mark.sprevak@ed.ac.uk

Predictive coding I: Introduction

Mark Sprevak
University of Edinburgh

19 June 2023

Predictive coding – sometimes also known as 'predictive processing', 'free energy minimisation', or 'prediction error minimisation' – is an exciting novel account of human cognition. It claims to offer a complete, unified theory of cognition that stretches all the way from cellular biology to phenomenology. However, the exact content of the view, and how it might achieve these ambitions, is not clear. This series of articles examines predictive coding and attempts to identify its key commitments and justification. The present article begins by focusing on possible confounds with predictive coding: claims that are often identified with predictive coding, but which are not predictive coding. These include the idea that the brain employs an efficient scheme for encoding its incoming sensory signals; that perceptual experience is shaped by prior beliefs; that cognition involves minimisation of prediction error; that the brain is a probabilistic inference engine; and that the brain learns and employs a generative model of the world. These ideas have garnered widespread support in modern cognitive neuroscience, but it is important not to conflate them with predictive coding.

1 Introduction

Predictive coding is a computational model of cognition. Like other computational models, it attempts to explain human thought and behaviour in terms of computations performed by the brain. It differs from more traditional approaches in at least three respects. First, it aspires to be *comprehensive*: it aims to explain, not just one domain of human cognition, but all of it – perception, motor control, decision making, planning, reasoning, attention, and so on. Second, it aims to *unify:* rather than explain cognition in terms of many different kinds of computation, it explains by appeal to a single, unified computation – one computational task and one computational algorithm are claimed to underlie all aspects of cognition. Third,

it aims to be *complete*: it offers not just part of the story about cognition, but one that stretches all the way from the details of neuromodulator release to abstract principles of rational action governing whole agents.[1]

However, understanding precisely what predictive coding says, and whether it can achieve these ambitions, is not straightforward. For one thing, the term 'predictive coding' means different things to different people.[2] For another, important features of the view, whatever its name, are liable to change or are underspecified in important respects. In this article and those that follow it, my aim is to sketch what predictive coding is, and how it might fulfil these ambitions.

I argue that predictive coding should be understood as a loose alliance of three claims. These claims, each of which may be precisified or qualified in variety of ways, are made at Marr's *computational*, *algorithmic*, and *implementation* levels of description.[3] At Marr's computational level, the claim is that the computational *task* facing the brain is to minimise sensory prediction error. At the algorithmic level, the claim is that the *algorithm* by which our brain attempts to solve this task involves the action of a hierarchical network of abstract prediction and error units. This network may be viewed, in a further step, as running a variational algorithm for approximate Bayesian inference. At Marr's implementation level, the claim is that the *physical resources* that implement the algorithm are primarily located in the neocortex: anatomically distinct cell populations inside neocortical areas implement distinct prediction and error units.

Each of these claims needs to be qualified in certain respects and supplemented by further details. Each needs to be stated more precisely and ideally associated with a quantitative mathematical formalisation. A path needs to be forged from the claims to supporting empirical evidence. Finally, one needs to show that the resultant model delivers the kinds of benefits originally promised – a comprehensive, unifying, and complete account of cognition. Different researchers within the predictive coding community have different opinions about how to do all this, and many details are currently left open. This means that the exact commitments of predictive coding are, to put it mildly, contentious. For these reasons, it is more accurate to think of predictive coding as a research programme rather than a mature theory that can be stated in full now. The aim of the research programme is to articulate and defend

[1] For examples of these broad claims, see Clark (2013); Clark (2016); Hohwy (2013); Friston (2009); Friston (2010).

[2] Some authors use 'predictive coding' to refer to only one aspect of the view: for example, to the efficient coding strategy described in Section 2, or to the algorithm described in Section 2 of Sprevak (forthcoming[b]). Some authors call the overall research programme 'predictive processing', 'prediction error minimisation', or 'free energy minimisation'. In what follows, I use the term 'predictive coding' to refer to the overall research programme.

[3] See Marr (1982), Ch. 1 for a description of these levels.

some sophisticated – likely heavily modified and precisified – descendent of the three claims above. As with any such programme, the merits of predictive coding should be judged in the round and, to some degree, prospectively: not just in terms of the raw predictive power and confirmation of what it says now, but also in terms of its future potential, and its ability to inspire and guide fruitful research.[4]

Before saying what predictive coding is, it is first helpful to say what it is not. In this article, I outline five ideas that are often presented alongside predictive coding, but which should be distinguished from predictive coding. In the three articles that follow, I focus primarily on the positive content of the view. These explore predictive coding's claims at Marr's computational, algorithmic, and implementation levels respectively (Sprevak, forthcoming[a]; Sprevak, forthcoming[b]; Sprevak, forthcoming[c]). As we will see, there are many ways in which its basic ideas may be elaborated and refined. My strategy is to present what, in my opinion, are the 'bare bones' of the approach. For readers new to this topic, I hope that this will provide you with a basic scaffold on which to build a more nuanced future understanding of the view based around further sources.[5]

For the remainder of this article, I focus on five ideas that feature prominently in many expositions of predictive coding, but which should be distinguished from predictive coding. These ideas are: (i) that the brain employs an efficient coding scheme; (ii) that perception has top-down, expectation-driven effects; (iii) that cognition involves minimisation of prediction error; (iv) that cognition is a form of probabilistic inference; (v) that cognition makes use of generative models. All these ideas are used by predictive coding but, I argue, they are also shared by a variety of other computational approaches. They do not reflect – taken either singly or jointly – what is distinctive about predictive coding's research programme. If one wishes to know what is special about predictive coding, these ideas, whatever their intrinsic

[4]The term 'research programme' is used here to highlight that the precise details, goals, and conditions of correct application of a scientific model are often not fully worked out in advance and are liable to change over time. It is not meant to indicate commitment to a specific philosophical understanding of a scientific research programme (e.g. that of Lakatos (1978) or Laudan (1977)). In what follows, I use 'framework', 'approach', 'view', 'account', 'theory', and 'model' interchangeably with 'research programme', with alternative uses flagged along the way.

[5]To help build that understanding, helpful reviews include Aitchison and Lengyel (2017); Friston (2003); Friston (2005); Friston (2009); Friston (2010); Kanai et al. (2015); Keller and Mrsci-Flogel (2018). For reviews that focus on the describing the mathematical and computational framework, see Bogacz (2017); Gershman (2019); Jiang and Rao (2022); Spratling (2017); Sprevak and Smith (forthcoming). For reviews that focus on the possible neural implementation, see Bastos et al. (2012); Jiang and Rao (2022); de Lange, Heilbron and Kok (2018); Kok and de Lange (2015). For reviews that focus on philosophical issues and possible applications to existing problems in philosophy, see Clark (2013); Clark (2016); Friston, Fortier and Friedman (2018); Hohwy (2013); Hohwy (2020); Metzinger and Wiese (2017); Roskies and Wood (2017).

value, can function as potential distractors. An important corollary of this point is that evidence for predictive coding does not necessarily accumulate from evidence that supports these more general ideas.

