A Realist Perspective on Bayesian Cognitive Science

Michael Rescorla

Abstract: *Bayesian cognitive science* constructs detailed mathematical models of perception, motor control, and many other psychological domains. The models postulate mental activity that approximately conforms to Bayesian norms. Often, the postulated activity is subpersonal. I defend a *realist* stance towards Bayesian cognitive science. Bayesian models vary widely in their scientific merit, but many are well-confirmed and explanatorily superior to non-Bayesian alternatives. We have good reason to believe that these models are approximately true. I argue that realism about Bayesian cognitive science offers significant explanatory advantages over a rival *instrumentalist* view, on which Bayesian models are predictive tools that we should not construe even semi-literally.

§1. Bayesian modeling of the mind

Bayesian decision theory is a mathematical framework for modeling inference and decision-making under uncertain conditions. It has two central notions: *subjective probability* (or *credence*), which reflects the degree to which an agent believes that a state of affairs obtains; and *utility*, which reflects the degree to which an agent desires an outcome. At the heart of the framework lie norms governing credence and utility:

- The *probability calculus axioms* constrain how to allocate credence at any moment.
- *Conditionalization* dictates how to reallocate credence in light of new evidence.

- *Expected utility maximization* dictates what the agent should do in light of her current credences and utilities: namely, choose the action that maximizes expected utility.

The Bayesian framework has proved remarkably fruitful within a wide range of disciplines, including statistics (Berger, 1993), philosophy (Earman, 1992), physics (Trotta, 2008), artificial intelligence (Thrun, Burgard, and Fox, 2006), and medical science (Ashby, 2006).

The Bayesian framework attained its modern form in the work of Ramsey (1931) and de Finetti (1937/1980), who conceived of it in normative terms. The goal was to capture how agents *should* proceed. Subsequent researchers have often deployed the Bayesian framework for descriptive ends (e.g. Arrow, 1971; Luce and Suppes, 1965), maintaining that it helps us describe actual humans in an idealized way. There are longstanding debates about how well it serves this purpose (Kahneman and Tversky, 1979).

The debate has recently been transformed by the advent of *Bayesian cognitive science*, which constructs detailed Bayesian models of *perception* (Knill and Richards, 1996), *motor control* (Wolpert, 2007), *causal reasoning* (Gopnik, et al., 2004), *social cognition* (Baker and Tenenbaum, 2014), *intuitive physics* (Battaglia, Hamrock and Tenenbaum, 2013; Sanborn, Masinghka, and Griffiths, 2013), *human and nonhuman navigation* (Madl et al., 2014; Petzschner and Glasauer, 2011), *natural language parsing* (Levy, Reali, and Griffiths, 2009), and many other psychological domains. Bayesian modeling postulates mental activity that approximately conforms to Bayesian norms. Often, the postulated activity is *subpersonal*: executed by a mental subsystem rather than the individual herself. For example, Bayesian perceptual psychology treats the perceptual system as executing a Bayesian inference from proximal sensory stimulations (such as retinal stimulations) to perceptual estimates of shape, size, color, and other distal conditions. The inferences are not consciously accessible. No amount of introspection or soul-searching will reveal the credences or credal transitions instantiated by one's perceptual system. The emerging picture is that unconscious Bayesian inference underlies many core mental phenomena, including perception.

Bayesian cognitive science elicits diverse critical reactions. Critics charge that it gives no insight into neural implementation mechanisms and so is unexplanatory (Jones and Love, 2011; Herschbach and Bechtel, 2011), or that it is vacuous because it can fit any dataset through artful setting of parameters (Bowers and Davis, 2012; Glymour, 2011), or that its putative explanations are flawed because they do not specify how initial credences arise (Orlandi, 2014), or that it does not accurately describe how the mind works even though it fits some experimental results (Block, 2018; Colombo and Seriès, 2012; Glymour, 2011), or that many mental phenomena violate Bayesian norms (Colombo, Elkin, and Hartmann, forthcoming). All these critics deny that we have good reason to postulate unconscious Bayesian inferences.

I favor a different line. Bayesian models vary widely in their scientific merit, but many strike me as non-vacuous, well-confirmed, and explanatorily superior to non-Bayesian alternatives. We have good reason to believe that these models are approximately true. As you can see, I favor a *realist* viewpoint towards Bayesian modeling of the mind. I have defended my realist viewpoint in previous writings (Rescorla, 2015a; Rescorla 2016a). Here I will clarify what the realist viewpoint involves, and I will bolster my earlier defenses of it. §2 reviews basic aspects of Bayesian perceptual psychology. §3 broadens attention to Bayesian modeling beyond perception. §§4-5 clarifies the goals and methods of Bayesian cognitive science. §6 rebuts some widely discussed objections to Bayesian modeling. §§7-9 favorably compare my realist perspective with an *instrumentalist* view on which Bayesian models are predictive tools that we should not construe even semi-literally.

§2. Perception as unconscious Bayesian inference

The phrase "unconscious inference" traces back to Helmholtz (1867), who highlighted an *underdetermination problem* endemic to perception. The perceptual system cannot directly access conditions in the distal environment. It has direct access only to proximal sensory stimulations, which underdetermine their distal causes. How does perception solve this underdetermination problem? How does it estimate distal conditions based upon proximal sensory input? Helmholtz hypothesized that proximal stimulations trigger an unconscious inference regarding the most likely distal cause of the stimulations. Bayesian perceptual psychology builds upon Helmholtz's approach, postulating an unconscious Bayesian inference from proximal stimulations to perceptual estimates (Knill and Richards, 1996; Rescorla, 2015a).

A Bayesian perceptual model features a *hypothesis space*, where each hypothesis h concerns some aspect of the distal environment. h might concern shape, size, color, etc. The *prior probability* $p(h)$ is the initial credence assigned to h. The *prior likelihood* $p(e \mid h)$ is the conditional credence in proximal sensory input e given h. Upon receiving input e, the perceptual system reallocates credence over the hypothesis space in accord with Conditionalization, computing the *posterior probability* $p(h \mid e)$: the conditional credence in h given e. Bayes's Theorem states that

$$p(h \mid e) = \eta \, p(h) \, p(e \mid h),$$

where η is a normalizing constant to ensure that probabilities sum to 1. Based upon the posterior, the perceptual system selects a privileged estimate $\hat{h}$ of distal conditions. Usually, although not always, $\hat{h}$ is selected through expected utility maximization, where the utility function reflects the cost of an incorrect perceptual estimate. The selected estimate $\hat{h}$ informs the final percept.

Bayesian perceptual psychology has produced numerous well-confirmed models of perceptual processing. An acclaimed example is the motion perception model offered by Weiss, Simoncelli, and Adelson (2002). The model assumes a "slow motion" prior, i.e. a prior probability that favors slow distal speeds. One notable feature of the model is that, when stimulus contrast is low, computation of the posterior assigns higher weight to the slow motion prior, resulting in a slower speed estimate. This explains the well-known *Thompson effect*: perceived speed is slower when stimulus contrast is low (Thompson, 1982). The model explains a range of additional motion illusions that had previously resisted explanation within a single unified framework.

Another good example is the object-tracking model offered by Kwon, Tadin, and Knill (2015). The model applies when a perceiver visually tracks an object covered with a textured pattern (e.g. a soccer ball) that rotates as it moves through space. The model divides time into discrete stages separated by interval Δt. At time t, the perceptual system receives retinal input e_t and on that basis estimates three distal variables: x_t, the object's position at time t; v_t^{obj}, the object's translational velocity at time t; and $v_t^{pattern}$, pattern motion within the object. To form these estimates, the perceptual system employs priors that treat objects as likely to decelerate over time. This generalizes the slow motion prior from (Weiss, Simoncelli, and Adelson, 2002). The priors also treat object motion as more likely than pattern motion. Finally, the priors enshrine reasonable assumptions about environmental dynamics, such as that

(*) $x_t = x_{t-1} + \Delta t \times v_{t-1}^{obj}$,

and about the interface between perceiver and environment. At time t, the perceptual system computes the posterior

$$p(x_t,\ v_t^{obj},\ v_t^{pattern} \mid e_1, e_2, \ldots, e_t)$$

and selects privileged estimates $\hat{x}_t$, $\hat{v}_t^{obj}$, and $\hat{v}_t^{pattern}$. The posterior at t and retinal input e_{t+1} jointly determine the posterior at t+1:

$$p(x_{t+1},\ v_{t+1}^{obj},\ v_{t+1}^{pattern} \mid e_1, e_2, \ldots, e_{t+1}).$$

The result is a sequence of posteriors, each based on its predecessor and on current retinal input.

The object-tracking model explains an impressive range of phenomena. Consider *motion-induced position shift* (MIPS): a stimulus with a moving pattern appears shifted in the direction of pattern motion. A video of a typical MIPS stimulus is available online.[1] According to the Bayesian model, MIPS reflects the perceptual system's attempt at disambiguating inherently ambiguous retinal input. Retinal texture motion is jointly caused by translational velocity v_t^{obj} and pattern motion $v_t^{pattern}$. The perceptual system must disentangle how much retinal texture motion is due to v_t^{obj} and how much to $v_t^{pattern}$. When stimuli appear in the center of the visual field, position estimates have low uncertainty and the visual system estimates v_t^{obj} and $v_t^{pattern}$ quite accurately. When stimuli appear in the periphery of the visual field, position estimates are relatively uncertain. Accordingly, the Bayesian model leans heavily upon its prior bias in favor of object motion, attributing retinal texture motion largely to v_t^{obj} rather than $v_t^{pattern}$. The dynamical assumption (*) then enforces a change in estimated position toward the direction of perceived motion: the MIPS effect. An immediate consequence is that MIPS magnitude should negatively correlate with perceived pattern speed. Kwon, Tadin, and Knill (2015) confirmed this prediction, manipulating positional uncertainty both by blurring the stimulus and by moving its position in the visual field. The model also explains *peripheral slowing*: perceived

[1] http://movie-usa.glencoesoftware.com/video/10.1073/pnas.1500361112/video-1.

pattern motion becomes slower as the stimulus moves towards the periphery, because the prior bias favoring object motion over pattern motion dominates.[2]

The object-tracking model has ten free parameters, reflecting detailed assumptions about sensory noise and environmental dynamics. Kwon, Tadin, and Knill (2015) used the MIPS stimulus to fit the free parameters to experimental data for individual subjects. The results matched the data quite well. Remarkably, the results also matched data for several additional motion illusions. Videos of these illusions are available online.[3] I highly recommend that you watch all the videos. The key point here is that parameters derived from video 1 yield extremely accurate predictions for videos 2-6. A single model *with a single set of parameters derived from video 1* accommodates the data for all 6 videos. This provides strong support for the object-tracking model. It shows that the model has the unifying power we expect from good explanations.

