Philosophy of Science (2025), **92**, 1215–1224
doi:10.1017/psa.2025.10119

PSA PHILOSOPHY OF SCIENCE ASSOCIATION
PHILOSOPHY OF SCIENCE

CONTRIBUTED PAPER

A New Foundation for Belief Updating?

Ibrahim Haydar

Department of History and Philosophy of Science, University of Pittsburgh, Pittsburgh, PA, USA
Email: ibh8@pitt.edu

(Received 22 January 2025; revised 28 May 2025; accepted 24 July 2025; first published online 10 September 2025)

Abstract

Eric Mandelbaum has reported some troubles with Bayesianism in cognitive science. He has brought some behavioral data to show that belief updating in humans is fundamentally Bayesian perverse. I argue that the behaviors that he seeks to explain do not undermine Bayesian accounts of belief updating and can instead be explained as idiosyncratic consequences of an appropriately bounded implementation of a Bayesian-normative belief-updating system.

1. Introduction

Bayesian inference is invoked ubiquitously in contemporary cognitive science, but is its ubiquitous invocation grounded in ubiquitous instantiation? Behavioral data in such disparate domains as sensorimotor learning (Körding and Wolpert 2004), perceptual tasks (Kersten et al. 2004), and conceptual acquisition (Goodman, Tenenbaum, and Gerstenberg 2015) are frequently found to be well described under Bayesian models. Must all cognitive capacities be Bayesian normative?

Eric Mandelbaum (2019) says no, denying that belief updating is Bayesian normative by claiming that people systematically exhibit two kinds of Bayesian-perverse belief-updating behaviors. The first behavior is what I call *middling maintenance*, whereby, despite evidence, people resist adjusting beliefs of low conviction (beliefs of a middling credal value). The second behavior is *polarized persistence*, whereby, despite evidence, beliefs of strong conviction become only more strongly held. Mandelbaum claims that no Bayesian account of belief updating can account for these behaviors and argues that belief updating must instead be governed by a "psychological immune system" (PsIS)—a system that functions to maintain one's self-image as "good, smart, and competent" (154). Mandelbaum's PsIS picture is one in which belief updating in humans is not guided toward truth seeking but is rather guided against cognitive dissonance and threats to one's self-image. Mandelbaum explains the two kinds of behaviors using the PsIS picture as follows: (1) Correcting one's weakly held beliefs is not worth the cost of learning that one was

© The Author(s), 2025. Published by Cambridge University Press on behalf of Philosophy of Science Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/psa.2025.10119 Published online by Cambridge University Press

wrong and (2) one's strongly held beliefs must not be corrected but instead reinforced in virtue of their proximity to one's sense of self.

This article is a twofold critique of the PsIS picture. I first argue that the picture clashes with most belief-updating behavior, which seems both Bayesian normative and in tension with PsIS. So, while the PsIS picture explains the Mandelbaumian pair of behaviors, this explanation comes at the explanatory expense of countless other belief-updating behaviors. I further argue that such an extravagant mechanism is not even needed to account for the Mandelbaumian pair, which I explain by appealing only to architectural features of the brain's evidence-gathering processes—processes that enjoy firm support in cognitive science.

The remainder of the article is structured as follows. Section 2 goes through Mandelbaum's (2019) arguments, data, and explanation. Section 3 raises global objections to PsIS. Section 4 sets up the alternative picture by exploring bounded implementations of Bayesianism. Section 5 uses the architecture to explain Mandelbaum's data.

2. Mandelbaum's data and explanation

To make the case that belief updating is fundamentally Bayesian perverse, Mandelbaum (2019) presents behavioral data of two kinds. The first kind is middling maintenance: instances in which people do not update beliefs of low conviction. The second kind is polarized persistence, whereby people persistently increase their conviction in a strongly held belief despite what the evidence says.

