**Bayesian Cognitive Science**

**Keywords** Bayesian brain hypothesis; Bayesian models of mental capacities; Testability of Bayesian models; Explanation and Bayesian models; Neural coding of uncertainty

**Summary** Bayesian cognitive science is a research programme that relies on modelling resources from Bayesian statistics for studying and understanding mind, brain and behaviour. Conceiving of mental capacities as computing solutions to inductive problems, Bayesian cognitive scientists develop probabilistic models of mental capacities and evaluate their adequacy based on behavioural and neural data generated by humans (or other cognitive agents) performing a pertinent task. The overarching goal is to identify the mathematical principles, algorithmic procedures and causal mechanisms, which enable cognitive agents to take uncertainty into account and weigh it appropriately in producing adaptive behaviour.

The appeal of Bayesian cognitive science derives from its transparency, normativity and unifying power. Resources from Bayesian statistics allow cognitive scientists to characterise transparently, in terms of random variables and probabilistic dependencies between them, the inductive problems that many mental capacities are presumed to solve. Given this characterisation, researchers can define an upper bound on one’s performance on that problem and use it as a benchmark for interpreting experimental results, formulating hypotheses about why and when deviations from optimal performance should be expected, and seeking explanations of mental capacities within one encompassing theoretical framework.

Although critics question the testability, explanatory power and plausibility of Bayesian models of mental capacities, Bayesian cognitive science has proved itself to be an incredibly fruitful research programme. Fuelled by empirical and theoretical results from psychology, neuroscience, philosophy, computer science, artificial intelligence and evolution, Bayesian cognitive science has advanced our understanding of why cognitive systems should keep track of uncertainty and use it to facilitate adaptive behaviour, clarified how the brain might encode information about the uncertainty of its stimuli and perform computations with probability distributions, and crystallized the insight that cognitive agents’ management of uncertainty might be grounded in predictive processes aimed at avoiding surprising exchanges with the environment.

**§1. Why Bayesian Cognitive Science**

Bayesian cognitive science belongs to a research tradition that tries to understand how cognitive agents make sense of the world in analogy with scientists’ statistical methods for data analysis, inference and decision making (Gigerenzer & Murray 1987; Hatfield 2002).

Bayesian statistical methods rely on Bayes’ theorem to compute and update probabilities in the light of new data. Bayes’ *theorem* is a mathematical statement that follows from the axioms of probability theory and the definition of conditional probability. Given two propositions *H* and *D*, Bayes’ theorem expresses the relationship between a conditional probability and its inverse in this way: *P*(*H*|*D*) = *P*(*D*|*H*) × *P*(*H*) / *P*(*D*), provided *P*(*D*) ≠ 0. Bayes’ theorem expressed in odds form is known as Bayes’ *rule*. Often confused with Bayes’ theorem, the *Bayesian rule of conditionalization* is a prescriptive norm that dictates how to reallocate probabilities given new data (or evidence), such that somebody’s new degree of belief that *H* is true after observing that *D* is the case should be the posterior probability of *H* given *D*, *viz*. *P*(*H*)new= *P*(*H*|*D*). Because Bayesian conditionalization can be viewed as an algorithmic procedure for updating beliefs, and also as a normative standard for what those beliefs *should* be, it is sometimes easy—as we will see below—to confuse normative and empirical interpretations of results in Bayesian cognitive science. But what does it mean to view the workings of the mind in analogy with Bayesian statistical methods?

It means to view the mind as storing knowledge about the causal structure of the world in the form of statistical, probabilistic models. Representing the uncertainty inherent to all its inputs probabilistically, the mind would execute the Bayesian rule of conditionalizationto continuously update those statistical models in the light of new inputs (aka “sensory data” or “sensory evidence”). Deploying statistical models of the world to produce and manage behaviour in the face of uncertainty, the mind would perform a great many cognitive functions (Colombo & Hartmann 2017), such as estimating perceptual properties of objects in the surroundings like their size, colour, temperature and smell (Simoncelli 2009), controlling action (McNamee & Wolpert 2019), making causal inferences (Gopnik et al. 2004; Körding et al. 2007), reasoning about new problems and learning how to solve them (Oaksford & Chater 2007; Perfors et al. 2011).

**§1.1 Uncertainty and the Bayesian brain hypothesis**

To survive and thrive, biological cognitive agents must effectively deal with uncertainty. Sources of uncertainty for biological cognitive agents include constant *change in their environment*, *noise*, which consists in random disturbances corrupting the quality of the information in the sensory data processed by the neural system (Faisal et al. 2008), and the *underdetermination* of perceptual states by sensory data, which means that for any stream of sensory data there are multiple causes that fit the data equally well but would generate different perceptual states—say, the two causal hypotheses “This object is convex” vs. “This object is concave” might fit the same pattern of retinal stimulations equally well at a certain time, but the perceptions of shape associated with those hypotheses are obviously different (Knill & Richards 1996).

If biological cognitive agents must grapple with uncertainty and Bayesian statistics is one—albeit not the only one (Colombo et al. 2021)—worked-out language for representing uncertainty and making inferences under uncertainty (Tenenbaum et al. 2011; Ghahramani 2015), it is appealing for cognitive scientists to use the tools and conceptual resources of Bayesian statistics to model and try to understand how biological cognitive agents deal with uncertainty.

The *Bayesian brain* *hypothesis* is the key empirical hypothesis in Bayesian cognitive science. It says that “the brain represents information probabilistically, by coding and computing with probability density functions or approximations to probability density functions” (Knill & Pouget 2004, 713). If this hypothesis is true, neurons do *not* compute single values (i.e., point estimates) of biologically relevant variables, but would instead compute multiple values of such variables along with their degree of uncertainty—for example, they would compute different values for the current ambient temperature along with their probabilities rather than one single best guess.

