# School of Informatics


## Informatics Research Review
## Evaluating the Bayesian Brain Framework in Cognitive Science

[redacted]

**January 2021**

**Abstract**

This literature review addresses the utility of the emerging Bayesian Brain framework to our understanding of the functionality of the brain in Cognitive Science. The paper explores the foundations and emergence of the Bayesian hypothesis, reviews studies supporting its value at the computational level of Marr's hierarchy of analysis, and reviews studies challenging its value, particularly at the algorithmic and implementational levels. The paper concludes by identifying future areas of work required to advance the Bayesian Brain Cognitive Science and Computational Neuroscience fields.

Date: Friday 29th January, 2021

**Supervisor:** [redacted]

# 1 Introduction

One of the major goals of computational neuroscience and cognitive science is to provide a unified framework for the overall functioning of the brain which can guide progress in our mechanistic understanding of the brain. David Marr's perspective [1] on levels of analysis for brain function is relevant to achieving this. Marr states that there are hierarchical levels of analysis for brain function with the highest being the computational level (i.e. what the brain is trying to do), the middle being the algorithmic level (how the system achieves the computational goal) and the lowest being implementation (how is it physically realized). He argues that understanding the brain requires this hierarchical approach and should begin at the computational level. This approach is described as top-down in contrast to the bottom-up approach often taken by neurobiologists. At the computational level many recent conceptions of the brain suggest that the brain's primary role is to generate appropriate actions through prediction and thereby minimizing the uncertainty that is inherent in the environment within which the brain is situated [2]. To do so, it is postulated that the brain uses generative models of its environment. Generative models in Cognitive Science are considered to be causal relationships that map probabilistic dependencies in the outside world [3]. This perspective lends itself well to a Bayesian framework where the brain's function at a computational level is that of a Bayesian predictive system which operates in order to reduce uncertainty. However, for the framework to be successful in guiding the development of cognitive science and computational neuroscience towards a true understanding of the function of neural systems and the mind, the framework must not simply exist in the computational realm but also demonstrate a capacity to further the research efforts at the algorithmic and implementation levels of analysis. This paper attempts to address the utility of the Bayesian Brain framework to cognitive science and computational neuroscience through the lens of Marr's levels of analysis by assessing the contribution of Bayesian theory to each level. The paper will first provide a summary of the Bayes theorem and the emergence of the Bayesian Brain theory. This is followed by a review of papers that attempt to provide a unified Bayesian brain framework at the computational level by focusing on papers positing Bayesian models from sensory processing to motor actions and higher cognitive capacity. This is followed by a critique of the utility of the Bayesian framework in developing a holistic understanding of the brain by reviewing papers that identify constraints of the utility of the models at the algorithmic and the implementation levels. This is followed by an overview of papers which focus on incorporating and addressing these criticisms and constraints and therefore provide support for the continued utility of the Bayesian Framework in providing insight into the understanding of the brain across all levels of analysis. The paper ends by identifying some unanswered problems in the field and exploring how the theory may be of further use going forward.

## 1.1 Background

Bayesian inference is a method of statistical inference using Bayes Theorem which is a mathematical theorem named after Thomas Bayes that is used to compute conditional probabilities [4, 5]. The Bayes equation is:

$$P(\mathbf{H}|\mathbf{E}) = \frac{P(\mathbf{E}|\mathbf{H})P(\mathbf{H})}{P(\mathbf{E})} \tag{1}$$

In Bayesian inference one updates the probability of a hypothesis as more information becomes available. The Bayesian approach to probability is distinct in that, instead of relying on frequency or propensity of some outcome, probability is interpreted as a reasonable expectation

which represents a state of knowledge [5]. In Bayesian inference one has a prior $P(\mathbf{H})$ and it is this prior which is updated given new evidence by multiplying the prior by the likelihood, which is the probability of the evidence given the hypothesis $P(\mathbf{E}|\mathbf{H})$ to form a posterior belief $P(\mathbf{H}|\mathbf{E})$. Therefore, under a Bayesian Brain framework, the brain is thought to be updating a posterior probability distribution for a given hypothesis by updating existing prior probabilities of the hypothesis given the evidence, and normalizing that over the sum (integral) of the hypotheses in the hypothesis space. Bayesian inference is optimal in the sense that it maximizes the probability of generating the correct hypothesis given the hypothesis space, observed evidence and prior probabilities.