For middling maintenance, Mandelbaum (2019) presents a series of studies (Anderson 1983) involving participants forming a belief about firefighters and risk aversion (either by considering made-up evidence or simply by picking a position and performing ad hoc justification), then being shown (fictitious) evidence contrary to whatever opinion they had formed. The studies found "stubborn adherence" to the beliefs initially formed, even when participants were confronted with a substantial amount of counterevidence. Mandelbaum also brings forth cases of overlooking evidence when forming a belief, such as rats overlooking audiovisual stimuli when forming beliefs about what is causing their nausea, despite being able to easily form associations between nausea and gustatory stimuli. These two pieces of data are supposed to demonstrate a lack of learning from evidence where a Bayesian-normative belief updating requires it. The first demonstrates an instance of clinging to a belief that's been formed, and the second demonstrates an instance of overlooking some evidence entirely.

For polarized persistence, Mandelbaum (2019) cites an influential study (Lord, Ross, and Lepper 1979) demonstrating that, when shown equivocal data on the death penalty, subjects with strong beliefs on the death penalty report strengthened belief beyond their initial stance, rather than a tempered belief, as would be expected by any Bayesian-governed belief updating.[1] Mandelbaum also provides a study

[1] The death penalty study faces some concerning features in attempts at replication. Studies (Miller et al. 1993; Monro and Ditto 1997) replicating the effect for the death penalty, affirmative action, and gay marriage have found that the results replicate only under the original experimental design—when subjects report, after the fact, how their attitudes regarding the issue in question have changed in light of the equivocal evidence. However, in trials with pre- and postexposure measurement of subject beliefs,

https://doi.org/10.1017/psa.2025.10119 Published online by Cambridge University Press

(Batson 1975) in which a subset of people who reported strong belief that Jesus is the son of God went on to accept a piece of evidence—an article that purports to show that the New Testament is fabricated—but reported an increase in their belief in the proposition it undermines, namely, that Jesus is the son of God.[2] These two pieces of data demonstrate a strict increase in the conviction of polarized beliefs, despite the nature of the evidence. This may happen through biased uptake of pro-attitudinal evidence or biased scrutiny of counterevidence, as in the first case, or even somehow in taking up contrary evidence, as in the second case, which seems irreconcilable with any picture of Bayesian updating. Thus Mandelbaum takes them to indicate a need for a new foundation for belief updating altogether.

To explain the preceding behaviors, Mandelbaum (2019) offers a picture where a PsIS is fundamental to belief updating in humans. The PsIS involves "constantly adjusting beliefs to ward off serious threats to one's sense of self" (152). The PsIS picture suggests that in belief updating, humans act to maintain their self-image as competent belief formers. Receiving counterattitudinal evidence hurts because learning that we are wrong threatens our idea that we form good beliefs. Under this picture, Bayesian perversity stems not from "mere errors in a system's processing but rather stem[s] from a system that is properly functioning" (151). And so, under this view, "belief updating itself is, at its core, deeply nonoptimal in a way that contravenes Bayes' rule" (144).

The PsIS explains the aforementioned behaviors as follows: Middling maintenance happens when beliefs enjoy stability in the face of counterevidence because they are not worth updating. They do not matter much, and because learning we were wrong hurts, our belief-updating system simply retains the old belief. Polarized persistence happens when strongly held beliefs persist with pro-attitudinal gains because, as beliefs that the holder identifies with, they enjoy the most protection by the PsIS.

3. Global argument against PsIS

I claim that PsIS simply does not account for most data on belief updating, even outright contradicting much of the evidence. A big problem with PsIS is a massive underspecification of when it kicks in. It seems that for each example Mandelbaum (2019) gives of middling beliefs not being updated for the sake of the marginal comfort of retaining beliefs, we can recall several instances of adjusting our beliefs about utterly trivial facts. Moreover, many of us have experienced tempering a closely held belief, such as a political or religious conviction. This is something that a naive version of PsIS predicts should never happen. Let me examine some studies in greater detail to further support these points.

polarization is not exhibited. This makes it more reasonable to consider the reporting of the subjects as some sort of group-support signal, rather than being indicative of their actual beliefs.