Over the past century, researchers have explored many alternative frameworks for explaining perception. The alternative frameworks are not nearly as explanatorily powerful as Bayesian perceptual psychology. That is why the Bayesian paradigm dominates contemporary scientific research into perception.

§3. Beyond perception

Inspired by the success of Bayesian perceptual psychology, cognitive scientists have offered Bayesian models for many other psychological domains, including the domains listed in §1 and many others besides.

---

[2] For a video of peripheral slowing, modify the URL from note 1 by replacing “video-1” with “video-2”.
[3] For the additional four videos, modify the URL from note 1 by replacing “video-1” with “video-3”, “video-4,” “video-5”, or “video-6”.

Outside perceptual psychology, the most impressive Bayesian models lie within *sensorimotor psychology*: the study of how we control our bodies to achieve our goals. Suppose I resolve to lift some cup to my mouth. For me to achieve this goal, my motor system must estimate the cup's size, location, and shape, along with the current configuration of my motor organs. The motor system deploys these estimates to select motor commands that promote my goal. On a Bayesian approach, motor commands are selected through expected utility maximization. The utility function rewards achievement of my goal (e.g. lifting the cup to my mouth) and penalizes energetic expenditure. Expectations are computed relative to current credences, which are sequentially updated based upon sensory input and efference copy of motor commands. Models of this kind have proved explanatorily successful (Todorov and Jordan, 2002; Wolpert, 2007; Wolpert and Landy, 2012; Rescorla, 2016a).

Beyond perception and motor control, matters become less straightforward. In my opinion, other areas of Bayesian cognitive science do not match Bayesian perceptual psychology and Bayesian sensorimotor psychology in explanatory power. However, Bayesian models in some other areas have been fairly successful. A good example is Bayesian modeling of intuitive physics (Battaglia, Hamrock and Tennenbaum, 2013). Bayesian models successfully predict intuitive physical judgments in a range of scenarios, such as judgments about whether a pile of objects will collapse. Sanborn, Masinghka, and Griffiths (2013) show that the Bayesian approach compares favorably with rival non-Bayesian models of intuitive physics.

Like most scientific research programs, Bayesian cognitive science varies in its achievements. Some Bayesian models are highly explanatory, others less so. We must evaluate individual models on a case-by-case basis.

§4. Goals and methods of Bayesian cognitive science

In a widely discussed critique of Bayesian cognitive science, Jones and Love (2011, p. 170) write: "the primary goal of much Bayesian cognitive modeling has been to demonstrate that human behavior in some task is rational with respect to a particular choice of Bayesian model." They coin the name *Bayesian fundamentalism* for this research agenda. They offer a series of arguments against Bayesian fundamentalism.

Jones and Love's critique has elicited numerous rejoinders (e.g. Chater, et al., 2011). The rejoinders amply demonstrate that few if any practicing scientists endorse Bayesian fundamentalism. Jones and Love are attacking a strawman position. Bayesian cognitive scientists do not aim to show that all or most mental processing conforms to Bayesian norms. They aim to construct well-confirmed explanations of mental and behavioral outcomes. They regard idealized Bayesian modeling as a good starting point for constructing such explanations. Ultimately, one must test each Bayesian model against the data. In some cases, actual performance may deviate slightly or dramatically from the model.

More explicitly, Bayesian cognitive scientists pursue the following methodology when studying a psychological task:

(i) *Use Bayesian decision theory to articulate a normative model of the task*. The model describes how an idealized Bayesian system would execute the task. In general, the model will contain various free parameters. For example, the Bayesian object-tracking model contains ten free parameters.

(ii) *Fit the normative model as well as possible to the data by specifying its free parameters*. Kwon, Tadin, and Knill (2015) fit the object-tracking model to the data from video 1 by specifying all free parameters.

(iii) *Examine how well the model with all details specified fits actual performance*. The object-tracking model with free parameters specified was an excellent fit for the data for video 1 and also for the additional videos 2-6.

The core methodology is to articulate a normative model and then fit any free parameters to the experimental data. The model serves as a benchmark. The goal is to evaluate how well actual psychological processing conforms to the benchmark. Human performance often, although not always, conforms quite well.

This norms-based methodology implicitly presupposes some degree of baseline approximate conformity to Bayesian norms. If mental activity never remotely conformed to Bayesian norms, or if it approximately conformed only in exceptional circumstances, then constructing idealized Bayesian models would not be a good use of scientific resources. Thus, Bayesian cognitive science enshrines a *methodological* commitment to some baseline level of approximate conformity.[4] Researchers who pursue Bayesian modeling presuppose that at least some mental processes at least approximately conform to Bayesian norms. Clearly, that methodological commitment falls far short of Bayesian fundamentalism as defined by Jones and Love. One can adopt (i)-(iii) as a fruitful methodology without aiming to establish that all, many, or any mental processes conform to Bayesian norms.

The norms-based methodology has been amply vindicated over the past few decades. It has produced successful explanations for numerous mental phenomena, especially perceptual and motor phenomena.

[4] In (Rescorla, 2016a), I described the methodological commitment as a "working hypothesis." That description tallies with the language one finds among some practicing Bayesian cognitive scientists (e.g. Stocker, 2018). Overall, though, I think that the phrase "methodological commitment" more accurately captures how Bayesian cognitive science operates.

The norms-based methodology does not necessitate *doctrinal* commitment to any putative law or generalization along the following lines:

> Perceptual activity approximately conforms to Bayesian norms, *ceteris paribus*.
>
> More perceptual processes approximately conform to Bayesian norms than do not.
>
> Mental activity approximately conforms to Bayesian norms, *ceteris paribus*.
>
> More mental processes approximately conform to Bayesian norms than do not.

For example, Bayesian perceptual models do not contain, presuppose, or entail any doctrine to the effect that most perceptual processes approximately conform to Bayesian norms. A methodological presupposition of baseline conformity underlies how Bayesian perceptual psychologists *search* for successful models, but no such presupposition figures in the resulting models. No theory espoused by Bayesian perceptual psychologists asserts baseline conformity to Bayesian norms as part of its content. An analogous diagnosis applies to other areas of Bayesian cognitive science. The science supplies Bayesian models of specific mental phenomena, not general pronouncements about the scope of Bayesian modeling. One can hold that certain mental processes approximately conform to Bayesian norms while conceding that *other* mental processes, perhaps *many other* mental processes, dramatically violate Bayesian norms.

The norms-based methodology raises several questions. Why has the methodology proved so successful? Are there *a priori* reasons to expect that the methodology would succeed, or did things simply turn out that way? Why might the methodology prove more successful when applied to some psychological domains (e.g. perception) than others (e.g. high-level reasoning)? In what way, if any, do evolutionary or developmental pressures impel certain psychological systems to conform at least approximately to Bayesian norms? These questions deserve sustained investigation. They connect with topics of longstanding philosophical interest, such as the relation between psychological description and normative evaluation (Davidson, 1980). There is room here for productive interchange between science and philosophy. Unfortunately, overemphasis on strawman positions such as Bayesian fundamentalism has tended to derail recent discussion. More fruitful inquiry should begin by accurately gauging the methodological and doctrinal commitments of Bayesian cognitive science.

§5. Implementing Bayesian inference

Bayesian models typically prescind from neural implementation details. The models posit *credal states* (assignments of subjective probabilities to hypotheses) and *credal transitions* (mental transitions among credal states). They do not say how the brain encodes credal states. Nor do they identify neural processes or mental computations that underlie credal transitions. A major research program in contemporary neuroscience aims to illuminate the neural basis of Bayesian inference (Pouget, et al., 2013). This research program has generated several interesting proposals about how neural tissue might implement credal states and transitions, but so far no proposal has emerged as well-confirmed.

Although we do not know how the brain implements credal states and transitions, we know quite a lot about possible ways that possible physical systems can implement credal states and transitions. As I will now discuss, Bayesian cognitive science draws upon this knowledge to refine the norms-based methodology presented in §4.

In principle, there are many different ways that a physical system can implement credal states. Here are three possible implementation strategies:

- *Explicit enumeration of probabilities*. A physical system can explicitly enumerate the credence assigned to each hypothesis. This implementation scheme is not feasible when the hypothesis space is infinite.
- *Parametric encoding of a probability distribution*. For example, a physical system can encode a Gaussian distribution by recording the distribution's mean and its variance. A parametric encoding scheme is only feasible for probability distributions, such as Gaussians, that are encodable through finitely many parameters. Most probability distributions are not finitely encodable.
- *Sampling*. Imagine a physical system that draws finitely many *samples* from the hypothesis space. For any hypothesis h, there is an objective probability $q(h)$ that the system will draw h. Call $q(h)$ a *sampling probability*. As several researchers have proposed (Fiser, et al., 2010; Icard, 2016; Sanborn and Chater, 2016), sampling probabilities can serve as subjective probabilities. Drawing hypotheses with a certain objective probability is one way of assigning credences to them. The system can implicitly encode a probability distribution via sampling probabilities.[5]

Parametric and sampling encodings both figure prominently in scientific applications of the Bayesian framework, such as within robotics (Thrun, Burgard, and Fox, 2006).

One might ask why these diverse physical implementations all count as credal states. What do the implementations have in common, such that they count as ways of attaching subjective probabilities to hypotheses? Answering that question would require answering a deep question: what is it to attach a subjective probability to a hypothesis? Unfortunately, no one knows the answer to the deep question. A large literature addresses the nature of credal states

[5] My formulations here and in my subsequent remarks about sampling require minor emendation when the hypothesis space is continuous, but not in any way that affects the overall thrust.

(Erikkson and Hájek, 2007), but the literature has yielded disappointing results.[6] Any good answer to the deep question must take as a starting point that there are diverse ways to instantiate credal states, including parametric and sampling encodings. All neuroscientific research into neural implementation of Bayesian inference begins from this starting point.

