

Explaining norms and norms explained

David Danks^a and Frederick Eberhardt^b

^a*Department of Philosophy, Carnegie Mellon University and Institute for Human & Machine Cognition Pittsburgh, PA 15213;* ^b*Institute of Cognitive and Brain Sciences, University of California, Berkeley, Berkeley, CA 94720 and Department of Philosophy, Washington University in St. Louis, St. Louis, MO 63130.*

ddanks@cmu.edu

<http://www.hss.cmu.edu/philosophy/faculty-danks.php>

fde@berkeley.edu

<http://www.phil.cmu.edu/~fde/>

Abstract: Oaksford & Chater (O&C) aim to provide teleological explanations of behavior by giving an appropriate normative standard: Bayesian inference. We argue that there is no uncontroversial independent justification for the normativity of Bayesian inference, and that O&C fail to satisfy a necessary condition for teleological explanations: demonstration that the normative prescription played a causal role in the behavior's existence.

We understand Oaksford & Chater (henceforth, O&C) (Oaksford & Chater 2007) as providing the following argument for Bayesian models as teleological explanations in psychology:

1. *Normative Justification*: The normativity of Bayesian inference can be established independently of empirical observations. (p. 31 of their book)
2. *Normative Force*: Appropriate normative principles lead to behavior conforming to those principles. (p. 33)
3. *Default Assumption*: People are pretty good at achieving their goals (i.e., they are “everyday rational”). (p. 19) Therefore, people's behavior is a guide to the appropriate normative principles. (p. 30)
4. *Empirical Results*: People's behavior conforms to the prescriptions of Bayesian models. (chaps. 6 & 7)

Conclusion: Therefore, Bayesian models explain *why* that behavior occurs. (Ch. 8)

The first three premises are *formal* claims that are instances of necessary parts of proper teleological explanations. Without independent justification (Premise 1), there is no normative power behind Bayesian inference. Without influence on behavior (Premise 2), normative principles provide no explanation. And if the normative prescription cannot be reasonably interpreted in terms of desirable human goals (Premise 3), then the explanation is implausible.

We suggest, however, that neither Premise 1 nor Premise 2 is supported by O&C's arguments.

Premise 1 is usually grounded in two standard arguments. Dynamic Dutch book arguments aim to establish Bayesian inference as the unique belief change procedure that avoids irrational, between-time inconsistencies, understood as acceptance of bets (potential, not actual) over time that result in guaranteed loss (Teller 1973). Standard forms of these arguments, however, also imply absurd prescriptions, including: reasoners should not protect themselves

against predictable irrationalities (Maher 1992); reasoners should not retrospect on their past beliefs (Levi 1988); and reasoners should never change their conditional commitments (Levi 1988; 2002). If these arguments are weakened to avoid these unwanted implications, then Bayesian inference is only one of infinitely many ways to avoid dynamic Dutch book; it has no particularly special status.

The second standard argument is long-run convergence: Roughly, any nonzero degree of belief in the truth will converge to 1 (using Bayesian inference) as one collects more evidence (Savage 1972), and no other reliable method always converges to the truth faster (Schulte 1999). However, the convergence arguments have the unreasonable requirement that the Bayesian reasoner be logically omniscient. There are also problems that can be solved by a naïve falsificationist, but which the Bayesian can solve only if she can “compute” uncomputable functions (Juhl 1993; Osherson et al. 1988). Long-run convergence thus cannot provide a conceptual justification for the normativity of Bayesian inference.

There might be other arguments for Premise 1, but we have not seen them, nor do O&C provide them. We agree with O&C that the normative principles underlying rationality may vary between situations. But independent justification of normativity must be provided for *whatever* principles are appropriate, or else one cannot have an adequate teleological explanation. The dependence of the normative principle on the situation is inversely correlated with the explanatory power provided by the normative principle. A situation-independent justification of Premise 1 (or some similar premise) is necessary for teleological explanations.

Our second worry is that, even if Premise 1 holds, O&C require only an empirical match between behavior and normative prescription. Their argument thus neglects the requirement for a teleological explanation that the normative principle must have played a causal role (ontogenetic,

phylogenetic, or both) in the behavior's existence or persistence. "Origin stories" are required for teleological explanation, but are never provided by O&C. Behavior B could be optimal for task T even though behavior B results from considerations that are independent of task T; B's optimality might be coincidental. In this case, the claim "Behavior B because it is optimal for task T" is an incorrect explanation, even though B conforms (empirically) to the normative prescription.

O&C seem to have two lines of response. First, if there is widespread alignment between the normative prescription and people's behavior, then it is arguably improbable that the behavior is only coincidentally optimal. Consequently, research confirming correspondence in a wide range of conditions is evidence in favor of the normative principle. This response fails, however, because widespread alignment is actually to be expected for Bayesian models, given the many "free parameters" in such models: hypothesis space, prior probabilities, utilities, likelihood functions, and various plausible computational approximations. In particular, for any evidence and any behavior (represented as a probability distribution over possible choices), there exists a prior probability distribution such that the predicted posterior distribution after inference on the evidence matches the observed behavior.¹ That is, for (almost) any psychological experiment, we know a priori that a Bayesian model will be capable of fitting the observed behavior, whatever it might be. Repeatedly developing *sui generis* Bayesian models for each task does not compensate for a lack of "origin stories," even if the models successfully predict.

O&C's second line of response is more promising: successful model predictions for experiments that vary features of one particular task make it plausible that the normative

¹ There are other technical assumptions (e.g., nonextremal likelihoods), but they are irrelevant to our core point.

constraints played a causal role in shaping the observed behavior. We support this line of response to the lack of “origin stories,” but are doubtful about how much support the evidence provides for the *normativity* of Bayesian inference. O&C here primarily cite manipulations of the base-rate (pp. 146ff, 178ff), and as they recognize, there are several competing models with similar normative appeal. Moreover, there is substantial controversy about which model provides the best fit of the data. There is a potential response, but it is currently only potential.

O&C admirably search for teleological explanations of human behavior. We are in complete agreement with them that such explanations are desirable, but we believe that their enterprise requires a stronger foundation. They have neither properly established Bayesian inference as a normative principle, nor shown – directly or indirectly – that the optimality of Bayesian inference (if it *is* optimal in some respect!) is actually a cause of people approximating those prescriptions. They offer only an existence proof – “Behavior *could* be caused by a *potentially* normative principle” – and such accounts are too weak to provide teleological explanations.

Acknowledgment:

The second author is funded by a fellowship from the James S. McDonnell Foundation.

References

- Juhl, C. F. (1993) Bayesianism and reliable scientific inquiry. *Philosophy of Science* 60:302–19.
- Levi, I. (1988) The demons of decision. *Monist* 70:193–211.
- Levi, I. (2002) Money pumps and diachronic books. *Philosophy of Science* 69:S235–47. Maher, P. (1992) Diachronic rationality. *Philosophy of Science* 59:120–41.
- Oaksford, M. & Chater, N. (2007) *Bayesian rationality: The probabilistic approach to human reasoning*. Oxford University Press.
- Osherson, D. N., Stob, M. & Weinstein, S. (1988) Mechanical learners pay a price for Bayesianism. *Journal of Symbolic Logic* 53:1245–51.
- Savage, L. J. (1972) *Foundations of statistics*. Dover.
- Schulte, O. (1999) The logic of reliable and efficient inquiry. *Journal of Philosophical Logic* 28:399–438.
- Teller, P. (1973) Conditionalization and observation. *Synthese* 26:218–38.