[2] It is worth noting that this study is somewhat dubious. It has a small sample size of people voicing acceptance of the article (twelve Christian believers and eight nonbelievers from one youth program for female high school students in New Jersey) and is nowhere meaningfully replicated. A series of experiments across more than 10,100 participants testing fifty-two potential similar cases found no instances of the behavior occurring (Wood and Porter 2019).

https://doi.org/10.1017/psa.2025.10119 Published online by Cambridge University Press

A recent study (Liu and Xu 2021) found that children can be made to recant several of their most core beliefs, including principles of intuitive physics learned in infancy, such as the principle that objects move continuously through space and time. Under the PsIS picture, there seems to be no reason to revise such a long-held belief, especially a belief that permeates daily experience as strongly as continuity. Moreover, PsIS predicts that such beliefs demand not only stubborn adherence but strengthened conviction to assuage any dissonance. Being wrong about something so commonplace is sure to generate discomfort for one's sense of self. Though, perhaps the PsIS picture is not meant to extend to children, and we can append Mandelbaum's (2019) picture with the idea that childhood learning is about developing a sense of self and a sense of assimilation with some community. Then, children will be more receptive to whatever is being taught to them so that they can fit in, and then once they cement into their adult senses of self, their PsISs will kick in and govern their belief updating.

It is hard, however, to see how such a tack-on would fare with other empirical results. For example, a study (Beam, Hutchens, and Hmielowski 2020) demonstrates that exposure to counterattitudinal information on Facebook leads to depolarization of political beliefs. This is, again, the opposite of what PsIS predicts. Political affiliation nowadays enjoys a status as the archetypal example of polarization on the basis of identity. However, the finding, conducted in the heat of the 2016 election, demonstrates precisely what an epistemically guided belief-updating system would predict, precisely the opposite of the PsIS prediction that polarized beliefs persist. Similar results abound in the political polarization literature. A study (Fishkin and Luskin 2005) found that deliberative processes in democracy often led to opinion change. Moreover, opinion changes tracked gains in information straightforwardly and not in the convoluted polarization-persistent way predicted by PsIS.

Another study (Bullock et al. 2013) found that when given assessments where truth is incentivized with small dollar amounts, participants who voice strong, polarized beliefs on the basis of political identity provide comparatively temperate answers. This study was replicated repeatedly to the same effect (Prior, Sood, and Khanna 2015). This gap between voiced beliefs and actual ones produced on moderately incentivized tests elicits a concern for the experimental design of the studies Mandelbaum (2019) cites for cases of polarized persistence. Less than calling into question the nature of belief updating, perhaps such studies merely demonstrate that people sometimes shout rallying cries to their groups.

This first prong against PsIS thus consists in a rejection of the claim that Mandelbaum's (2019) data indicate systemic Bayesian perversity in belief updating. It seems that the behaviors Mandelbaum presents, even taken at face value, are idiosyncratic, rather than inherent to belief updating. Given how Mandelbaum's data stack against other, more familiar data concerning belief updating makes it far more likely the case that the data are better explained by something other than belief updating, such as something like group signaling.

In sum, to claim that PsIS is fundamental to belief updating goes against the vast majority of data and experience corresponding to belief updating. A theory so strongly in discord with intuition ought to earn its hard-swallowed status through its explanatory power and scientific merit, but PsIS does not.

https://doi.org/10.1017/psa.2025.10119 Published online by Cambridge University Press

4. Boundedly rational Bayesian sampling

Mandelbaum's (2019) argument is impactful only if it is an argument against an appropriately bounded Bayesian belief-updating process. Bayesian belief updating simpliciter is simply not on the table. For one, it is computationally intractable. Belief updating that is literally Bayesian would require computing the probabilities of countless hypotheses given the evidence, a computational task that outstrips any real capacities. Even if we restrict ourselves to a finite hypothesis space, the number of parameters that need to be accounted for in representing joint probability distributions grows exponentially with hypotheses. Therefore, if Mandelbaum is arguing that belief updating is not literally Bayesian, then he is trivially correct, and his argument is far less interesting. So let us consider Mandelbaum's claims against bounded versions of Bayesianism. I will show that pre-Bayesian considerations, required by any reasonably bounded implementation of Bayesianism, determine which evidence is admitted in belief updating—and such considerations explain the behaviors that Mandelbaum presents.

