Can resources save rationality? “Anti-Bayesian” updating in cognition and perception

Eric Mandelbaum\textsuperscript{a}, Isabel Won\textsuperscript{c}, Steven Gross\textsuperscript{b} and Chaz Firestone\textsuperscript{c}

\textsuperscript{a}Baruch College, CUNY Graduate Center, Department of Philosophy, New York, NY10016; \textsuperscript{b}Department of Philosophy, Johns Hopkins University, Baltimore, MD 21218; and \textsuperscript{c}Department of Psychological & Brain Sciences, Johns Hopkins University, Baltimore, MD 21218
emandelbaum@gc.cuny.edu
iwon1@jhu.edu
sgross11@jhu.edu
chaz@jhu.edu
http://ericmandelbaum.com
http://perception.jhu.edu
https://sites.google.com/site/grosssteven/
doi:10.1017/S0140525X19001717, e16

Abstract

Resource rationality may explain suboptimal patterns of reasoning; but what of “anti-Bayesian” effects where the mind updates in a direction opposite the one it should? We present two phenomena – belief polarization and the size-weight illusion – that are not obviously explained by performance- or resource-based constraints, nor by the authors’ brief discussion of reference repulsion. Can resource rationality accommodate them?

Resource rationality takes seemingly irrational behaviors and reframes them as rational or optimal given other constraints on agents. For example, anchoring-and-adjustment and overestimating extreme events turn out be “rational” after all, by reflecting the rational allocation of cognitive resources. Thus, even for such classically irrational phenomena, “the resulting train of thought eventually converges to the Bayes-optimal inference” (p. 38).

In such cases, reasoners fall short of perfectly rational updating, and it is illuminating that resource- and performance-based constraints can accommodate such suboptimal reasoning. But what about cases where the mind behaves not merely suboptimally, but rather against the norms of Bayesian inference? Here, we explore cases where the mind is moved by prior knowledge in precisely the reverse direction of what a rational analysis would recommend. These cases are not merely suboptimal, but rather “anti-Bayesian,” for actively defying Bayesian norms of inference. We consider two such phenomena: belief polarization and sensory integration (Fig. 1). Can resource rationality handle them?

First, belief polarization: Receiving evidence contrary to your beliefs should soften those beliefs, even if ever-so-slightly. But, this isn’t what actually happens when the beliefs in question are central to one’s identity – in belief polarization, contrary or disconfirming evidence causes more extreme beliefs, not more moderate ones. A classic example was vividly documented by Festinger et al. (1956): Cult members who predict the world will end on some date – but who then see that date come and go with no
cataclysm – end up strengthening their beliefs in the cult’s tenets, not softening them. In other words, credible evidence against their worldview only makes them hold that worldview more strongly – directly defying Bayesian inference norms.

The same phenomenon can be found under laboratory conditions. For example, one study exposed people who believe that Jesus is the Son of God to a (fake) news article reporting that archeologists had unearthed carbon-dated letters from the New Testament authors; the letters said the Bible was fraudulent and that its authors knew Jesus was not divinely born (Batson 1975). Subjects who did not believe the article’s content left their beliefs about Jesus unchanged; but, fascinatingly, subjects who did believe the article’s content ended up strengthening their belief that Jesus was the Son of God. In other words, when one lifts them up, the beliefs against one’s beliefs can cause those beliefs to strengthen (Batson 1975), whereas Bayesian norms of inference recommend softening those beliefs. (B) In the size-weight illusion, one is shown two objects of different sizes but equal weights; when one lifts them up, the smaller one feels illusorily heavier than the larger one (Buckingham 2014; Charpentier 1891; Won et al. 2019). In other words, ambiguous sensory data about which of two objects is heavier is resolved “against” one’s prior expectations, rather than in favor of one’s priors as recommended by Bayesian norms of inference. Can resource rationality accommodate such paradigmatically “irrational” phenomena?

