOUVEN AND SCHUPBACH (2015).

	k	LL	AIC	ΔAIC	BIC	ΔBIC
MMO	6	282.34	−552.68	82.37	−531.32	68.12
MMOAB	15	329.13	−628.26	6.79	−574.85	24.59
MMOD	10	327.53	−635.05	0.00	−599.44	0.00

Note: k is the number of parameters and LL the log-likelihood. ΔAIC is the difference in AIC value with the model with lowest AIC value, and similarly for ΔBIC.

Bayesians, at least for those holding that Bayesian norms also achieve greater predictive accuracy than competing norms. After all, again by the principal principle, Douven and Schupbach's participants, after each draw, should have set their degree of belief that urn A had been selected equal to whatever the objective conditional probability was that that urn had been selected, given all registered draws at that point. It is not just that this failed to materialize. The bigger problem is that, while from a Bayesian perspective it should appear puzzling why the participants' judgments of explanatory goodness reliably helped to predict their degrees of belief, the previously mentioned rival norm of inference to the best explanation *predicts* that to happen: according to that norm, judgments of explanatory goodness *should* guide our belief updating.

III. JUST NOISE?

Perhaps Bayesians can avoid this problem. Recent work by Costello and Watts shows that a number of experimental findings that have been reported in the literature as potentially undermining descriptive Bayesianism may in fact be perfectly compatible with the correctness of that position, properly qualified. The same may hold true for the data from Douven and Schupbach's study.

At the root of Costello and Watts' work [5], [6] is the utterly plausible contention that people make random errors in estimating probabilities, including estimating conditional probabilities. For example, on their account we would estimate the conditional probability that a student will pass a given exam on the supposition that he or she studies hard by sampling from memory students who worked hard for an exam and then counting among those the ones who passed the exam. But this process is error-prone: our memories are not fully dependable, and we may thus misremember a student who worked hard but failed to have passed the exam, or the other way around; indeed, we may even misremember a student to have worked hard for an exam. Costello and Watts [6, p. 9] make the general assumption that "events have some chance $d < 0.5$ of randomly being read incorrectly." They demonstrate that, for some probabilistic identities, such errors tend to cancel out, whereas for others they do not or even get compounded. They muster an impressive amount of evidence for their hypothesis and show how that helps to explain a number of well-known biases, including order effects in sequential probability judgments [24] and the celebrated conjunction fallacy [30].

The aim of Costello and Watts is not to show that these and related biases are unproblematic, contrary to what the mainstream believes. Rather, it is to identify their source, namely, the noise present in the process by which we determine probabilities. Thus—the claim is—the said biases do not refute descriptive Bayesianism, given that this position is not committed to the assumption that human memory is failsafe.[1] Because "our reasoning processes are necessarily subject to noise" [5, p. 131], any alternative to descriptive Bayesianism faces the same problem that it can only be accurate up to the biasing effects of sampling errors.

Costello and Watts do not consider deviations from Bayes' rule.[2] But they are right to remark that their results present a challenge that goes beyond non-Bayesian explanations of the biases they did consider:

> random noise in reasoning can cause systematic biases in people's responses even when people are using the rational reasoning processes of standard frequentist probability estimation. To demonstrate conclusively that people are using heuristics, researchers must show that observed biases cannot be explained as the result of systematic effects caused by random noise. [5, p. 132]

Surely the same is true for attempts to explain biases by reference to mechanisms other than heuristics, like for instance attempts to attribute the findings summarized in the previous section to the influence of explanatory considerations on people's updating their degrees of belief. Might

[1] To forestall misunderstanding, "Bayesianism" is understood here as being committed to Bayes' rule being the only rational update rule. (The term is used somewhat loosely by both psychologists and philosophers, and different authors may use it to designate slightly different positions; see [15] and [11].) Thus understood, it is neutral regarding the exact nature of conditional probabilities, so in particular regarding the question of whether these are to be interpreted in terms of betting dispositions, relative frequencies, as introspectively given, or some combination of these possibilities. To see that a commitment to Bayes' rule does not *ipso facto* commit one to the aforementioned possibilities, note that Bayes' rule only relates conditional degrees of belief at one point in time to unconditional degrees of belief at a later point in time, while being silent on where the former come from. This terminological issue is important, inasmuch as Costello and Watts [5, p. 132] seem to understand the Bayesian position as being incompatible with the idea that conditional probabilities can be interpreted as relative frequencies. See also footnote 2.