**§1.2 Rationality, computational analysis and optimality**

Bayesian cognitive science is appealing also because of its *normativity*. This normativity has three dimensions. First, Bayesian statistics offers a framework for modelling *norms of epistemic rationality* *for ideal agents*, where ideally rational agents’ degrees of belief (i.e., their “credences”) that certain propositions are true ought to be probabilistically coherent at any point in time and be updated given the evidence by following the rule of Bayesian conditionalization (Staffel 2022).

Justifications for the idea that conditionalization and probabilistic coherence are norms of ideal rationality can be pragmatic, epistemic or evolutionary. If your degrees of belief were not probabilistically coherent—say, you believe with 0.6 probability that a coin will land heads and believe also with 0.6 probability that the same coin will land tails—or were not updated in line with Bayesian conditionalization, then you can be “Dutch booked,” which means there is a set of bets you would consider fair, but that guarantee your monetary loss (Hájek 2008). Moreover, by complying with Bayesian conditionalization and probabilistic coherence, you can reasonably be expected to have higher degrees of belief in truths and lower degrees of belief in falsehoods (Pettigrew 2019). Finally, theoretic and empirical results indicate that it is often evolutionarily advantageous to update one’s beliefs by conditionalization (Dall et al. 2005; Okasha 2013).

A second dimension of the normativity of a Bayesian approach is associated with the notion of *rational analysis* at the *computational level* (Marr & Poggio 1977; Anderson 1991). Such analyses aim to identify the relationship between a mental capacity, the problem this capacity is presumed to solve, its inputs and outputs, and the environment (Danks 2008). If a rational analysis specifies what problem is being solved by an agent’s mental capacity, what input-output function the capacity computes and why computing that function is appropriate in a certain environment, this analysis provides researchers with a hypothesis about the *goal* of the capacity. This hypothesis could be informed and evaluated based on ethological evidence about the adaptive value for the agent of pursuing that goal in various environments. In turn, hypotheses at the computational level constrain the search for both the *algorithmic procedures* and *representational structures* underlying a mental capacity and the *causal mechanisms* implementing those procedures and structures in actual cognitive agents (Griffiths, Vul & Sanborn 2012; Zednik & Jäkel 2016).

Bayesian norms of rationality and computational-level analyses are often confused with a third normative notion, namely: *Bayesian optimality* (Rahnev & Denison 2018). A mental capacity is Bayes-optimal just in case it makes inferences that minimize the posterior expected value of some loss function, where a loss function specifies the cost of making a wrong inference. Optimality is thus not a fixed property of a Bayesian system, but one that is relative to a loss function and a model of a certain task the system tries to solve (Simoncelli 2009). While humans and other agents have been found to come close to Bayes-optimal performance in several motor and perceptual tasks (Kersten, Mamassian & Yuille 2004; Trommershäuser, Körding & Landy 2011; Rescorla 2016), this finding has improved understanding of when and why *sub*-optimal performance is to be expected, and how sub-optimal Bayesian inferences might depend on variability in the environment, computational demands of approximations to Bayesian inferences and the varying reliability of agents’ sensors (Beck et al. 2012; Acerbi et al. 2014).

**§1.3 Unification**

A Bayesian approach enjoys a great degree of *unifying power*, in the sense that Bayesian statistics provides cognitive scientists with an encompassing mathematical language for developing many different models that cast many different mental capacities as each computing solutions to inductive problems.

Bayesian cognitive science contrasts with a “bag-of-tricks” approach to cognitive science, where different mental phenomena require a particular kind of model and explanation in a case-by-case fashion (Ramachandran 1985). While a “bag-of-tricks” approach suggests, for example, that empirical data about how motion perception works require different kinds of models and explanations in different environmental conditions and for different types of motion, a Bayesian approach allows researchers to use the same kind of model of motion perception for explaining that humans and other cognitive agents have an a priori preference for slower velocities and that visual perception integrates streams of sensory data (i.e., sensory inputs) with this slow-speed prior according to their relative reliability (i.e., degree of uncertainty). Using a Bayesian model with a single free parameter, Weiss et al. (2002) could fit a wide range of empirical data about motion perception in different situations, make predictions about visual illusions and offer a unified account in terms of slow-speed prior and sensory inputs with varying reliability, which would explain, for example, why automobile drivers tend to speed up in the fog.

This is not to say that there are no significant differences between different Bayesian models of different mental capacities. But despite these differences, all Bayesian models represent their targets in terms of random variables and probabilistic dependencies between them, which change as new data are observed in accordance with Bayesian conditioning (or some suitable approximation). Although this common mathematical structure might be taken to provide us with “the foundation for universal laws of cognition—principles that we expect to hold true for intelligent organisms of any kind, anywhere in the universe” (Griffiths et al. 2012, 418), it is contentious that adequate explanations of cognitive phenomena should appeal to any “universal law” and that Bayesian unification reveals one kind of causal mechanism underlying many different mental capacities (Danks 2014; Colombo & Hartmann 2017; Kaplan & Hewitson 2021).

**§2. Testing Bayesian models in cognitive science**

Empirical tests of Bayesian models of mental capacities generally involve a comparison between experimental participants’ behaviour and/or neural activity in a task and the simulated behaviour of different models of that task and/or variation in the values of specific variables in the models. When a Bayesian model simulates individual participants’ behaviour accurately, showing similar patterns of errors, or when variation in the variables of the model meaningfully correlates with variation in neural activity, this constitutes relevant evidence for evaluating the empirical adequacy of the model compared to alternative models of the same task. Although this is standard scientific practice, some researchers criticize Bayesian cognitive science for telling us untestable “just-so stories” (Jones & Love 2011; Bowers & Davis 2012) and relying on models that are so flexible to accommodate any data (Glymour 2007; Marcus & Davis 2013).