There is a long history of viewing the brain as extracting sensory information from the world in a probabilistic manner in order to develop an internal model of the outside world. The first to suggest this was Hermann Helmotz in the 1860s [6]. Herman argued that the perceptual system has direct access to only those of our senses which are incapable of fully capturing the physical world. Therefore, the brain must compute unconscious inference which determines the most likely physical causes of the sensory stimulation [7]. As indicated above, Bayesian Inference provides an optimal way of computing the unconscious inference and generating an internal model in the form of a posterior probability through the combination of the existing internal model of the world and the incoming sensory information. Hence the Bayesian brain hypothesis represents a mathematically well-defined normative framework which captures the ideas of unconscious inference and is optimal given the constraints of uncertainty necessarily generated by imperfect senses.

While the above is the definition of Bayesian Inference in a mathematical sense. In this paper I will often follow the readily used definition of Bayesian in Cognitive Science literature where the term Bayesian acts as a placeholder for a set of interrelated problem solving mechanisms which are unified by the use of 1) uncertainty which measures degrees of belief, 2) degrees of belief ought to satisfy the axioms of probability and, 3) degrees of belief represented by deterministic probabilities ought to be updated in light of new information, typically through the use of conditionalization. [8]

The general methodology pursued at the computational level by Bayesian Cognitive Scientists is to 1) Use Bayesian theory to develop a model of a task based on how an idealized Bayesian system would execute a task. The model would contain free parameters dependent on the task in question .2) Fit the model's free parameters to maximize how well the model captures the data, and 3) compare the model's performance with that of humans on the specified task [7].

While the computational and algorithmic levels of analysis are agreed upon by Bayesian Cognitive Scientists, there is a divergence of opinion regarding whether the Bayesian Brain framework should be viewed through an instrumentalist or a realist lens. An instrumentalist view sees scientific models as useful devices to predict observable outcomes concerning a given system. Realists, on the other hand, consider good models to be those which pick out component entities and activities in the system being explored [7, 9]. These different perspectives have implications for how the computational level findings should guide research at the algorithmic and implementation levels [1] and will therefore play a role in the critical analysis of the theory and its utility going forward.

# 2 Literature Review

## 2.1 Bayesian Brain evidence at Marr's Algorithmic level

### 2.1.1 Support for the Bayesian Brain in sensory tasks

Perhaps the strongest support for the Bayesian Brain theory lies in the realm of perception. Under a Bayesian framework, sensory perception operates within a hypothesis space of causes of the sensory input received. Each cause is a potential hypothesis in the set. The prior probability is therefore the initial belief in the occurrence of a given sensory hypothesis $P(\mathbf{h})$. Upon receiving sensory input or evidence h, the areas of the brain related to sensory processing and perception compute the posterior probability $P(\mathbf{h}|\mathbf{e}) = \mathbf{n}P(\mathbf{e}|\mathbf{h})P(\mathbf{h})$ where$n$ is the normalizing constant. Our perception is then determined by the posterior distribution over sensory hypothesis. There have been numerous studies demonstrating the success of modelling sensory perception as Bayesian inference. One of the key strengths of the Bayesian models of sensory processing lies in their ability to explain sensory illusions which often elude accounts of sensory processing which posit that the brain directly represents the surrounding physical world [10]. These models crucially rely on the existence of a prior p(h) being utilized by the brain which systematically weights sensory evidence so as to skew the posterior probability of sensory hypotheses. One study which demonstrated this was the paper entitled "motion illusions as optimal percepts" [11] which argues that the existing theories of how the brain integrates motion of visual stimuli such as Intersection of constraints (IOC), Vector Average (VA) or Feature Tracking (FT), "lack predictive power as each appears to be beneficial in certain restricted domains and lack the capacity to explain a number of well-known motion illusions" [11]. However, by implementing a Bayesian estimator model with an explicit prior which favours slow movements in the environment this simple prior led to the model reproducing the motion illusion effect whereby humans perceive object motions to be slower in low contrast environments [11]. By capturing human specific perceptual oddities, the model suggests that humans also rely on sensory priors which systematically bias perception and lead to sensory illusions. A second paper [12] argues for the brain being appropriately modelled as Bayesian based on the fact that, under the assumptions of a Bayesian prior and independent Gaussian noise, humans performed a visual-haptic discrimination task similarly to a maximum likelihood estimate given by:

$$\hat{S} = \sum_i w_i \hat{S}_i \tag{2}$$

$$w_i = \frac{\frac{1}{\sigma_i^2}}{\sum_j \frac{1}{\sigma_j^2}} \tag{3}$$

Where is $\hat{S}$ is the sensory input from a given sense modality (i.e. vision or touch)

