Forms of prediction in the nervous system

Christoph Teufel[1*] and Paul C. Fletcher[2, 3, 4*]

[1] Cardiff University Brain Research Imaging Centre (CUBRIC), School of Psychology, Cardiff University, Cardiff, UK

[2] Department of Psychiatry, University of Cambridge, Cambridge, UK

[3] Cambridgeshire and Peterborough NHS Foundation Trust, Cambridge, UK

[4] Wellcome Trust-MRC Institute of Metabolic Science – Metabolic Research Laboratories, Cambridge Biomedical Campus, Cambridge, UK

*email:teufelc@cardiff.ac.uk; pcf22@cam.ac.uk

Abstract| The idea that predictions shape how we perceive and comprehend the world has become increasingly influential in the field of systems neuroscience. It also forms an important framework for understanding neuropsychiatric disorders, which are proposed to be the result of disturbances in the mechanisms through which prior information influences perception and belief, leading to the production of sub-optimal models of the world. There is a widespread tendency to conceptualize the influence of predictions exclusively in terms of 'top-down' processes, whereby predictions generated in higher-level areas exert their influence on lower-level areas within an information-processing hierarchy. However, this excludes from consideration the predictive information embedded in the 'bottom-up' stream of information processing. We describe evidence for the importance of this distinction and argue that it is critical for the development of the predictive processing framework and, ultimately, for an understanding of the perturbations that drive the emergence of neuropsychiatric symptoms and experiences.

[H1] Introduction

Biological organisms use sensory inputs to uncover the structure of their surroundings, in order to create a representation of their environment. Such a representation is crucial for an agent **[G]** to regulate its interaction with the world (**BOX 1**) [1]. However, sensory inputs are ambiguous and noisy and it is believed that the creation of an accurate representation of the environment therefore also requires prior information[2-4]. Predictions **[G]** derived from such prior information are thought to help to resolve the ambiguity in current or future sensory signals and allow inference about the external causes of inputs. According to this predictive processing framework prediction that is based on prior information about the world is a key feature of brain function.

This idea has become increasingly influential across the fields of human psychophysics[5,6], primate electrophysiology[7-10], cognitive and computational neuroscience[3,11-14] and clinical neuroscience[15-22], where it offers new perspectives on disturbances in perception, belief, and action. However, it is frequently unclear exactly what is meant by 'prediction' and there is general imprecision in how the term is conceptualised and used. Almost ubiquitously[3,10,23-27], prediction is considered in terms of higher-level processes acting in a top-down manner on mechanisms lower in the information-processing hierarchy. Here, we describe evidence that challenges this exclusively top-down view and argue for a framework that acknowledges that many forms of predictive information are embedded within the nervous system as constraints on bottom-up processing.

In this Perspective, we argue that it is important to distinguish between two types of regularity, or 'patterns', in the world that together form the basis for predictions in the nervous system. We introduce a fundamental distinction between spatiotemporally global or constant, non-hierarchical regularities and spatiotemporally local regularities that depend on context and are thus hierarchical. We provide evidence to show that predictions based on prior knowledge of these two types of regularity are mechanistically distinct: they are associated with two different forms of information processing (bottom-up and top-down). Finally, we consider, with examples, the importance of distinguishing these two forms of prediction. While we believe that it is likely that further sub-divisions will emerge as the

3

field matures, we see this primary distinction —which is based on the form rather than the content of predictions —as an important initial step towards a more comprehensive appreciation of the diversity of mechanisms underpinning predictive processing in health and disease.

[H1] The nature of predictions

Prediction figures prominently in information theory **[G]** and Bayesian decision theory **[G]** [28]. Bayesian models have been particularly important in advancing our understanding of brain function, formalising the idea that perceptual and cognitive inference **[G]** does not exclusively rely on current inputs but is shaped by predictions that are based on so-called priors **[G]**, or background information about the structure of the environment. It is important to emphasise that Bayesian decision theory provides a normative framework[29]: that is, it allows the researcher to specify how an agent should use current inputs and prior information to maximise a specified utility **[G]**, given the information to which it has access, but it is agnostic as to the precise mechanical implementation of this process[29]. Just as a map may detail an optimal route without suggesting the best means of transport, the Bayesian decision framework is concerned with the overall objective of predictions rather than the details of how predictions are implemented in the nervous system.

The concept of predictive coding **[G]** has been immensely influential in shaping how we think about neural information-processing, both in health and disease[3,13,14,30-32] (**FIG. 1** and **BOX 2**) and has inspired some of the most detailed mechanistic formalisations of prediction in neuroscience. A range of different predictive coding models have been proposed[14]. These are computationally similar, but make very different assumptions regarding the neural implementation (**BOX 2**). Nevertheless, most predictive coding models have promoted a conceptualisation of prediction exclusively in terms of top-down processing[2,3,13,30].