**§2.1 Bayesian just-so stories?**

The criticism that Bayesian models provide us with untestable “just-so stories” about how the mind works assumes that all Bayesian models of the mind are aimed at offering rational analyses at the computational level. Making this assumption, critics claim that Bayesian modelling is unconstrained by relevant empirical knowledge about psychological processes, neural mechanisms, evolution and the structure of the environment. They conclude that Bayesian modelling delivers only speculative stories, devoid of empirical content, about the supposedly rational workings of the mind (Jones & Love 2011; Bowers & Davis 2012).

The problem with this criticism is that not all Bayesian models are aimed at providing rational analyses and not all Bayesian models are formulated as just rational analyses at the computational level (Hahn 2014). Many Bayesian models are presented as empirical hypotheses about the workings of specific mental capacities and are not meant to underwrite any claim about their rationality or optimal functioning. Such models posit algorithmic procedures and representations, which cognitive systems might use to solve problems they face. The idealizations and empirical assumptions of these models are informed by modellers’ goals as well as by available data from psychology, neuroscience or evolution (Griffiths et al. 2012; Tauber et al. 2017). So, actual Bayesian modelling is often constrained by pertinent empirical and theoretical considerations and does not boil down to untestable storytelling disguised in mathematics, which one can use to explain any behaviour as optimal or rational. The “just-so-story” criticism fails to do justice to actual aims and modelling practices in Bayesian cognitive science (Chater et al. 2011).

**§2.2 Overly flexible and underdetermined?**

The criticism that Bayesian models are difficult to test can be motivated by the observation that Bayesian modelling allows for probabilistic representations with diverse shape and structure, and algorithms with different degrees of complexity, domain-specificity and energetic costs (Griffiths et al. 2010). Given this flexibility, critics notice that Bayesian cognitive scientists do not have an overarching theory, which may guide them to develop non-arbitrary models of mental capacities, and then conclude that, with a suitable choice of prior probabilities, likelihoods and loss function, a Bayesian modeller can fit virtually any dataset, turning the Bayesian approach to cognition into a vacuous exercise in post-hoc modelling of observed data (Glymour 2007; Marcus & Davis 2013).

One problem with this criticism is that the choice of prior probabilities, likelihoods and loss functions in a Bayesian model of a given capacity is often dependant on the task being modelled, where experimental participants are assumed to rely on a certain amount of information or data—for example, visual and tactile information about the size of an object—to produce one among a set of possible responses—for example, a button press about whether the perceived size of an object is greater than a given measure (for one prominent example, see Ernst & Banks 2002). Like in any other experiment, Bayesian cognitive scientists have often control over the data available to their experimental participants, the degree of uncertainty of such data—for example, the degree of blur of the visual information about the size of an object—the set of possible responses, and the consequences of a certain response in the task—for example, whether participants receive some reward/punishment for their responses. So, the choice of priors, likelihoods and loss function in a Bayesian model of a given capacity is often justified by the researchers’ goals and the experimental task used to pursue such goals.

Another problem with the “overly-flexible” criticism is that Bayesian cognitive scientists are sometimes interested in estimating the prior probabilities humans (or other cognitive systems) might be using in a given task—for example, to find out whether the visual system in humans has actually a “slow-speed” prior—and priors estimated in one task often generalize to other tasks (see Sotiropoulos & Seriès 2015 for examples). This means that the choice of priors in Bayesian models of mental capacities is often based on previous findings in a non-arbitrary way.

Another criticism motivated by the flexibility of Bayesian modelling is that non-Bayesian models can often be constructed for the same task that share some or all of the empirical implications of a given Bayesian model (Bowers & Davis 2012). No matter how impressive a Bayesian model is in fitting a dataset, this success should not favour the Bayesian model over any alternative model that could also successfully fit the same dataset.

The problem with this criticism is that it is unclear Bayesian modelling suffers from distinctive and widespread problems of underdetermination. That is, it is unclear that, for many mental capacities, there are equally well-confirmed, genuinely alternative, non-Bayesian models. Even if, say, a lookup table model, which maps inputs to outputs to approximate some function in a task, could be construed to approximate Bayesian performance in a task, that does not necessarily mean that the lookup table model shares all the empirical implications of any Bayesian model in that task or across different tasks (Maloney & Mamassian 2009).

While some Bayesian models are insufficiently specified for experimental evaluation (Eberhardt & Danks 2011), this issue highlights that good modelling practice, in Bayesian cognitive science like in any other modelling approach to mind, should involve sound methodologies for experimental design, model fitting, model comparison, sensitivity and robustness analysis, and so on. Bayesian cognitive science does not suffer from widespread, distinctive problems of confirmation.

**§3. What we can learn about minds and brains with Bayes**

If Bayesian models of mental capacities *are* testable and *do* enjoy some degree of empirical success, what do we learn from them? Does successful Bayesian modelling of mental capacities support the Bayesian brain hypothesis? Does it show that the brain *is* Bayesian? To address these questions, we should clarify the computational limits of the algorithmic processes posited by Bayesian models and the sort of evidence relevant to evaluate the Bayesian brain hypothesis.

**§3.1 Intractable Bayes**

Bayesian modelling of mental capacities is typically successful in highly controlled, simple, psychophysical tasks, where participants generally adjust their behaviour systematically, as a function of trial-to-trial manipulations of the reliability of sensory cues (Trommershäuser et al. 2011). Not all results in such tasks are in line with the predictions of Bayesian models (Rahnev & Denison 2018), but the main challenge to the Bayesian approach is a large body of experimental findings about judgement and decision making, suggesting that the mind—particularly when it comes to “high-level” cognitive capacities—hardly ever works in a Bayesian fashion (Gilovich et al. 2002; Mandelbaum 2019; Williams 2019).