The optimal estimate (estimate with least cumulative variance) is given by a sum of the sensors weighted by their normalized reciprocal variance. In the experiment, subjects were first asked to determine height differences between visual and haptic stimuli independently, as well as using both visual and haptic senses together, and the threshold for discerning the difference was measured. Noise was then added such that it reduced the reliability of input from a given sensory modality and the changes in threshold matched the ideal maximum likelihood estimate predicted by the model by systematically moving towards the discrimination threshold of the less noisy sense proportional to the noise (variance) provided to the other sensory input [12]. Finally, there is evidence to support the brain's use of Bayesian strategies in sensorimotor learning, therefore

extending the framework from a purely sensory model of brain function to the motor domain. Traditionally motor operations have been characterized as input-output relationships which do not take into account the probabilistic nature of either the task or the sensory input [13]. However, in this paper, users were asked to reach for a target while receiving visual feedback halfway through the movement. The visual feedback was distorted (i.e laterally displaced) by a distance defined by a Gaussian and bimodal distribution. Consistent with the brain using a Bayesian strategy, the brain appears to systematically represent the prior distribution and uncertainty of visual feedback and combine them in a manner consistent with performance-optimizing Bayesian processes. On each movement, the lateral shift was randomly drawn from a Gaussian prior distribution with a mean displacement of 1cm to the right and a standard deviation of 0.5cm. is referred to as the true prior. [13] The visual feedback was provided briefly midway through the movement and the feedback was systematically distorted. Participants consistently showed an increased reliance on the prior distribution as sensory feedback became less reliable.

### 2.1.2 Support for the Bayesian Brain in Cognitive Tasks

The evidence supporting the Bayesian Brain hypothesis at the computational level is most compelling for sensory and sensorimotor processes. However, more recently there has been increasing use of Bayesian models which account for cognitive processes such as concept learning. Humans have a remarkable ability to learn new concepts which they can acquire often with just one example and utilize in rich ways, such as correctly generalizing to new novel items, indicating that they have developed a presentation of the concept boundaries [14]. The paper entitled "Human-level concept learning through probabilistic program induction" demonstrates that these remarkable capabilities can begin to be captured under a Bayesian framework. To do so, the authors introduced Bayesian Program learning (BPL). This program is a generative model which can sample new types of concepts by combining sub parts in new ways. Each sub part is itself a generative model over primitives which creates a hierarchy of generative models. The program captures the human ability to 'learn to learn' [15] and thereby creates rich concepts by combining simpler primitives through developing hierarchical priors, which means that previous experience with related concepts facilitates learning of new concepts [14]. The model and humans were shown a single character of an alphabet and asked to select another example of the character from a set of distinct characters drawn by a typical drawer. The model had an error rate of 3.3% while humans had a similar error rate of 4.5%. Both had significantly lower error rates than deep convolutional networks with an error rate of 13%. This suggests that human concept learning can be well characterized using optimal Bayesian models. This lends credence to a unified view of the brain's computational level function (i.e the 'what' level of Marr's analysis) being to perform Bayesian inference to update and utilize priors in tandem with sensory evidence, and that the brain can therefore be thought of as being Bayes optimal, given the constraints of neuronal noise and suboptimal sensory input [16].

## 2.2 Critiques of the Bayesian Brain Theory

To explore the major critiques of the Bayesian Brain hypothesis at the computational level, and to relate the criticisms to the question of the utility of the theory to Cognitive Science developments at the 'implementation' and 'algorithmic' levels, the paper "Bayesian Just-So Stories in Psychology and Neuroscience" [17]. will be reviewed. The major points in this paper are that, under a theory of evolution by natural selection, animals and systems have evolved to be