This consideration follows a tradition in the rationality literature known as *bounded* or *resource-relative rationality* (Griffiths, Lieder, and Goodman 2015; Lieder and Griffiths 2020). Bounded rationality takes the stance that an agent cannot be adequately evaluated as rational without considering the constraints (time, space) under which the agent operates. This means that though some people's decisions may appear irrational by the abstract dictates of logic and probability theory, they are often rational within the constraints of their limited information and cognitive abilities. In other words, bounded rationality suggests that people's rationality is limited or "bounded" by their ability to process information and make decisions. Therefore, in considering whether Mandelbaum's (2019) behaviors illustrate whether belief updating is Bayesian perverse, we must consider whether they fundamentally contravene any bounded form of Bayesianism.

A prominent form of bounded Bayesianism in cognitive science and in statistical computing is approximating a Bayesian posterior distribution with Monte Carlo methods, which involve computing using samples from a distribution as opposed to using the distribution itself. One important class of Monte Carlo methods for Bayesian approximation are Markov chain Monte Carlo (MCMC) sampling procedures. MCMC is the exploration of a sample space by drawing random samples and using them to construct a Markov chain, where each sample is selected according to a conditional transition probability that depends on the previous sample. After repeating this process many times, the Markov chain approximates the distribution underlying the sample space without directly solving the equations that define it. One prominent form of MCMC can be used to approximate a posterior Bayesian distribution by selecting transition probabilities according to the so-called Metropolis–Hastings decision procedure.

Sampling models provide the best psychological fleshing out of Bayesianism (Vul et al. 2014; Sanborn, Griffiths, and Navarro 2006). Monte Carlo methods have been shown to efficiently approximate statistical inferences (Sanborn, Griffiths, and Navarro 2010). They have been argued to provide an optimal balance between time and accuracy (Lieder, Griffiths, and Goodman 2012). Moreover, Monte Carlo methods map naturally onto the brain's stochastic neural dynamics (Buesing et al. 2011).

https://doi.org/10.1017/psa.2025.10119 Published online by Cambridge University Press

The Markov property additionally fits nicely with well-understood features of memory retrieval and mental simulation—when retrieving or simulating an item, the mind is primed to "jump" to the next one (Chater et al. 2020). And the broader family of sampling procedures includes our best accounts of psychological evidence aggregation, with unparalleled success on the description of how humans make basic binary judgments (Ratcliff et al. 2016). They are used widely across cognitive neuroscience and experimental psychology, and they thus constitute a unique theoretical object in the study of the mind—indeed, most other models in the field remain constrained to the description of a particular phenomenon in question.

Bayesian sampling is immediately equipped to address a host of traditional challenges to Bayesian rationality (Chater et al. 2020; Sanborn and Chater 2016). For example, the famous availability heuristic from Kahneman and Tversky's heuristics and biases research program can be straightforwardly explained using features of sampling models. The availability heuristic describes the tendency of people to overestimate the probabilities of events or items that are more readily available in memory and simulation—for example, that people judge words of the form *-ing* as more common than words ending with *–n-*, even if words of the first category are a proper subset of words of the second. Bayesian sampling explains the availability heuristic as follows: The easier retrieval cue of the first prompt leads to easier generation of samples in memory and simulation. Moreover, these samples prime the subject to draw nearby samples. Thus the subject experiences words of the form *-ing* as overrepresented in samples, especially when compared to items with harder retrieval cues, such as words of the form *–n-*.