A broader perspective on the nature of prediction is offered by early work in the field of cybernetics, which prefigured much of the thinking within the current predictive processing

framework. Note that here we use 'predictive processing' as a general term that encompasses predictive coding, the latter being one specific form of predictive processing[33]. The term cybernetics derives from the Greek term for 'steersman'[34], which captures the idea that a successful agent must control the effects of its environment in a particular way: not necessarily by constraining that environment but by adaptively responding to changes in its relevant parameters. The steersman does not control the breezes, tides, and currents but makes adjustments that minimise their effects on the boat's desired course.

Prominent among cyberneticists, W. R. Ashby formulated the 'Law of Requisite Variety' and the 'Good Regulator Theorem' (**BOX 1**), two complementary principles that are germane to our considerations[1,35]. These principles offer a useful perspective on the nature of prediction. They imply that considering the structure of relevant environmental influences has the potential to provide important and principled insights into fundamental design features of the agent (for similar ideas, see REFs [3,36]). Taking this idea further, we believe that a consideration of the regularities in the environment that impact on, and must thus be modelled by, the agent, provides an opportunity to elucidate the form of the predictions that are required.

Different types of environmental regularities can be categorized according to the spatiotemporal scales over which they impact on the agent. We suggest that some regularities are spatiotemporally global, that is, they are not limited to specific spatial or temporal locations,) and are relevant for each encounter between agent and environment. They are independent of contextual factors and are therefore non-hierarchical: that is, the existence of the regularity is immutable and is not dependent on the context or on other states of the environment. It can therefore be hypothesised that modelling such regularities requires that the agent possesses similarly context-independent, non-hierarchical predictions. Below, we review a growing body of evidence from which we conclude that prior knowledge allowing the prediction of spatiotemporally global regularities is embedded in the structure of, and thereby affects and constrains, bottom-up information-processing(**FIG. 2a-d**). These prior-based influences are predictive in the sense that they are estimations of relevant (context-independent) aspects of the agent's environment that are not predicated on current sensory input. We hypothesise that they act automatically and

ineluctably on every stimulus that an agent encounters and they determine and shape our interaction with the world at all times.

The notion of global predictive constraints, though not new[37], is neglected in current predictive processing accounts, which concern themselves primarily with a different form of prediction, one that relates to spatiotemporally local regularities[2,3,10,13,23-26,30]. These regularities are present, and impact on the agent, only in specific contexts. Owing to their context-dependency, they are nested within a hierarchical structure, in which the current environmental state determines their presence or relevance. In modelling these regularities, the agent's brain must mirror their characteristics. This can be achieved through a hierarchical top-down processing system: higher-level information processing mechanisms extract the current context and feed the resulting prediction back to lower-level units to modulate earlier processing[2,3,10,13,23-26,30] (**FIG. 2e-g**).

Both types of regularity-to-prediction mappings have their equivalence in the cybernetic steersman analogy: the material and shape of the boat are constant because the core properties of the medium in which the steersman must travel are constant. By contrast, the deployment of the structural features of the boat must change in a context-dependent manner to meet the challenges created by changing features of the environment (wind, tide, current). In short, the agent has unchanging features, which regulate the unchanging influences of its world, but also context-dependent features, which mirror and regulate the context-dependent features of its world. In the next section, we outline the evidence supporting this distinction.

Henceforward, we use the term 'constraint' to refer to prior information that relates to context-independent regularities and, consequently, forms the basis for context-independent predictions. The term is derived from the computational vision field[37], where it refers to a similar idea. It intuitively captures the notion that the structure of the nervous system forces information processing to proceed along predetermined paths. The trajectory of these paths is an estimation of the agent's environment based on prior information and is, thus, predictive. Note that we do not argue for a broadening of the definition of the term 'prediction'. Rather, we argue that a consistent application of a computational definition of

this term leads us to consider constraints on bottom-up processing as being predictive. Contrasting with 'constraints', we use the term 'expectation' to refer to prior information relating to context-dependent regularities. Attentional top-down influences are functionally different from the predictive processing that is our focus here (**BOX 3**), and we therefore do not discuss them in detail.

[H1] Regularities and predictions

[H2] Context-independent regularities and constraints

The natural world seems highly varied. Yet, surprisingly, images of most natural scenes — from Alpine meadows to Mediterranean coastlines — show a large degree of similarity in their general statistical properties: in the distribution of orientations of local edges, the shapes of contours and the positions of objects, for example[38,39]. Sensory systems exploit these regularities to maximise the amount of information they encode, to optimise performance and to minimise metabolic cost[38]. A growing number of studies exploring the neural implementation of the integration of sensory evidence with knowledge-based predictions suggest that prior information about global, context-independent regularities is implicitly embedded within the structure of information-processing mechanisms[40-43]. For instance, the distribution of orientation in natural images is not uniform: vertical and horizontal orientations are overrepresented[42] (**FIG. 2a,b**). As is evident in perceptual biases towards the cardinal axes, and in higher sensitivity of neurons to stimuli oriented close to these axes, observers exploit this non-uniformity when perceiving local orientation (**FIG. 2c,d**) [42]. Critically, this constraint is thought to be implicitly embedded within the structure of primary visual cortex (V1): electrophysiological work in animals and fMRI in humans suggests that neurons tuned to the cardinal orientations are overrepresented in V1 and have narrower tuning functions than those tuned to other orientations[44,45]. These structural inhomogeneities implicitly represent prior information of the orientation statistics in natural scenes and provide a means for Bayesian inference to be performed in the absence of explicit representation of a prior in a top-down hierarchy[42]. Other regularities in the basic attributes of the environment, such as the speed at which objects move, have also been suggested to be implicitly represented by embedded constraints in the form of inhomogeneities of neuronal densities and tuning functions in relevant neural populations[46].

