Post-Hoc Rationalism in Science

Eric Luis Uhlmann

HEC Paris

(Commentary on Jones and Love, "Bayesian Fundamentalism or Enlightenment? On the Explanatory Status and Theoretical Contributions of Bayesian Models of Cognition")

ABSTRACT: 60

MAIN TEXT: 983

REFERENCES: 670

ENTIRE TEXT: 1,799

CONTACT:

Eric Luis Uhlmann
HEC Paris - School of Management
Management and Human Resources Department
1, Rue de la Libération
78351 Jouy-en-Josas
France
Tel: 33 (0)1 39 67 97 44
E-mail: eric.luis.uhlmann@gmail.com

Abstract

In advocating Bayesian Enlightenment as a solution to Bayesian Fundamentalism, Jones and Love rule out a broader critique of rationalist approaches to cognition. However, Bayesian Fundamentalism is merely one example of the more general phenomenon of Rationalist Fundamentalism: the tendency to characterize human judgments as rational and optimal in a post hoc manner, after the empirical data are already known.

Post-Hoc Rationalism in Science

Jones and Love are right to criticize what they term Bayesian Fundamentalism as not empirically grounded, uninformed by psychological data, open to multiple rational accounts of a task or decision, and conducive to post-hoc explanations. However in advocating Bayesian Enlightenment as a solution, they appear to rule out a broader critique of rationalist approaches to human cognition. Specifically, Bayesian Fundamentalism is one example of the more general phenomenon of Rationalist Fundamentalism: the tendency to characterize a given judgment as rational and optimal in a post hoc manner, after the empirical data are already known. Few researchers would argue human behavior is perfectly optimal and rational. However, a desire to see the human mind as operating rationally, and the use of post-hoc justifications to reach this conclusion, suggests we should be skeptical of after-the-fact "rational" explanations.

Decades of empirical studies show people are strongly motivated to see themselves as rational and objective (for reviews, see Armor, 1999; Pronin, Gilovich, & Ross, 2004; Pyszczynski & Greenberg, 1987; Ross & Ward, 1996). Decision makers engage in motivated reasoning and psychological rationalizations designed to preserve this "illusion of objectivity" (Armor, 1999; Pronin, Lin, & Ross, 2002), for instance changing their definition of what an optimal judgment is after the fact (Dunning & Cohen, 1992; Epstein, Lipson, Holstein, & Huh, 1992; Kunda, 1987; Norton, Vandello, & Darley, 2004; Uhlmann & Cohen, 2005). Evidence that general psychological processes are not rational or optimal represents a threat to this cherished illusion. Fundamentalist resistance to evidence of human irrationality further stems from economics and

related disciplines, in which optimality and the maximization of utility are widely perceived as necessary assumptions about human behavior.

A rationalist defense can involve constructing a post-hoc Bayesian account of an empirical finding predicted a priori from theories grounded in psychological limitations and motives. Consider the phenomenon of biased assimilation, in which participants rate a scientific study that supports their political beliefs (e.g., about the deterrent effects of capital punishment) as methodologically superior to a study that refutes their beliefs (Lord, Ross, & Lepper, 1979). The cognitive-rationalist interpretation is that decision makers are simply making Bayesian inferences, taking into account subjective probabilities (e.g., their prior political beliefs) when evaluating new evidence. However, further findings contradict the claim biased assimilation is merely the product of Bayesian inferences. For instance, individuals whose positive self-image is affirmed are less likely to exhibit biased assimilation (Cohen, Aronson, & Steele, 2000; see also Dunning, Leuenberger, & Sherman, 1995; Sherman & Cohen, 2002). This is consistent with the idea that biased information processing stems from a motivated desire to dismiss evidence that threatens valued beliefs, and by extension the self (Sherman & Cohen, 2006; Steele, 1988). When a decision maker is feeling good about herself there is less need to be biased. In addition, would-be parents who believe day care is bad for children, but plan to use day care themselves (and therefore desire to conclude that day care is just as good as home care) show biased assimilation in favor of day care (Bastardi, Uhlmann, & Ross, in press). What decision makers *desire* to be true seems to trump what they *believe* to be factually true— the ostensive basis for any Bayesian inferences.

As Jones and Love point out, one of the most problematic aspects of rational models is how little attention can be paid to whether the assumptions of the statistical model correspond to what actually is actually going on in people's heads as they engage in a task or make a decision. I once debated an economist who argued micro-level psychological data on what goals people pursue in the dictator game are irrelevant: the material self interest account *must* be true if people's offers correspond to the predictions of the statistical model. However it is dangerous to assume that because a rational statistical model can mimic or reproduce a pattern of data, the underlying psychological process is a rational one. That a computer can mimic some of the outputs of human thought does not necessarily mean the mind functions in the same way as a computer.

The last defense of post-hoc rationalism is to swap normative models of rationality entirely. In other words, researchers can speculate post-hoc as to what alternative goals decision-makers may have been pursuing in order to preserve the view participants were acting rationally. Never mind the goals to optimize material outcomes or achieve accuracy: judgmental biases can be defined as "rational" because they preserve the decision maker's personal self-image, psychological well-adjustment, public reputation, cherished religious beliefs, desire to punish norm violators, existential goals, likelihood of survival in ancestral environments, or even the happiness of their marriage (Cosmides & Tooby, 1994; Hamilton, 1980; Krueger & Funder, 2004; Lerner & Tetlock, 1999; Tetlock, 2002; Tetlock, Kristel, Elson, Green, & Lerner, 2000; Tetlock, Visser, Singh, Polifroni, Elson, Mazzocco, & Rescober, 2007).