Another heuristic that is similarly addressed by considering bounded Bayesianism is the anchoring heuristic, where, when people are given a fictitious "anchor," or initial guess, for some quantity they are asked to estimate, they tend to choose a number closer to that anchor than they otherwise would, and this effect holds for quite ridiculous anchors as well. But a component of MCMC is that the constructed Markov chain depends intimately on the starting point, so a biased starting point reflects in the samples (Chater et al. 2020).

Thus structural features of bounded Bayesianism are able to address a host of challenges traditionally raised against literal, computationally intractable Bayesianism. Similarly, while the Mandelbaumian behaviors contravene any literal Bayesian belief-updating system, I will show that they can be addressed by structural features of boundedly Bayesian belief updating. Recall the concern raised at the end of the last section: that for Mandelbaum's (2019) data to bear on the Bayesian normativity of belief updating, we require reason to believe that failure occurred to update a belief according to evidence that is accepted. So let us discuss how evidence gathering might be implemented in a realistic, bounded Bayesianism.

Primitive psychological evidence aggregation processes are well characterized by sampling models. Many cognitive processes, including action selection and belief formation, can be thought of as fast, evidence-based decisions, and sampling models enjoy wide-ranging and convergent support as the models of such fast, evidence-based decisions (Forstmann, Ratcliff, and Wagenmakers 2016). An elusive feature of sampling models is the tremendous domain specificity that they exhibit. Somehow, each task well described by sampling models chooses relevant sample spaces from which to sample evidence. We'll use the term *sampling bucket* for such a sampling space.

https://doi.org/10.1017/psa.2025.10119 Published online by Cambridge University Press

Although the relevance-deciding factors that dictate the formation of sampling buckets remain poorly understood, I make two assumptions about them, which are common to the cognitive scientific contexts in which sampling models are employed, and are moreover common and fruitful assumptions across the cognitive sciences. These features are (1) *context sensitivity* (that sampling buckets are specially formed for a given task) and (2) *reinforcement* (that part of the way that sampling buckets are formed or chosen is by their past successes). As I discuss further in the next section, attention to sampling buckets offers a straightforward explanation for confirmation bias—the larger class of phenomena to those presented by Mandelbaum (2019)—because it has been found that previously forming beliefs results in disproportionate weighing of old evidence over new evidence (Stone et al. 2022). This leads to more familiar, confirmatory evidence being simply overrepresented in samples.

Let me summarize the key points of this section: Mandelbaum's (2019) argument is interesting only if it states that no bounded Bayesianism can account for the data he provides, given the massive complexity of literal Bayesianism. Moreover, all forms of bounded Bayesianism feature a question of which evidence will be used in approximating the posterior distribution, because it is computationally intractable to use literally everything as data. It is these, pre-Bayesian, evidence selection procedures that can explain Mandelbaum's data, not as symptoms of fundamentally Bayesian-perverse belief updating, but as idiosyncratic structural features of a bounded Bayesian-normative process.

5. Revisiting the data

Equipped with an understanding of sampling and evidence aggregation, let us revisit the data Mandelbaum (2019) presents to see if the evidence demonstrates a need for a non-Bayesian foundation for belief updating. First, let us recall the study in which participants were first asked to form an initial opinion and then were found to be in stubborn adherence to that opinion in light of counterevidence. How might a Bayesian sampling implementation explain this bad behavior? First, we recall that a bounded Bayesianism requires a pre-Bayesian method for selecting the sampling bucket from which evidence will be sampled. A reasonable assumption is that sampling buckets are reinforced: Part of how they are formed is how useful they have been in previous, similar cases—in this case, the formation of the initial belief. Therefore pro-attitudinal evidence will be overrepresented in sampling, and this manifests in a bias toward the initial belief. An interesting empirical question that falls out of this view is what the time frame would be for new evidence to gain familiarity that allows it to be comparatively sampled with the old, confirmatory evidence. This account enjoys grounding in cognitive science, because similar arbitration methods are used in other evidence-based processes. For example, in value-based decision-making, it has been found that arbitration between value assignments in a given situation depends on the previous success of those valuation systems in analogous circumstances (Haas 2018). Therefore a pre-Bayesian process of selecting relevant evidence on the basis of previous use explains this first behavior.