[2] That is, they do not consider deviations from what is here called "Bayes' rule," which is an update rule. They do consider what they call "Bayes rule identities," by which they mean the most direct consequences of the standard ratio definition of conditional probability, such as that $\Pr(A \mid B)\Pr(B) = \Pr(B \mid A)\Pr(A)$, or that $\Pr(A \mid B) = \Pr(B \mid A)\Pr(A) \div \Pr(B)$. Costello and Watts [5] refer to the latter as "Bayes' rule." That is fine, the terminology not being very consistent in the Bayesian literature. Note, however, that their Bayes' rule is actually a theorem of probability theory (and is therefore more often referred to as "Bayes' theorem") and does not concern any updating of probabilities, time not being a parameter in probability theory.

that ostensible influence not also be in reality attributable to estimation errors?

First, there are some a priori reasons to doubt that Douven and Schupbach's results are due to the kind of sampling noise that figures in Costello and Watts's account. As said, that noise is supposed to arise from the fact that we estimate probabilities, including conditional probabilities, by sampling and counting from memory, and in that process sometimes make mistakes. While plausible in general, it is to be recalled that in the experiment reported in [10], participants at all times had the drawing of the urns with their contents in front of them, and were allowed to consult this memory aid as often as they wanted. At a minimum, their estimation of conditional probabilities would seem to have carried a lesser risk of being affected by noise than if these conditional probabilities had been the result of sampling from memory.

Second, adopting Costello and Watts' proposal to account for the deviations from Bayes' rule recorded by Douven and Schupbach would raise an explanatory challenge. For then where would the predictive superiority of the explanatory models come from? Given that, in Costello and Watts' model, the noise is supposed to be *random*, how could it moderate people's probability judgments precisely in such a way that their explanatory judgments would come out as highly significant predictors in the mixed models analysis Douven and Schupbach conducted? There does not appear to be anything in Costello and Watts' model that could account for a close link between the (putative) random noise at work in people's probability estimates and those same people's judgments of explanatory goodness.

But raising a priori doubts about the applicability of Costello and Watts' proposal to the data at issue will only take us so far. After all, while it is uncontroversial that human memory is error prone, researchers have identified various sources of noise in the nervous system that can have behavioral consequences, including causing information processing slips [17]. So, random noise may corrupt the estimation of conditional degrees of belief even if that estimation does not involve any sampling from memory. And for the Bayesian purpose of explaining away the seeming influence of explanatory considerations, the exact provenance or nature of the noise are immaterial. Indeed, while Costello and Watts' assumption that the errors they measured are due to memory glitches may be plausible, there is nothing in their papers to exclude that those errors are not actually, or not partly, of a different origin.[3]

Thus, a better way to answer the question of whether Costello and Watts' model can account for Douven and Schupbach's findings is to try and predict the subjective updated probabilities reported by the latter authors not by the objective conditional probabilities but instead by the "noisy" version of that predictor, transformed according to a formula given by Costello and Watts.

[3]Nor would it matter to their overall conclusion. See the citation given earlier in this section, where they refer to random noise in reasoning in general.

To do so, we first defined a function that takes Costello and Watts' error parameter d as input and outputs the transformed predictor O (the objective conditional probabilities), to be designated as $f(\mathrm{O}, d)$, for the given value of d. The function f is based on Equation 17 in [6], according to which

$$\begin{aligned}\Pr_*(A \mid B) \;=\; &\big((1-2d)^2 \Pr(A \wedge B) \;+\\ &d(1-2d)\big(\Pr(A)+\Pr(B)\big)+d^2\big) \;\div\\ &\big((1-2d)\Pr(B)+d\big),\end{aligned}$$

where $\Pr_*(A \mid B)$ is the noisy estimate of the probability of A conditional on B, and with the noise parameter $d \in [0, .5)$. Notice that if $d = 0$, indicating that there is no noise, then $\Pr_*(A \mid B) = \Pr(A \wedge B)/\Pr(B) = \Pr(A \mid B)$.

We used the `MixedModels.jl` package [1] for the scientific computing language Julia [4] to fit, for d going from 0 to 0.5 in increments of 0.005, mixed models like the ones specified above except that the new models have the noisy objective conditional probabilities, $f(\mathrm{O}, d)$, rather than the untransformed ones, O, as a predictor. If the deviations from Bayes' rule reported in Douven and Schupbach are due to the interference of noise in registering frequencies, in the manner of Costello and Watts, then we should find that for some values of d the corresponding Bayesian model with $f(\mathrm{O}, d)$ as its only predictor does best, and adding judgments of explanatory goodness as predictors should not lead to any improvement.

Figure 1 shows the AIC values of all the models that were fit. It is immediately obvious that replacing O by $f(\mathrm{O}, d)$, for any admissible value of d, only led to worse fit and that in particular for no value of d did the Bayesian model—or at least the model that does not take judgments of explanatory goodness into account—come close, in terms of model fit, to the models that do take such judgments into account (whether directly or indirectly, via the difference in goodness between the two urn hypotheses). Comparison in terms of BIC values led to qualitatively identical results. (See this Jupyter notebook for details; the data from [10], which are required to run the notebook, can be downloaded here.)