The literature on predictive coding is vast and rapidly developing. In what follows, I ignore many interesting developments, proposals, and applications. My description is also inevitably partisan: there is too much disagreement within the primary literature to be able to characterise the view in a wholly uncontroversial way. If you disagree with my description, I hope that what I say at least provides a foil by which to triangulate your own views.

In both the present article and those that follow, I only consider predictive coding as a theory of subpersonal cognitive processing. I do not consider how its computational model might be adapted or extended to account for personal-level thought or conscious experience. Explaining conscious experience with predictive coding is a relatively recent development that has gained some traction in philosophy. However, it is a project that assumes we have a prior understanding of what predictive coding's computational model is. That question is the focus of this review.[6]

2 Efficient neural coding

A key idea that predictive coding employs is that the brain's coding scheme for storing and transmitting sensory information is, in a certain sense, efficient. The relevant form of efficiency is quantified by the degree to which the brain compresses incoming sensory information (measured in terms of Shannon information theory). To compress information, the sensory system should aim to transmit only what is 'new' or 'unexpected' or 'unpredicted' relative to its expectations. If the sensory system embodies certain assumptions about its incoming sensory data, these would enable it to predict certain 'bits' of that incoming sensory stream. This means that fewer bits would need to be stored or transmitted inwards from the sensory boundary, yielding a potential reduction in the costs of the brain physically storing and transmitting that data from the sensory organs to the rest of the brain. The more accurately the brain's internal assumptions reflect its incoming sensory stream, the less information would need to be stored or transmitted inwards from the sensory periphery. All that would need to be sent inwards would be an error signal – what is new or unexpected with respect to those predictions. A similar idea underlies coding schemes that allow electronic computers to store and transmit images and videos across the Internet (e.g. JPEG or MPEG).

[6]For examples of work that applies predictive coding's computational model to explain conscious experience, see Clark (2019); Clark (2023); Dolega and Dewhurst (2021); Hohwy (2012); Kirchhoff and Kiverstein (2019); Seth (2017); Seth (2021).

The notion that our brains use a sensory coding scheme that is efficient in this respect dates back at least to the work of Attneave (1954) and Barlow (1961). They argued that the brain uses a compressing, 'redundancy reducing' code for encoding sensory information based partly on the grounds that neurons in the early visual system have a limited physical dynamic range: the action potentials they send inwards to cortical centres are precious and should not be squandered.[7] Predictive coding adopts the same basic perspective, but elevates it to a universal design principle: not only the early stages of perception, but every aspect of cognition, should be viewed as an attempt to compress the incoming sensory data. To this, predictive coding adds a range of further assumptions about (i) the particular algorithm by which the incoming sensory data are compressed; (ii) how assumptions used for sensory compression are updated during learning; (iii) where physically in the brain all this takes place.

Predictive coding has rather unusual views about how compression of sensory signals works – see (i)–(iii) above. It also holds the rather extreme position that sensory compression is the brain's *only* goal. As Barlow made clear in his later work, even if one thinks that reducing redundancy in sensory data is one thing that the brain does, it is not obvious that it is the only thing it does. In some circumstances, it may pay the brain *not* to compress:

> The point Attneave and I failed to appreciate is that the best way to code information depends enormously on the use that is to be made of it … if you simply want to transmit information to another location, then redundancy-reducing codes economizing channel capacity are what you need … But the brain is not just a communication system, and we now need to survey cases where compression is not the best way to exploit statistical structure. (Barlow, 2001, p. 246).

One can appreciate Barlow's point by considering what would count as an 'efficient' coding scheme for image data on an electronic computer. If all one wishes to do is to transmit an image across a low-bandwidth channel (e.g. a slow Internet connection), then compressing it using a redundancy reducing code (like JPEG) might be a good solution, since it would reduce the number of physical signals one would need to send. Similarly, if one only wishes to store the image on a hard disk drive, then compressing it would mean less physical resources would be required

[7] See Simoncelli and Olshausen (2001); Sterling and Laughlin (2015); Stone (2018) for reviews of efficient coding in the sensory system.

for that storage.[8] However, if one wishes to *transform* the image or *perform an inference* over it, then a code like JPEG, which aims at redundancy reduction, may not be the best or most efficient solution. Compressed data are typically harder to work with. If one asks one's computer to rotate an image 23 degrees clockwise, the machine will generally not attempt to transform a compressed encoding of the image data. Instead, it will switch to a version of the data with known redundancy (e.g. a two-dimensional array of RGB values at X, Y pixel locations). Image processing algorithms defined over this encoding of the data tend to be shorter, simpler, and faster than those over their compressed counterparts.[9] Uncompressed images have extra structure, and that structure can make the job of an algorithm that operates on them easier, even if it adds extra overhead to store or transmit.[10]

If the only things that matter to the brain during cognition are the transmission and storage costs of incoming sensory data, then it may make sense for the brain to aim to maximally compress that sensory data. However, if speed, simplicity, and ease of inference matter too, then it may make sense to add or preserve redundant structure.[11] An 'efficient' coding scheme, therefore, cannot simply be equated with a coding scheme that aims at maximal redundancy reduction.

It is common for contemporary work on efficient sensory coding to acknowledge this point.[12] Predictive coding, in its strongest and purest form, has rather extreme views here: it equates compression of incoming sensory data with efficient coding, and it claims that the entire brain (not just certain areas in the sensory cortex) is devoted to this compression; it also claims that the sensory compression is accomplished by a specific algorithm and representational scheme. Predictive coding employs the idea of efficient coding, but that idea is not unique to predictive coding. Similarly, although evidence for efficient coding in, e.g. early stages in the visual cortex, may

[8] Other coding schemes such as wavelet-based codes (Usevitch, 2001) or deep neural networks (Bühlmann, 2022; Toderici et al., 2016) would outperform JPEG in these respects. However, these schemes tend to impose even higher computing burdens than JPEG if one wishes to decode or transform an image.

[9] This is an instance of a more general common trade-off that is in computer science between optimising for time and optimising for space. Compressing data saves space, but generally it has an adverse effect on the number of computing cycles required to do inference on that data to accomplish certain tasks. You might have experienced this trade-off any time you waited for a '.zip' archive to uncompress before working on its contents.

[10] A related point is that uncompressed data are more resistant to noise during storage and transmission.

[11] Gardner-Medwin and Barlow (2001) list examples in which adding redundancy to sensory signals produces faster and more reliable inference over sensory data.

[12] For example, Simoncelli and Olshausen (2001) suggest that the nature of the downstream task a cognitive system faces in a specific context should be considered when measuring the overall efficiency of a coding scheme, and not merely the degree of compression of the incoming sensory signal (p. 1210).

be compatible with predictive coding, it is also compatible with a range of other, arguably less bold proposals about the role of efficient coding in cognition.