The grouping of individual features in the environment is also often characterised by global and context-independent regularities. For instance, the contours that define the visual boundaries of objects follow certain regularities. When extracting contours from an image, the human visual system uses prior information of typical contour shape to group local information into larger units. This grouping mechanism is called an 'association field', which results in a grouping of local orientation that closely match the statistical regularities of contours in the world[47-49]. Such contours are critical for defining image features and objects, and animal studies suggest that the selectivity of horizontal connections between neurons in early retinotopic cortices might play an important role in establishing the association field[50,51]. For instance, a recent animal study of the 'silent' surround – the part of visual space in which a stimulus is insufficient to trigger activity in a neuron by itself but can modulate its activity –of orientation-tuned V1 neurons suggests that the horizontal connectivity structure in V1 is spatially laid out in a pattern highly similar to the psychophysical association field in humans [51]. The horizontal connectivity pattern in V1 therefore provides a plausible neurophysiological mechanism for the implementation of predictions regarding contours based on prior knowledge of environmental structure. These structurally embedded constraints ensure that local orientation information is integrated into contours – which are important to separate objects from their background – in line with prior knowledge of the structure of the world. These context-independent mechanisms of contour integration are likely to be complemented by additional flexible top-down influences[8].

Even complete objects show certain basic regularities that are context-independent. For instance, contours belonging to objects tend to be convex relative to the object[52]. While a range of mechanisms[53], including top-down processes[54,55], are involved in the visual system's separation of a figure from its background, there is evidence to suggest an important role for context-independent predictions that are embedded as a convexity constraint on bottom-up processing. Feedforward and feedback neural connections terminate in different layers of V1[56,57], providing an opportunity to distinguish top-down modulation from bottom-up processes (which include horizontal influences). Using laminar recordings in macaque monkeys, it has been demonstrated that horizontal connectivity in

V1 plays an important role in early aspects of figure–ground separation[58]. Importantly, computational models suggest that the specificity of the facilitatory and inhibitory horizontal connections between neuronal circuits in V1 allows them to implement predictions based on a convexity constraint that exploits context-independent object structure[59,60].

Another example of context-independent regularities relates to the fact that specific object types are often found in highly predictable locations. Because of the structured way in which agents interact with the environment, regularities in world-centred coordinates often translate into regularities in the visual field (i.e., in retinotopic coordinates). For instance, grass and carpets are typically found in the lower half of our visual field, faces and text in the centre and tree-top canopies in the upper half. Intriguingly, recent evidence using population receptive-field mapping in humans suggests that predictions regarding these positional regularities are embedded within the receptive field properties of high-level visual cortex neurons that are tuned to specific object categories[61,62]: specifically, their receptive fields are biased towards the locations of the visual field in which the preferred object category is typically found[62]. For instance, word-selective neuronal populations exhibit receptive fields that are small, biased towards central vision, and extend more horizontally than vertically in English speakers[61]. Similar correspondences between receptive field properties in neurons tuned to certain categories of objects and the typical location of these objects in visual space can be found in faces and scenes[63].

Overall, the evidence described above serves to illustrate that information relating to time- and space-invariant statistical regularities of environmental properties is implicitly encoded in stable, structural components of the information-processing system. The resulting predictions thus act on, or constrain, bottom-up information processing. Here, we have largely focussed on perceptual processing, but similar examples of context-independent constraints can be found in learning and other cognitive domains. For example, it has been shown that not all environmental regularities are learned equally well by all organisms[64]. Rather, there seem to be constraints on the readiness to form associations that might reflect an embedded model based on environmental regularities relevant for a given organism. A fascinating recent example comes from a study that exposed two groups of

Drosophila to experimental environments, in which a visual or an olfactory cue, respectively, was a reliable predictor of an aversive chemical stimulus[65]. After 40 generations, the insects had evolved into two lines with a readiness to form an association with the respective relevant cue. Embedding prediction in constraints on bottom-up information processing allows the organism to maximise information content and performance, while keeping metabolic costs at a minimum[41]. It is therefore not surprising that algorithms used in artificial intelligence exploit similar context-independent regularities, and embed them within the structure of artificial networks[31].

