Cognition 142 (2015) 299–311

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

The role of explanatory considerations in updating

Igor Douven[a,*], Jonah N. Schupbach[b,1]

[a] Sciences, Normes, Décision, Paris-Sorbonne University, France
[b] Department of Philosophy, University of Utah, United States

ARTICLE INFO

Article history:
Received 4 July 2014
Revised 27 April 2015
Accepted 29 April 2015
Available online 9 June 2015

Keywords:
Explanation
Updating
Inference
Probability
Bayesianism

ABSTRACT

There is an ongoing controversy in philosophy about the connection between explanation and inference. According to Bayesians, explanatory considerations should be given weight in determining which inferences to make, if at all, only insofar as doing so is compatible with Strict Conditionalization. Explanationists, on the other hand, hold that explanatory considerations can be relevant to the question of how much confidence to invest in our hypotheses in ways which violate Strict Conditionalization. The controversy has focused on normative issues. This paper investigates experimentally the descriptive question of whether judgments of the explanatory goodness of hypotheses do play a role when people revise their degrees of belief in those hypotheses upon the receipt of new evidence. We present the results of three experiments that together strongly support the predictive superiority of the explanationist position.

1. Introduction

Human learning involves the modifying ("updating") of degrees of confidence (or "credences") over time in response to new information. A crucial question thus facing any account of human learning is what factors and principles determine when and to what extent we change our credences. The Bayesian response to this question is that, upon learning that some proposition is undoubtedly true, an agent's updated credences ought to match his or her prior credences (taken just prior to learning that new information) conditional on the supposition that same proposition holds true. In other words, where credences are formally represented as probabilities, posterior (updated) probabilities are determined by prior conditional probabilities. This idea gets officially canonized in the following rule:

Strict Conditionalization: Upon learning $A \in \mathcal{A}$ and nothing else between times t_1 and t_2, an agent's credences are to be updated so as to satisfy the equality $\Pr_{t_2}(B) = \Pr_{t_1}(B|A)$ for all propositions $B \in \mathcal{A}$ (provided $\Pr_{t_1}(A) > 0$).

Here, $\mathcal{A}$ is an algebra of propositions over which the probability measures $\Pr_{t_1}$—representing the agent's credences at t_1—and $\Pr_{t_2}$—representing the agent's credences at t_2—are defined, and $\Pr_{t_1}(B|A)$ designates the conditional probability of B given A at t_1. Adherence to Strict Conditionalization commits Bayesians to the claim that prior conditional credences *alone* determine posterior credences. An ongoing controversy in philosophy points to an alternative theory, however. In debate over the confirmation-theoretic status of explanatory considerations, so-called explanationists hold that judgments of the explanatory goodness of hypotheses are directly relevant to the question of how much confidence we should invest in those hypotheses. Moreover, this claim is often taken to imply that explanatory considerations influence our credences in ways not captured by Strict Conditionalization.[2] By contrast, Bayesians hold that explanatory considerations have no home in confirmation theory, at least not in any way that might conflict with Strict Conditionalization (Salmon, 2001; van Fraassen, 1989).

* Corresponding author.
E-mail addresses: igor.douven@paris-sorbonne.fr (I. Douven), jonah.n.schupbach@utah.edu (J.N. Schupbach).

[1] The authors contributed equally to this article.

[2] Some explanationists have proposed ways in which explanatory considerations might influence posterior probabilities *purely via* prior conditional probabilities, thus making explanation's influence on updating compatible with Strict Conditionalization (see, e.g., Okasha (2000), Lipton (2004), and Weisberg (2009); see Douven (2011) for a critical discussion of this approach). This is akin to the proposal, made by some psychologists (e.g., Lombrozo, 2007), that priors may be informed by considerations of simplicity. Such an approach provides a gloss on, rather than alternative to, Bayesianism. In this paper, we reserve the label "explanationism" for approaches that are incompatible with Bayesianism.

http://dx.doi.org/10.1016/j.cognition.2015.04.017
0010-0277/

The philosophical controversy has focused on normative issues—particularly the issue of whether explanatory considerations ought to be given special confirmatory weight in the logic of updating. This paper instead presents three experiments that collectively aim to investigate the respective *descriptive* merits of Bayesianism and explanationism. For our purposes, when taken as descriptive theories, we understand these positions as involving the following general claims:

Bayesianism: People actually update their credences in accordance with Strict Conditionalization.

Explanationism: In updating their credences, people take into account their explanatory judgments in a way not already captured by Strict Conditionalization.