When studying how the brain implements credal states and transitions, we must grapple with the *intractability* of Bayesian computation (van Rooij et al., 2019). A computation is *tractable* if it can be implemented by a physical system with limited memory and computing time.[7] In certain special cases, computing the posterior from the priors is a tractable task. For example, if the prior probability and the prior likelihoods are Gaussian, then the posterior will also be Gaussian, and its mean and variance are easily computable. In general, though, computing the posterior from the priors is an intractable task. Consider again Bayes's Theorem:

$$p(h \mid e) = \eta\, p(h)\, p(e \mid h).$$

Multiplying together $p(h)$ and $p(e \mid h)$ is easy. But computing the normalizing constant η requires summation (or integration) over the hypothesis space, which is in general an intractable task. Expected utility calculations are also in general intractable. The intractability of Bayesian computation poses a significant problem for Bayesian cognitive science, because the brain can only implement tractable computations.

Bayesians across a range of disciplines have intensively studied how a physical system with limited computational resources can *approximately* execute intractable Bayesian

[6] Most discussants tie credal states to personal-level psychological capacities. For example, de Finetti (1937/1980) cites the gambling odds that an agent would accept, while Davidson (1980) explores how an idealized interpreter would measure an agent's credences based upon the agent's linguistically-revealed preferences. Both strategies tie credences to sophisticated personal-level activities, such as gambling or linguistic communication. For that reason, they do not directly apply to the credal states studied within most of Bayesian cognitive science. In particular, the credal states postulated by Bayesian perceptual psychology are *subpersonal*. The person cannot access them. They do not serve as fodder for gambling, linguistic communication, or other sophisticated personal-level activities.

[7] C*omputational complexity theory* studies the distinction between tractable and intractable computation. For discussion of computational complexity theory and its relation to cognitive science, see (van Rooij et al., 2019).

computations. Computer scientists, engineers, and statisticians offer various schemes for approximating idealized Bayesian computations in tractable fashion. Cognitive scientists enlist these schemes to construct psychological models (Sanborn, 2017). The approximation schemes employed within Bayesian cognitive science generally fall into two main categories:

- *Variational algorithms* approximate the posterior using a probability distribution drawn from a nicely behaved family (e.g. Gaussian distributions). The basic idea is to pick the distribution from this family that is “closest” to the actual posterior. Picking the “closest” distribution from a nicely behaved family is often a much more tractable task than computing the actual posterior. The literature offers various measures for “closeness” of probability distributions.
- *Sampling algorithms* approximate the posterior by drawing finitely many samples from the hypothesis space. Rather than compute the actual posterior, the system responds to new evidence by altering its sampling probabilities in accord with the sampling algorithm. The new sampling probabilities serve as the system’s new subjective probabilities.

Variational and sampling approximations both feature credal transitions that violate Conditionalization but that approximately satisfy it:

- *Variational approximation*. Suppose that a system begins with prior $p(h)$ and that the ideal Bayesian response to input e is to compute a posterior $p(h \mid e)$. Suppose that the system instead computes some probability distribution $q(h)$ drawn from a “nice” family of distributions, such as Gaussians. By updating credences to $q(h)$ rather than $p(h \mid e)$, the system violates Conditionalization. However, when the variational approximation algorithm is well-chosen, $q(h)$ will be quite “close” to $p(h \mid e)$.

- *Sampling approximation*. Suppose that the system begins with prior $p(h)$ and then responds to input e by instantiating a sampling probability $q(h)$ over the hypothesis space, as dictated by some sampling algorithm. $q(h)$ serves as the new credence assigned to hypothesis h. By updating credences to $q(h)$ rather than $p(h \mid e)$, the system violates Conditionalization. However, when the sampling algorithm is well-chosen, $q(h)$ will approximate $p(h \mid e)$. More precisely, the objective probability $q(h)$ that the system samples h becomes (roughly) proportional to $p(h \mid e)$ as the number of samples increases. In that sense, the approximation asymptotically approaches the posterior as the number of samples approaches infinity.

In both cases, credal transitions roughly conform to Conditionalization. The result is a new credal state that approximately, although not exactly, matches the posterior. This new credal state can inform subsequent approximate Bayesian computation, including both inference and decision-making. For example, in the case of sampling algorithms, the system can approximate expected utility maximization by computing over the samples rather than the actual posterior.

Bayesian cognitive scientists deploy variational and sampling approximation schemes to study the mind. They thereby emend the norms-based methodology (i)-(iii) isolated in §4. They begin with a normative model of the psychological task, and they supplement the model with a second model that tractably approximates it. They then try to fit the second model to the data as well as possible by fixing any free parameters. The goal is to evaluate how well actual human performance conforms to the tractable approximation (Griffiths, Vul, and Sanborn, 2012).

This revised methodology has achieved notable successes. Sampling algorithms look especially promising (Sanborn and Chater, 2016).[8] To illustrate, consider *binocular rivalry*.

[8] In a series of publications (e.g. Friston and Stephan, 2007), Friston and collaborators pursue a version of the variational approximation scheme. Throughout his voluminous writings, Friston provides virtually no serious

When conflicting images are presented to the two eyes, the usual result is that the percept toggles between the images. Gershman, Vul, and Tenenbaum (2012) suggest that binocular rivalry results from a sampling approximation to Bayesian inference. Their main idea is that the perceptual system estimates whether stimulation of some retinal patch reflects distal conditions or whether it should be discarded as an outlier (e.g. it should be discarded if the retinal patch is damaged, or if an occluder blocks that portion of one's visual field). They delineate a computationally intractable Bayesian model that estimates distal conditions via an outlier estimation process. They also present a sampling approximation to the intractable model. According to the sampling approximation, the perceptual system draws hypotheses with objective probability approximately proportional to the posterior. When the two eyes receive conflicting images, multiple hypotheses have relatively high posterior probability. As a result, the percept fluctuates even though retinal stimulation does not change. Gershman, Vul, and Tenenbaum show that their model can explain a variety of phenomena, such as the distribution of time intervals between perceptual switches. The model, which is much more explanatorily powerful than alternative models of binocular rivalry, nicely illustrates the potential payoff from approximation schemes for idealized Bayesian inference.

§6. Objections to Bayesian modeling

Bayesian cognitive science has elicited many objections. This section addresses some popular objections, focusing especially on their force against Bayesian perceptual psychology.

§6.1 Violation of Bayesian norms

---

empirical support for his favored version of the variational scheme. Clark (2015) and Hohwy (2014) enthusiastically promote Friston's approach while neglecting much better confirmed sampling approximations.

A persistent criticism of Bayesian modeling is that human subjects often violate Bayesian norms. Over a series of enormously influential publications, Kahneman and Tversky argued that personal-level reasoning routinely flouts the probability calculus axioms and that personal-level decision-making routinely flouts expected utility maximization (e.g. Kahneman and Tversky, 1979; Tversky and Kahneman, 1983). Some researchers hold that *subpersonal* mental activity also routinely violates Bayesian norms. For example, Morales et al. (2015) present experimental evidence that perceptual processing can violate Conditionalization when the perceiver does not fully attend to the stimulus. Colombo, Elkin, and Hartmann (forthcoming) and Rahnev and Denison (2018) adduce further perceptual phenomena that they deem anti-Bayesian.[9]

I respond that Bayesian cognitive science does not regard Bayesian norms as anything like universal psychological laws. The goal is not to establish that all mental processes conform to Bayesian norms. The goal is to investigate the extent to which various mental processes conform to Bayesian norms and, in cases where they closely conform, to construct good explanations on that basis. Bayesian cognitive scientists can happily say that some mental processes conform to Bayesian norms while others do not. Bayesian perceptual psychologists can happily say that some perceptual processes conform to Bayesian norms while others do not.

Of course, if it turned out that no mental activity conformed even approximately to Bayesian norms, then Bayesian modeling of the mind would not be an empirically fruitful enterprise. However, many mental processes conform quite well. This is especially true of

[9] There is some evidence that the Thompson effect can reverse at high speeds (Thompson, Brooks, and Hammett, 2006): in other words, that fast-moving stimuli can appear to move even faster at low contrast. This reversed Thompson effect is inconsistent with the (Weiss, Simoncelli, and Adelson, 2002) model and with other standard Bayesian models of motion perception. However, the reversed Thompson effect does not replicate very reliably (Sotiropoulos, Seitz, and Seriès, 2014). Moreover, Bayesians could in principle accommodate the effect by suitably altering the prior (Stocker and Simoncelli, 2006). Thus, it is an open question how much of a challenge the reversed Thompson effect poses to Bayesian modeling of perception.

perception, where human performance closely approximates the Bayesian ideal across a wide range of cases and circumstances (Shen and Ma, 2016).

In some cases, allegations that a mental phenomenon violates Bayesian norms have turned out to be ill-founded. Consider the *size-weight illusion*: when you lift two objects of equal weight but different size, the smaller object seems larger. At first, the size-weight illusion looks like an anti-Bayesian effect, because it flouts a prior expectation that larger objects are heavier. Colombo, Elkin, and Hartmann (forthcoming) and Rahnev and Denison (2018) cite the illusion as evidence against a Bayesian approach to perception. But Peters, Ma, and Shams (2016) show that the illusion naturally arises from a Bayesian model that estimates relative *densities*. As this example illustrates, the Bayesian framework has repeatedly shown itself flexible enough to accommodate apparent anomalies. By adopting a sufficiently sophisticated Bayesian model, it often turns out to be possible for Bayesians to accommodate perceptual phenomena that initially look anti-Bayesian (Wei and Stocker, 2015). In similar fashion, many of the perceptual phenomena catalogued by Rahnev and Denison (2018) may eventually prove explicable by sufficiently sophisticated Bayesian models (Stocker, 2018).[10]

Even when a mental process dramatically violates Bayesian norms, Bayesian modeling may shed light upon it (Griffiths et al., 2012). In many cases, we can explain anti-Bayesian phenomena by pursuing the emended methodology sketched in §5: construct an idealized Bayesian model along with a tractable approximation to the idealized Bayesian model; evaluate how well the tractable approximation fits actual human performance. For example, human

[10] Anderson, O'Vari, and Barth (2011) discuss a motion stimulus that causes a percept as of a highly improbable illusory contour. They claim that the contour illusion is anti-Bayesian, on the grounds that a Bayesian perceptual system should not select a highly improbable hypothesis. Colombo, Elkin, and Hartmann (forthcoming) and Rahnev and Denison (2018) concur. However, there are Bayesian models that allow one to select a highly improbable hypothesis, so the contour illusion taken on its own has little force against the Bayesian program (Fleming, 2011). That being said, the illusion is an intriguing one, and it merits closer study by Bayesian perceptual psychologists.

cognition often exhibits *order effects*: the order in which evidence is received impacts judgment. Order effects violate Bayesian norms. Nevertheless, Sanborn et al. (2010) use a sampling approximation model to explain order effects arising in categorization. Similarly, Levy, Reali, and Griffiths (2009) use a sampling approximation model to explain non-ideal parsing of "garden path" syntactic structures. In both examples, and in many others, the idealized Bayesian model figures crucially as a basis for the tractable approximation model and a benchmark against which we can compare human performance. Human performance deviates from the benchmark due to limited computational resources.