While researchers disagree about the correct interpretation of these apparent deviations from Bayesian rationality (Gigerenzer 1991; Kahneman & Tversky 1996; Samuels, Stich & Bishop 2002; Elqayam & Evans 2011), these findings, at least, highlight that the algorithmic procedures posited by Bayesian models are often too demanding in terms of time and energy requirements to be implemented by resource-bounded minds like ours (Binmore 2007; Brighton & Gigerenzer 2008; Kwisthout & Van Rooij 2013). The worry is that Bayesian procedures do not tractably scale up to the sort of complex problems cognitive agents encounter in their environment. If Bayesian procedures cannot tractably find solutions for real-world problems, then, even acknowledging the successes of Bayesian modelling in psychophysical tasks in constrained laboratory conditions, we cannot conclude these successes warrant the conclusion that the mind is Bayesian.

One reply to this tractability challenge is that the mind might perform computations that *approximate* exact Bayesian inference. Models positing such approximations are sometimes empirically successful and would also be computationally tractable (Icard 2008; Gershman, Horvitz, & Tenenbaum 2015; Lieder & Griffiths 2020). Two problems with this reply, however, are that it is unclear in what sense such approximations work in a Bayesian way (Griffiths et al. 2012), and, more substantially, that approximate “Bayesian” algorithms are often also computationally too demanding and may not scale up to tractably address complex real-world problems (Kwisthout, Wareham, & van Rooij 2011). It would be too quick to conclude that the mind is Bayesian from experimentally successful, approximate “Bayesian” algorithms.

A different reply to the tractability challenge emphasises that biological cognitive agents are environmentally embedded. In accounting for how their mental capacities work, more explanatory weight should be placed on the environment compared to algorithmic procedures and probabilistic representations for (approximate) Bayesian inferences. If the wiring of the visual system reflects, or is attuned to, the statistics of the environment, then cognitive agents’ perceptual systems do not need to *represent* those statistics or make inferences by implementing algorithms that manipulate probabilistic representations (Orlandi 2014; Bruineberg & Rietveld 2014), which would undercut the tractability challenge, since being attuned to the structure of our eco-niches enable successful interaction with the environment without any computational cost.

While identifying the statistical properties of the environment can surely advance understanding of the design features of evolved perceptual systems as well as constrain particular Bayesian models of perception (Girshick, Landy & Simoncelli 2011), one problem with this reply is that it can in fact be *mal*-adaptive for cognitive agents if their perceptual systems reflect environmental regularities (or are “ecologically tuned”), since they would risk overfitting such regularities, making successful interaction with the environment *less* likely (Feldman 2013; Brighton & Gigerenzer 2008).

**§3.2 Where’s Bayes in the brain?**

As we saw, much of the empirical evidence in Bayesian cognitive science comes from behavioural and psychophysical experimental tasks. While this type of evidence falls short of warranting the conclusion that the brain is Bayesian (Danks 2008; Colombo & Seriès 2012), it has stimulated the development of artificial neural network models for implementing Bayesian inference (Deneve 2008), as well as the search for electrophysiological evidence relevant to evaluate the Bayesian brain hypothesis (Doya et al. 2006; Pouget et al. 2013; Ma & Jazayeri 2014).

Recall that this is the key empirical hypothesis behind Bayesian cognitive science. This hypothesis does *not* say that perceptual experiences can feel uncertain or that people can assign some degree of uncertainty that their judgements based on perceptual evidence are true (Denison 2017). The hypothesis does *not* say that neural activity is stochastic or noisy either (Griffiths et al. 2012). And the hypothesis does *not* say that the brain, as a whole, is a Bayesian mechanism. Rather, the commitment of the Bayesian brain hypothesis is that neurons take account of the uncertainty of their inputs, and thus their activity does not compute only one “best estimate” (i.e., point values) of biologically relevant variables. Instead, neurons perform computations that are sensitive to the degree of uncertainty of their inputs, giving more weight to more reliable (less uncertain) inputs.

The core insight of the Bayesian brain hypothesis is then that the nervous system is fundamentally adapted to manage environmental variability, underdetermination of perceptual states by sensory input (or ambiguity in the sensory input) and internal noise; and this management requires some means of representing uncertainty neurally and using representations of uncertainty in the neural computations underlying mental capacities.

There are several accounts that formalize this core insight in an empirically evaluable way. One proposal is the *probabilistic population code* account, which says that different functions of neural activity encode the *parameters* defining the posterior probability distributions of task-relevant variables (Ma et al. 2006). A different proposal is the *neural* *sampling code* account, which says that neural activity encodes *samples* from the posterior distributions of unobserved variables in generative models of the sensory inputs (Fiser et al. 2010).

These two proposals are consistent with data concerning neural activity (Fetsch et al. 2012; Berkes et al. 2011), but there is substantial debate regarding which is the best candidate for how the nervous system takes account of the uncertainty of different variables, whether there must be a single neural code for uncertainty that can generalize to all mental capacities and whether we are warranted to believe that any empirically successful Bayesian model is true even though we do not know how the brain assigns probabilities to variables and performs probabilistic inferences (Rescorla 2020). Clever experimental paradigms, more sophisticated tools for recording and manipulating neural activity and a better understanding of the nature and scope of the achievements of Bayesian cognitive science will help us address these issues.

**§4. Conclusions**

Bayesian cognitive science is an incredibly fruitful and rich research programme. If anything, it has crystallized the insight that cognitive systems’ management of uncertainty might be grounded in predictive processes aimed at avoiding surprising exchanges with the environment (Clark 2013; Hohwy 2013; Friston 2019). This entry has concentrated on its motivations, purposes, explanatory power, testability and plausibility, but there are several other philosophically interesting questions raised by Bayesian cognitive science, issues about the rationality of perception (Siegel 2016), the cognitive penetrability of perception (Lupyan 2015), representationalism (Gładziejewski 2016), the modularity of the mind (Drayson 2017), nativism (Perfors 2012; Colombo 2018), conscious experience (Clark et al. 2019; Nave 2021), and more.