The question arises as to the origins of these embedded constraints. In biological organisms, it is often conceptually and methodologically difficult to tease apart the contribution of phylogeny and ontogeny in these embodied processes. A detailed discussion of this question goes beyond the scope of this paper, but the existing evidence suggests there is no one-size-fits-all explanation. It seems most likely that the neural circuits underlying constraints are established by an interaction between the shaping of the developing nervous system in response to sensory stimulation during sensitive periods and a phylogenetically-determined predisposition[66,67]. However, in some extreme examples in both humans and animals, constraints appear to be independent of sensory experience and to be determined by a genetically-defined blueprint[68,69]. It might also be possible for constraints to emerge in response to consistent experience during adulthood.

[H2] Context-dependent regularities and expectations

Many regularities in an agent's environment are context-dependent (**FIG. 2e-g**). For instance, a forest walk makes an encounter with woodland birds more likely than an encounter with a wader. Identification of the bird species in turn leads to a high-level representation that predicts other features — which might not be available at the time of identification — such as the presence of a specific type of beak[70]. The beak determines lower-level regularities such as the presence of specific contours or oriented edges in a specific part of visual space. In this example, context-dependency refers to the fact that the regularities of local, low-level features of the input are determined by its higher-level aspects, or by information that is independent of this input. Thus, in addition to its context-

independent regularities, our environment is further structured in a hierarchical and nested manner: higher-level aspects of the environment induce, or determine regularities at a lower level. We hypothesise that such context-dependent environmental regularities must be paralleled by similarly hierarchical and nested information-processing mechanisms in the brain. We suggest that this is achieved by the implementation of top-down influences within a hierarchy of processing steps, such that higher-level processes extract contextual information, derive predictions, and feed them back to modulate earlier aspects of perceptual processing.

Top-down processes have been characterised at many different levels of the cortical hierarchy. In the ornithological example above, the visual scene context determines which animal is likely to be encountered. There is substantial evidence to suggest that the brain uses such scene–object dependencies to aid and modulate object perception[71]. For instance, objects presented in their typical scene contexts are identified faster and more accurately[72]. This facilitation is thought to be based on predictions that are rapidly derived from the scene in high-level context-specific cortices and are fed back to shape lower-level object representations[73]. Once an object's rough outline is segmented and separated from the scene, predictions about regularities at a smaller scale are derived, leading to a highly dynamic interaction between the processing of local features and the representation of the segmented object[5]. As is to be expected, given the hierarchical and nested nature of the environment, electrophysiological evidence in primates suggests that top-down influences filter down the information processing hierarchy in a highly specific manner, reaching even some of the earliest levels of information processing in subcortical structures such as the lateral geniculate nucleus[74,75].

In the examples discussed above, the information from which predictions are derived is largely provided by the sensory input itself. However, a small but growing body of literature suggests that top-down effects that mirror environmental regularities go far beyond those that are input-based. For example, the expert knowledge of field ornithologists allows them to detect and identify bird species within a split second. Psychophysical and neuroimaging evidence has indeed shown that specific object-knowledge plays an important role in the segregation of a figure from its background, exerting its influence via top-down

modulation[54,55]. Moreover, detailed psychophysical studies suggest that expectations about specific object properties and even semantic meaning, which are both acquired prior to encountering the object and are thus stored in high-level memory systems, can flexibly shape the properties of early visual feature-detectors by top-down modulation[6,76-78] (**FIG. 2e-g** and **BOX 2**). These psychophysical findings are consistent with neuroimaging studies showing that prior object-knowledge, which is represented in a distributed network including high-level frontal and parietal areas, dynamically interacts with visual processing in early retinotopic areas[79,80]. Interestingly, both psychophysical[81] and neuroimaging studies[11] suggest that merely the expectation of specific stimulus properties activates feature-specific templates in early visual cortices.

Social interactions in humans, and other social animals, are highly context-sensitive[82] and several studies indicate that predictions derived from prior social knowledge are an important source of top-down influences on information processing. For instance, psychophysical evidence suggests that the human brain uses knowledge of context-dependent regularities of social interactions in a top-down manner to guide processing of motion patterns generated by other people[83]. Even factors such as the mental states attributed to another person, for example the intention to initiate a movement, can have top-down effects that influence how social input is processed by early sensory processes [84-87].

Repeated exposure to the same or similar sensory stimulation also affects sensory processing and perception, a phenomenon often called adaptation. A number of different effects are subsumed under this term, and their mechanisms are not well understood[24,88-91]. Current models largely reject passive 'neural fatigue' as an explanation and regard adaptation as a set of active processes[88,89]. However, there is no consensus on whether adaptation is underpinned by predictive processing[92-95]. Effects such as the reduced neural response to repeated or predictable stimuli, or the closely related enhancement in response to unpredictable stimuli, are thought to be linked to context-dependent predictions[24,96] based on top-down processing[93,96-100]. However, the picture is highly complex, since bottom-up processes have also been shown to contribute to adaptation effects[91,94]. Experimental work linking adaptation to bottom-up processing under constant viewing conditions, but to

top-down processing under variable viewing conditions[101,102] suggests the intriguing possibility of two separate mechanisms: a context-independent constraint that acts on bottom-up processing and context-dependent predictions underpinned by top-down processing that take effect in changeable environments.