3 Top-down, expectation-driven effects in perception

Top-down, expectation-driven effects in perception are instances in which an agent's prior beliefs systematically affect that agent's perceptual experience. Top-down, expectation-driven effects are sometimes presented as a hallmark feature of predictive coding. Predictive coding's computational model is thought to imply that perception is top-down or expectation-laden: 'What we perceive (or think we perceive) is heavily determined by what we know' (Clark, 2011). Evidence for top-down effects in perception is also thought to support predictive coding's computational model: we should give higher credence to predictive coding's computational proposal based on observation of behavioural evidence of top-down effects in perception.[13]

However, the relationship between predictive coding and top-down, expectation-driven effects in perception is more complex and less direct.

For one thing, top-down effects in perception are standardly defined in terms of a relationship between an agent's *personal-level* states: what an agent *believes* affects their *perceptual experience*.[14] Predictive coding, at least in the first instance, only makes a claim about the agent's subpersonal computational states and processes. The 'top' and 'bottom' in predictive coding's computational model refer, as we will see, to subpersonal computational states. 'High-level' neural representations (implemented deep in the cortical hierarchy) are assumed to have a 'top-down' influence on 'low-level' representations (implemented in the early sensory system). How this kind of subpersonal 'top-down effect' relates to personal-level top-down effects observed in psychology is presently unclear.

One might argue that personal-level top-down effects require *some* subpersonal information flow from high-level cognitive centres to low-level sensory systems. However, it is difficult to know what can be inferred from such an observation. Not every piece of subpersonal information posited by predictive coding's computational model features in the contents of either personal-level belief or perceptual

[13] For examples of this kind of reasoning, see Clark (2013), p. 190; Lupyan (2015).

[14] See characterisations in Macpherson (2012); Firestone and Scholl (2016). One could also define a 'top-down effect' in terms of how various high-level states in predictive coding's subpersonal computational model change the subject's non-intentionally characterised behaviour (e.g. button presses by a subject during a psychophysics experiment). Such a claim would plausibly fall within the scope of predictive coding's model, but it does not have an obvious bearing on top-down effects characterised as a relationship between personal-level thought and perceptual experience. Thanks to Matteo Colombo for this point.

experience. Only a tiny fraction of that subpersonal information appears to be present at the personal level. For predictive coding to say something specific about the existence or character of top-down effects at the personal level, it would need to say *which* aspects of that subpersonal information give rise to *which* personal-level states (beliefs and perceptual contents). These assumptions – connecting content at the subpersonal level to content at the personal level – are not currently to be found within predictive coding's computational model. Ideas about these connections have been proposed, but exactly how subpersonal states inside the computational model map onto personal-level beliefs and perceptual experiences remains a highly speculative matter.[15] Absent auxiliary assumptions about those connections, however, it is simply unclear how predictive coding's computational architecture bears, or if it bears at all, on personal-level top-down effects observed in psychology.[16]

A second complication is that positing top-down subpersonal information flow inside a computational model is not a characteristic that is unique to predictive coding. Almost any plausible computational model of cognition is likely to claim, as predictive coding does, that information flows both 'upwards' (from lower-level sensory systems to high-level cognitive centres) and 'downwards' (from high-level cognitive centres to lower-level sensory systems). As Ira Hyman observed in his introduction to the reprinting of Neisser's classic textbook: 'Cognitive psychology has been and always will be an interaction of bottom-up and top-down influences'.[17] This observation could be made even of nominally 'bottom-up' cognitive models, such as the account of vision proposed by Marr (1982). These models might appear to discount the role of top-down processes, but this is not because they hold that top-down influences do not exist or are unimportant, but rather because they are not necessary to explain a particular phenomenon of interest.[18] It is standard practice, across a wide variety of possible computational architectures, to invoke top-down information flow to account for endogenous attention, semantic priming, and to explain how the brain handles ambiguity, noise, and uncertainty in its sensory input.[19] The mammalian brain contains a huge number of cortical 'backward'

[15]For critical discussion of this point with respect to Seth (2021)'s proposals about personal-level experience, see Sprevak (2022).

[16]See Macpherson (2017); Drayson (2017) for further development of this line of argument. They suggest that predictive coding's computational model is compatible with *no* top-down effects occurring at the personal level at all.

[17]Neisser (2014), p. xvi.

[18]For example, Marr (1982): '… top-down information is sometimes used and necessary … The interpretation of some images involves more complex factors as well as more straightforward visual skills. This image [a black-and-white picture of a Dalmatian] devised by R. C. James may be one example. Such images are not considered here.' (pp. 100–101).

[19]See Gregory (1997); Poeppel and Bever (2010); Yuille and Kersten (2006). Firestone and Scholl (2016) suggest that endogenous attention requires subpersonal top-down information flow inside a computational model (p. 14).

connections that carry signals 'downwards' from cortical centres to peripheral sensory areas, suggesting a significant computational role for top-down information flow in the brain. Firestone and Scholl (2016) observe that there are many additional external routes by which high-level cognitive centres influence processing in low-level sensory systems – e.g. the decision to 'shut one's eyes', causing one's eyelids to close, changes low-level sensory inputs, systematically affecting the contents of states in subpersonal low-level sensory systems.[20] Advocates of predictive coding suggest that their model has a special relationship with top-down, expectation-driven effects observed at the personal level. The difficulty with this is to explain why predictive coding's specific set of top-down computational pathways is *uniquely* or *best* suited to explain these effects.

To be clear, predictive coding's computational model is *compatible* with personal-level top-down effects in perception occurring; it is also broadly *suggestive* that such effects would occur. What is not clear is that it is better suited to account for these effects than any number of other models that also incorporate recurrent elements or loops that facilitate top-down information flow. Predictive coding might have an affinity with top-down, expectation-driven effects in perception, but it is not clear that it is unique in this respect. It is also not clear how evidence for the mere existence of top-down effects at the personal level offers selective support to the view.

4 Minimising prediction error

A common paradigm in contemporary artificial intelligence (AI) is to characterise learning and inference in terms of minimising prediction error. During learning, an AI system might attempt to change its internal parameters to better predict its training data. During inference, an AI system might search for values of its variables that would result in it generating predictions that minimise its prediction error – that are as close to some 'ground truth' as possible.[21] Different types of AI system might vary in the types of data they try to predict, the mathematical model they use for prediction, or the methods they use to find values of either their parameters or variables to minimise prediction error.[22] A computational system's prediction

[20] Dennett (1991) argues that these kinds of external 'virtual wires', which loop into the environment, can enable sophisticated forms of top-down information processing, including those characteristic of rational thought (pp. 193–199).

[21] For example, see Bishop (2006), pp. 1–12 and Hohwy (2013), pp. 42–46.