'good enough' or 'better than', as there is no selective pressure for any progress beyond this, and therefore any claims of optimality within the constraints of neuronal noise and imperfect senses flies in the face of these established evolutionary constraints. This therefore suggests that non-Bayesian models are better suited to characterize the function of the brain. However, the major specific critique in the paper refers to the flexibility in prior choice available to cognitive Bayesian modelers. In reference to the paper discussed above, "motion illusions as optimal percepts" the authors correctly cite that the use of the prior in the Bayesian model was not grounded in any empirical data [11] and, due to the simplicity of the update rule in Bayesian inference, the prior distribution has significant influence over the model. As noted in the background section Bayesian models are configured to best match the data. Therefore, Bowers claims that, without an empirical basis to set the prior Bayesian, models can become so flexible as to essentially be meaningless. From a realist perspective of the Bayesian brain, the ad-hoc nature of the priors means that there is little evidence to motivate an exploration of direct implementation of Bayesian components at the algorithmic or implementation levels. Even from an instrumentalist perspective, despite not needing to be directly instantiated at the algorithmic or implementation level, the loss of predictive power of the model, due to effectively over-fitting the data, means that there need not be any compelling reason to believe that Bayesian perspectives should place constraints on the lower level which require that the theories lead to outcomes that can be modelled by a Bayesian system. The paper entitled "Bayes in the Brain—On Bayesian Modelling in Neuroscience" [9] which promotes an instrumentalist view of the Bayesian Brain theory acknowledges these issues and states that "in general, if the aim of Bayesian modelling is to acquire knowledge about underlying mechanisms of perception, then the criterion for choosing the prior should include some 'ecological' consideration since neural processing is influenced by the statistical properties of the environment." The Bayesian conception also faces the critique that, despite being portrayed as Bayes optimal based on modelling experiments from sensory tasks to cognitive tasks, humans display alarming systematic probability reasoning errors [18] such as the conjunction fallacy which occurs when humans rate the conjunction of two items as being more probable than the individual constituents alone [19].

Direct instantiation of Bayesian Brain hypothesis at the algorithmic and implementation levels faces further critiques. It has been proven that, in the complex hypothesis space within which the brain exists, marginalization, which requires integration over the entire hypothesis space, is computationally intractable [20, 21]. Given that the brain has finite memory and computational power, it is therefore impossible that the brain implements precise Bayesian inference. While instrumentalist accounts of the Bayesian Brain do not necessitate that the form of the implementation is captured in the algorithmic model, nevertheless, as argued by Colomb and Series [9], transition from an instrumentalist perspective of Bayesian models to a realist perspective, requires that three issues need to be addressed. These are: "(i) How might neurons represent uncertainty? (ii) How might they represent probability distributions? (iii) How might they implement different approximations to Bayesian inference?" This means that, for the Bayesian brain hypothesis to live up to its potential to provide a unified framework of the function of the brain, and for Bayesian modelling experiments to have utility in uncovering these general principles, there must be computationally tractable and neuronally implementable mechanisms for Bayesian computations. In the following section, the paper will explore the leading theories for plausible implementation of the algorithmic level Bayesian models.

## 2.3 The Bayesian framework at Implementation and Algorithmic Levels

To provide a unified framework of brain function across Marr's levels of analysis [1], and to prove useful in constraining and defining further neurobiological and computational neuroscience research at the implementation level, there must be feasible explanations for how Bayesian models can be implemented in the brain, given the fundamental computational constraints to precise Bayesian inference described above. This is an active area of current research which will ultimately determine the true utility of the Bayesian Brain hypothesis, given the goals of computational neuroscience and cognitive science to address and understand brain function across Marr's three levels of analysis. To review the current status of the problem, this section will be split into three parts each dealing with a major approach the field has taken to address the plausibility of implementation. The two major approaches are Predictive Coding and Bayesian Approximation. As stated above there necessarily also must be evidence for neurons encoding probabilities and therefore current research in that area is also addressed

### 2.3.1 Predictive Coding

Predictive coding, often called predictive processing, has been heralded as facilitating neuronally plausible and tractable Bayesian inference [22, 23]. It assumes that the brain continuously tries to predict its sensory inputs based on a hierarchical structure of hypotheses of the world. In this conception, the prediction is considered to be 'top down' whereby higher order hypotheses make predictions about the hypotheses below, and eventually make predictions about sensory input [24]. The contention is that, at each level in the hierarchy, the level updates its hypotheses about the world when the top-down prediction mismatches the bottom-up information. The efficiency of the model derives from the fact that, if predictions from higher levels to lower levels are correct, there is no need for spikes to be transmitted and therefore efficiency is improved [25]. This model is demonstrated in retinal ganglion cells which take a weighted mean of the signals in neighbouring cells to predict the current light intensity at the center of the target cell. The cell then transmits the difference between the predicted light intensity and the measured light intensity minimizing the range of outputs transmitted by the centre and increasing efficiency [25, 26]. This is therefore clearly a biologically plausible mechanism which could be implemented in the brain. Further it can be shown that, under the assumption of Gaussian distributions, Bayesian sequential updating of beliefs can be computed using the following equation where $\mu$ is mean and $\pi$ is precision

$$\mu_{\text{posterior}} = \mu_{\text{prior}} + \frac{\pi_{\text{likelihood}}}{\pi_{\text{posterior}}}(x - \mu_{\text{prior}}) \tag{4}$$