In summary, there are numerous instances in which the predictability of the environment is context-dependent. Under such circumstances, the predictive information embedded in constraints on bottom-up processing described in the previous section is unhelpful. Rather, the optimal behaviour of an agent will depend upon its ability to deploy predictions that can flexibly modulate information processing via top-down processes. It is this second form of prediction that is the sole focus of current predictive processing models.

4. Implications and applications

One might argue that there is no need to distinguish different forms of prediction because all ultimately serve the same purpose[103]: to facilitate inference about the state of the world and thereby optimise an organism's interaction with it. This point is reasonable when efforts are directed towards high-level, functional descriptions of behaviour, as is the case for many models concerned with optimality[29]. Such models provide an invaluable benchmark against which to evaluate an agent's performance from a functional perspective.

However, the predictive processing framework frequently makes an additional mechanistic commitment: the default assumption is that predictive processes are mediated by top-down mechanisms[10,23-26]. As we show above, this view is incomplete: predictive information can be implemented in the brain in at least two broad forms. If a mechanistic understanding is our goal, a correction to the current unitary view is essential. We see a number of ways in which recognising this distinction enhances and extends the value of the predictive processing framework.

[H2] Linking computational models to mechanisms
If a computational approach only recognises top-down predictions, it risks a disconnect between modelling and mechanistic insight, which in turn can impede progress generated

by the feedback loop between models and empirical research. Examining predictive processing accounts of illusions[25,104-108] illustrates this problem. 'Illusions' generally arise from sensory inputs that are deliberately manufactured to violate the predictions of perceptual systems. In current predictive processing accounts, illusions are therefore conceptualised as resulting from top-down processing. For instance, the classic Cornsweet effect[109] – in which two equiluminant patches separated by a central graded section appear to differ in terms of luminance – has been treated as an illustration of top-down influences, relating to prior beliefs about spatial gradients of luminance and reflectance[25] and has been simulated using a network explicitly implementing top-down predictions[104]. However, though it might be modulated by higher-level influences[110], there is evidence to suggest that most of the effect is due to predictive information that is embedded in early subcortical or even retinal processes: the Cornsweet effect can be theoretically linked to the receptive field structure of retinal ganglion cells[111], is strongly correlated with signals recorded in the lateral geniculate nucleus[112] (to which retinal ganglion cells project) and has been demonstrated to arise from monocular neurons[112], suggesting that it is of subcortical origin. More generally, a major challenge for the top-down processing account of illusions is the finding that the neural circuits responsible for their emergence can be independent of any prior experience: congenitally or early-onset blind individuals experience certain illusions, such as the Müller-Lyer illusion, immediately after eye-sight restoring surgery[68].

We do not contest that 'illusions' such as the Cornsweet effect are experienced because of predictive processing. Importantly, however, we argue that many (but not necessarily all) of these phenomena are better explained by predictive information realised in context-independent constraints on bottom-up processing rather than top-down modulation. To provide another example, a number of 'illusions' are related to the light-from-above prior[113], which we discuss in detail in the next section. The disconnect between empirical evidence and model demonstrates that an explanation might have descriptive validity at a computational level but be misguided at the mechanistic level. Interestingly, it has been shown that even Rao and Ballard's seminal predictive coding framework can be reformulated in such a way that predictions are implemented by lateral inhibition rather than feedback connections[14,30,114].

Clearly we need a much tighter integration of mechanistic insight and computational modelling as well as a move towards greater precision in distinguishing different forms of predictions at different levels of granularity. For instance, one recent finding tentatively supports the view that context-dependent predictions might share a common source and context-independent predictions that might rely on inbuilt constraints share another source[115], but much more detail is required. We see the recognition that predictive information can be implemented in two broad forms (constraints and expectations) as a first step that may help to inspire models that retain the computational benefits of predictive processing but are mechanistically more precise and more powerful in their capacity to elucidate neural mechanisms. For example, these models might stimulate investigations of the implications of a single computational principle being implemented by different mechanisms at the systems or behavioural level.

[H2] Elucidating interactions between predictions

Within hierarchical predictive processing accounts, predictions are generated at a number of different levels of the information-processing hierarchy. Predictions at different levels are thought to interact with each other via top-down mechanisms to ensure that, ultimately, all predictions are mutually consistent[13,14,30]. However, it is likely that bottom-up constraints and top-down expectations might interact in a fundamentally different way. Since constraints will remain largely unaltered by short-term changes in expectations, constraints and expectations may affect the same processes but will not directly influence each other to become mutually aligned.

The 'light-from-above prior' provides an instructive example. The direction from which light hits a visual scene and the resulting shading provides information that the human visual system uses to infer object shape[113]. In the absence of explicit information about the position of the light-source, human observers judge object shape in a way that suggests that the visual system implicitly predicts that the light comes from above[113]. Interestingly, this prediction can be modified through experience[116]: after training in which observers receive feedback indicating that the position of the light-source has shifted, the visual system's predictions move towards the new location. The conventional predictive processing account of this phenomenon suggests that the top-down predictive information has been

updated[117]. However, this is inconsistent with findings that the new prior is specific to the laboratory context[118,119], as are priors arising from other learned cues [120]. An alternative interpretation, which is consistent with this context-specificity and with electrophysiological evidence[121], is that the light-from-above prior is implemented as a constraint on bottom-up processing, and is unchanged by short-term experience. The experimental training, rather than changing the original constraint, produces a novel, context-specific expectation that light has shifted[118,119].