How about the evidence Mandelbaum (2019) points to that suggests that we overlook evidence, such as in the case of rats not learning relationships between audiovisual stimuli and gustatory responses? This is addressed by the context

https://doi.org/10.1017/psa.2025.10119 Published online by Cambridge University Press

sensitivity associated with evidence aggregation. When forming a belief about a particular kind of phenomenon, such as a belief about what might be causing nausea, evidence that is gustatory in nature will be selected in the formation of a sampling bucket. This is sensible—it is the same kind of bias that deters me from scrutinizing the arrangement of books on an interviewer's bookshelf in an attempt to discern how an ongoing job interview is going. This is another instance in which, as finite agents, there is a need to select which evidence will be considered in updating a belief.

The third bit of Mandelbaumian data presented earlier is the case in which polarized beliefs become yet more closely held in the face of equivocal evidence. This is yet again an instance in which the sampling bucket overrepresents bits of evidence that were previously useful––that is, bits of evidence that contributed to the initial formation of the polarized belief. An interesting potential investigation suggested by this picture is considering when it is that sampling buckets form equivocally in the face of equivocal evidence and lead to tempered beliefs, which seems to happen more often than polarized persistence. The point, however, is that this behavior does not demonstrate a foundation to belief updating that is irreconcilable with a bounded Bayesian picture. It, moreover, does not even seem to be a systematically exhibited behavior, as raised in the global considerations against PsIS, so it is better explained as an idiosyncratic side effect of Bayesian approximation than it is explained by PsIS.

Last, Mandelbaum (2019) brings the odd phenomenon of backfiring: cases in which people report belief in a bit of evidence and, at the same time, report strengthened belief in the proposition undermined by that evidence. Is it true, as Mandelbaum states, that the effect is fundamentally devastating to any account of bounded Bayesian belief updating? Although it is true that no Bayesian process itself could exhibit this behavior, a bounded Bayesian implementation involves pre-Bayesian processing that may give rise to it. In particular, we have seen that sampling buckets are context sensitive and that the formation of a sampling bucket depends on previous iterations. Therefore there is a gap between vocalizing acceptance of a piece of evidence and having that evidence available in sampling buckets, and moreover having that piece of evidence adequately represented in the sampling buckets from which evidence is sampled in the evaluation of strongly reinforced beliefs. Because the sampling buckets of strongly held beliefs will enjoy robustness due to repeated use, we can expect polarized persistence to occur in cases in which the sampling bucket has not formed with the latest data.

This second prong against the PsIS picture concerns the Bayesian perversity of Mandelbaum's (2019) data. Even if we grant that the data concerned belief updating, any appropriately bounded Bayesian-normative process must be incompatible with the data if it is to show that belief updating is systemically Bayesian perverse. Any bounded Bayesian process is faced with the subtask of selecting relevant evidence. Minimal assumptions about pre-Bayesian evidence aggregation can explain Mandelbaum's data.

To sum, I have argued (1) that the PsIS picture clashes with most belief-updating behavior and (2) that the view is unnecessary to explain the alleged Bayesian-perverse behaviors of interest, because they can be explained through appeal to Bayesian-normative processes that enjoy firm support in cognitive science. The data do not provide strong reason to think belief updating fundamentally Bayesian perverse.

https://doi.org/10.1017/psa.2025.10119 Published online by Cambridge University Press

Acknowledgments. Special thanks to Chandra Sripada. Thanks also to Eissa Haydar and Caitlin Mace.