**References**

Acerbi, L., Vijayakumar, S., & Wolpert, D. M. (2014). On the origins of suboptimality in human probabilistic inference. *PLoS Computational Biology*, 10(6), e1003661.

**Investigates whether sub-optimal probabilistic inference depends on the complexity of priors and provides evidence suboptimality might be due to learning complex priors rather than computing with them**

Anderson, J. R. (1991). Is human cognition adaptive? *Behavioral and Brain Sciences, 14*, 471-484.

**Describes a methodology called “rational analysis” in the context of Bayesian modelling of cognition**

Beck, J. M., Ma, W. J., Pitkow, X., Latham, P. E., & Pouget, A. (2012). Not noisy, just wrong: the role of suboptimal inference in behavioral variability. *Neuron*, *74*(1), 30-39.

**Argues that sub-optimal probabilistic inference rather than internal noise explains behavioural variability**

Berkes, P., Orbán, G., Lengyel, M., & Fiser, J. (2011). Spontaneous cortical activity reveals hallmarks of an optimal internal model of the environment. *Science*, *331*(6013), 83-87.

**Develops a Bayesian model using sampling-based representations of uncertainty and tests this model analysing visual cortical activity of ferrets during development**

Binmore, K. (2007). Rational decisions in large worlds. *Annales d'Economie et de Statistique*, 25-41.

**Argues that we need to look beyond Bayesian decision theory for understanding how we make rational decisions under uncertainty in a “large world”**

Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin, 138,* 389–414. doi:10.1037/a0026450

**Criticizes Bayesian cognitive science as offering hard-to-test, “just-so stories” about mind and behaviour**

Brighton H., & Gigerenzer, G. (2008). Bayesian brains and cognitive mechanisms: harmony or dissonance? In N. Chater & M. Oaksford (Eds.), *The Probabilistic Mind: Prospects for Bayesian Cognitive Science*. New York: Oxford University Press, 189–208.

**Critically examines the “metaphor” of the probabilistic mind, arguing that a probabilistic mind would suffer from fundamental problems of tractability**

Bruineberg, J., & Rietveld, E. (2014). Self-organization, free energy minimization, and optimal grip on a field of affordances. *Frontiers in human neuroscience*, *8*, 599.

**Provides an account of perception in the context of the “free energy principle” in terms of (dis)attunement between environmental dynamics and internal dynamics, rather than in terms of probabilistic inference**

Chater, N., Goodman, N. D., Griffiths, T. L., Kemp, C., Oaksford, M., & Tenenbaum, J. B. (2011). The imaginary fundamentalists: The unshocking truth about Bayesian cognitive science. *Behavioral and Brain Sciences, 34,* 194–196. doi:10.1017/S0140525X11000239

**Replies to Jones & Love (2011) clarifying that Bayesian cognitive scientists have different aims and are interested in models at multiple levels of analysis**

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181-204.

**Synthesises many results and ideas from the cognitive sciences to formulate and support a Predictive Processing account of mind and brain**

Clark, A., Friston, K. J., & Wilkinson, S. (2019). Bayesing qualia: Consciousness as inference, not raw datum*. Journal of Consciousness Studies*,26(9-10), 19–33.

**Clarifies the relevance of Bayesian accounts of the mind for addressing questions about conscious experience**

Colombo, M. (2018). Bayesian cognitive science, predictive brains, and the nativism debate. *Synthese*, *195*(11), 4817-4838.

**Clarifies the relevance of Bayesian accounts of the mind for the controversy between nativists and empiricists**

Colombo, M., Elkin, L., & Hartmann, S. (2021). Being Realist about Bayes, and the Predictive Processing Theory of Mind. *The British Journal for Philosophy of Science*. 72, 185-220.

**Argues that a positive epistemic attitude toward the content of our best Bayesian models is unwarranted**

Colombo, M., & Hartmann, S. (2017). Bayesian Cognitive Science, Unification, and Explanation. *The British Journal for Philosophy of Science*, 68, 451-484.

**Argues that Bayesian modelling unifies cognitive science by showing that different mental capacities share the same mathematical structure rather than one underlying causal mechanism in the brain**

Colombo, M., & Seriès, P. (2012). Bayes in the Brain. On Bayesian Modelling in Neuroscience. *The British Journal for Philosophy of Science*, 63, 697-723.

**Argues that the success of Bayesian modelling in psychophysics does not entail that the brain is Bayesian**

Dall, S. R., Giraldeau, L. A., Olsson, O., McNamara, J. M., & Stephens, D. W. (2005). Information and its use by animals in evolutionary ecology. *Trends in ecology & evolution*, *20*(4), 187-193.

**Offers a framework based on statistical decision theory for analysing animal information use in evolutionary ecology**

Danks, D. (2014). *Unifying the mind: Cognitive representations as graphical models*. Cambridge, MA: The MIT Press.

**Argues that various sets of empirical data about causal cognition and categorization can be explained in a unified fashion by a cognitive architecture grounded in Bayesian graphical models**

Danks, D. (2008). Rational analyses, instrumentalism, and implementations. In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Prospects for Bayesian cognitive science*. Oxford: Oxford University Press, 59–75.

**Clarifies the methodology of “rational analysis” and its explanatory power**

Deneve, S. (2008). Bayesian spiking neurons I: inference. *Neural computation*, *20*(1), 91-117.

**Develops an artificial neural network model, where the dynamics of spiking neurons can be interpreted as Bayesian inference in time**

Denison, R. N. (2017). Precision, Not Confidence, Describes the Uncertainty of Perceptual Experience: Comment on John Morrison's “Perceptual Confidence”. *Analytic Philosophy*, *58*(1), 58-70.