Our hypothesis about the interactions between different forms of prediction leads to an interesting prediction: that when information from constraints and expectation interacts, the former might never be fully overwritten by the latter. This stands in contrast to the purely top-down predictive processing account, wherein the ultimate aim is to ensure that all predictions are mutually consistent. Furthermore, our account suggests that experimental manipulations of top-down processing should differentially affect the newly acquired prediction but might leave the original prior intact. This differential effect should furthermore be observable at a neural level. For example, we would predict that neuroimaging experiments would show the effects of short-term learning of a shifted light-source location in higher processing areas, while neural signatures of early, bottom-up processes thought to underpin the constraint[121] would remain unchanged.

This form of interaction between two different forms of prediction could help to strike an optimal balance between robustness and flexibility, allowing for context-related modification of the expression of embedded constraints without altering either the constraints themselves or their expression outside the narrow confines of this context. Interestingly, the interaction appears not to be common to all organisms: even after extended experience, chickens perceive object shape in a way that suggests their visual system assumes light to come from above[69]. From birth, it appears, chickens have an immutable bottom-up constraint predicting light from above and are unable to acquire a context-dependent expectation to modulate these inbuilt predictions, powerfully illustrating that predictive information can be entirely decoupled from an individual's experience with the world.

[H2] Identifying mechanisms to inform clinically-relevant questions

Clinical practice and research in psychiatric and neurological illnesses are hampered by the fact that heterogeneous symptom clusters may be classed within the same diagnostic categories while, conversely, different diagnoses may be underpinned by overlapping neurophysiological disturbances[122,123]. Therefore, research often eschews standard diagnostic categories to focus on single symptoms, seeking a narrower but deeper understanding of the mechanisms by which these symptoms arise. The predictive processing framework has been an important part of this enterprise[15-17,82,106]. But its focus on one form of prediction has constrained the extent to which it can account for the diversity of symptoms.

In particular, models that are mechanistically misguided or under-developed will face problems, even if they correctly capture symptoms at the computational level. Instances in which different mechanistic disturbances underlie two apparently similar clinical patterns, provide an illustration of this issue. For example, a current challenge in neuropsychiatry concerns the clinical overlap — but underpinning neurobiological distinction — between autoantibody-mediated psychosis and other forms of psychosis (such as that found in schizophrenia) [124]. Increasing attention has been drawn to the significant proportion of people who present with psychotic experiences and are found to have IgA antibodies to NMDA receptors in their serum or cerebro-spinal fluid. These antibodies may be, but are not necessarily, the underlying cause of the psychosis[124]. Since the treatment for antibody-mediated psychosis – immunotherapy – should only be administered when clinically indicated, clinicians are faced with a decision that demands a mechanistic rather than a computational understanding. Put more simply, although a predictive processing model may be agnostic to mechanisms and still provide an adequate high-level explanation for psychotic symptoms, practical clinical considerations, such as treatment selection, demand a mechanistic account.

As questions relating predictive processing to clinical symptoms and syndromes become more specific, a comprehensive perspective on the different forms of prediction, and their interactions, will yield important insights. One important question is whether different types of predictive information can compensate for each other, and what the implications of such

compensation are at the systems and behavioural level. Hallucinations, perceptual experiences that occur in the absence of an external stimulus, provide a useful illustration of this idea. Hallucinations are associated with a range of different psychiatric and neurological disorders[125-127], are experienced by a surprisingly large number of healthy individuals[128] and are also associated with drugs, both therapeutic and recreational[129,130]. They are a key feature of schizophrenia where they have been hypothesised to occur as a consequence of over-reliance on top-down, predictive processing[15,17,105,131,132]. Paradoxically, however, the predictive processing framework has also been used to derive the opposite conclusion about the nature of the imbalance in psychosis — that is, that it is underpinned by an under-reliance on top-down processing – both clinically[107] and as a part of the psychedelic drug experience[129]. This idea has been partly inspired by the relative resistance of people with psychosis to illusions [105-108] (but see REF [133]), which are generally conceptualised as arising from top-down processing. To resolve this apparent inconsistency, it has recently been speculated that people with hallucinations exhibit under-weighting of top-down processing early in the hierarchy (conferring illusion-resistance) and over-weighting of top-down processing higher up in the hierarchy (conferring hallucination-proneness)[105,106]. However, this account fails to specify what is considered to be low or high levels of the processing hierarchy. Furthermore, as we discuss above, the empirical evidence does not support the notion that all illusions are due to top-down processing based on prior experience.