[22] Note that a 'prediction' need not be about the future. A prediction is an estimate concerning something that the system does not already know. In principle, a prediction might concern what happened in the past, what is happening in the present, or what will happen in the future. For a helpful review of the relevant notion of prediction, see de Lange, Heilbron and Kok (2018), p. 766, Box 2 and Forster (2008).

error might also be measured in a variety of ways. A common formalisation is the mean-squared error – the average of the squares of the differences between prediction values and the true values of the relevant data.[23]

The logical space of computational systems that aim to minimise their prediction error is vast. One can get some idea of the diversity of that space by opening up any current textbook on machine learning or statistics.[24] A simple example of such a system is one that attempts to perform simple linear regression on its training data. Here, minimising prediction error reduces just to fitting a straight-line model to the training data and using that straight-line model to make predictions about unseen cases. Learning consists in finding the value of two parameters (slope and y-intercept) that would define a straight line that minimises mean-squared error over the training data. Classical statistics contains many algorithms for finding those values (e.g., ordinary least squares). Deep neural networks provide much more complicated examples of computational systems that aim to minimise their prediction errors. Learning for a deep neural network typically consists in finding the values of not just two, but millions or billions of internal parameters. Learning algorithms like backpropagation are commonly used to find these values. During inference, a prediction generated by a deep neural network might involve performing a long sequence of mathematical operations over many variables in an effort to yield a value is as close to the ground truth as possible.

Predictive coding suggests that the brain, like many other computational systems, aims to minimise its prediction error. What distinguishes predictive coding is that it makes specific claims about the *data*, *model*, and *algorithm* used in this task; a distinctive claim is also made about the *role* of this instance of prediction error minimisation within the brain's wider cognitive economy.

Regarding the *data*, predictive coding claims that the brain aims to minimise its prediction error over incoming *sensory data*. This should be distinguished from approaches that claim that the brain attempts to minimise prediction error over

[23] Strictly speaking, AI systems aim to minimise a *cost function*, which combines prediction error with other factors. A common cost function is the prediction error plus the sum of the squares of the model's parameters. The latter serves as regularisation term that penalises more complex models. For discussion, see Russell and Norvig (2010), pp. 709–713.

[24] For example, Bishop (2006); MacKay (2003); Barber (2012); Matloff (2017).

other forms of data, such as *reward* signals.[25] The mathematical *model* the brain uses to generate its predictions is encoded in an abstract hierarchical network containing prediction and error units linked by weighted connections. The network is similar to a connectionist neural network, although the way in which the units are connected and the overall topology of the network differs from those commonly used in deep learning. The *algorithm* that adjusts the internal parameters of the hierarchical network during learning also differs. Deep learning networks tend to use a version of backpropagation, whereas predictive coding uses a Hebbian learning algorithm.[26] Finally, a special *role* is assigned to prediction error minimisation in cognition. Predictive coding holds that this instance of prediction error minimisation is not just one among many tasks undertaken by the brain, but its only or fundamental mode of operation.

It is commonplace to appeal to prediction error minimisation inside a computational model of cognition. What marks out predictive coding as unusual is that it proposes that cognition involves prediction error minimisation over a specific dataset, with a specific mathematical model, and using a specific learning and inference algorithm. It also proposes that this form of prediction error minimisation is the only or fundamental objective of cognition. Evidence for the bare existence of prediction error signals in the brain, although it may be compatible with predictive coding, is also likely to be compatible with any number of alternative models. Such evidence, if it is to tell in favour of predictive coding, would need to settle the nature, role, and function of any observed prediction error signals: Are the observed predictions about upcoming sensory signals? How were these predictions generated? How are they revised in light of new evidence? What is their role across different cognitive tasks?

5 Cognition as a form of probabilistic inference

Brains receive noisy, incomplete, and sometimes contradictory information via their sensory organs. They need to weigh this information rapidly and integrate it with (sometimes conflicting) background knowledge in order to reach a decision and generate behaviour. Probabilistic models of cognition provide a broad framework

[25] There are a wide range of computational models of learning and decision-making that attribute the goal of minimising prediction error over reward signals to the brain (Niv and Schoenbaum, 2008; Schultz, Dayan and Montague, 1997). Although these models bear a family resemblance to predictive coding, advocates of predictive coding are generally clear that the two approaches are distinct (Friston, 2009). See Friston, Schwartenbeck et al. (2013); Schwartenbeck et al. (2015) for an attempt to show that minimising reward prediction error can be formalised as minimising an expected free-energy measure if one redefines utilities/preferences as a special kind of probability distribution.

[26] See Sprevak (forthcoming[b]), Section 2.3.

by which to understand how brains do this. According to such a model, brains do not represent the world in purely categorical way (e.g. 'the person facing me is my father'), but instead represent multiple rival possibilities (e.g. 'the person facing me is my father, my uncle, his cousin, …') along with some measure of uncertainty regarding those outcomes.[27] Computational models typically formalise this by ascribing mathematical *subjective probability distributions* to brains. These probability distributions measure the brain's degree of confidence in a range of different possibilities.[28] Cognitive processing is then modelled as a series of operations in which one subjective probability distribution conditions, or updates, another. The exact manner in which this happens may vary between different models. In principle, cognitive processing may maintain this probabilistic character until the moment the brain is forced to plump for a specific outcome in action (e.g. the agent is required to respond 'yes'/'no' in a forced-choice task).

A particularly influential example of this computational approach is the *Bayesian brain hypothesis*.[29] This suggests that Bayes' rule, or some approximation to it, describes how the brain combines and updates its probability distributions.[30] Because exact Bayesian inference is computationally intractable, advocates of the Bayesian brain hypothesis generally assume that the brain implements some version of approximate Bayesian inference. Approximate Bayesian inference can be achieved in many ways, the most popular of which tend to fall into two camps: *sampling algorithms* (which trace the path of multiple categorical samples through inference to create an empirical distribution that approximates the true Bayesian posterior) and *variational algorithms* (which change the parameters of some simpler, more computationally tractable probability distribution in order to try to find a posterior

[27] For examples, see Chater, Tenenbaum and Yuille (2006); Danks (2019).

[28] The subjective probabilities in question are formally handled in a similar manner to subjective probabilities in classical formulations of Bayesianism – i.e. as degrees of belief or credences of some reasoning agent (de Finetti, 1990; Ramsey, 1990). However, unlike in traditional treatments, these representational states need not be ascribed to the entire agent; they may be ascribed to subpersonal parts of the agent (e.g. to individual brain regions, neural populations, or single neurons) (for example, see Deneve, 2008; Pouget et al., 2013). For discussion of how the concept of subjective probability has been extended to apply to subpersonal parts of agents, see Icard (2016); Rescorla (2020).

[29] Chater and Oaksford (2008); Knill and Pouget (2004).