**Distinguishes the notions of “confidence” and “precision” in the context of probabilistic modelling of perception and argues that perceptual uncertainty should be understood in terms of perceptual precision**

Doya, K., Ishii, S., Pouget, A., & Rao, R. P. N. (Eds.). (2006). *Bayesian Brain: Probabilistic Approaches to Neural Coding*. MIT Press

**Explores in the light of theoretical and empirical results how Bayesian concepts can be used for studying and understanding the neural coding of uncertain information, as well as for interpreting neurobiological data**

Drayson, Z. (2017). Modularity and the predictive mind. In T. Metzinger & W. Wiese (Eds.), *Philosophy and predictive processing*. Frankfurt am Main, Germany: MIND Group. Retrieved from https://predictive-mind.net/papers/modularity-and-the-predictive-mind

**Clarifies the relevance of Bayesian accounts of the mind for the debate about the modularity of the human cognitive architecture**

Eberhardt, F., & Danks, D. (2011). Confirmation in the cognitive sciences: The problematic case of Bayesian models. *Minds and Machines*, *21*(3), 389-410.

**Argues that key problems in the testability of Bayesian models depend on insufficient model specification, and illustrates this point with the example of ‘‘probability matching’’ used as confirmation**

Elqayam, S., & Evans, J. S. B. (2011). Subtracting “ought” from “is”: Descriptivism versus normativism in the study of human thinking. *Behavioral and Brain Sciences*, *34*(5), 233-290.

**Criticizes the idea that human thinking reflects normative principles like Bayesian norms**

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429-433.

**A prominent example of an experiment showing that humans can integrate different sources of information in a statistically optimal way**

Faisal, A. A., Selen, L. P. J., & Wolpert, D. M. (2008). Noise in the Nervous System. *Nature Reviews Neuroscience*, 9, pp. 292-303.

**Explains various sources of uncertainty for biological agents and suggests how the nervous systems might rely on Bayesian principles to counter detrimental effects of noise and uncertainty**

Feldman, J. (2013). Tuning your priors to the world. *Topics in cognitive science*, *5*(1), 13-34.

**Argues that tuning one’s priors to environmental statistics is often maladaptive due to overfitting**

Fetsch, C. R., Pouget, A., DeAngelis, G. C., & Angelaki, D. E. (2012). Neural correlates of reliability-based cue weighting during multisensory integration. *Nature neuroscience*, *15*(1), 146-154.

**Investigates the neural basis of sensory cue integration in monkeys performing a heading discrimination task and shows how their experimental findings are consistent with probabilistic population coding of uncertainty**

Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. *Trends in cognitive sciences*, *14*(3), 119-130.

**Formulates and motivates a sampling code for how neural systems might represent uncertainty**

Friston, K. (2019). Beyond the desert landscape. In M. Colombo, E. Irvine, & M. Stapleton (Eds.), *Andy Clark and his critics.* Oxford University Press, 174–190.

**Formulates and motivates from first principles in statistical physics a free energy principle for life and mind**

Gershman, S. J., Horvitz, E. J., & Tenenbaum, J. B. (2015). Computational rationality: A converging paradigm for intelligence in brains, minds, and machines. *Science*, *349*(6245), 273-278.

**Proposes a notion of computational rationality for bounded agents, which could promote cross-disciplinary exchange on questions, tools and results**

Ghahramani, Z. (2015). Probabilistic Machine Learning and Artificial Intelligence. *Nature*, *521*(7553), 452–459.

**Explains why and how the language of probability constitutes a powerful modelling framework for understanding what learning is in the context of machine learning, artificial intelligence and robotics**

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond “heuristics and biases”. *European review of social psychology*, *2*(1), 83-115.

**Argues that apparent “errors” in probabilistic reasoning need not be interpreted as signs of “irrationality”**

Gigerenzer, G., & Murray, D. J. (1987). *Cognition as intuitive statistics*.Hillsdale, NJ: Erlbaum.

**A historically informed account of how methods for statistical inference can guide theory construction and discovery in psychology**

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge university press.

**Reviews psychological evidence in the “heuristics & biases” programme indicating that judgement and decision making under uncertainty does not rest on Bayesian processes, but on simple heuristics**

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature neuroscience*, *14*(7), 926-932.

**Provides empirical evidence that statistical models for orientation that human observers might use for perceptual estimation of the orientation of objects match relevant environmental statistics**

Gładziejewski, P. (2016). Predictive coding and representationalism. *Synthese*, *193*(2), 559-582.

**Focuses on predictive coding accounts of how neural systems might implement uncertain information and argues for a representationalist interpretation of such accounts**

Glymour, C. (2007). Bayesian Ptolemaic psychology. In W. Harper & G. Wheeler (Eds.), *Probability and inference: Essays in Honor of Henry E. Kyburg, Jr*., London: Kings College Publishers, 123–41.

**Argues that Bayesian psychology does not predict or explain observed data, but merely accommodates them**

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A Theory of Causal Learning in Children: Causal Maps and Bayes Nets. *Psychological Review, 111*(1), 3–32.

**Offers an empirically informed, testable computational theory of children’s causal learning in terms of Bayesian networks**

Griffiths, T. L., Chater, N., Norris, D. & Pouget, A. (2012). How the Bayesians got their beliefs (and what those beliefs actually are): comment on Bowers and Davis (2012). *Psychol. Bull.* 138, 415–22.

**Replies to Bowers & Davis’s (2012) “just-so-story” criticism clarifying the motivations behind Bayesian modelling and its unique merits**

Griffiths, T. L., Vul, E., & Sanborn, A. N. (2012). Bridging levels of analysis for probabilistic models of cognition. *Psychological Science*, *21*, 263–268.

**Surveys how Bayesian models can span different levels of analysis and suggests various Monte Carlo algorithms to link Bayesian models to both psychological and neural processes**

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in cognitive sciences*, *14*, 357-64.

**Contrasts “top-down” probabilistic modelling of cognition with “bottom-up” connectionist models and emphasises the greater degree of flexibility and transparency of probabilistic modelling compared to connectionism**

Hahn, U. (2014). The Bayesian boom: good thing or bad?. *Frontiers in psychology*, *5*, 765.