§6.2 Falsifiability

Many critics complain that the Bayesian framework is vacuous or unfalsifiable (Bowers and Davis, 2012; Glymour, 2011). The framework allows us to postulate any priors we like, so one may greet an apparently anti-Bayesian phenomenon by insisting that suitably different priors would accommodate the phenomenon. This immunity to falsification may seem highly suspect. As Anderson, O'Vari, and Barth (2011, p. 495) put it: "The set of Bayesian models is infinite; at most, only a particular combination of priors, likelihoods, and utility functions can be rejected. This renders the perception as Bayesian inference claim untestable at best, and to the extent that it is always possible to find some combination of priors, likelihood, and utility that can generate any data set, it becomes meaningless."

I reply that worries about falsifiability rest upon a problematic conception of scientific theory choice. Kuhn (1962) argues convincingly that mature scientific theorizing usually operates within a *paradigm*, such as heliocentric astronomy or evolution by natural selection. The paradigm includes commitments so general or abstract that they resist direct empirical test.

For example, Newton's three laws taken on their own have little if any empirical content. They are not "testable" or "falsifiable." Only when one supplements them with additional principles, such as the law of universal gravitation, does experimental testing become possible. Scientists accept a scientific paradigm not because they have directly tested it but rather because it has proved strikingly successful in explaining certain phenomena. Scientists then develop the paradigm so as to bring it into better contact with data. In many cases, the paradigm eventually proves able to explain anomalies that it initially struggled to accommodate (e.g. it took sixty years for Newtonian physics to explain the observed motion of the moon's apogee). If enough recalcitrant anomalies accumulate, then scientists may ultimately replace the paradigm with a new one --- as when physicists replaced Newtonian physics with relativity theory.

The progress of Bayesian cognitive science fits the foregoing Kuhnian template. Researchers work within the framework sketched in §§4-5. They accept the framework because it has achieved striking explanatory successes (e.g. the motion estimation model). They develop the framework by constructing models of specific mental phenomena, including apparently anti-Bayesian phenomena such as the size-weight illusion. The individual models are empirically tested. The framework itself is not amenable to direct empirical testing (Griffiths, et al., 2012), any more than Newton's laws are amenable to direct empirical testing. If enough recalcitrant anomalies accumulate, another framework may ultimately prove more attractive.[11] So far, the Bayesian program has fared quite well in handling apparent anomalies.

[11] Colombo, Elkin, and Hartmann (forthcoming) observe that there are other mathematical frameworks besides Bayesian decision theory for modeling inference and decision-making under uncertainty. They suggest that cognitive scientists should explore these alternative frameworks as theories of mental activity. At present, though, no alternative framework has achieved anything approaching the massive success of the Bayesian framework, especially as applied to perception and motor control. That situation might of course change, but current evidence in the perceptual and motor domains strongly favors the Bayesian framework.

Worries about falsifiability are particularly inapt as applied to Bayesian cognitive science, whose core commitments are methodological rather than doctrinal. The core methodology is to assess how well mental activity conforms to Bayesian norms. The science does *not* incorporate general laws to the effect that all or most mental activity is Bayesian. In particular, Bayesian perceptual psychology does not incorporate any general law to the effect that all perceptual processing is unconscious Bayesian inference. The science only incorporates a methodological commitment to constructing and testing Bayesian models of mental activity. This commitment is not the sort of thing that one can "test" through direct confrontation with empirical evidence. The only meaningful "test" of the commitment is whether it produces explanatorily fruitful individual models. To date, it has passed that test most impressively.

§6.3 Ad hoc priors

Another worry about Bayesian modeling is that priors are chosen in ad hoc fashion (Glymour, 2011; Jones and Love, 2011). The Bayesian framework allows us to select any priors we please. The extreme flexibility makes Bayesian modeling look to some critics more like an exercise in curve-fitting than a source of genuine explanations.

This complaint may apply to some Bayesian modeling, but it does not apply to Bayesian perceptual psychology. In typical Bayesian perceptual models, the general form of the priors is well-motivated by environmental statistics or established psychophysics. That general form usually suffices for qualitatively accurate predictions. Of course, quantitative accuracy requires curve-fitting. However, parameters fit to one task often generalize to other tasks. In the object-tracking model, for example, parameters fit to individual performance for video 1 also match performance for videos 2-6. I submit that Bayesian perceptual psychology offers genuine explanations, not just *ad hoc* redescriptions of the data. Similar points apply to Bayesian sensorimotor psychology and at least some other areas of Bayesian cognitive science.

§6.4 Where do priors and hypotheses come from?

A common criticism of Bayesian models is that they *postulate* priors without explaining how the priors arise (Orlandi, 2014, p. 91). A related criticism is that Bayesian models postulate a hypothesis space over which priors are defined, rather than explaining how the hypothesis space is chosen (Orlandi, 2014, p. 91; Orlandi, 2016).

Both criticisms conflate *incomplete* theories with *unexplanatory* theories. Explanation must begin somewhere. 19th century chemists explained numerous properties of molecules by postulating that atoms form chemical bonds, but they did not know how chemical bonds arise. Darwin explained speciation and the observed fossil record by postulating evolution through natural selection, but the mechanisms underlying heredity remained mysterious to him. In each case, the explanation was incomplete and widely recognized as such. In each case, the incompleteness served as an impetus for future research. In each case, further developments helped fill the explanatory gap. In each case, the incomplete theory already offered powerful explanations. The same goes for Bayesian cognitive science. A Bayesian model presupposes priors and a hypothesis space, so its explanations are incomplete. Future research should try to fill the gap. Even in its present incomplete state, the Bayesian framework already offers powerful explanations of many psychological phenomena.

Orlandi (2016, p. 335) writes: “If we are seeking to explain how we derive a single percept from underdetermined stimuli, then we cannot leave aside the question of how the hypothesis space is limited. This would amount to trading the original mystery with a new, similar mystery.” In most cases, it not so mysterious why a given hypothesis space is operative. In the object-tracking model, for example, the perceptual system seeks to estimate three variables: x_t (position), v_t^{obj} (translational velocity), and $v_t^{pattern}$ (pattern motion). Accordingly, it employs a hypothesis space composed of all possible values for x_t, v_t^{obj}, and $v_t^{pattern}$. No other hypothesis space would be as appropriate, given the estimation task in which the perceptual system is engaged. One might ask why the perceptual system engages in that particular estimation task. One might also ask how the perceptual system is able to represent x_t, v_t^{obj}, and $v_t^{pattern}$ in the first place. These are interesting questions. Answering them would require significant progress within both philosophy and psychology regarding the phylogeny, ontogeny, and metaphysics of perceptual representation. Even absent such progress, the Bayesian model taken on its own already illuminates perceptual object-tracking. It isolates crucial explanantia (the priors) and specifies how they causally influence estimation of x_t, v_t^{obj}, and $v_t^{pattern}$.

§6.5 Mechanisms

Bayesian models do not say how the brain encodes priors. They do not identify neural processes or mental computations that underlie (approximate) Bayesian inference. Thus, Bayesian modeling does not address the mechanisms through which the brain implements credal states and transitions. Some critics argue on that basis that Bayesian models do not provide good explanations (Jones and Love, 2011; Herschbach and Bechtel, 2011).

This criticism assumes that good explanations must reveal mechanisms that produce the explanandum. Scientific practice offers numerous counterexamples: successful scientific explanations that are not remotely mechanistic (Rescorla, 2018). For example, the ideal gas law helps us explain the pressure exerted by a gas upon a container by isolating causally relevant factors (volume, number of moles of the gas, and temperature) and describing in systematic terms how those factors causally influence temperature. The ideal gas law does not specify underlying physical mechanisms, yet even so it is explanatory. It provides a *non-mechanistic causal explanation* of temperature. Similarly, a Bayesian perceptual model isolates causally relevant factors (the priors) and describes in systematic terms how they influence the percept. The model does not specify underlying neural mechanisms through which the priors influence the percept. Nevertheless, it is explanatory. It provides a *non-mechanistic causal explanation* of perceptual estimation (Rescorla, 2018).[12]

I acknowledge that mechanistic details often *improve* a scientific explanation. For example, statistical mechanics improves upon the ideal gas law by describing a gas in mechanistic terms as a collection of tiny interacting particles. Cognitive scientists hope to improve Bayesian modeling by identifying neural implementation mechanisms for credal states and transitions (Pouget, et al., 2013). A satisfying theory of neural implementation mechanisms would doubtless enhance the explanatory power of non-mechanistic Bayesian models. Even lacking mechanistic details, many non-mechanistic Bayesian models are already well-confirmed and highly explanatory.

§7. Realism versus instrumentalism about Bayesian cognitive science

I have argued that many Bayesian models yield satisfying explanations. I will now argue that we should regard these models as approximately *true*. Thus, I will be defending a realist perspective on Bayesian cognitive science. This section introduces realism along with a

[12] See (Woodward, 2003; Woodward, 2018) for general discussion of causal explanation, including non-mechanistic causal explanation.

competing instrumentalist approach. §8 critiques some prominent anti-realist arguments. §9 highlights key explanatory advantages that realism offers over instrumentalism.

I presuppose a broadly *scientific realist* viewpoint. Scientific realism has been a central topic within philosophy of science for decades, coming in many versions with many arguments pro and con (Chakravartty, 2017). The intuitive idea behind most versions is that explanatory success is a *prima facie* guide to truth or approximate truth. Scientific realists recommend some kind of positive attitude towards the approximate truth of explanatorily successful scientific theories. My discussion will not hinge upon how exactly one formulates scientific realism.