Importantly, there are two distinct ways in which explanationism could prove descriptively superior to Bayesianism. Firstly, explanatory considerations may have a significant role in a descriptive account of updating *in addition to* conditional probabilities. The implication here would be that the Bayesian model is on the right track, but ultimately in need of explanationist augmentation. Secondly, explanatory judgments may have a significant role in a descriptive account of updating *in place of* conditional probabilities. Here, the supported model would be one that abandons the Bayesian model altogether, *replacing*—rather than augmenting—conditional probabilities with explanatory judgments.

Our investigation aims to shed light on the following specific questions:

Q1. How do Bayesianism and explanationism compare with regard to their descriptive adequacy? Do judgments of the explanatory goodness of hypotheses play an essential role in updating in a way that is incompatible with the Bayesian doctrine?
Q2. If explanatory judgments are found to have such a role, do conditional probabilities retain an important influence in updating alongside such judgments?
Q3. What sort of explanatory judgments in particular (if any) factor into updating?

Past studies suggest that people's updates deviate from Strict Conditionalization (e.g., Phillips & Edwards, 1966; Robinson & Hastie, 1985; Zhao, Crupi, Tentori, Fitelson, & Osherson, 2012). There is also empirical work showing that explanatory considerations do have an impact on people's beliefs (e.g., Douven & Verbrugge, 2010; Khemlani & Johnson-Laird, 2011, 2012, 2013; Koehler, 1991; Lombrozo, 2006, 2007; Lombrozo & Carey, 2006; Pennington & Hastie, 1992). And while of late a broadly Bayesian approach to learning has come to resonate among many working in cognitive psychology (e.g., Baratgin, Over, & Politzer, 2013; Elqayam & Evans, 2013; Oaksford & Chater, 2007; Oaksford & Chater, 2013; Over, 2009) and cognitive neuroscience (e.g., Doya, Ishii, Pouget, & Rao, 2006, Friston and Stephan, 2007, Hohwy, 2013), we are not aware of any research in those areas that could be said to favor Strict Conditionalization over some probabilistic version of explanationism. Indeed, Oaksford and Chater (2013:374) conclude their discussion of the issue of belief change in the context of the new Bayesian paradigm in the psychology of reasoning with the remark that "it is unclear what are the rational probabilistic constraints on dynamic inference." At any rate, a systematic empirical comparison of Bayesianism and explanationism is, to the best of our knowledge, not yet to be found in the literature. With the following three experiments, we aim to begin filling this lacuna.

2. Experiment 1

Experiment 1 was originally conducted with an eye toward comparing different probabilistic measures of explanatory goodness (the results of this comparison were presented in Schupbach, 2011). It was only later realized that the data gathered in the experiment might also shed light on questions Q1–Q3. For the purposes of comparing measures of explanation, subjective probability judgments (credences) as well as objective probabilities were used to calculate values of explanatory goodness according to the various measures, which were then compared with participants' judgments of explanatory goodness. Here, our goal is different. Instead of aiming to determine which measure of explanation best predicts actual judgments of explanatory goodness, we are interested in the role (if any) that such judgments play in updating credences. To answer this question, we reanalyzed the data from the earlier experiment. We begin by summarizing the experiment.

2.1. Participants

Twenty-six students from the University of Pittsburgh were individually interviewed as part of the study. In return for their participation, they received \$10 each. The mean age of the participants was 20 years ($SD = 2$). Twelve of the participants were females.

2.2. Materials and procedure

Experiment 1's materials and method were based closely upon those used by Phillips and Edwards (1966)—and more recently by Tentori, Crupi, Bonini, and Osherson (2007). Two opaque urns were used in the interview, each containing 40 balls, but one urn ("urn A") containing 30 black balls and 10 white ones, and the other urn ("urn B") containing 15 black balls and 25 white ones. At the beginning of the interview, each participant was shown both urns and informed of their respective contents. The participant was also given a visual representation of these contents, which he or she was allowed to consult at any point during the interview.

The decision of which urn to use throughout the remainder of the interview was next decided via an actual flip of a fair coin. Each participant saw that the coin flip determined the choice of urn; however, whether the chosen urn was A or B was concealed from the participant. Ten balls were then randomly drawn from the chosen urn without replacement. Throughout the interview, the drawn balls were lined up in front of the participant, in the order in which they had been drawn.

After each ball was drawn, the participant was asked first to judge the explanatory goodness, in light of the evidence so far, of the hypothesis that urn A had been selected (H_A), and then to do the same for the hypothesis that urn B had been selected (H_B). The participant had to answer these questions by making a mark on a continuous scale with range $[-1, 1]$. The scale included five interpreted points labeled at equal distances, the leftmost label reading that the hypothesis at issue was an extremely poor explanation of the evidence so far, the rightmost 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.

After these questions had been answered, the participant was asked how likely it was in his or her judgment that urn A had been selected, given the color of the ball or balls that so far had been drawn. The same question was asked concerning urn B, but here the participant was notified that the answer to this question and