**Emphasises the diversity of Bayesian modelling in cognitive science and explains how different Bayesian models aim to advance different scientific goals**

Hájek, A. (2008) Arguments for − or against − probabilism? *British Journal for the Philosophy of Science* 59:793–819.

**Surveys “Dutch book arguments” for the idea that ideally rational agents ought to comply with Bayesian norms**

Hatfield, G. (2002). Perception as unconscious inference. In D. Heyer & R. Mausfeld (Eds.), *Perception and the physical world: Psychological and philosophical issues in perception*. John Wiley & Sons, Ltd., New York, pp. 115-143.

**Offers a historically informed philosophical review of the idea of perception as unconscious inference**

Hohwy, J. (2013). *The predictive mind*. Oxford University Press.

**Develops an account of the mind as a hypothesis-testing mechanism that aims to minimize the error of its predictions about the sensory inputs it receives from the world**

Icard, T. F. (2018). Bayes, bounds, and rational analysis. *Philosophy of Science*, *85*(1), 79-101.

**Explains the computational limits of probabilistic inference and proposes a notion of rationality appropriate for computationally bounded Bayesian agents**

Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences, 34,* 169–188. doi:10.1017/S0140525X10003134

**Criticizes Bayesian cognitive science as aiming merely at displaying behaviour as optimal or rational**

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions: A reply to Gigerenzer’s critique. *Psychological Review*, 103, 582-591.

**Responds to Gigerenzer (1991) explaining how human probabilistic judgements are prone to large and systematic biases**

Kaplan, D. M., & Hewitson, C. L. (2021). Modelling Bayesian computation in the brain: unification, explanation, and constraints. In F. Calzavarini & M. Viola (Eds.). *Neural Mechanisms. Studies in Brain and Mind*, vol 17. Springer, Cham, 11-33.

**Replies to Colombo & Hartmann (2017) arguing that we need a more refined account of Bayesian unification and its relevance for mechanistic explanation**

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, *55*, 271-304.

**Reviews relevant evidence in visual perception and explains why Bayesian modelling helps us understand how vision handles complexity, ambiguity and underdetermination in object perception**

Knill, D. C., & Richards, W. (Eds.). (1996). *Perception as Bayesian inference*. Cambridge University Press.

**Reviews theoretical motivations and empirical applications of the idea of perception as Bayesian inference**

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712-719.

**States and motivates the Bayesian brain hypothesis**

Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., & Shams, L. (2007). Causal inference in multisensory perception. *PLoS one*, *2*(9), e943.

**Develops and evaluates various Bayesian and non-Bayesian models of multisensory cue combination to study causal inference in perception**

Kwisthout, J., & van Rooij, I. (2013). Bridging the gap between theory and practice of approximate Bayesian inference. *Cognitive Systems Research*, 24, 2-8.

**Clarifies three notions of “approximation” in the context of tractability challenges to Bayesian cognitive science**

Kwisthout, J., Wareham, T., & van Rooij, I. (2011). Bayesian intractability is not an ailment that approximation can cure. *Cognitive Science*, *35*(5), 779-784.

**Shows that claims of “tractable approximability of intractable (Bayesian) models” of cognition are unjustified, and showcases complexity-theoretic tools to assess the intractability of Bayesian models**

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*.

**Develops a modelling framework for integrating rational principles of probabilistic inference with realistic cognitive constraints and shows how some examples of irrational behaviour can be explained within this framework**

Lupyan, G. (2015). Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems. *Review of philosophy and psychology*, *6*(4), 547-569.

**Clarifies the relevance of Bayesian accounts of the mind for the debate about the cognitive penetrability of perception**

Ma, W. J., & Jazayeri, M. (2014). Neural coding of uncertainty and probability. *Annual review of neuroscience*, *37*, 205-220.

**Reviews empirical results concerning the ways neural systems might represent uncertainty and probability**

Ma, W. J., Beck, J. M., Latham, P. E., & Pouget, A. (2006). Bayesian inference with probabilistic population codes. *Nature Neuroscience*, *9*(11), 1432–1438.

**Formulates and motivates a probabilistic population code for how neural systems might represent uncertainty**

Maloney, L. T., & Mamassian, P. (2009). Bayesian decision theory as a model of human visual perception: Testing Bayesian transfer. *Visual neuroscience*, *26*(1), 147-155.

**Discusses potential problems of underdetermination for Bayesian models of perception and suggests experimental criteria termed “transfer criteria” to address these problems**

Mandelbaum, E. (2019). Troubles with Bayesianism: An introduction to the psychological immune system. *Mind & Language*, *34*(2), 141-157.

**Reviews evidence from belief acquisition and change indicating that the mind is not (approximately) Bayesian, but that it aims at maintaining a “psychological immune system”**

Marcus, G. F., & Davis, E. (2013). How robust are probabilistic models of higher-level cognition? *Psychological Science*, 24, 2351–2360.

**Criticizes Bayesian cognitive science as a futile exercise in post-hoc modelling**

Marr, D. & Poggio, T. (1977). From Understanding Computation to Understanding Neural Circuitry. *Neurosciences Research Progress Bulletin*, 15, 470–88.

**Introduces distinct levels of analysis for understanding neural computation**

McNamee, D., & Wolpert, D. M. (2019). Internal models in biological control. *Annual review of control, robotics, and autonomous systems*, *2*, 339-364.

**Reviews empirical evidence and theoretical results indicating that animals’ ability to construct and use internal probabilistic models of the motor system and environmental dynamics facilitates adaptive behaviour**

Nave, K. (2021). Visual experience in the predictive brain is univocal, but indeterminate. *Phenomenology and the Cognitive Sciences*, 1-25.