Scientific realism entails that, when a Bayesian model of a mental process is explanatorily successful, we have reason to regard the model as approximately true. Approximate truth of the model requires that there exist credal states roughly like those described by the model and that mental activity transit between those states roughly as described by the model. For example, a Bayesian perceptual model posits credal states that causally interact with one another and with sensory inputs, yielding perceptual estimates. The Bayesian motion estimation model posits three credal states: a prior probability, a prior likelihood, and a posterior. The Bayesian object-tracking model posits a sequence of credal states $p(x_t, v_t^{obj}, v_t^{pattern} \mid e_1, e_2, \ldots, e_t)$, computed in response to sequential sensory input. These two models are explanatorily successful, and they invoke credal states and transitions in an essential way. Assuming a scientific realist viewpoint, we have strong reason to believe that the models are approximately true. Thus, we have strong reason to believe that motion estimation and object-tracking deploy credal states that interact in approximate accord with Bayesian norms.

My realism is a realism about specific Bayesian models. Realists about Bayesian cognitive science do not undertake a commitment to endorsing all Bayesian models of the mind, any more than scientific realists undertake a commitment to endorsing all scientific theories. One must examine the details of a particular Bayesian model to see whether the model is well-confirmed. One must evaluate how it compares with rival models, whether its predictive successes are attributable entirely to curve-fitting, and so on. When a model passes successfully through this confirmatory crucible, we have reason to accept it as at least approximately true. Some but not all Bayesian models pass the test.

My realist perspective extends straightforwardly to psychological models that approximate idealized Bayesian inference through tractable computations. When such a model is explanatorily powerful, we have reason to regard it as at least approximately true. For example, the empirical success of the binocular rivalry model provides good reason to believe that perception instantiates computations at least roughly like those posited by the model.

A realist approach to Bayesian cognitive science contrasts with an *instrumentalist* approach, on which Bayesian models are predictively useful devices that do not accurately depict psychological reality. Colombo and Seriès (2012, p. 714) endorse instrumentalism: "Bayesian models should be understood as no more than toolboxes for making predictions and systematizing data." Block (2018, p. 6) agrees: "the best attitude towards the Bayesian formalism is an 'as if' or instrumentalist attitude." Block focuses on perception. He urges that, even when the perceptual system transits from proximal input to percept *as if* executing a Bayesian inference, we should not conclude that it *actually* executes a Bayesian inference. Orlandi (2014) also develops a broadly instrumentalist view of Bayesian perceptual psychology. She says that, when perceptual psychologists talk about "priors," we should not interpret this talk too literally. We should not posit causally efficacious credal states. We should instead view priors as "biases" or "simple constraints" that are "wired" into the perceptual system (pp. 82-83).

In my opinion, instrumentalism about Bayesian cognitive science is no more plausible than instrumentalism regarding physics, chemistry, biology, or any other successful science. Just as the explanatory success of physics provides evidence for gravity, or the explanatory success of chemistry provides evidence for the chemical bond, or the explanatory success of biology provides evidence for evolution by natural selection, so does the explanatory success of Bayesian cognitive science provide evidence for credal states and transitions across a range of psychological domains. In particular, the striking explanatory success of Bayesian perceptual psychology provides strong evidence for subpersonal credal states figuring in perception. The rest of the paper defends this realist viewpoint.[13]

§8. Arguments against realism

The literature offers numerous arguments against realism regarding Bayesian cognitive science. Many arguments, including those critiqued in §6, question the scientific value of Bayesian modeling. Other arguments, some of which I will now address, concede that Bayesian cognitive science has scientific value but deny that we should accept Bayesian models as even approximately true.

§8.1 Idealization and approximate truth

Colombo and Seriès (2012) motivate instrumentalism by observing that Bayesian models often incorporate false idealizing assumptions. For example, the Bayesian motion estimation

[13] Dennett (1987) develops a broadly instrumentalist approach to psychological explanation. According to Dennett, psychological explanation involves taking the "intentional stance" towards a subject, without any commitment regarding the subject's actual mental states. Dennett applies his instrumentalist approach both to personal-level and subpersonal psychological explanation. Hornsby (2000) and McDowell (1994) agree with Dennett regarding the subpersonal level, but they favor a more realist approach to the personal level. See (Rescorla, 2016a) for critical discussion of Dennett.

model posits Gaussian priors. Gaussian priors are convenient because they enable us to derive an elegant closed-form expression for the posterior. However, there is strong evidence that the velocity prior has "heavier tails" than a Gaussian (Stocker and Simoncelli, 2006). More generally, Bayesian cognitive scientists often select a prior for mathematical convenience rather than psychological realism. If a Bayesian model avowedly incorporates false idealizing assumptions, how can we construe the model as literally true?

To evaluate this argument, note first that false idealization assumptions are pervasive throughout science. The real world is complex. When modeling it, scientists deliberately introduce simplifying distortions so as to achieve mathematical or analytic tractability. A physicist might assume a frictionless plane; a population geneticist might assume an infinitely large population; and so on. Successful theories routinely incorporate idealizing assumptions known to be false or even impossible. McMullin (1985) calls these assumptions *Galilean idealizations*. Scientists introduce Galilean idealizations quite self-consciously, hoping that future research will yield more accurate models. To use Weisberg's (2007) example: quantum chemistry at first gave only highly approximate descriptions of wave functions for virtually all molecules, but with time it gave much more accurate descriptions. Scientific disciplines often take idealized models as a starting point and then gradually produce more accurate models by eliminating or reducing Galilean idealization.

Scientific realists accommodate Galilean idealization by invoking *approximate truth*. A scientific theory that incorporates false idealizing assumptions is not true, but it may be approximately true. When a theory is explanatorily successful, realists hold that we have *prima facie* reason to believe that it is approximately although not perhaps literally true. Subsequent research aims to discover a more accurate theory.

The same goes for Bayesian cognitive science. When a Bayesian model incorrectly postulates a Gaussian prior, this is (and is typically advertised as) a deliberate distortion introduced to simplify calculation and analysis. A model that incorporates an incorrect prior cannot be literally true. If the model exhibits striking empirical success, then we have strong reason to believe that it is approximately true. Scientists can improve upon the model by describing the actual prior more accurately, as Stocker and Simoncelli (2006) did for the velocity prior.[14] A similar progression has transpired within Bayesian modeling of *sensory cue combination*, which initially postulated idealized Gaussian priors and subsequently identified less idealized priors (Trommershäuser, Körding, and Landy, 2011; Rescorla, forthcoming). Thus, scientific realists can accommodate Galilean idealization within Bayesian perceptual psychology in the same way that they accommodate Galilean idealization within other scientific disciplines.

One might challenge the realist appeal to approximate truth. No one has ever explained in general, satisfying terms what it is for a theory to be "approximately true." However, this worry seems no more troublesome for Bayesian cognitive science than for any other science. I submit that Galilean idealization poses no greater a challenge to realists about Bayesian cognitive science than it does to scientific realists more generally. As already noted, the present paper assumes a broadly scientific realist viewpoint.

§8.2 Explicit enumeration of credences?

Block (2018) critiques realism about Bayesian perceptual psychology. He discusses a version of realism on which "priors and likelihoods (and utilities)… are represented explicitly in

[14] Sotiropoulos, Seitz, and Seriès (2014) revise the (Stocker and Simoncelli, 2006) model to include a pre-processing step that models speed tuning in the visual cortex. This pre-processing step influences the stimulus measurement e that serves as input to Bayesian computation. The resulting model fits experimentally observed interactions between speed and contrast quite well.

perceptual systems" (p. 8). Block notes that there are many ways to implement or approximately implement a Bayesian model besides explicit enumeration of probabilities. He mentions sampling as one alternative implementation strategy. He says that we have no reason to favor an explicit encoding scheme over a sampling implementation. He concludes that we should adopt an instrumentalist rather than a realist construal of Bayesian perceptual psychology.

I agree with Block that we have no reason to postulate explicit enumeration of probabilities by the perceptual system. In most perceptual tasks, the distal variable being estimated has infinitely many possible values. The hypothesis space is infinite, so explicit enumeration of credences is not an option. It is unlikely that explicit enumeration of probabilities plays an important role in perception or in any almost any other psychological domain.[15]

I dispute Block's suggestion that these observations undermine a realist construal of Bayesian modeling. Realists about Bayesian perceptual psychology claim that the perceptual system instantiates credal states and that, in certain cases, transitions among those states conform approximately to Bayesian norms. Realists do *not* claim that the perceptual system explicitly enumerates credences. Realism allows that the perceptual system may employ a parametric encoding, a sampling encoding, or some other encoding. In rejecting explicit enumeration of credences, Block is not rejecting realism.[16]

In one passage, Block hints that genuine Bayesian inference requires explicit enumeration of probabilities. He writes (p. 7):

---

[15] Orlandi (2016, p. 334, p. 336) suggests that the perceptual system must somehow limit the hypothesis space so as to render it finite. I see no need for any such finitary limitation. A physical system can allocate credence over an infinite (indeed, uncountably infinite) hypothesis space. For example, it can parametrically encode a Gaussian distribution. This is how numerous models from Bayesian perceptual psychology work, including the Bayesian object-tracking model. Few models in Bayesian perceptual psychology feature a finite hypothesis space, aside from tinker-toy models deployed for heuristic purposes by introductory expositions.

[16] Block attributes to me "a realist version of Bayesianism in which priors are explicitly represented" (p. 8). I do not endorse the attributed view either in the two papers that he cites, (Rescorla, 2015a; 2015b), or in any other work. Several of my previous writings have extensively discussed sampling implementation of credal states (Rescorla, 2009; Rescorla, 2012).

> What would show that something that deserves to be called Bayesian inference actually occurs in perception? In the most straightforward implementation, there would have to be perceptual representations of prior probabilities for alternative hypotheses, perceptual representations of likelihoods and some process that involves something that could be described as multiplication of these values.

The "most straightforward implementation" emphasized by Block plays virtually no role in scientific applications of Bayesian decision theory, because virtually all scientific applications feature an infinite hypothesis space. A physical system can execute Bayesian inferences even though it does not employ anything resembling Block's "most straightforward implementation." When priors are Gaussian, for example, the system can conform to Conditionalization by updating the mean and variance of the probability distribution. Such a system does not explicitly enumerate probabilities, let alone multiply probabilities together.

Block's discussion rests upon an overly narrow conception of what it is for a physical system to instantiate credal states. The discussion saddles realists with an implausible commitment (explicit enumeration of credences) that they do not and should not accept.

§8.3 Approximate Bayesian inference

Block claims that computational intractability poses a challenge to realism about Bayesian modeling (p. 8): "A major problem with realist theories in which Bayesian inference literally takes place in the brain is that the kind of Bayesian computations that would have to be done are known to be computationally intractable… So, any realist version of Bayesianism will have to tell us what exactly is supposed to be involved in the computations."

