The Bayesian brain: What is it and do humans have it?

Dobromir Rahnev

School of Psychology, Georgia Institute of Technology, Atlanta, GA 30332.

Abstract

It has been widely asserted that humans have a "Bayesian brain." Surprisingly, however, this term has never been defined and appears to be used differently by different authors. I argue that Bayesian brain should be used to denote the realist view that brains are actual Bayesian machines and point out that there is currently no evidence for such a claim.

In his target article, Brette criticized the claim that people have a "Bayesian brain." This term has been widely adopted to describe the nature of the human brain (Friston 2012; Knill & Pouget 2004; Sanborn and Chater 2016). Surprisingly, however, there is no agreed-upon definition of the term. Two rather informal definitions have been offered. First, Knill and Pouget (2004) describe the "Bayesian coding hypothesis" as follows: "the brain represents sensory information probabilistically, in the form of probability distributions"; second, according to Friston (2012), the "Bayesian brain says that we are trying to infer the causes of our sensations based on a generative model of the world." Neither of these definitions even mentions Bayesian computations, which, one may expect, should be central to the idea of a Bayesian brain. So, what then, is exactly meant by the "Bayesian brain?"

Any model of Bayesian computation contains at a minimum a set S of known stimuli, a set r of internal responses, and a known generative model $P(r|S)$ of the response generated by each stimulus. Bayes' theorem is used to invert the generative model to compute a likelihood function that is then combined with a prior $P(S)$ to obtain a posterior distribution. The result can be used to inform a forthcoming action or simply the percept of the observer.

A *Bayesian* brain must be implementing such Bayesian computations on some level. One can distinguish between two possible views here (Block 2018). The "as if" view holds that the brain does not necessarily literally have a generative model and does not literally use Bayes' theorem to derive a likelihood function. Instead, the computations performed by the brain can be seen "as if" it performs these operations. The "realist" view, on the other hand, holds that a generative model, a likelihood function, and a prior are actually represented in the brain and that the computations performed are literally the computations required by Bayes' theorem. Unfortunately, most authors do not necessarily commit to one or the other interpretation and, in some cases, appear to make different theoretical commitments in different papers.

rahnev@psych.gatech.edu.

Importantly, the "as if" view is typically expressed at Marr's "computational level" with no commitment to brain implementation (Griffiths et al. 2012). Consequently, using the term "Bayesian brain" in an "as if" sense appears almost contradictory because this usage is explicitly *not* about what happens in the brain. Thus, if the "Bayesian *brain*" is really a claim about the brain, then it has to be reserved for the realist view that the brain literally implements the components of Bayesian computation.

Is there evidence for the claim that humans have a Bayesian brain in the realist sense? No direct evidence has been presented to date. Instead, what is usually offered is an indirect argument from behavior. For example, Knill and Pouget (2004) motivated the view that brains are Bayesian by "the myriad ways in which human observers behave as optimal Bayesian observers" (p. 712). The problem is that this argument ignores the fact that findings of suboptimality are at least as common as findings of optimality (Rahnev and Denison 2018). Even more importantly, Bayesian optimality can be achieved by non-Bayesian algorithms (Ma 2012), and thus. such findings do not imply that brain computations are literally Bayesian.

In fact, as Brette eloquently explains, there are many reasons to doubt that brains are literally implementing Bayesian computations. Here, I formalize some of the issues examined by Brette and discuss some additional problems.

First, as pointed out by Brette, the internal response depends on more than just the stimulus of interest. Instead, the internal response to, for example, a tilted bar is better described not as $P(r|S)$ but as $P(r|S, \Theta)$, where Θ is a set of variables that affect neural firing, including the color of the bar, the color of the background, the size of the bar, the level of illumination, contrast, attention, arousal, metabolic state, and so forth. Dozens of such "confounding" variables can easily be present in any real-world situation. Inverting this generative model necessitates the integration (i.e., marginalization) over all possible values of all of these variables. For many forms of the assumed internal response, this computation is infeasible in real brains.

Second, as also discussed by Brette, Bayesian computations depend on the existence of a well-defined response r. However, brain activity is a dynamic, recurrent, never-ending string of action potentials. It is unclear how the Bayesian brain isolates "the response" to any given stimulus to perform the necessary Bayesian computations.

Third, an even more insidious problem that Brette did not examine in the context of the Bayesian brain is that a realist Bayesian brain must already know the set S of possible stimuli and the generative model $P(r|S)$ for each stimulus. However, the brain has to first learn both the stimuli in the world and their associated generative models. A truly Bayesian brain would thus form a probability distribution over the stimuli and generative models, which goes against current models that assume the existence of a predefined set S of stimuli.

Finally, a central tenet of the Bayesian brain – that the brain represents and computes with full probability distributions – has only been supported by theoretical proposals of how this *could* be achieved. Recent empirical research has, however, challenged this tenet (Yeon and Rahnev 2019).

The idea of the "Bayesian brain" has gained popularity perhaps not despite but because of the fact that it has never been clearly defined. This ambiguity shields it from criticism but it also robs it from any chance of contributing to scientific progress. To be useful, the term should be defined according to its plain meaning of a realist view where the brain literally represents the different components of Bayesian computations and researchers should present evidence for it that goes beyond "some behavior is close to optimal." Until then, the "Bayesian brain" should be seen for what it is: a theoretical possibility fully divorced and shielded from the empirical reality.

References

Block N (2018) If perception is probabilistic, why does it not seem probabilistic? Philosophical Transactions of the Royal Society B Biological Sciences 373:20170341 Available at: http://rstb.royalsocietypublishing.org/lookup/doi/10.1098/rstb.2017.0341.

Friston K (2012) The history of the future of the Bayesian brain. Neuroimage 62:1230–33. Available at: http://www.ncbi.nlm.nih.gov/pubmed/22023743 [Accessed March 23, 2019.] [PubMed: 22023743]

Griffiths TL, Chater N, Norris D & Pouget A (2012) How the Bayesians got their beliefs (and what those beliefs actually are): Comment on Bowers and Davis (2012). Psychological Bulletin 138:415–22 Available at: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0026884. [PubMed: 22545687]

Knill DC & Pouget A (2004) The Bayesian brain: The role of uncertainty in neural coding and computation. Trends in Neurosciences 27(12):712–719 Available at: http://www.ncbi.nlm.nih.gov/pubmed/15541511 [Accessed July 10, 2014]. [PubMed: 15541511]

Ma WJ (2012) Organizing probabilistic models of perception. Trends in Cognitive Sciences 16:511–18 Available at: http://www.ncbi.nlm.nih.gov/pubmed/22981359 [Accessed March 2, 2013.] [PubMed: 22981359]

Rahnev D & Denison RN (2018) Suboptimality in perceptual decision making. Behavioral and Brain Sciences 41:E223.

Sanborn AN & Chater N (2016) Bayesian brains without probabilities. Trends in Cognitive Sciences 20:883–93. Available at: 10.1016/j.tics.2016.10.003. [PubMed: 28327290]

Yeon J & Rahnev D (2019) The nature of the perceptual representation for decision making. bioRxiv:537068 Available at: https://www.biorxiv.org/content/10.1101/537068v1 [Accessed March 5, 2019.]