**Explores and clarifies the connection between the idea of experience being probabilistic and the notion of phenomenological indeterminacy identified by Edmund Husserl**

Oaksford, M. & Chater, N. (2007). *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*, Oxford: Oxford University Press.

**Argues that the right approach to the psychology of reasoning is probability theory rather than logic, because rationality should be defined in terms of the capacity to reason about uncertainty**

Okasha, S. (2013). The evolution of Bayesian updating. *Philosophy of Science*, *80*(5), 745-757.

**Provides an evolutionary argument for Bayesian conditionalization**

Orlandi, N. (2014). *The Innocent Eye: Why Visio is not a Cognitive Process*. Oxford, NY: Oxford

University Press.

**Develops the “Embedded View” theory of vision grounded in the crucial roles that the environment plays in perception**

Perfors, A. (2012). Bayesian models of cognition: What's built in after all? *Philosophy Compass*, *7*(2), 127-138.

**Focusing on the notion of hypothesis space, explains how hierarchical Bayesian models can account for puzzles concerning concept acquisition**

Perfors, A., Tenenbaum, J. B., Griffiths, T. L., & Xu, F. (2011). A tutorial introduction to Bayesian models of cognitive development. *Cognition*, *120*(3), 302-321.

**An introduction to Bayesian approaches to cognitive development during childhood**

Pettigrew, R. (2019). Epistemic Utility Arguments for Probabilism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*, URL = <https://plato.stanford.edu/archives/win2019/entries/epistemic-utility/>.

**Surveys “epistemic utility arguments” for the idea that ideally rational agents ought to have probabilistically coherent beliefs and comply with Bayesian norms**

Pouget, A., Beck, J. M., Ma, W. J., & Latham, P. E. (2013). Probabilistic brains: knowns and unknowns. *Nature neuroscience*, *16*(9), 1170-1178.

**Reviews evidence and challenges for various accounts of how the brain might implement probabilistic representations and probabilistic inferences**

Rahnev, D., & Denison, R. N. (2018). Suboptimality in perceptual decision making. *Behavioral and Brain Sciences*, 41, Article e223, 1-66.

**Reviews evidence of sub-optimal performance in perceptual tasks, distinguishes methodological and explanatory uses of the notion of “Bayesian optimality” and discusses hypotheses about the origins of suboptimal perception**

Ramachandran, V. S. (1985). The neurobiology of perception. *Perception,* *14*, 97–103.

**Distinguishes three theories of perception and suggests that perception may best be understood as a “bag of tricks”**

Rescorla, M. (2020). A Realist Perspective on Bayesian Cognitive Science. In A. Nes and T. Chan (eds.), *Inference and Consciousness*. New York: Routledge, pp. 40-73.

**Defends the view that the success of a Bayesian model gives us reason to believe the model is approximately true**

Rescorla, M. (2016). Bayesian sensorimotor psychology. *Mind & Language*, *31*(1), 3-36.

**Defends a realist view toward the content of our best Bayesian models in sensorimotor psychology**

Samuels R, Stich S, & Bishop M (2002). Ending the rationality wars: how to make disputes about human rationality disappear. In Elio R. (Ed.), *Common sense, reasoning, and rationality*. Oxford: Oxford University Press, pp. 236–268.

**Argues that disagreements about the nature and scope of human rationality in the light of empirical results in the psychology of judgement and decision making have been greatly exaggerated**

Siegel, S. (2016). *The rationality of perception*. Oxford University Press.

**Argues that perceptual experiences themselves and the processes that produce them can be rational or irrational**

Simoncelli, E. P. (2009). Optimal Estimation in Sensory Systems. In M. S. Gazzaniga (Ed*.*), *The Cognitive Neurosciences*, Volume 4, Cambridge, MA: MIT Press, pp. 525–35.

**Surveys how perceptual systems might perform optimal estimation**

Sotiropoulos, G., & Seriès, P. (2015). Probabilistic inference and Bayesian priors in visual perception. In G. Cristobal, L. Perrinet, & M. S. Keil (Eds.), *Biologically inspired computer vision: Fundamentals and applications* (pp. 203–220). New York: Wiley.

**Reviews studies in human psychophysics focusing on the role of Bayesian priors in visual perception**

Staffel, J. (2022). Bayesian Norms and Non-Ideal Agents. In M.-A. Lasonen & C. M. Littlejohn (Eds.), *The Routledge Handbook of the Philosophy of Evidence*. New York: Routledge.

**Reviews various Bayesian norms of rationality and examines how they might apply to boundedly rational agents**

Tauber, S., Navarro, D. J., Perfors, A., & Steyvers, M. (2017). Bayesian models of cognition revisited: Setting optimality aside and letting data drive psychological theory. *Psychological Review, 124*(4), 410–441.

**Distinguishes normative and descriptive uses of Bayesian modelling and illustrates them with specific case studies**

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279-1285.

**Describes how the modelling language of statistical inference and structured probabilistic representations can shed light on the nature and origins of human thought**

Trommershäuser, J., Körding, K., & Landy, M. S. (Eds.). (2011). *Sensory cue integration*. Oxford University Press.

**Surveys Bayesian models, experimental paradigms and empirical evidence concerning sensory cue integration within and across sensory modalities**

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature neuroscience*, *5*(6), 598-604.

**Develops a Bayesian model of visual motion perception that can account for a wide variety of psychophysical phenomena**

Williams, D. (2021). Epistemic Irrationality in the Bayesian Brain. *The British Journal for the Philosophy of Science*, 72(4), 913-938.

**Puts into focus four “sources of epistemic irrationality” in the context of experimental results in tension with predictions from Bayesian models of cognition**

Zednik, C., & Jäkel, F. (2016). Bayesian reverse-engineering considered as a research strategy for cognitive science. *Synthese*, *193*(12), 3951-3985.

**Explains various strategies Bayesian cognitive scientists use to “reverse engineer” the mind and bridge computational, algorithmic and implementation levels of analysis**