[30] Bayesian updating is far from the only option for handling inference under uncertainty. Plenty of rules and heuristics do not fit the Bayesian norms but still generate adaptive behaviour (Bowers and Davis, 2012; Colombo, Elkin and Hartmann, 2021; Eberhardt and Danks, 2011; Rahnev and Denison, 2018). Rahnev (2017) considers the possibility that brains do not store full probability distributions, but store only a few categorical samples or summary statistics (e.g. variance, skewness, kurtosis) and use these partial probabilistic measures to generate adaptive behaviour.

distribution that is close to the true Bayesian posterior).[31] Both types of Bayesian inference algorithm are common in commercial applications of machine learning. Proponents of the Bayesian brain hypothesis do not agree about whether the brain uses a sampling method, a variational method, or something else entirely.[32]

Predictive coding is an example of a probabilistic model of cognition and, specifically, of the Bayesian brain hypothesis. Advocates of predictive coding hold that the brain uses a variational method for approximate Bayesian inference. If a range of further simplifying assumptions are made, the problem of Bayesian inference can be described as a problem of minimising sensory prediction error.[33] If the numerical values in predictive coding's hierarchical algorithm are interpreted as the means and variances of multivariate Gaussian probability distributions, then that algorithm can be understood as performing an instance of variational Bayesian inference.[34] At the implementation level, predictive coding suggests that these key numerical values, which fix the brain's subjective probability distributions, are encoded by the firing rates of certain cell populations in the neocortex, and the strength of the synaptic connections between those populations encodes the degree to which these subjective probability distributions condition one another.[35]

One might endorse a probabilistic model of cognition, or the Bayesian brain hypothesis, while rejecting some or all of these assumptions. For example, one might not accept that a single probabilistic model underlies every aspect of cognition, or that the subjective probability distributions in the brain are always Gaussian, or that the brain uses the specific version of variational Bayesian inference proposed by predictive coding, or that the brain's subjective probability distributions are encoded in the neocortex.[36] Predictive coding is one of many possible examples of a probabilistic model of cognition. Evidence for probabilistic inference in the brain, without further elaboration, is likely to be compatible with any number of other alternatives.

[31] For an introduction to sampling methods (e.g. Markov chain Monte Carlo methods or particle filtering), see Bishop (2006), Ch. 11. For an introduction to variational methods, see Bishop (2006), Ch. 10.

[32] For exploration of the idea that the brain uses a sampling method, see Fiser et al. (2010); Griffiths, Vul and Sanborn (2012); Hoyer and Hyvärinen (2003); Moreno-Bote, Knill and Pouget (2011); Sanborn and Chater (2016); Sanborn and Chater (2017). Predictive coding is an example of a view that holds that the brain uses a variational method for approximate Bayesian inference.

[33] Sprevak (forthcoming[a]), Section 8; Sprevak and Smith (forthcoming).

[34] Sprevak (forthcoming[b]), Section 5.

[35] Sprevak (forthcoming[c]), Section 3.

[36] Aitchison and Lengyel (2017) consider how predictive coding's proposals might be changed if its algorithm for variational Bayesian inference were replaced with a sampling algorithm (pp. 223–224).

6 Cognition uses a generative model

A generative model is a special kind of representation that describes how observations are produced by unobserved ('latent') variables in the world. If a generative model were supplied with the information that your best friend enters the room, it might predict which sights, sounds, smells you would experience. At the highest level of abstraction, one might conceive of a generative model as a black box that takes as input a hidden state of the world and that yields as output sensory data that are likely to be observed. It is widely thought that generative models – in particular, probabilistic generative models – play an important role in cognition. This is for at least three reasons.

First, a generative model could help the brain to distinguish between changes to its sensory data that are *self-generated* and *externally generated*. When our eyes move, our sensory input changes. How does the brain know which changes are due to movement of our sensory organs and which are due to movement of external objects in the environment? von Helmholtz (1867) proposed that our brain makes a copy of its upcoming motor plans and uses this copy (the 'efference copy') to predict how its plans are likely to affect incoming sensory data. A generative model (the 'forward motor model') predicts the likely sensory consequences of a planned movement (e.g. how sensory data would be likely to change if the eyeballs rotate). These predictions are then fed back to the sensory system and 'subtracted away' from incoming sensory data. This would allow the brain to compensate for changes its own movement introduces into its sensory data stream.[37]

Second, a generative model would help the brain to overcome inherent latency, noise, and gaps in its sensory data. If you execute a complex, rapid motion – e.g. a tennis serve – your brain needs to have accurate, low-latency sensory feedback. It needs to know where your limbs are, how its motor plan is unfolding, whether any unexpected resistance is being met, and how external objects (like the tennis ball) are moving. Due to the limits of the brain's physical hardware, this sensory feedback is likely to arrive late, with gaps, and with noise. A generative model would help the brain to alleviate these problems by regulating motor control based, not on actual sensory feedback, but on expected sensory feedback. When the incoming sensory data do arrive, the brain could integrate them into motor control in a way that takes into account any background information it has about bias, noise, and uncertainty concerning that sensory data. Franklin and Wolpert (2011) argue that this would enable the brain to make optimal use of its sensory input during motor control – 'optimal' in the sense that the brain would make use of all its available information.[38]

[37] Keller and Mrsci-Flogel (2018), pp. 424–425. Blakemore, Frith and Wolpert (1999) use a model of this kind to explain why it is difficult to tickle yourself.

[38] Grush (2004); Körding and Wolpert (2004); Körding and Wolpert (2006); Rescorla (2018).

Third, if a generative model takes a probabilistic form, it could in principle be inverted to yield a *discriminative* model.[39] Discriminative models are of obvious utility across many areas of cognition. A discriminative model tells the cognitive system, given some sensory signal, which state(s) of the world are most likely to be responsible for its observations.[40] Discriminative models are needed in visual perception, object categorisation, speech recognition, detection of causal relations, and social cognition. A discriminative model and a generative model facilitate inference in opposite directions: a discriminative model enables one to infer from sensory data to the value of latent variables; a generative model enables one to infer from the value of latent variables to sensory observations. The latter form of inference might not initially appear to be useful in the context of a discriminative problem, but if the system applies Bayes' theorem, a generative model can be flipped to create a discriminative model. This may be a computationally attractive strategy because sometimes generative models are easier to learn, more compact to represent, and less liable to break as background conditions change.[41] In AI, a common strategy for solving a discriminative problem is to first learn a generative model and then invert it using Bayes' theorem. This strategy is frequently suggested as the way in which the brain tackles discriminative problems in cognition.[42]

A generative model is a common component in a modern computational model of cognition. A generative model's content and structure, the methods by which it is updated, and how it might be physically implemented in the brain, might be filled out in many ways, including ways that depart substantially from those suggested by predictive coding. In the context of predictive coding, it is common to assume that a single probabilistic generative model is employed across all aspects of cognition.

[39]Bayes' theorem is $P(Y \mid X) = P(Y \mid X)P(Y)/P(X)$, and follows from standard axioms and definitions of probability theory. Bayes' rule (referenced in Section 5) says that an agent's subjective probabilities should be updated using Bayesian conditionalisation, $P_{t+1}(Y) = P_t(Y \mid X)$; its justification does not follow from the axioms of probability (Strevens, 2017).