It has been argued that the heuristics-and-biases approach to cognition is itself biased, in the direction of attributions to irrationality (Krueger & Funder, 2004). However despite its

shortcomings the heuristics-and-biases research program is at least based on a priori theoretical hypotheses. There are few cases of "post-hoc irrationalism" in which robust empirical effects predicted a priori by Bayesian or otherwise rationalist models are redefined post-hoc as due to motives such as the need for self esteem or control.

Although Bayesian Enlightenment, as advocated by Jones and Love, is a major improvement on Bayesian Fundamentalism, it is still subject to post-hoc rationalism. An interface between Bayesian or otherwise rationalist models and data on psychological processes leaves plenty of room for the former to distort interpretations of the latter. A wealth of evidence indicates human beings are subject to a powerful illusion of rationality and objectivity they are strongly motivated to maintain and which influences their perceptions of scientific data. Researchers are also human beings. It would be remarkable indeed if scientists were immune to the empirical phenomena we study.

References

Armor, D.A. (1999). *The illusion of objectivity: Bias in the belief in freedom from bias.* Doctoral dissertation, University of California Los Angeles.

Bastardi, A., Uhlmann, E.L., & Ross, L. (in press). Wishful thinking: Belief, desire, and the motivated evaluation of scientific evidence. *Psychological Science.*

Cohen, G. L., Aronson, J., & Steele, C. M. (2000). When beliefs yield to evidence: Reducing biased evaluation by affirming the self. *Personality and Social Psychology Bulletin, 26,* 1151–1164.

Cosmides, L., & Tooby, J. (1994). Origins of domain specificity: The evolution of functional organization. In L. A. Hirschfeld & S. Gelman (Eds.), *Mapping the mind*: *Domain specificity in cognition and culture* (pp. 85–116). New York: Cambridge University Press.

Dunning, D., & Cohen, G.L. (1992). Egocentric definitions of traits and abilities in social judgment. *Journal of Personality and Social Psychology, 63,* 341-355.

Dunning, D., Leuenberger, A., & Sherman, D.A. (1995). A new look at motivated inference: Are self-serving theories of success a product of motivational forces? *Journal of Personality and Social Psychology, 69,* 58-68.

Epstein, S., Lipson, A., Holstein, C., & Huh, E. (1992). Irrational reactions to negative outcomes: Evidence for two conceptual systems. *Journal of Personality and Social Psychology, 62,* 328–339.

Hamilton, V. L. (1980). Intuitive psychologist or intuitive lawyer: Alternative models of the attribution process. *Journal of Personality and Social Psychology, 39,* 767–773.

Krueger, J. I., & Funder, D. C. (2004). Towards a balanced social psychology: Causes, consequences and cures for the problem-seeking approach to social behavior and cognition. *Behavioral and Brain Sciences, 27*, 313-327.

Kunda, Z. (1987). Motivated inference: Self-serving generation and evaluation of causal theories. *Journal of Personality and Social Psychology, 53*, 37-54.

Lerner, J. S., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin, 125*, 255–275.

Lord., C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology, 37*, 2098-2109.

Norton, M.I., Vandello, J.A., & Darley, J.M. (2004). Casuistry and social category bias. *Journal of Personality and Social Psychology, 87*, 817-831.

Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: Perceptions of bias in self versus others. *Psychological Review, 111*, 781-799.

Pronin, E., Lin, D.Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. *Personality and Social Psychology Bulletin, 28*, 369-381.

Pyszczynski, T., & Greenberg, J. (1987). Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model. In L. Berkowitz (Ed.), *Advances in experimental social psychology*, Vol. 20. (pp. 297-340). San Diego, CA: Academic Press.

Ross, L., & Ward, A. (1996). Naive realism in everyday life: Implications for social conflict and misunderstanding. In E.S. Reed & E. Turiel (Eds.), *Values and knowledge* (pp. 103-135). Hillsdale, NJ: Lawrence Erlbaum.

Sherman, D. K., & Cohen, G. L. (2002). Accepting threatening information: Self-affirmation and the reduction of defensive biases. *Current Directions in Psychological Science, 11*, 119-

123.

Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defense: Self-affirmation theory. In M. P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. 38, pp. 183-242). San Diego, CA: Academic Press.

Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 21, pp. 261-302). New York: Academic Press.

Tetlock, P.E. (2002). Social functionalist frameworks for judgment and choice: Intuitive politicians, theologians, and prosecutors. *Psychological Review, 109,* 451-471.

Tetlock, P.E., Kristel, O., Elson, B., Green, M., & Lerner, J (2000). The psychology of the unthinkable: Taboo trade-offs, forbidden base rates, and heretical counterfactuals. *Journal of Personality and Social Psychology, 78*, 853-870.

Tetlock, P. E., Visser, P., Singh, R., Polifroni, M., Elson, B., Mazzocco, P., & Rescober, P. (2007). People as intuitive prosecutors: The impact of social control motives on attributions of responsibility. *Journal of Experimental Social Psychology, 43*, 195-209.

Uhlmann, E.L., & Cohen, G.L. (2005). Constructed criteria: Redefining merit to justify discrimination. *Psychological Science, 16*, 474-480.