$$\pi_{\text{posterior}} = \pi_{\text{prior}} + \pi_{\text{liklihood}} \tag{5}$$

Where the last term in equation (4) $x - \mu_{\text{prior}}$ is interpreted as a prediction error and where $\mu_{(\text{prior})}$ is the prediction of what the new measurement x will be. Hence Bayesian inference can be implemented by iteratively adjusting predictions with the prediction error that is produced by each experiment. The learning rate represented by $\frac{\pi_{\text{likelihood}}}{\pi_{\text{posterior}}}$ captures the Bayesian logic of weighting the update and evidence more heavily when prior knowledge is low, and relying on the prior more heavily when either the likelihood is low or the prior is large. Importantly it can also be demonstrated that this update expression can be shown to be applicable beyond the univariate gaussian case [27].

An example of the explanatory power of the predictive coding algorithm is its explanation of binocular rivalry in vision [28]. When presented with a stimulus of a house to one eye and a face to the other, we are being presented with a stimulus with low prior probability. Hence either a house or a face is predicted and we experience switching perceptions of the house and face: we do not see a blended version. However, the actual stimulus does not match what was predicted and therefore a prediction error is triggered and again, taking prior probability and prediction error into account, the percept will switch to the opposite stimulus [28].

### 2.3.2 Bayesian Approximation

A second common approach to arguing how Bayesian models could be implemented at the neuronal level, and simultaneously to respond to the critique that people often make systematic errors in their probability reasoning, is to appeal to approximation. The intractability of Bayesian inference in computer science has led to a number of approximation algorithms being developed. The two most popular of these are arguably variational and sampling approximations [7]. This review focuses on sampling algorithms, the most common of which are the Markov Chain Monte Carlo Methods. In the paper "Bayesian Brain without Probabilities" [18], the author argues that the brain should be thought of as a Bayesian Sampler which, rather than computing Bayesian inferences, instead samples randomly from the distribution and therefore approximates the posterior distribution. In the limit, this produces optimal inference, but this is self-evidently impossible, and therefore the systematic errors in probabilistic judgements made by people are an expected by-product of the sampler missing peaks of the probability distribution and provides credence to an account of the brain utilizing a sampling algorithm. A second paper arguing for the utility of sampling as approximation of Bayesian inference situates the argument in a resource rational framework (optimizing decisions given limited resources) by arguing that there is a trade-off between deliberation time to make more optimal decisions and the cost of spending more time and energy deliberating [29]. The authors model the trade-off using a sampling agent in a binary choice and find that, assuming that the unknown distribution is gaussian, it takes very few samples to get within a close approximation of the actual distribution. Therefore, it is often advantageous to make decisions from relatively few samples, so saving time and energy. The authors then extend this analysis to choices of number 4, 8, 16, and 32 and show that, despite error rates increasing as the number of choices increases, and assuming that sampling is costly in time and energy, the ideal trade-off is still relatively few samples. One issue with this formulation is that the choice of cost for sampling is arbitrary and it is therefore difficult to derive ecologically relevant conclusions despite a compelling argument for the virtue of using less samples in constrained computational settings.

## 2.4 Critiques of Algorithmic and Implementation level approaches

This section will attempt to challenge the veracity of the potential approaches taken to posit plausible accounts for how the Bayesian Framework at the computational level can be translated into the algorithmic and implementation levels. While predictive coding is commonly argued to provide a computationally tractable and neuronally plausible account of how the brain could implement Bayesian Inference, this view has been questioned by the paper entitled "Computational Resource Demands of a Predictive Bayesian Brain" [24] The paper argues that, despite the belief that processing only the prediction error significantly increases the tractability of inference, complexity analyses of the sub-computations involved in predictive processing utilizing the generative models on structured representation of the type argued to be required for

cognitive processes based on Bayesian modelling [30], indicate that the majority are intractable, due to the sub-components themselves being NP-hard. This sharply contrasts with the claims made previously that predictive processing is in fact tractable. However, all of these previous claims are based on simplifications of the representations of the model, for example assumptions of independence of variables or gaussian probability distributions [24]. Yet many of these simplification restrictions eliminate the capacity to form the rich, structured representations which are argued to be key in Bayesian models of cognition and hence are often utilized by Bayesian modellers [30]. This would therefore negate the capacity of predictive coding as it presently stands to provide an algorithmic level and neuronally implementable account of the increasing success of Bayesian models of higher order cognitive tasks such as concept learning [14].