References

Anderson, Craig A. 1983. "Abstract and Concrete Data in the Perseverance of Social Theories: When Weak Data Lead to Unshakeable Beliefs." *Journal of Experimental Social Psychology* 19 (2):93–108. https://doi.org/10.1016/0022-1031(83)90031-8.

Batson, C. Daniel. 1975. "Rational Processing or Rationalization? The Effect of Disconfirming Information on a Stated Religious Belief." *Journal of Personality and Social Psychology* 32 (1):176. https://doi.org/10.1037/h0076771.

Beam, Michael A., Myiah J. Hutchens, and Jay D. Hmielowski. 2020. "Facebook News and (De) Polarization: Reinforcing Spirals in the 2016 US Election." In *Digital Media, Political Polarization and Challenges to Democracy*, edited by Maureen Beaufort, 26–44. Routledge. https://doi.org/10.4324/9780429243912-3.

Buesing, Lars, Johannes Bill, Bernhard Nessler, and Wolfgang Maass. 2011. "Neural Dynamics as Sampling: A Model for Stochastic Computation in Recurrent Networks of Spiking Neurons." *PLOS Computational Biology* 7 (11):e1002211. https://doi.org/10.1371/journal.pcbi.1002211.

Bullock, John G., Alan S. Gerber, Seth J. Hill, and Gregory A. Huber. 2013. *Partisan Bias in Factual Beliefs About Politics*. Report w19080. National Bureau of Economic Research. https://doi.org/10.3386/w19080.

Chater, Nick, Jian-Qiao Zhu, Jake Spicer, Joakim Sundh, Pablo León-Villagrá, and Adam Sanborn. 2020. "Probabilistic Biases Meet the Bayesian Brain." *Current Directions in Psychological Science* 29 (5):506–12. https://doi.org/10.1177/0963721420954801.

Fishkin, James S., and Robert C. Luskin. 2005. "Experimenting with a Democratic Ideal: Deliberative Polling and Public Opinion." *Acta Politica* 40:284–98. https://doi.org/10.1057/palgrave.ap.5500121.

Forstmann, Birte U., Roger Ratcliff, and E.-J. Wagenmakers. 2016. "Sequential Sampling Models in Cognitive Neuroscience: Advantages, Applications, and Extensions." *Annual Review of Psychology* 67 (1):641–66. https://doi.org/10.1146/annurev-psych-122414-033645.

Goodman, Noah D., Joshua B. Tenenbaum, and Tobias Gerstenberg. 2015. "Concepts in a Probabilistic Language of Thought." In *The Conceptual Mind: New Directions in the Study of Concepts*, edited by Eric Margolis and Stephen Laurence, 623–54. MIT Press. https://doi.org/10.7551/mitpress/9383.003.0035.

Griffiths, Thomas L., Falk Lieder, and Noah D. Goodman. 2015. "Rational Use of Cognitive Resources: Levels of Analysis Between the Computational and the Algorithmic." *Topics in Cognitive Science* 7 (2):217–29. https://doi.org/10.1111/tops.12142.

Haas, Julia. 2018. "An Empirical Solution to the Puzzle of Weakness of Will." *Synthese* 195 (12):5175–95. https://doi.org/10.1007/s11229-018-1712-0.

Kersten, Daniel, Pascal Mamassian, and Alan Yuille. 2004. "Object Perception as Bayesian Inference." *Annual Review of Psychology* 55 (1):271–304. https://doi.org/10.1146/annurev.psych.55.090902.142005.

Körding, Konrad P., and Daniel M. Wolpert. 2004. "Bayesian Integration in Sensorimotor Learning." *Nature* 427 (6971):244–47. https://doi.org/10.1038/nature02169.

Lieder, Falk, and Thomas L. Griffiths. 2020. "Resource-Rational Analysis: Understanding Human Cognition as the Optimal Use of Limited Computational Resources." *Behavioral and Brain Sciences* 43:e1. https://doi.org/10.1017/s0140525×1900061x.