[40]A discriminative model estimates the probability of a latent variable, Y, given an observation, x, i.e. $P(Y \mid X = x)$. A generative model is defined either as the likelihood function, i.e. the probability of an observation, X, given some hidden state of the world, y, $P(X \mid Y = y)$; or, as the full joint probability distribution, $P(X, Y)$. The difference between these rarely matters in practice as the joint probability distribution equals the product of the likelihood and the system's priors over those unobserved states, $P(X, Y) = P(X \mid Y)P(Y)$, and both likelihood and priors are needed to invert the model under Bayes' theorem.

[41]The reasons why generative models sometimes provide these advantages are complex and depend partly on the contingent way our world happens to be structured. For a brief intuitive explanation, see Russell and Norvig (2010), pp. 497, 516–517.

[42]See Bishop (2006), Ch. 4 on creating discriminative classifiers using generative models. See Chater and Manning (2006); Kriegeskorte (2015); Poeppel and Bever (2010); Tenenbaum et al. (2011); Yuille and Kersten (2006) for various proposals about how the brain uses generative models to answer discriminative queries in cognition.

This generative model is claimed to have a specific structure, content, and to be implemented in a specific way in the brain. Someone might believe that generative models play some role in cognition, but not accept any of this. They might hold that multiple distinct generative models exist in the brain in relative functional isolation from each other – e.g., there might be a domain-specific generative model dedicated to motor control.[43] They might hold that the brain does not use a generative model to solve every discriminative problem – the brain might sometimes attempt to learn a discriminative model of a domain directly, or employ some other, entirely non-model-based strategy.[44] They might disagree about the content of the generative model or how the generative model is physically implemented in the brain.[45] Predictive coding does not have a monopoly on generative models. Evidence for the existence of a generative model in cognition, such as that cited above, does not selectively support predictive coding over alternative proposals.

7 Conclusion

The aim of this paper is to separate five influential ideas about cognition from predictive coding. Many philosophers first encounter these ideas in the context of predictive coding. However, it is important to recognise that those ideas exist in a broader intellectual landscape and they are employed by approaches that have little or nothing to do with predictive coding. Accepting one or more of these ideas does not constitute an endorsement of predictive coding. Similarly, evidence that supports one or more of the ideas should not be taken as evidence for predictive coding. If one wants to understand what is unique about predictive coding, one needs to disentangle it from these ideas.

Why not hold that 'predictive coding' refers to some non-specific, broad synthesis of the five ideas? The problem is that the five ideas described are employed, not just in isolation, but also in combination, by almost every current approach in computational modelling. There are also good reasons internal to the predictive coding research programme to resist such a move. Advocates of predictive coding are keen to stress, not only that their view is credible, but also that it is novel and that it faces genuine jeopardy with respect to empirical evidence. If these claims are to be taken seriously, one would need to show (i) that the view departs from plausible

[43] Wolpert, Ghahramani and Flanagan (2001); Grush (2004) suggest this. They also suggest that this motor model is not implemented in the neocortex but in the cerebellum.

[44] Ng and Jordan (2002) consider conditions under which it is more efficient to learn a discriminative model of a domain than learning a generative model first and then inverting it. Raina et al. (2003); Lasserre, Bishop and Minka (2006) examine a range of hybrid discriminative-generative approaches to inference.

[45] See Sprevak (forthcoming[b]), Section 2.5; Sprevak (forthcoming[c]), Section 6.

rivals; and (ii) that it is not so anodyne as to be consistent with any likely empirical evidence. To this end, Clark warns against interpreting predictive coding as an 'extremely broad vision'; it should be interpreted as a 'specific proposal' (Clark, 2016, p. 10). Hohwy observes that there is often an ambiguity which renders presentations of predictive coding 'both mainstream and utterly controversial' (Hohwy, 2013, p. 7). He argues that in order for it to meaningfully make contact with empirical evidence, it should be understood as a specific detailed proposal (Hohwy, 2013, pp. 7–8).[46]

In what follows, I argue that what distinguishes predictive coding from contemporary rivals is a combination of three claims, each of which may be precisified or qualified in various ways. These claims concern how cognition works at Marr's *computational*, *algorithmic*, and *implementation* levels.

It is worth tempering what follows with a cautionary note. As mentioned earlier, the content of predictive coding is in no way a settled matter. Researchers disagree about which features of the view are essential, whether the model truly has a universal scope, whether the computational, algorithmic, and implementation level claims should be combined, and the exact form that each claim should take. Cutting across the disagreement, however, is a simple and bold picture of the mind, its abstract computational structure, and its physical implementation. This somewhat idealised vision has inspired many researchers, and it will be the focus of the next three papers.

Bibliography

Aitchison, L. and M. Lengyel (2017). "With or without you: Predictive coding and Bayesian inference in the brain". In: *Current Opinion in Neurobiology* 46, pp. 219–227.

Attneave, F. (1954). "Informational aspects of visual perception". In: *Psychological Review* 61, pp. 183–193.

Barber, D. (2012). *Bayesian Reasoning and Machine Learning*. Cambridge: Cambridge University Press.

Barlow, H. B. (1961). "Possible principles underlying the transformations of sensory messages". In: *Sensory Communication*. Ed. by W. A. Rosenblith. Cambridge, MA: MIT Press, pp. 217–234.

— (2001). "Redundancy reduction revisited". In: *Network: Computation in Neural Systems* 12, pp. 241–253.

[46] Colombo (2017) argues that Clark still sometimes interprets predictive coding as a broad vision.

Bastos, A. M., W. M. Usrey, R. A. Adams, G. R. Mangun, P. Fries and K. Friston (2012). "Canonical microcircuits for predictive coding". In: *Neuron* 76, pp. 695–711.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York, NY: Springer.

Blakemore, S. -J., C. D. Frith and D. M. Wolpert (1999). "Spatio-temporal prediction modulates the perception of self-produced stimuli". In: *Journal of Cognitive Neuroscience* 11, pp. 551–559.

Bogacz, R. (2017). "A tutorial on the free-energy framework for modelling perception and learning". In: *Journal of Mathematical Psychology* 76, pp. 198–211.

Bowers, J. S. and C. J. Davis (2012). "Bayesian just-so stories in psychology and neuroscience". In: *Psychological Bulletin* 128, pp. 389–414.

Bühlmann, M. (2022). "Stable Diffusion Based Image Compression". URL: https://pub.towardsai.net/stable-diffusion-based-image-compresssion-6f1f0a399202.

Chater, N. and C. D. Manning (2006). "Probabilistic models of language processing and acquisition". In: *Trends in Cognitive Sciences* 10, pp. 335–344.

Chater, N. and M. Oaksford, eds. (2008). *The Probabilistic Mind: Prospects for Bayesian Cognitive Science*. Oxford: Oxford University Press.

Chater, N., J. B. Tenenbaum and A. Yuille (2006). "Probabilistic models of cognition: Conceptual foundations". In: *Trends in Cognitive Sciences* 10, pp. 287–291.

Clark, A. (2011). "What scientific concept would improve everybody's cognitive toolkit?" In: *Edge*. last checked 10 March 2022. URL: https://www.edge.org/response-detail/10404.