The second major critique focuses on the sampling and approximation algorithmic level views. This aims at dispelling the belief that simple appeals to approximations can solve intractability and therefore provide credence to the realist view of the Bayesian brain. While Bayesian inference approximation algorithms have been developed and often work quite well, these algorithms are restricted to being tractable only on a restricted subset of input domains [31] which cannot account for the generality of the models proposed by Bayesian modelers of higher cognitive processes at the computational level [30, 31]. Therefore, before appeals to approximations can justify realist accounts of the Bayesian Brain hypothesis and provide a route towards the Bayesian Brain providing an account of brain function across Marr's levels of analysis, work must be done in understanding how, and if, the brain has ecologically or psychologically defined constraints on connectivity and input domains to facilitate tractable approximations.

Finally, it is worth noting that the Bayesian Brain hypothesis in large part began due to probabilistic inference representing an optimal mechanism for operating under uncertainty. However, if tractability issues lead Bayesian Brian proponents to ever greater appeals to approximations, there is no guarantee that these approximations remain in any sense more optimal than other strategies for operating under uncertainty, as tractable approximations can in theory be arbitrarily less accurate than what would be expected of the ideal model [8, 32]. As a result, this in turn undercuts the computational level justification of the Bayesian brain hypothesis as providing an optimal solution to operating under uncertainty as in fact the brain is not capable of operating at that level of optimality.

## 3 Limitations and Future Directions

The studies discussed above provide insight into the utility of the Bayesian Brain framework for furthering the fields of Cognitive Science and Computational Neuroscience in understanding of the brain across Marr's three level of analysis. While this paper approaches the topic by constraining itself to an analysis utilizing Marr's approach, the Bayesian Brain field appears unclear as regards where the Bayesian Brain hypothesis stands in the hierarchy. This paper has taken the view that the Bayesian Brain represents a computational level concept of what the brain is attempting to achieve, and this position is endorsed by a number of researchers [27, 33]. However, the literature also defines the theory as functional at the algorithmic level [9]. This illustrates the need for the field to more effectively and consistently utilize Marr's levels of analysis if concerted progress is to be made. Furthermore, the field lacks genuine integration across the levels of analysis with computational level approaches often disregarding the need to ground the parameters of their models in empirical evidence from neurobiology. This makes it difficult for papers exploring the Bayesian Brain hypothesis at one level of Marr's analysis to develop connections and testable predictions based on work performed at another level. There

is therefore often no relationship between research aimed at Marr's computational level, where Bayesian modellers have enjoyed success, and research at the implementation level which has, by comparison, struggled to make significant progress. This disconnection is most obvious in the complexity of the models developed for higher cognitive capacities at the computational level with seeming disregard for the paucity of credible implementation accounts to match the models. Finally, the field seems to struggle with lack of clarity as regards an instrumentalist vs realist account of the Bayesian Brain theory. This contrast is striking with prominent advocates supporting both realist accounts [7, 16] and instrumentalist accounts [9]. However, which perspective is taken drastically affects the approach to current research as well as future directions of the field. A realist perspective creates a potentially limiting dogma where researchers at the lower levels of Marr's hierarchy take the Bayesian Inference account literally and therefore seek to find direct mechanistic accounts of the proposed computational level model in a form similar to that of the model. However, an instrumentalist account creates no more than a predictive framework for the outcomes (behaviour) of the algorithmic and implementation levels but remains ambiguous as regards how this effect could come about at the lower levels, potentially increasing the freedom for researchers at these levels to make meaningful headway unconstrained by dogma. It is the contention of this paper that, as it currently stands, there is little concrete evidence to support a realist account of the Bayesian Brain theory and therefore little evidence that it provides a unifying account of brain function across Marr's three levels of analysis. This is not to say that the Bayesian Framework is not useful at a computational level and that it cannot help to guide research towards a unified understanding of brain function across Marr's hierarchy, but rather that, as the field stands, there is little definitive evidence to support a direct instantiation of the mechanistic processes involved in Bayesian Brain accounts at the neuronal level.

Future works in the Bayesian Brain cognitive science field should endeavour to bridge the gap between Marr's levels of analysis. Bayesian computational modelling must ground itself in mechanistic empirical neuronal realities as far as possible so that the models produced may lead to tractable and accurate mechanistic and algorithmic explanations for those models derived at the computational level. However, it is important to appreciate that there is considerable utility in the computational level models as they currently stand, particularly in the area of computational psychiatry which is a relatively new field. Bayesian models of cognition can account for many psychiatric disorders through altered prior weights and belief updating. Examples of this include computational models of schizophrenia and autism [34, 35]. A particularly compelling research direction is therefore to explore how models of behaviour and sensory integration of healthy subjects can be systematically used to observe divergent responses in patients, and therefore provide a quantitative and systematic approach to diagnosis and classification.