Lieder, Falk, Thomas L. Griffiths, and Noah Goodman. 2012. "Burn-In, Bias, and the Rationality of Anchoring." *Advances in Neural Information Processing Systems* 4:2690–98.

Liu, Rongzhi, and Fei Xu. 2021. "Revising Core Beliefs in Young Children." *Proceedings of the Annual Meeting of the Cognitive Science Society* 43.

Lord, Charles G., Lee Ross, and Mark R. Lepper. 1979. "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence." *Journal of Personality and Social Psychology* 37 (11):2098. https://doi.org/10.1037//0022–3514.37.11.2098.

Mandelbaum, Eric. 2019. "Troubles with Bayesianism: An Introduction to the Psychological Immune System." *Mind and Language* 34 (2):141–57. https://doi.org/10.1111/mila.12205.

Miller, Arthur G., John W. McHoskey, Cynthia M. Bane, and Timothy G. Dowd. 1993. "The Attitude Polarization Phenomenon: Role of Response Measure, Attitude Extremity, and Behavioral Consequences of Reported Attitude Change." *Journal of Personality and Social Psychology* 64 (4):561. https://doi.org/10.1037//0022-3514.64.4.561.

https://doi.org/10.1017/psa.2025.10119 Published online by Cambridge University Press

Monro, Geoffrey D., and Peter H. Ditto. 1997. "Biased Assimilation, Attitude Polarization, and Affect in Reactions to Stereotype-Relevant Scientific Information." *Personality and Social Psychology Bulletin* 23 (6):636–53. https://doi.org/10.1177/0146167297236007.

Prior, Markus, Gaurav Sood, and Kabir Khanna. 2015. "You Cannot Be Serious: The Impact of Accuracy Incentives on Partisan Bias in Reports of Economic Perceptions." *Quarterly Journal of Political Science* 10 (4):489–518. https://doi.org/10.1561/100.00014127.

Ratcliff, Roger, Philip L. Smith, Scott D. Brown, and Gail McKoon. 2016. "Diffusion Decision Model: Current Issues and History." *Trends in Cognitive Sciences* 20 (4):260–81. https://doi.org/10.1016/j.tics.2016.01.007.

Sanborn, Adam N., and Nick Chater. 2016. "Bayesian Brains Without Probabilities." *Trends in Cognitive Sciences* 20 (12):883–93. https://doi.org/10.1016/j.tics.2016.10.003.

Sanborn, Adam N., Thomas L. Griffiths, and Daniel J. Navarro. 2006. "A More Rational Model of Categorization." https://cocosci.princeton.edu/tom/papers/rational1.pdf.

Sanborn, Adam N., Thomas L. Griffiths, and Daniel J. Navarro. 2010. "Rational Approximations to Rational Models: Alternative Algorithms for Category Learning." *Psychological Review* 117 (4):1144. https://doi.org/10.1037/a0020511.

Stone, Caleb, Jason B. Mattingley, and Dragan Rangelov. 2022. "On Second Thoughts: Changes of Mind in Decision-Making." *Trends in Cognitive Sciences* 26 (5):419–31. https://doi.org/10.1016/j.tics.2022.02.004.

Vul, Edward, Noah Goodman, Thomas L. Griffiths, and Joshua B. Tenenbaum. 2014. "One and Done? Optimal Decisions from Very Few Samples." *Cognitive Science* 38 (4):599–637. https://doi.org/10.1111/cogs.12101.

Wood, Thomas, and Ethan Porter. 2019. "The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence." *Political Behavior* 41:135–63. https://doi.org/10.1007/s11109-018-9443-y.

Cite this article: Haydar, Ibrahim. 2025. "A New Foundation for Belief Updating?" *Philosophy of Science* 92 (5):1215–1224. https://doi.org/10.1017/psa.2025.10119

https://doi.org/10.1017/psa.2025.10119 Published online by Cambridge University Press