Recognising different forms of prediction offers an alternative perspective on this paradox, which might add (or be an alternative) to the existing explanation: a weakening of embedded constraints would confer resistance to illusions and, to compensate for the resulting reduced influence of constraints, would also enhance the relative impact of top-down influences that produce the hallucinatory experiences[17]. Moreover, since hallucinations occur in multiple disorders with varying pathologies, it is likely that a predictive processing account that encompasses both embedded constraints and context-dependent expectations, as well as the interaction between them, provides important explanatory potential. On the one hand it offers perspectives on the lower level perceptual changes that may accompany symptoms in psychotic illness and that are often neglected[134,135] in favour of more complex disturbances and their associated high-level explanations. On the other hand it may provide deeper understanding of how complex

symptoms such as hallucinations, which are similarly defined in psychosis, dementia, or sensory disturbance[127], may (despite superficial similarities) be explained in terms of very different underlying mechanisms. These advantages may transfer well to considerations of the multiple pharmacological manipulations that produce psychotic experiences: the variable effects produced by a number of drugs have already been recognised to demand a consideration of both top-down and bottom-up effects[130] and a readiness, therefore, to consider predictions in their distinct forms will enrich this theoretical perspective.

[H2] Relating an agent's neural toolbox to environmental statistics

We have suggested that characteristics in the environment underpinning predictions are usefully classed as context-independent or context-dependent regularities, and that these map onto constraints and expectations, respectively. A move towards acknowledging the importance of these fundamentally different forms of prediction may inspire a more general extension of the predictive processing framework than we have outlined in the previous sections. In particular, the aim of linking an agent's information processing to environmental statistics has been a guiding principle in work on natural scene statistics[38] and offers, we suggest, a similar, principled framework for guiding research in predictive processing. For instance, analysis of environmental regularities might allow us to predict and explain the extent to which predictions are mediated by top-down processes or by constraints on bottom-up processing, or why differences exist in the susceptibility of predictive processes to context-dependent, short-term changes. More generally, formal analyses of relevant regularities in an organism's environment might prove useful in developing a unifying framework that is able to explain why an organism's predictions in different domains might come in different forms. This idea provides a means of linking the predictive processing framework to insights from sensory ecology and evolutionary theory[136].

[H1] Conclusions

The principle underlying our argument is simple: critical design features of an agent's information-processing mechanisms can be understood by observing the structure of its environment. This basic idea is not new and drove the thinking of early cognitive

biologists[137] and cyberneticists[34], as well as current work on natural scene statistics[38]. Here, we have argued that applying this principle to the role of prediction in brain function points to a fresh and useful perspective. To achieve a deeper understanding of the brain as a 'prediction machine'[33], we suggest that it is necessary to recognise that prediction in the nervous system comes in different forms. Immutable regularities of an environment are mirrored in corresponding context-independent predictive mechanisms that act on bottom-up processing, mechanisms that we have referred to as constraints. Fluctuating, context-dependent regularities in the world likewise point towards flexible, context-dependent predictive mechanisms implemented by top-down processes, which we refer to as expectations. As predictive processing develops and is applied to ever more specific, mechanistically-based questions, other distinctions may usefully be made but we argue that the one suggested here is primary and fundamental.

While we have explored these ideas mainly with respect to neural information processing, a comprehensive view of the role of prediction in information-processing must ultimately include an appreciation that the whole organism interacts with its environment[138]. Predictive information is present not only in neural mechanisms but in the morphology of the organism as a whole, as is illustrated in numerous examples of sensory and behavioural ecology. Ultimately, the predictive processing framework should aim to incorporate the many different ways in which prediction is part of biological information-processing in order to offer more comprehensive insights into how we interact with our world in health and disease.