I disagree. Realists about a scientific theory need not specify exactly how the theory applies to specific cases. Realists about evolution by natural selection are not obliged to say how exactly a given species evolved. Realists about Bayesian cognitive science are not obliged to say how exactly a given mental process approximately implements Bayesian computation. In each case, it is scientific progress to discover as many details as possible. Even absent the desired progress, one can maintain a well-justified realist attitude towards the theory's key elements. One need not know how exactly an organism evolved through natural selection to believe with strong justification that it did so. One need not know exactly how the perceptual system approximately implements Bayesian inference to believe with strong justification that it does so.

Still, it may seem that any appeal to *approximate* Bayesian inference blurs the border between realism and instrumentalism, perhaps even draining realism of all content. If mental activity implements a tractable approximation rather than an idealized Bayesian model, then in what sense does the activity count as Bayesian? Once realists concede that mental activity violates Bayesian norms, what distinguishes their position from the instrumentalist view that mental activity proceeds *as if* it executes Bayesian inferences? In Block's words (2018, p. 8): "[W]hat is the difference between approximate implementation of Bayesian inference and behaving roughly as if Bayesian inference is being implemented…? Until this question is answered, the jury is out on the dispute between realist and anti-realist views."

I respond that there is a huge difference between systems that approximately implement Bayesian inference and systems that merely behave as if they do. A system of the former type instantiates credal states that interact in rough conformity to Bayesian norms. A system of the latter type may simulate a system of the former type, but it does not instantiate credal states that interact in approximate accord with Bayesian norms. For example, Maloney and Mamassian

(2009) note that a physical system can simulate certain simple Bayesian perceptual inferences by using a look-up table, without executing anything like a mental transition among credal states. The system merely looks up what perceptual estimate it should output in response to a given sensory input. Such a system does not approximately implement a Bayesian inference. It does not even instantiate credal states.

Any Bayesian model describes a mapping from inputs to outputs. For example, a Bayesian perceptual model describes a mapping from proximal sensory inputs to perceptual estimates.[17] A Bayesian model posits credal states that mediate the mapping from inputs to outputs. The simplest models posit three mediating credal states:

the prior probability $p(h)$

the prior likelihood $p(e \mid h)$

a credal state $q(h)$ that results from the prior probability, the prior likelihood, and input e.

In an idealized Bayesian model, $q(h)$ is the posterior $p(h \mid e)$. In an approximation model, $q(h)$ may only approximate $p(h \mid e)$. From a realist perspective, the credal states $p(h)$, $p(e \mid h)$, and $q(h)$ are genuine mental states that causally impact the transition from inputs into outputs. The two priors combined with input e cause credal state $q(h)$, which in turn causes a "decision," such as selection of a privileged hypothesis $\hat{h}$ (in the case of perceptual estimation) or selection of a motor command (in the case of motor control). Thus, input-output mappings are mediated by a causal structure that conforms approximately to Bayesian norms. Some Bayesian models, such as the object-tracking model, describe a richer causal structure that embeds additional credal states.

Instrumentalists are neutral about the causal structure that mediates between inputs and outputs. From an instrumentalist perspective, input-output mappings might just as well be

[17] The mapping may be stochastic rather than deterministic. Due to sensory noise, the same proximal input does not yield the same perceptual estimate on each occasion. There are various ways for Bayesian models to capture stochastic variation, such as incorporating a noise term that corrupts expected utility maximization.

mediated by a look-up table. So realists are far more committal than instrumentalists about the mental processes that mediate between inputs and outputs.

The concept of "approximation" is vague. No doubt there are borderline cases: cases where it is indeterminate whether a system "approximately implements" Bayesian inference. In practice, most naturally arising cases fall determinately on one or another side of the border. Variational and sampling schemes are clear-cut cases of approximate Bayesian inference. A look-up table is a clear-cut case in which approximate Bayesian inference does not occur. Realists contend that some mental activity falls on the approximate Bayesian inference side of the border. Instrumentalists say that we have no reason to think so. This is substantive and well-defined disagreement, even though it hinges upon a vague concept. Many useful concepts are vague (Williamson, 1994). The vagueness of "approximation" does not forestall constructive debate between realists and instrumentalists.

§8.4 Scientific practice

Block (2018, p. 8) cites scientific practice to support his instrumentalist construal of Bayesian modeling. He notes that some Bayesian cognitive scientists claim only to be addressing input-output mappings, without any commitment as to the causal structure that mediates between inputs and outputs. These researchers offer a Bayesian model as an ideal solution to a problem faced by the mind, such as estimating distal conditions based on proximal sensory input, or parsing an utterance's syntactic structure, or choosing a motor command that promotes one's goals. The researchers aim to assess how closely humans match the ideal solution, not to discover how humans actually solve the problem. Why should we regard a Bayesian model as approximately true when its own creators decline to do so?

I reply that the dispute between realism and instrumentalism is not about what scientists believe, any more than the dispute between Platonism and nominalism regarding mathematical entities is about what mathematicians believe. The dispute is about what we have *reason* to believe. Some practicing Bayesian cognitive scientists are indeed agnostic about the postulates of their own models, just as some mathematicians are agnostic about whether mathematical entities exist. In neither case does the agnosticism militate against a realist construal. The issue is whether agnosticism is well-grounded, not whether any practitioners partake in it.

In any event, Block neglects important aspects of current scientific practice that fit much better with a realist construal than with an instrumentalist construal. As noted in §5, a major strand in current neuroscience is the search for neural mechanisms that approximately implement Bayesian inference. This search presupposes a realist perspective on credal states and transitions. It presupposes that credal states posited by Bayesian models are genuine mental states instantiated by the brain, not just useful fictions. Investigating how priors are realized in the brain would be a fool's errand if there were no priors. Instrumentalists must reject as misguided all ongoing research into neural mechanisms of approximate Bayesian inference.

§9. Explanatory advantages of realism

I contend that realism offers key explanatory advantages over instrumentalism. I focus on perception, but my arguments generalize to explanatorily successful Bayesian models of non-perceptual domains. I offer two arguments: *the argument from altered priors* and *the argument from iterated inference*.

The argument from altered priors proceeds from the observation that priors can change in response to sensory input. For example, Sotiropoulos, Seitz, and Seriès (2011) investigated whether one can alter the slow motion prior employed during motion estimation. They repeatedly exposed subjects to fast-moving stimuli. In response to this experimental manipulation, the velocity prior shifted so as to favor faster speeds. Motion estimation changed accordingly: the same stimulus looked to be moving faster after the manipulation than it did before the manipulation. As a general matter, one can experimentally manipulate both prior probabilities (Adams, Graf, and Ernst, 2004; Ernst, 2007) and prior likelihoods (Sato and Kording, 2014; Sato, Toyoizumi, and Aihara, 2007; Seydell, Knill, Trommershäuser, 2010) deployed by the perceptual system.[18]

When priors change, there is a change in the mapping from proximal sensory inputs to perceptual estimates. Realists can offer a principled explanation for why the mapping changes as it does. If the perceptual system approximately executes a Bayesian inference based upon prior $p(h)$, then changing the prior from $p(h)$ to $p^*(h)$ will cause approximate execution of a Bayesian inference based upon $p^*(h)$. The original prior $p(h)$ induces one mapping Γ from sensory inputs to perceptual estimates, while the new prior $p^*(h)$ induces a different mapping Γ^*. Realists can describe in systematic terms how different environmental statistics yield different priors and thereby induce different mappings from sensory inputs to perceptual estimates.

Instrumentalists cannot offer nearly so satisfying an explanation. The fact that a system behaves *as if* executing a Bayesian inference from prior $p(h)$ gives no reason expect that any given experimental manipulation would cause the system to behave *as if* executing a Bayesian inference from some other prior $p^*(h)$. From an instrumentalist perspective, there is no reason why a given experimental manipulation should cause mapping Γ^* to replace mapping Γ. Instrumentalists can augment their theory by *stipulating* that the experimental manipulation

[18] In (Rescorla, 2015b), I mistakenly claimed that (Beierholm, Quartz, and Shams, 2009) is an example where the prior likelihood changes.

causes mapping Γ^* to replace mapping Γ. Clearly, though, the augmented theory does not explain in a principled way why Γ changes as it does. Similarly, Orlandi can say that the perceptual system is "wired" one way and then becomes "wired" a different way in response to changing environmental statistics, but her account yields no principled explanation for why one "wiring" replaces another (Rescorla, 2015b).

Here we see a stark contrast between realist and instrumentalist construals of Bayesian modeling. From a realist perspective, priors are genuine, causally efficacious mental states. We can say what would happen if we were to hold one mental state fixed while varying another --- e.g. if we were to hold the prior likelihood fixed while varying the prior probability. We thereby explain why certain experimental manipulations impact perceptual processing as they do. Instrumentalists cannot offer a comparably satisfying explanation, because they deny that the Bayesian model describes genuine mental states that causally influence perceptual processing. Instrumentalists hold that talk about priors is simply a way of summarizing the mapping from proximal inputs to perceptual estimates. Because they invest the Bayesian model with no psychological reality beyond the mapping Γ itself, they have no theoretical resources to explain why certain experimental manipulations change the mapping one way rather than another.

So goes the argument from altered priors. I offered a version of the argument in (Rescorla, 2015a; 2015b). Block responds:

> I find this argument unconvincing because whatever it is about the computations of a system that simulates the effect of represented priors… might also be able to simulate the effect of change of priors. Without a comparison of different mechanisms that can accomplish the same goal, the argument for realism is weak.

Block mentions sampling as an example of how a system might "simulate the effect of change of priors." However, we saw in §8 that sampling is consistent with a realist construal of Bayesian modeling. Sampling algorithms studied in cognitive science do not *simulate* approximate Bayesian inference. They *implement* approximate Bayesian inference. A sampling implementation does not just "simulate the effect of change of priors." The system instantiates priors, and those priors can change in response to suitable experimental manipulations, yielding a different mapping from sensory inputs to perceptual estimates. Sampling is an illustration of my argument, not a counterexample to it.