In short, not only is there a priori little reason to believe that Costello and Watts' proposal might also hold the key to understanding the deviations from Bayesian updating to be found in the data from [10], but there is also no support from the data.

IV. Conclusion

We had a second look at data that, at face value, undermine descriptive Bayesianism, according to which people accommodate new information by updating their degrees of belief via Bayes' rule, that is, by making their former conditional degrees of belief, on the supposition that the information they acquired holds true, into their new unconditional degrees of belief. The point of this endeavor was to see whether a new proposal used by Costello and Watts in an attempt to explain away some long-known

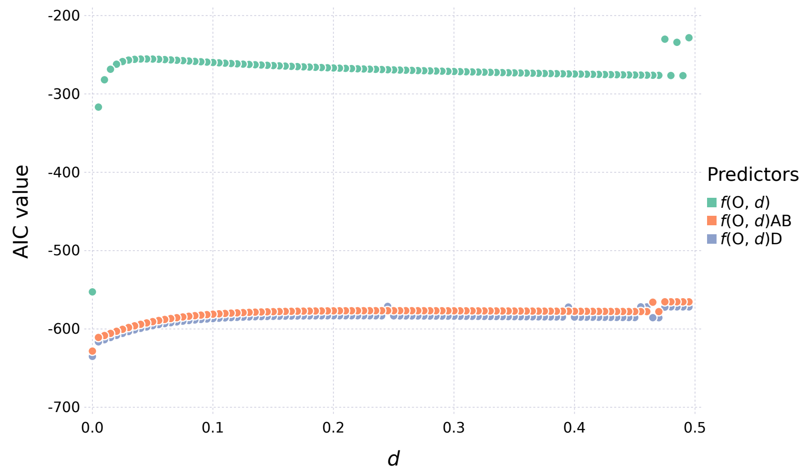

Figure 1. AIC values of the models with the same predictors as MMO, MMOAB, and MMOD, respectively, except that O is replaced by $f(\mathrm{O}, d)$, for d going from 0 to 0.5 in steps of 0.005.

probabilistic biases could also account for the deviations from Bayes' rule.

Costello and Watts' proposal relied on the idea that people arrive at their conditional degrees of belief by sampling from memory, and the further—by itself eminently plausible—idea that by doing so they are susceptible to making random errors. We thus considered a great number of transformations of the objective conditional probabilities that, from a descriptive Bayesian standpoint, should have been able to best account for the participants' belief changes in Douven and Schupbach's [10] experiment, and we checked whether one or more of those transformations did in fact best account for those belief changes, notably better than the models that also attended to the participants' judgments of explanatory goodness. The answer turned out to be negative.

Thus, the evidence for holding that the way people change their degrees of belief is influenced by explanatory considerations still stands. Whether this is evidence for a bias—something regrettable even if perhaps unavoidable—or for people's updating their degrees of belief rationally, just not according to the rationality criteria advocated by Bayesians, is a question we explicitly sidestepped here.

It is to be emphasized that nothing said in the above jeopardizes Costello and Watt's proposal. First, these authors are concerned to show that data that have been taken to be evidence for people's relying on heuristics rather than *probability theory* can actually be interpreted as being evidence for people's relying on probability theory while being subject to sampling errors. And Bayes' rule, as understood in this paper, is an update rule, that is, a rule connecting degrees-of-belief functions at different points in time, and as such is not part of probability theory (see note 2). Indeed, there are versions of inference to the best explanation that comply perfectly with the axioms of probability theory ([9], [12], [19], [20]).

Second, Costello and Watts nowhere claim that their proposal is meant to apply to each and every deviation of any norm proposed by Bayesians. As mentioned, they point out that the success of their proposal in explaining away some probabilistic biases raises a challenge for anyone claiming to have discovered another such bias, but that leaves open the possibility that the challenge can be met. In the present case, it could.[4]

References

[1] D.M. Bates, `MixedModels.jl`. Julia package version v0.17.0, https://github.com/dmbates/MixedModels.jl, doi = 10.5281/zenodo.1211642, 2018.

[2] B. Bes, S. Sloman, C.G. Lucas, and E. Raufaste, "Non-Bayesian inference: Causal structure trumps correlation," *Cognitive Science*, vol. 36, pp. 1178–1203, 2012.

[3] K.P. Burnham and D.R. Anderson, *Model selection and multi-model inference: A practical information-theoretic approach*. Berlin: Springer, 2002.

[4] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, "Julia: A fresh approach to numerical computing," *SIAM Review*, vol. 59, pp. 65–98, 2017.