predictable consequence of dissonance theory – “the psychological immune system” (Gilbert et al. 1998) – applied to one’s values. What is mysterious is why this should occur in a Bayesian mind – even one constrained by “resources.” Belief polarization is irrational not because people are insufficiently moved by evidence, but rather because people are moved in the direction opposite the one they should be. And, importantly, these patterns cannot be explained by biased attitudes toward the evidence’s source. For example, Bayesian models of milder forms of belief polarization (e.g., Jern et al. 2014) suggest that subjects infer that contrary evidence must have come from unreliable sources (e.g., biased testimony); but this seems inapplicable to the above cases, where the sources are either nature itself (e.g., the world failing to end), or evidence the subject has actively accepted (e.g., news articles they endorsed).

Indeed, “anti-Bayesian” updating is widespread, occurring even in basic perceptual processes. When we have prior expectations about new and uncertain sensory data, rational norms of inference say we should interpret such data with respect for those priors; “people should leverage their prior knowledge about the statistics of the world to resolve perceptual uncertainty” (p. 40). But, sensory integration frequently occurs the opposite way. Consider the size-weight illusion, wherein subjects see two equally weighted objects – one large and one small – and then lift them both to feel their weight. Which feels heavier? We “should” resolve the ambiguous haptic evidence about which object is heavier in favor of our priors; but instead, the classic and much-replicated finding is that we experience the smaller object as heavier than the equally-weighted larger object (Buckingham 2014; Charpentier 1891). This too is “irrational” – not for falling short of Bayesian norms of inference, but for proceeding opposite to them, because we resolve the ambiguous sensory evidence – two equally weighted objects – against the larger-is-heavier prior, not in favor of it (Brayanov & Smith 2010; Buckingham & Goodale 2013). Indeed, this backwards pattern of updating is so strong that it can produce outcomes that are not merely odd or improbable, but even “impossible” (Won et al. 2019): If subjects are shown three boxes in a stack – Boxes A, B, and C – such that Box A is heavy (250 g) but Boxes B and C are light (30 g), then subjects who lift Box A alone and then Boxes A +B+C together report that Box A feels heavier than Boxes A+B+C – an “impossible” experience of weight (because a group could never weigh less than a member of that group).

How can a “rational” account – even a resource-rational one – explain this? Lieder and Griffiths accommodate other sensory “repulsion” effects (Wei & Stocker 2015; 2017), but that modeling work seems inapplicable to the size-weight illusion. And whereas the original size-weight illusion could perhaps have a tortuous Bayesian explanation (Peters et al. 2016), Won et al.’s modification seemingly cannot: First, it’s unclear if previous models of simultaneous lifting apply to Won et al.’s temporally-extended case; but second, there is just no logical chain of reasoning that should end with A alone being heavier than A+B+C together.

More generally: What are the principles that lead to perverse “anti-Bayesian” updating? Perhaps resource rationality wasn’t intended to cover all cases (in which case it is not an “Imperial Bayesian” theory; Mandelbaum 2019). But, the problem isn’t merely that there are counterexamples to resource rationality, but rather that these are predictable, law-like counterexamples that do not reflect performance constraints between interacting mental processes. Indeed, when it comes to these more entrenched patterns, even “resources” may not save rationality.

Commentary/Lieder and Griffiths: Resource-rational analysis

---

![Figure 1](https://www.cambridge.org/core/core/jhoh.html)
References

[The letters “a” and “r” before author’s initials stand for target article and response references, respectively]


References/Lieder and Griffiths: Resource-rational analysis


Gigerenzer, G., Todd, P. M. & ABC Research Group, pp. 80–110. Oxford University Press. [aFL]


References/Lieder and Griffiths: Resource-rational analysis


Svenson, S. (2013) Irrationality: The enemy within. *Pinter & Martin Ltd*. [aFL]


Downloaded from https://www.cambridge.org/core. Johns Hopkins University, on 11 Mar 2022 at 15:19:20, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/S0002828851000641X