# 4 Summary & Conclusion

The Bayesian Brain theory in Cognitive Science was developed in response to the observation of the complexity of the environment in which the brain operates and the inability of the senses to convey sufficient information to accurately describe and represent that complexity. The conclusion was that the brain must be performing inference to reduce this uncertainty. Bayesian Inference represents an optimal mechanism through which an ideal agent can reason probabilistically to reduce its uncertainty. The Bayesian Brain theory therefore argues that the brain operates to reduce its uncertainty by utilizing Bayesian Inference. This approach has been gaining in popularity in Cognitive Science and Computational Neuroscience as a framework through which the brain's functionality can be explained [22]. This review has

attempted to analyze the utility of a Bayesian Brain framework to the goal in Cognitive Science and Computational Neurosciences of describing the function and operation of the brain across Marr's three hierarchy levels, computational, algorithmic and implementation.

To achieve this goal, the paper first critically analyzed the theory's contributions at the computational level of Marr's hierarchy. Studies reviewed demonstrated that Bayesian modelling could replicate visual illusions [11], sensory integration [12], sensorimotor functions [13] and higher cognitive processes in humans [14] suggesting that humans are nearly Bayes optimal across a number of domains. However, these views of the utility of the Bayesian framework have been challenged by other studies which argue that these findings were skewed by ad-hoc priors in the Bayesian models [17], and emphasis that that the models are not useful at the algorithmic or implementation level as precise Bayesian Inference is computationally intractable [20, 21, 31]. The review then critically analyzed papers which posited ways in which the Bayesian Framework could nevertheless be made compatible with algorithmic and implementation level accounts of brain function [18, 22, 29]. The review concludes that future work in the Bayesian Brain cognitive science field should endeavour to bridge the gap between Marr's levels of analysis, since Bayesian computational modelling must be grounded in mechanistic neuronal realities if a holistic understanding of brain functionality is to be advanced.

# References


[1] Marr, D. n.d. "Vision (1982) A Computational Investigation into the Human Representation and Processing of Visual Information." WH Freeman, New York.

[2] Friston, Karl. 2010. "The Free-Energy Principle: A Unified Brain Theory?" Nature Reviews. Neuroscience 11 (2): 127–38.

[3] Love, Bradley C., Michael Ramscar, Thomas L. Griffiths, and Matt Jones. 2015. "Generative and Discriminative Models in Cognitive Science." In CogSci. https://cogsci.mindmodeling.org/2015/papers/0013/paper0013.pdf.

[4] Bayes, Thomas. 1763. "LII. An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, FRS Communicated by Mr. Price, in a Letter to John Canton, AMFR S." Philosophical Transactions of the Royal Society of London, no. 53: 370–418.

[5] MacKay, David J. C., and David J. Mac. 2003. Information Theory, Inference and Learning Algorithms. Cambridge University Press.

[6] Patton, Lydia. 2018. "Hermann von Helmholtz." In The Stanford Encyclopedia of Philosophy, edited by Edward N. Zalta, Winter 2018. Metaphysics Research Lab, Stanford University. https://plato.stanford.edu/archives/win2018/entries/hermann-helmholtz/.

[7] Rescorla, Michael. 2019. "A Realist Perspective on Bayesian Cognitive Science." In Inference and Consciousness, 40–73. Routledge.

[8] Colombo, Matteo, Lee Elkin, and Stephan Hartmann. 2020. "Being Realist about Bayes, and the Predictive Processing Theory of Mind." The British Journal for the Philosophy of Science, November, 000–000.

[9] Colombo, Matteo, and Peggy Seriès. 2012. "Bayes in the Brain—On Bayesian Modelling in Neuroscience." The British Journal for the Philosophy of Science 63 (3): 697–723.

[10] Gibson, James J. 2002. "A Theory of Direct Visual Perception." Vision and Mind: Selected Readings in the Philosophy of Perception, 77–90.

[11] Weiss, Yair, Eero P. Simoncelli, and Edward H. Adelson. 2002. "Motion Illusions as Optimal Percepts." Nature Neuroscience 5 (6): 598–604.

[12] Ernst, Marc O., and Martin S. Banks. 2002. "Humans Integrate Visual and Haptic Information in a Statistically Optimal Fashion." Nature 415 (6870): 429–33.

[13] Körding, Konrad P., and Daniel M. Wolpert. 2004. "Bayesian Integration in Sensorimotor Learning." Nature 427 (6971): 244–47.