[5] F. Costello, and P. Watts, "People's conditional probability judgments follow probability theory (plus noise)," *Cognitive Psychology*, vol. 89, pp. 106–133, 2016.

[6] F. Costello and P. Watts, "Invariants in probabilistic reasoning," *Cognitive Psychology*, vol. 100, pp. 1–16, 2018.

[7] D. Danks, "Probabilistic models," in M. Colombo and M. Sprevak (eds.), *Routledge handbook of the computational mind*. London: Routledge, in press.

[8] I. Douven, "Inference to the best explanation made coherent," *Philosophy of Science*, vol. 66, pp. S424–S435, 1999.

[9] I. Douven, "Inference to the best explanation: What is it? And why should we care?" in T. Poston and K. McCain (eds.), *Best explanations: New essays on inference to the best explanation*, pp. 4–22. Oxford: Oxford University Press, 2017.

[10] I. Douven and J.N. Schupbach, "The role of explanatory considerations in updating," *Cognition*, vol. 142, pp. 299–311, 2015.

[11] I. Douven and P. Mirabile, "Best, second-best, and good-enough explanations: How they matter to reasoning," *Journal of Experimental Psychology: Language, Memory, and Cognition*, in press.

[12] I. Douven and S. Wenmackers, "Inference to the best explanation versus Bayes' rule in a social setting," *British Journal for the Philosophy of Science*, vol. 68, pp. 535–570, 2017.

[13] F. Eberhardt and D. Danks, "Confirmation in the cognitive sciences: The problematic case of Bayesian models," *Minds and Machines*, vol. 21, pp. 389–410, 2011.

[4] I am greatly indebted to Mike Oaksford and to an anonymous referee for valuable comments on a previous version.

[14] W. Edwards, “Conservatism in human information processing,” in B. Kleinmuntz (ed.), *Formal representation of human judgment*, pp. 17–52. New York: Wiley, 1968.
[15] S. Elqayam and J. St. B. T. Evans, “Rationality in the new paradigm: Strict versus soft Bayesian approaches,” *Thinking & Reasoning*, vol. 19, pp. 453–470, 2013.
[16] B. Eva and S. Hartmann, “Bayesian argumentation and the value of logical validity,” *Psychological Review*, in press.
[17] A. A. Faisal, L. P. J. Selen, and D. M. Wolpert, “Noise in the nervous system,” *Nature Reviews Neuroscience*, vol. 9, pp. 292–303, 2008.
[18] B. Fischhoff and and S. Lichtenstein, “Don’t attribute this to Reverend Bayes,” *Psychological Bulletin*, vol. 85, pp. 239–243, 1978.
[19] D. H. Glass, “Coherence measures and inference to the best explanation,” *Synthese*, vol. 157, pp. 275–296, 2007.
[20] D. H. Glass, “Inference to the best explanation: Does it track truth?” *Synthese*, vol. 185, pp. 411–427, 2012.
[21] D. K. Lewis, “A subjectivist’s guide to objective chance,” in R. C. Jeffrey (ed.), *Studies in inductive logic and probability*, vol. 2, pp. 263–293. Berkeley CA: University of California Press, 1980.
[22] P. Lipton, *Inference to the best explanation* (2nd ed.). London: Routledge, 2004.
[23] T. Lombrozo, “Explanatory preferences shape learning and inference,” *Trends in Cognitive Sciences*, vol. 20, pp. 748–759, 2016.
[24] D. W. Moore, “Measuring new types of question-order effects: Additive and subtractive,” *Public Opinion Quarterly*, vol. 66, 80–91, 2002.
[25] M. Oaksford and N. Chater, *Bayesian rationality*. Oxford: Oxford University Press, 2007.
[26] N. Pennington and R. Hastie, “Reasoning in explanation-based decision making,” *Cognition*, vol. 49, pp. 123–163, 1993.
[27] L. D. Phillips and W. Edwards, “Conservatism in a simple probability inference task,” *Journal of Experimental Psychology*, vol. 72, pp. 346–354, 1966.
[28] D. A. Schum, I. L. Goldstein, and J. F. Southard, (1966). “Research on a simulated Bayesian information-processing system,” *IEEE Transactions on Human Factors in Electronics*, vol. 7, pp. 37–48, 1966.
[29] J. B. Tenenbaum, C. Kemp, T. L. Griffiths, and N. D. Goodman, “How to grow a mind: Statistics, structure, and abstraction,” *Science*, vol. 331, pp. 1279–1285, 2011.
[30] A. Tversky and D. Kahneman, “Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment,” *Psychological Review*, vol. 90, pp. 293–315, 1983.
[31] B. C. van Fraassen, *Laws and Symmetry*. Oxford: Oxford University Press, 1989.