— (2013). "Whatever next? Predictive brains, situated agents, and the future of cognitive science". In: *Behavioral and Brain Sciences* 36, pp. 181–253.

— (2016). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford: Oxford University Press.

— (2019). "Consciousness as generative entanglement". In: *The Journal of Philosophy* 116, pp. 645–662.

— (2023). *The Experience Machine: How Our Minds Predict and Shape Reality*. London: Allen Lane.

Colombo, M. (2017). "Review of Andy Clark, *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*". In: *Minds and Machines* 27, pp. 381–385.

Colombo, M., L. Elkin and S. Hartmann (2021). "Being realist about Bayes, and the predictive processing theory of mind". In: *The British Journal for the Philosophy of Science* 72, pp. 185–220.

Danks, D. (2019). "Probabilistic models". In: *The Routledge Handbook of the Computational Mind*. Ed. by M. Sprevak and M. Colombo. Routledge, pp. 149–158.

de Finetti, B. (1990). *Theory of Probability*. Vol. 1. New York, NY: Wiley & Sons.

De Lange, F. P., M. Heilbron and P. Kok (2018). "How do expectations shape perception?" In: *Trends in Cognitive Sciences* 22, pp. 764–779.

Deneve, S. (2008). "Bayesian spiking neurons I: Inference". In: *Neural Computation* 20, pp. 91–117.

Dennett, D. C. (1991). *Consciousness Explained*. Boston, MA: Little, Brown & Company.

Dolega, K. and J. E. Dewhurst (2021). "Fame in the predictive brain: a deflationary approach to explaining consciousness in the prediction error minimization framework". In: *Synthese* 198, pp. 7781–7806.

Drayson, Z. (2017). "Modularity and the predictive mind". In: *Philosophy and Predictive Processing*. Ed. by T. Metzinger and W. Wiese. 10.15502/9783958573031: MIND Group. DOI: 10.15502/9783958573024.

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

Firestone, C. and B. J. Scholl (2016). "Cognition does not affect perception: Evaluating the evidence for "top-down" effects". In: *Behavioral and Brain Sciences* 39, E229.

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

Forster, M. (2008). "Prediction". In: *The Routledge Companion to Philosophy of Science*. Ed. by S. Psillos and M. Curd. London: Routledge, pp. 405–413.

Franklin, D. W. and D. M. Wolpert (2011). "Computational mechanisms of sensorimotor control". In: *Neuron* 72, pp. 425–442.

Friston, K. (2003). "Learning and inference in the brain". In: *Neural Networks* 16, pp. 1325–1352.

— (2005). "A theory of cortical responses". In: *Philosophical Transactions of the Royal Society of London, Series B* 360, pp. 815–836.

Friston, K. (2009). "The free-energy principle: a rough guide to the brain?" In: *Trends in Cognitive Sciences* 13, pp. 293–301.

— (2010). "The free-energy principle: A unified brain theory?" In: *Nature Reviews Neuroscience* 11, pp. 127–138.

Friston, K., M. Fortier and D. A. Friedman (2018). "Of woodlice and men: A Bayesian account of cognition, life and consciousness: An interview with Karl Friston". In: *ALIUS Bulletin* 2, pp. 17–43.

Friston, K., P. Schwartenbeck, T. FitzGerald, M. Moutoussis, T. Behrens and R. J. Dolan (2013). "The anatomy of choice: active inference and agency". In: *Frontiers in Human Neuroscience* 7, p. 598.

Gardner-Medwin, A. R. and H. B. Barlow (2001). "The limits of counting accuracy in distributed neural representations". In: *Neural Computation* 13, pp. 477–504.

Gershman, S. J. (2019). "What does the free energy principle tell us about the brain?" In: *Neurons, Behavior, Data Analysis, and Theory*.

Gregory, R. L. (1997). "Knowledge in perception and illusion". In: *Philosophical Transactions of the Royal Society of London, Series B* 352, pp. 1121–1128.

Griffiths, T. L., E. Vul and A. N. Sanborn (2012). "Bridging levels of analysis for probabilistic models of cognition". In: *Current Directions in Psychological Science* 21, pp. 263–268.

Grush, R. (2004). "The emulator theory of representation: Motor control, imagery, and perception". In: *Behavioral and Brain Sciences* 27, pp. 377–442.

Hohwy, J. (2012). "Attention and conscious perception in the hypothesis testing brain". In: *Frontiers in Psychology* 3, pp. 1–14.

— (2013). *The Predictive Mind*. Oxford: Oxford University Press.

— (2020). "New directions in predictive processing". In: *Mind and Language* 35, pp. 209–223.

Hoyer, P. O and A. Hyvärinen (2003). "Interpreting neural response variability as Monte Carlo sampling of the posterior". In: *Advances in Neural Information Processing Systems 15*. Ed. by S. Becker, S. Thrun and K. Obermayer. Cambridge, MA: MIT Press, pp. 277–284.

Icard, T. (2016). "Subjective probability as sampling propensity". In: *Review of Philosophy and Psychology* 7, pp. 863–903.

Jiang, L. P. and R. P. N. Rao (2022). "Predictive coding theories of cortical function". In: *Oxford Research Encyclopedia of Neuroscience*. Ed. by S. M. Sherman. DOI: 10.1093/acrefore/9780190264086.013.328.

Kanai, R., Y. Komura, S. Shipp and K. Friston (2015). "Cerebral hierarchies: predictive processing, precision and the pulvinar". In: *Philosophical Transactions of the Royal Society of London, Series B* 370, p. 20140169.

Keller, G. B. and T. D. Mrsci-Flogel (2018). "Predictive processing: A canonical cortical computation". In: *Neuron* 100, pp. 424–435.

Kirchhoff, M. D. and J. Kiverstein (2019). *Extended consciousness and predictive processing*. Abingdon: Routledge.

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

Kok, P. and F. P. de Lange (2015). "Predictive coding in the sensory cortex". In: *An Introduction to Model-Based Cognitive Neuroscience*. Ed. by B. U. Forstmann and E.- J. Wagenmakers. New York, NY: Springer, pp. 221–244.

Körding, K. P. and D. M. Wolpert (2004). "Bayesian integration in sensorimotor learning". In: *Nature* 427, pp. 244–247.

— (2006). "Bayesian decision theory in sensorimotor control". In: *Trends in Cognitive Sciences* 10, pp. 319–326.

Kriegeskorte, N. (2015). "Deep neural networks: A new framework for modeling biological vision and brain information processing". In: *Annual Review of Vision Science* 1, pp. 417–446.

Lakatos, I. (1978). *The Methodology of Scientific Research Programmes: Philosophical Papers, Vol. 1*. Ed. by J. Worrall and G. Currie. Cambridge: Cambridge University Press.

Lasserre, J. A., C. M. Bishop and T. P. Minka (2006). "Principled hybrids of generative and discriminative models". In: *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. New York, NY: IEEE Computer Society, pp. 87–94.