[14] Lake, Brenden M., Ruslan Salakhutdinov, and Joshua B. Tenenbaum. 2015. "Human-Level Concept Learning through Probabilistic Program Induction." Science 350 (6266): 1332–38.

[15] Braun, Daniel A., Carsten Mehring, and Daniel M. Wolpert. 2010. "Structure Learning in Action." Behavioural Brain Research 206 (2): 157–65.

[16] Knill, David C., and Alexandre Pouget. 2004. "The Bayesian Brain: The Role of Uncertainty in Neural Coding and Computation." Trends in Neurosciences 27 (12): 712–19.

[17] Bowers, Jeffrey S., and Colin J. Davis. 2012. "Bayesian Just-so Stories in Psychology and Neuroscience." Psychological Bulletin 138 (3): 389–414.

[18] Sanborn, Adam N., and Nick Chater. 2016. "Bayesian Brains without Probabilities." Trends in Cognitive Sciences 20 (12): 883–93.

[19] Tentori, Katya. 2020. "What Can the Conjunction Fallacy Tell Us about Human Reasoning?" https://doi.org/10.31234/osf.io/yrh5f.

[20] Cooper, Gregory F. 1990. "The Computational Complexity of Probabilistic Inference Using Bayesian Belief Networks." Artificial Intelligence 42 (2): 393–405.

[21] Shimony, Solomon Eyal. 1994. “Finding MAPs for Belief Networks Is NP-Hard.” Artificial Intelligence 68 (2): 399–410.

[22] Clark, Andy. 2013. “Whatever next? Predictive Brains, Situated Agents, and the Future of Cognitive Science.” The Behavioral and Brain Sciences 36 (3): 181–204.

[23] Blokpoel, Mark, Johan Kwisthout, and Iris van Rooij. 2012. “When Can Predictive Brains Be Truly Bayesian?” Frontiers in Psychology 3 (November): 406.

[24] Kwisthout, Johan, and Iris van Rooij. 2020. “Computational Resource Demands of a Predictive Bayesian Brain.” Computational Brain  Behavior 3 (2): 174–88.

[25] Aitchison, Laurence, and Máté Lengyel. 2017. “With or without You: Predictive Coding and Bayesian Inference in the Brain.” Current Opinion in Neurobiology 46 (October): 219–27.

[26] Srinivasan, M. V., S. B. Laughlin, and A. Dubs. 1982. “Predictive Coding: A Fresh View of Inhibition in the Retina.” Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character. Royal Society 216 (1205): 427–59.

[27] The MIT Press. n.d. “Computational Psychiatry.” Accessed January 29, 2021. https://mitpress.mit.edu/books/computational-psychiatry-1.

[28] Hohwy, Jakob, Andreas Roepstorff, and Karl Friston. 2008. “Predictive Coding Explains Binocular Rivalry: An Epistemological Review.” Cognition 108 (3): 687–701.

[29] Vul, Edward, Noah Goodman, Thomas L. Griffiths, and Joshua B. Tenenbaum. 2014. “One and Done? Optimal Decisions from Very Few Samples.” Cognitive Science 38 (4): 599–637.

[30] Tenenbaum, Joshua B., Charles Kemp, Thomas L. Griffiths, and Noah D. Goodman. 2011. “How to Grow a Mind: Statistics, Structure, and Abstraction.” Science 331 (6022): 1279–85.

[31] Kwisthout, Johan, Todd Wareham, and Iris van Rooij. 2011. “Bayesian Intractability Is Not an Ailment That Approximation Can Cure.” Cognitive Science 35 (5): 779–84.

[32] Icard, Thomas F. 2018. “Bayes, Bounds, and Rational Analysis.” Philosophy of Science 85 (1): 79–101.

[33] Harkness, Dominic L., and Ashima L. Keshava. 2017. “Moving from the What to the How and Where – Bayesian Models and Predictive Processing.” https://doi.org/10.15502/9783958573178.

[34] Valton, Vincent, Liana Romaniuk, J. Douglas Steele, Stephen Lawrie, and Peggy Seriès. 2017. “Comprehensive Review: Computational Modelling of Schizophrenia.” Neuroscience and Biobehavioral Reviews 83 (December): 631–46.

[35] Sevgi, Meltem, Andreea O. Diaconescu, Lara Henco, Marc Tittgemeyer, and Leonhard Schilbach. 2020. “Social Bayes: Using Bayesian Modeling to Study Autistic Trait-Related Differences in Social Cognition.” Biological Psychiatry 87 (2): 185–93.