In principle, a system that simulates Bayesian inference from prior $p(h)$ might respond to certain experimental manipulations by simulating Bayesian inference from a new prior $p^*(h)$. But a system that simulates Bayesian inference from prior $p(h)$ need not so respond. Instrumentalists must offer a principled explanation for why the mapping Γ changes as it does. Maybe they will eventually do so. Maybe they will eventually provide a compelling alternative to the realist explanation. However, it is true of virtually *any* explanation that we may eventually discover a compelling alternative explanation. One does not undermine an abductive inference by noting the mere possibility that a compelling alternative explanation may someday emerge.

As an especially vivid illustration of the argument from altered priors, consider an experiment performed by Adams, Graf, and Ernst (2004). The experiment targets perceptual estimation of shape based upon shading cues. In typical humans, shape estimation relies upon a prior probability over possible directions for the light source. The prior assigns a relatively high probability to lighting directions that are overhead and slightly to the left. Adams, Graf, and Ernst (2004) manipulated the light-from-overhead prior by exposing subjects to deviant visual-haptic stimuli indicating a shifted direction for the light source. The prior changed accordingly, inducing an altered mapping from shading cues to shape-estimates. Moreover, the very same experimental manipulation altered performance in a *separate* perceptual task that required subjects to estimate which side of an oriented bar was lighter than the other.

Why does an experimental manipulation in one task (shape estimation) affect performance in a separate task (lightness estimation)? From a realist perspective, the answer is straightforward: both tasks deploy a common light-from-overhead prior; the experimental manipulation affects performance in both tasks by altering that prior. Instrumentalists seem unable to offer a comparably satisfying explanation. From an instrumentalist viewpoint, there is no reason to expect that an experimental manipulation of the mapping from sensory inputs to shape-estimates will *also* change the mapping from sensory inputs to lightness-estimates. For example, imagine a system that simulates Bayesian shape estimation using a look-up table and that simulates Bayesian lightness estimation using a separate look-up table. Let us stipulate that the system can respond to experimental manipulations in a perceptual task by altering the appropriate look-up table: e.g. it responds to deviant stimuli in the shape estimation task by altering the look-up table used for shape estimation. Let us stipulate that the system will change the mapping from shading cues to shape-estimates so as to simulate a change in the light-from-overhead prior. Our stipulations do not entail that the system *also* changes the look-up table used for lightness estimation. There is no principled reason to expect that a change in the look-up table used for shape estimation will correlate with a change in the look-up table used for lightness estimation. More generally, there is no principled reason why a system that simulates a changed prior in the shape estimation task should also simulate a changed prior in the lightness estimation task. Thus, realism offers a satisfying explanation where instrumentalism does not.

The argument from iterated inference is similar to the argument from altered priors, but it only applies to a restricted range of Bayesian models. It applies to models, such as the object-tracking model, that postulate sequential Bayesian inferences based upon evolving credal states.

The object-tracking model postulates a sequence of credal states $p(x_t, v_t^{obj}, v_t^{pattern} \mid e_1, e_2, \ldots, e_t)$, yielding perceptual estimates $\hat{x}_t$, $\hat{v}_t^{obj}$, and $\hat{v}_t^{pattern}$. Credal state $p(x_t, v_t^{obj}, v_t^{pattern} \mid e_1, e_2, \ldots, e_t)$ and sensory input e_{t+1} jointly determine the next credal state $p(x_{t+1}, v_{t+1}^{obj}, v_{t+1}^{pattern} \mid e_1, e_2, \ldots, e_{t+1})$. Thus, the model postulates sequential credal states that interact with sensory input to influence perceptual estimation. Each credal state $p(x_t, v_t^{obj}, v_t^{pattern} \mid e_1, e_2, \ldots, e_t)$ induces a mapping Γ_{t+1} from sensory input e_{t+1} to perceptual estimates $\hat{x}_{t+1}$, $\hat{v}_{t+1}^{obj}$, and $\hat{v}_{t+1}^{pattern}$. Intuitively: the system's current credences regarding position and motion determine how it will estimate position and motion based upon the next sensory input it receives. A different credal state $p^*(x_t, v_t^{obj}, v_t^{pattern} \mid e_1, e_2, \ldots, e_t)$ would induce a *different* mapping $\Gamma_{t+1}{}^*$. If we interpret the model realistically, we can explain in a systematic way why mapping Γ_{t+1} rather than mapping $\Gamma_{t+1}{}^*$ occurs. We regard the sequence of mappings $\Gamma_1, \Gamma_2, \ldots, \Gamma_t, \ldots$ as resulting from a fixed Bayesian estimator that updates credences based upon sensory inputs $e_1, e_2, \ldots, e_t, \ldots$ We can say that the sequence $\Gamma_1, \Gamma_2, \ldots, \Gamma_t, \ldots$ reflects a sequence of causally relevant credal states governed by a fixed Bayesian dynamics. Instrumentalists offer no comparable explanation. They cannot explain why, when the mapping at t is Γ_t, the mapping at t+1 is Γ_{t+1} rather than $\Gamma_{t+1}{}^*$. From an instrumentalist perspective, one sequence of mappings is no more to be expected than any other. Realists can offer a principled explanation for the sequence of mappings. Instrumentalists cannot.

It is not helpful to gloss credal states in terms of biases, constraints, wirings, or other similar locutions. There is of course a sense in which the perceptual system during the object-tracking task is wired one way at t (corresponding to mapping Γ_t) and then becomes wired a different way at t+1 (corresponding to mapping Γ_{t+1}). However, the "wiring" at t reflects a fleeting credal state that occurs at t and loses psychological relevance shortly thereafter. The "wiring" does not belong to the fixed architecture of perceptual processing. Sequential perceptual estimates $\hat{x}_t$, $\hat{v}_t^{obj}$, and $\hat{v}_t^{pattern}$ derive from an underlying sequence of credal states $p(x_t, v_t^{obj}, v_t^{pattern} \mid e_1, e_2, \ldots, e_t)$. Talk about biases, constraints, and wirings obscures this underlying causal structure. By acknowledging the causal structure, we reap explanatory dividends that appear otherwise unavailable.

The realist arguments I have provided are not decisive. A committed instrumentalist could insist that scientific theories are not in the business of accurate description and that we should construe even the most successful theory as nothing but a useful predictive device. Short of embracing a full-blown instrumentalist stance towards all scientific theorizing, I see little motivation for an instrumentalist construal of Bayesian perceptual psychology.

Although this section has focused on perception, my arguments generalize to other well-confirmed Bayesian models. The argument from altered priors applies whenever priors can change. The argument from iterated inference applies to any Bayesian model that postulates a sequence of inferences based upon evolving credal states --- such models are especially common in Bayesian sensorimotor psychology. Taken together, the two arguments show that realism offers notable explanatory advantages over instrumentalism.

§10. Conclusion

Helmholtz proposed that unconscious mental processes can resemble familiar conscious activities such as inference and decision-making. Bayesian cognitive science vindicates

Helmholtz's proposal through explanatorily powerful, well-confirmed models of perception, motor control, and other domains. Bayesian modeling establishes the existence of subpersonal mental computations that are inaccessible to consciousness yet that share theoretically crucial properties with personal-level credal inference. The computations incorporate transitions among credal states. In some cases, the transitions conform closely to the Bayesian ideal. In other cases, the transitions tractably approximate the intractable Bayesian ideal. How remarkable that the Bayesian paradigm, originally conceived with normative aspirations, has proved such a fertile source of empirical insights!

Acknowledgments

I presented versions of this material at the 2019 Annual Meeting of the Society for Philosophy and Psychology and at a 2019 Workshop on Current Topics in Cognitive Science at Ruhr Universität Bochum. I thank the audience members on both occasions for their helpful feedback. I also thank Steven Gross and Anders Nes for comments that improved the chapter. My research was supported by a fellowship from the National Endowment for the Humanities. Any views, findings, conclusions, or recommendations expressed in this publication do not necessarily reflect those of the National Endowment for the Humanities.

References

Adams, W., Graf, E., and Ernst, M. (2004). Experience can change the "light-from-above" prior. *Nature Neuroscience*, 7, 1057-1058.

Anderson, B., O'Vari, J., and Barth, H. (2011). Non-Bayesian contour synthesis. *Current Biology*, 21, 492-496.

Arrow, K. (1971). *Essays on the Theory of Risk-Bearing*. Chicago: Markham.

Ashby, D. (2006). Bayesian statistics in medicine: A 25 year review. *Statistics in Medicine*, 25, 3589-3631.

Baker, C., and Tenenbaum, J. (2014). Modeling human plan recognition using Bayesian theory of mind. In G. Sukthankar, R. P. Goldman, C. Geib, D. Pynadath, D., and H. Bui (Eds.), *Plan, Activity, and Intent Recognition: Theory and Practice* (pp. 177-204). Waltham: Morgan Kaufmann.

Battaglia, P. W., J. B. Hamrick, and J. B. Tenenbaum. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, 110, 18327–18332.

Beierholm, U., Quartz, S., and Shams, L. (2009). Bayesian priors are encoded independently from likelihoods in human multisensory perception." *Journal of Vision*, 9, 1-9.

Berger, J. (1983). *Statistical Decision Theory and Bayesian Analysis*. 2nd ed. New York: Springer Verlag.

Block, N. (2018). If perception is probabilistic, why does it not seem probabilistic?. *Philosophical Transactions of the Royal Society B*, 373, 20170341.

Earman, J. (1992). *Bayes or Bust?*. Cambridge: MIT Press.

Bowers, J., and Davis, C. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, 138, 389-414.

Chakravartty, A. (2017). Scientific realism. In E. Zalta (Ed.), *The Stanford Encyclopedia of*

Philosophy. Summer 2017. https://plato.stanford.edu/archives/sum2017/entries/scientific-realism/.

Chater, N., Goodman, N., Griffiths, T. Kemp, C., Oaksford, M., and Tenenbaum, J. (2011). The imaginary fundamentalists: The unshocking truth about Bayesian cognitive science. *Behavioral and Brain Sciences*, 34, 194-196.

Clark, A. (2015). *Surfing Uncertainty*. Oxford: Oxford University Press.

Colombo, M., Elkin, L., and Hartmann, S. (Forthcoming). Being realist about Bayes and the predictive processing theory of mind. *The British Journal for the Philosophy of Science*.

Colombo, M., and Seriès, P. (2012). Bayes on the brain --- on Bayesian modeling in neuroscience. *The British Journal for the Philosophy of Science*, 63, 697-723.

de Finetti, B. (1937/1980). Foresight. Its logical laws, its subjective sources. Rpt. in H. E. Kyburg, Jr. and H. E. Smokler (Eds.), *Studies in Subjective Probability* (pp. 94-158). Huntington: Robert E. Krieger.