Laudan, L. (1977). *Progress and Its Problems*. Berkeley, CA: University of California Press.

Lupyan, G. (2015). "Cognitive penetrability of perception in the age of prediction: Predictive systems are penetrable systems". In: *Review of Philosophy and Psychology* 6, pp. 547–569.

MacKay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge: Cambridge University Press.

Macpherson, F. (2012). "Cognivite penetration of colour experience: Rethinking the issue in light of an indirect mechanism". In: *Philosophy and Phenomenological Research* 84, pp. 24–62.

— (2017). "The relationship between cognitive penetration and predictive coding". In: *Consciousness and Cognition* 47, pp. 6–16.

Marr, D. (1982). *Vision*. San Francisco, CA: W. H. Freeman.

Matloff, N. (2017). *Statistical Regression and Classification*. Boca Raton, FL: CRC Press.

Metzinger, T. and W. Wiese (2017). "Vanilla PP for philosophers: A primer on predictive processing". In: *Philosophy and Predictive Processing*. Ed. by T. Metzinger and W. Wiese. Frankfurt am Main: MIND Group. DOI: 10.15502/9783958573024.

Moreno-Bote, R., D. C. Knill and A. Pouget (2011). "Bayesian sampling in visual perception". In: *Proceedings of the National Academy of Sciences* 108, pp. 12491–12496.

Neisser, U. (2014). *Cognitive Psychology*. Englewood Cliffs, NJ: Prentice-Hall.

Ng, A. Y. and M. I. Jordan (2002). "On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes". In: *Advances in Neural Information Processing Systems 14*. Ed. by T. G. Dietterich, S. Becker and Z. Ghahramani. Cambridge, MA: MIT Press, pp. 841–848.

Niv, Y. and G. Schoenbaum (2008). "Dialogues on prediction errors". In: *Trends in Cognitive Sciences* 12, pp. 265–272.

Poeppel, D. and T. G. Bever (2010). "Analysis by synthesis: A (re-)emerging program of research for language and vision". In: *Biolinguistics* 4, pp. 174–200.

Pouget, A., J. M. Beck, W. J. Ma and P. E. Latham (2013). "Probabilistic brains: Knows and unknowns". In: *Nature Neuroscience* 16, pp. 1170–1178.

Rahnev, D. (2017). "The case against full probability distributions in perceptual decision making". bioRxiv:108944. DOI: 10.1101/108944.

Rahnev, D. and R. N. Denison (2018). "Suboptimality in perceptual decision making". In: *Behavioral and Brain Sciences* 41, pp. 1–66.

Raina, R., Y. Shen, A. McCallum and A. Y. Ng (2003). "Classification with hybrid generative/discriminative models". In: *Advances in Neural Information Processing Systems 16*. Ed. by S. Thrun, L. K. Saul and B. Schölkopf. Cambridge, MA: MIT Press, pp. 545–552.

Ramsey, F. P. (1990). *Philosophical Papers*. Ed. by D. H. Mellor. Cambridge: Cambridge University Press.

Rescorla, M. (2018). "Motor computation". In: *The Routledge Handbook of the Computational Mind*. Ed. by M. Sprevak and M. Colombo. London: Routledge, pp. 424–435.

— (2020). "A Realist Perspective on Bayesian Cognitive Science". In: *Inference and Consciousness*. Ed. by A. Nes and T. Chan. London: Routledge, pp. 40–73.

Roskies, A. L. and C. C. Wood (2017). "Catching the prediction wave in brain science". In: *Analysis* 77, pp. 848–857.

Russell, S. and P. Norvig (2010). *Artificial Intelligence: A Modern Approach*. 3rd ed. Upper Saddle River, NJ: Pearson.

Sanborn, A. N. and N. Chater (2016). "Bayesian brains without probabilities". In: *Trends in Cognitive Sciences* 20, pp. 883–893.

— (2017). "The sampling brain". In: *Trends in Cognitive Sciences* 21, pp. 492–493.

Schultz, W., P. Dayan and P. R. Montague (1997). "A neural substrate of prediction and reward". In: *Science* 275, pp. 1593–1599.

Schwartenbeck, P., T. FitzGerald, C. Mathys, R. J. Dolan and K. Friston (2015). "The dopaminergic midbrain encodes the expected certainty about desired outcomes". In: *Cerebral Cortex* 25, pp. 3434–34445.

Seth, A. K. (2017). "The cybernetic brain: From interoceptive inference to sensorimotor contingencies". In: *Philosophy and Predictive Processing*. Ed. by T. Metzinger and W. Wiese. Frankfurt am Main: MIND Group. DOI: 10.15502/9783958570108.

— (2021). *Being You: A New Science of Consciousness*. London: Faber & Faber.

Simoncelli, E. P. and B. A. Olshausen (2001). "Natural image statistics and neural representation". In: *Annual Review of Neuroscience* 24, pp. 1193–1216.

Spratling, M. W. (2017). "A review of predictive coding algorithms". In: *Brain and Cognition* 112, pp. 92–97.

Sprevak, M. (2022). "Understanding phenomenal consciousness while keeping it real". In: *Philosophical Psychology*. DOI: 10.1080/09515089.2022.2092465.

— (forthcoming[a]). "Predictive coding II: The computational level". In: *TBC*.

— (forthcoming[b]). "Predictive coding III: The algorithmic level". In: *TBC*.

— (forthcoming[c]). "Predictive coding IV: The implementation level". In: *TBC*.

Sprevak, M. and R. Smith (forthcoming). "An introduction to predictive processing models of perception and decision-making". In: *Topics in Cognitive Science*.

Sterling, P. and S. Laughlin (2015). *Principles of Neural Design*. Cambridge, MA: MIT Press.

Stone, J. V. (2018). *Principles of Neural Information Theory: Computational Neuroscience and Metabolic Efficiency*. Sebtel Press.

Strevens, M. (2017). "Notes on Bayesian Confirmation Theory". last checked 15 June 2022. URL: http://www.strevens.org/bct/BCT.pdf.

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

Toderici, G., D. Vincent, N. Johnston, S. J. Hwang, D. Minnen, J. Shor and M. Covell (2016). "Full resolution image compression with recurrent neural networks". arXiv:1608.05148. DOI: 10.48550/arXiv.1608.05148.

Usevitch, B. E. (2001). "A tutorial on modern lossy wavelet image compression: foundations of JPEG 2000". In: *IEEE Signal Processing Magazine* 18, pp. 22–35.

von Helmholtz, H. (1867). *Handbuch der physiologischen Optik*. Hamburg und Leipzig: Leopold Voss.

Wolpert, D. M., Z. Ghahramani and J. R. Flanagan (2001). "Perspectives and problems in motor learning". In: *Trends in Cognitive Sciences* 5, pp. 487–494.

Yuille, A. and D. Kersten (2006). "Vision as Bayesian inference: analysis by synthesis?" In: *Trends in Cognitive Sciences* 10, pp. 301–308.