Davidson, D. (1980). *Essays on Actions and Events*. Oxford: Clarendon Press.

Dennett, D. (1987). *The Intentional Stance*. Cambridge: MIT Press.

Erikkson, L., and Hájek, A. (2007). What are degrees of belief?. *Studia Logica*, 86, 183-213.

Ernst, M. (2007). Learning to integrate arbitrary signals from vision and touch. *Journal of Vision*, 7, 1-14

Fiser, J., Berkes, P., Orbán, G., and Lengyel, M. (2010). Statistically optimal perception and learning: From behavior to neural representations. *Trends in Cognitive Science*, 14, 119–130.

Fleming, R. (2011). Visual perception: Bizarre contours go against the odds. *Current Biology*, 21, R259-261.

Friston, K., and Stephan, K. (2007). Free energy and the brain. *Synthese*, 159, 417-458.

Gershman, S., Vul, E., and Tenenbaum, J. (2012). Multistability and perceptual inference. *Neural Computation*, 24, 1-24.

Glymour, C. (2011). Osiander's psychology. *Behavioral and Brain Sciences*, 34, 199-200.

Gopnik, A., Glymour, G., Sobel, D., Schulz, L., and Kushnir, T. (2004). A theory of causal learning in children: Causal maps and Bayes nets. *Psychological Review*, 111, 3-32.

Griffiths, T., Chater, N., Norris, D., and Pouget, A. (2012). How Bayesians got their beliefs (and what those beliefs really are): Comment on Bowers and Davis (2012). *Psychological Bulletin*, 138, 415-422.

Griffiths, T., Vul, E., and Sanborn, A. (2012). Bridging levels of analysis for probabilistic cognition. *Current Directions in Psychological Science*, 21, 263-268.

Helmholtz, H. von. (1867). *Handbuch der Physiologischen Optik*. Leipzig: Voss.

Herschbach, M., and Bechtel, W. (2011). Relating Bayes to cognitive mechanisms. *Behavioral and Brain Sciences*, 34, 202-203.

Hornsby, J. (2000). Personal and sub-personal: A defense of Dennett's early distinction. *Philosophical Explorations*, 1, 6-24.

Hohwy, J. (2014). *The Predictive Mind*. Oxford: Oxford University Press.

Icard, T. (2016). Subjective probability as sampling propensity. *The Review of Philosophy and Psychology*, 7, 863-903.

Jones, M., and Love, B. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contribution of Bayesian models of cognition. *Behavioral and Brain Sciences*, 34, 169-188.

Kahneman, D., and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263-291.

Knill, D., and Richards, W. (Eds.). (1996). *Perception as Bayesian Inference*. Cambridge: Cambridge University Press.

Kuhn, T. (1962). *The Structure of Scientific Revolution*. Chicago: University of Chicago Press.

Kwisthout, J., Wareham, T., and van Rooij, I. (2011). Bayesian intractability is not an ailment that approximation can cure. *Cognitive Science*, 35, 779-784.

Kwon, O.-S., Tadin, D., Knill, D. (2015). Unifying account of visual motion and position perception. *Proceedings of the National Academy of Sciences*, 112, 8142-8147.

Levy, R., Reali, F., and Griffiths, T. 2009. Modeling the effects of memory on human online sentence processing with particle filters." In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou (Eds.), *Advances in Neural Information Processing Systems* 21 (pp. 937-944). La Jolla: NIPS Foundation.

Luce, R. D., and Suppes, P. (1965). Preference, utility, and subjective probability. In R. D. Luce, R. Bush, and E. Galanter (Eds.), *Handbook of Mathematical Psychology*, vol. iii (pp. 249-410). New York: Wiley.

McDowell, J. (1994). The content of perceptual experience. *Philosophical Quarterly*, 44, 190-205.

McMullin, E. (1985). Galilean idealization. *Studies in History and Philosophy of Science Part A*, 16, 247-273.

Madl, T., Franklin, S., Chen, K., Montaldi, D., and Trappl, R. (2014). Bayesian integration of information in hippocampal place cells. *PloS One* 9, e89762.

Maloney, L., and Mamassian, P. (2009). Bayesian decision theory as a model of human visual perception: Testing Bayesian transfer. *Visual Neuroscience*, 26, 147-155.

Morales, J., Solovey, G., Maniscalco, B., Rahnev, D., de Lange, F., and Lau, H. (2015). Low attention impairs optimal incorporation of prior knowledge in perceptual decisions. *Attention, Perception, and Psychophysics*, 77, 2021-2036.

Orlandi, N. (2014). *The Innocent Eye: Why Vision is not Cognitive Process*. Oxford: Oxford University Press.

---. (2016). Bayesian perception is ecological perception. *Philosophical Topics*, 44, 327-351.

Peters, M., Ma, W. J., and Shams, L. (2016). The size-weight illusion is not anti-Bayesian after all. *PeerJ*, 4, e2124.

Petzschner, F., and Glasauer, S. (2011). Iterative Bayesian estimation as an explanation for range and regression effects: A study on human path integration. *Journal of Neuroscience*, 31, 17220-17229.

Pouget, A., Beck, J., Ma., W. J., and Latham, P. (2013). Probabilistic brains: Knowns and unknowns. *Nature Neuroscience*, 16, 1170–1178.

Ramsey, F. P. (1931). Truth and probability. In R. B. Braithwaite (Ed.), *The Foundations of Mathematics and Other Logical Essays* (pp. 156-198). London: Routledge and Kegan.

Rahnev, D., and Denisov, R. (2018). Suboptimality in perceptual decision making. *Behavioral and Brain Sciences*, 41, e223.

Rescorla, M. (2009). Cognitive maps and the language of thought. *The British Journal for the Philosophy of Science*, 60, 377-407.

---. (2012). How to integrate representation into computational modeling, and why we should. *Journal of Cognitive Science*, 13, 1-38.

---. (2015a). Bayesian perceptual psychology. In M. Matthen (Ed.), *The Oxford Handbook of the*

Philosophy of Perception (pp. 694-716). Oxford: Oxford University Press.

---. (2015b). Review of Nico Orlandi's *The Innocent Eye: Why Vision is not a Cognitive Process*. *Notre Dame Philosophy Reviews*, January 2015.

---. (2016a). Bayesian sensorimotor psychology. *Mind and Language*, 31, 3-36.

---. (2018). An interventionist approach to psychological explanation. *Synthese*, 195, 1909-1940.

---. (Forthcoming). Perceptual co-reference. *The Review of Philosophy and Psychology*.

Sanborn, A. (2017). Types of approximation for probabilistic cognition: Sampling and variational. *Brain and Cognition*, 112, 98-101.

Sanborn, A., and Chater, N. (2016). Bayesian brains without probabilities. *Trends in Cognitive Science*, 20, 883-893.

Sanborn, A., Griffiths, T., and Navarro, D. (2010). Rational approximations to rational models: Alternative algorithms for category learning." *Psychological Review*, 117, 1144-1167.

Sanborn, A., Masinghka, J., and Griffiths, T. (2013). Reconciling intuitive physics and Newtonian mechanics for colliding objects. *Psychological Review*, 120, 411-437.

Sato, Y., and Kording, K. (2014). How much to trust the senses: Likelihood learning. *Journal of Vision*, 14, 1-13.

Sato, Y., Toyoizumi, T., and Aihara, K. (2007). Bayesian inference explains perception of unity and ventriloquism aftereffect: Identification of common sources of audiovisual stimuli. *Neural Computation*, 19, 3335-3355.

Seydell, A., Knill, D., and Trommershäuser, J. (2010). Adapting internal statistical models for interpreting visual cues to depth. *Journal of Vision*, 10, 1-27.

Shen, S., and Ma., W.-J. (2016). A detailed comparison of optimality and simplicity in perceptual decision-making. *Psychological Review*, 123, 452-480.

Sotiropoulos, G., Seitz, A., Seriès, P. (2011). Changing expectations about speed alters perceived motion direction. *Current Biology*, 21, R883-R884.

---. (2014). Contrast dependency and prior expectations in human speed perception. *Vision Research*, 97, 16-23.

Stocker, A. (2018). Credo for optimality. *Behavioral and Brain Sciences*, 41, e244.

Stocker, A., and Simoncelli, E. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience*, 4, 578-585.

Thompson, P. (1982). Perceived rate of movement depends on stimulus contrast. *Vision Research*, 22, 377-380.

Thompson, P., Brooks, K., and Hammett, S. (2006). Speed can go up as well as down at low contrast: implications for models of motion perception. *Vision Research*, 46, 782-786.

Thrun, S., Burgard, W., and Fox, D. (2006). *Probabilistic Robotics*. Cambridge: MIT Press.

Todorov, E., and Jordan, M. (2002). Optimal feedback control as a theory of motor coordination. *Nature Neuroscience*, 5, 1226-1235.

Trommershäuser, J., Körding, K., and Landy, M. (Eds.). (2011). *Sensory Cue Integration*. Oxford: Oxford University Press.

Trotta, R. (2008). Bayes in the sky: Bayesian inference and model selection in cosmology. *Contemporary Physics*, 49, 71-104.

Tversky, A., and Kahneman, D. 1983. "Extension versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment." *Psychological Review* 90: pp. 293-315.

van Rooij, I., Blokpoel, M., Kwisthout, J., and Wareham, T. (2019). *Cognition and\ Intractability*. Cambridge: Cambridge University Press.

Wei, X.-X., and Stocker, A. (2015). A Bayesian observer model constrained by efficient coding can explain “anti-Bayesian” percepts. *Nature Neuroscience*, 18, 1509-1517.

Weisberg, M. (2007). Three kinds of idealization. *The Journal of Philosophy*, 104, 639-659.

Weiss, Y., Simoncelli, E., and Adelson, E. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, 5, 598-604.

Williamson, T. (1994). *Vagueness*. New York: Routledge.

Wolpert, D. (2007). Probabilistic models in human sensorimotor control. *Human Movement Science*, 26, 511-524.

Wolpert, D., and Landy, M. (2012). Motor control is decision-making. *Current Opinion in Neurobiology*, 22, 1-8.

Woodward, J. (2003). *Making Things Happen*. Oxford: Oxford University Press.

---. (2018). Explanation in neurobiology: An interventionist perspective. In D. Kaplan (Ed.), *Integrating Psychology and Neuroscience: Prospects and Problems*. Oxford: Oxford University Press.