1. Conant, R. C. & Ashby, W. R. Every good regulator of a system must be a model of that system. *International Journal of Systems Science* **1,** 89–97 (1970).
2. Lee, T. S. & Mumford, D. Hierarchical Bayesian inference in the visual cortex. *J. Opt. Soc. Am. A* **20,** 1434–1448 (2003).
3. Friston, K. The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* **11,** 127–138 (2010).
4. Kersten, D., Mamassian, P. & Yuille, A. Object Perception as Bayesian Inference. *Annu. Rev. Psychol.* **55,** 271–304 (2004).
5. Neri, P. Object segmentation controls image reconstruction from natural scenes. *Plos Biol* **15,** e1002611 (2017).
6. Teufel, C., Dakin, S. C. & Fletcher, P. C. Prior object-knowledge sharpens properties of early visual feature-detectors. *Sci Rep* **8,** 1–12 (2018).
7. Liang, H. *et al.* Interactions between feedback and lateral connections in the primary visual cortex. *Proc. Natl. Acad. Sci. USA* **114,** 8637–8642 (2017).
8. Li, W., Piëch, V. & Gilbert, C. D. Learning to link visual contours. *Neuron* **57,** 442–451 (2008).
9. Nurminen, L., Merlin, S., Bijanzadeh, M., Federer, F. & Angelucci, A. Top-down feedback controls spatial summation and response amplitude in primate visual cortex. *Nature Communications* **9,** 1–13 (2018).
10. Keller, G. B. & Mrsic-Flogel, T. D. Predictive Processing: A Canonical Cortical Computation. *Neuron* **100,** 424–435 (2018).
11. Kok, P., Failing, M. F. & de Lange, F. P. Prior Expectations Evoke Stimulus Templates in the Primary Visual Cortex. *J Cognitive Neurosci.* **26,** 1546–1554 (2014).
12. Muckli, L. *et al.* Contextual Feedback to Superficial Layers of V1. *Curr. Biol.* **25,** 2690–2695 (2015).
13. Rao, R. & Ballard, D. H. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* **2,** 79–87 (1999).
14. Spratling, M. W. A review of predictive coding algorithms. *Brain Cogn* **112,** 92–97 (2017).
15. Powers, A. R., Mathys, C. & Corlett, P. R. Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. *Science* **357,** 596–600 (2017).
16. Fletcher, P. C. & Frith, C. D. Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat. Rev. Neurosci.* **10,** 48–58 (2009).
17. Teufel, C. *et al.* Shift toward prior knowledge confers a perceptual advantage in early psychosis and psychosis-prone healthy individuals. *Proc. Natl. Acad. Sci. USA* **112,** 13401–13406 (2015).
18. Homan, P. *et al.* Neural computations of threat in the aftermath of combat trauma. *Nat Neurosci* **22,** 470–476 (2019).
19. Schutter, D. J. L. G. A Cerebellar Framework for Predictive Coding and Homeostatic Regulation in Depressive Disorder. *Cerebellum* **15,** 30–33 (2015).
20. Barrett, L. F., Quigley, K. S. & Hamilton, P. An active inference theory of allostasis and interoception in depression. *Phil. Trans. R. Soc. B* **371,** 20160011 (2016).
21. Palmer, C. J., Seth, A. K. & Hohwy, J. The felt presence of other minds: Predictive

processing, counterfactual predictions, and mentalising in autism. *Conscious. Cogn.* **36,** 376–389 (2015).

22. Reichert, D. P., Seriès, P. & Storkey, A. J. Charles Bonnet Syndrome: Evidence for a Generative Model in the Cortex? *PLoS Comput Biol* **9,** e1003134 (2013).

23. Bar, M. The proactive brain: memory for predictions. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **364,** 1235–1243 (2009).

24. Summerfield, C. & de Lange, F. P. Expectation in perceptual decision making: neural and computational mechanisms. *Nat. Rev. Neurosci.* **15,** 745–756 (2014).

25. Lupyan, G. & Clark, A. Words and the World: Predictive Coding and the Language-Perception-Cognition Interface. *Curr. Dir. Psychol.* **24,** 279–284 (2015).

26. Gilbert, C. D. & Li, W. Top-down influences on visual processing. *Nat. Rev. Neurosci.* **14,** 350–363 (2013).

27. Hochstein, S. & Ahissar, M. View from the Top. *Neuron* **36,** 791–804 (2002).

28. Berger, J. O. *Statistical Decision Theory and Bayesian Analysis*. (Springer, 1985). doi:10.1029/2004GL021870/full

29. Kording, K. Decision Theory: What 'Should' the Nervous System Do? *Science* **318,** 606–610 (2007).

30. Huang, Y. & Rao, R. P. N. Predictive coding. *Wiley Interdisciplinary Reviews: Cognitive Science* **2,** 580–593 (2011).

31. Lee, T. S. The Visual System's Internal Model of the World. *Proc. IEEE* **103,** 1359–1378 (2015).

32. Aitchison, L. & Lengyel, M. With or without you: predictive coding and Bayesian inference in the brain. *Curr. Opin. Neurobiol.* **46,** 219–227 (2017).

33. Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain. Sci.* **36,** 181–253 (2013).

34. Pickering, A. *The Cybernetic Brain*. (The University of Chicago Press, 2010).

35. Ashby, W. R. Requisite Variety and Its Implications for the Control of Complex Systems. *Cybernetica* **2,** 83–99 (1958).

36. Mumford, D. On the computational architecture of the neocortex. *Biol. Cybern.* **66,** 241–251 (1992).

37. Marr, D. *Vision*. (The MIT Press, 1982).

38. Geisler, W. S. Visual Perception and the Statistical Properties of Natural Scenes. *Annu. Rev. Psychol.* **59,** 167–192 (2008).

39. Torralba, A. & Oliva, A. Statistics of natural image categories. *Network: Comput. Neural Syst* **14,** 391–412 (2003).

40. Shi, L. & Griffiths, T. L. in *Advances in Neural Information Processing Systems NIPS* (eds. Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C. K. I. & Culotta, A.) **22,** 1669–1677 (2009).

41. Ganguli, D. & Simoncelli, E. P. Efficient Sensory Encoding and Bayesian Inference with Heterogeneous Neural Populations. *Neural Comp.* **26,** 2103–2134 (2014).

42. Girshick, A. R., Landy, M. S. & Simoncelli, E. P. Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nat Neurosci* **14,** 926–932 (2011).

43. Fischer, B. J. & Peña, J. L. Owl's behavior and neural representation predicted by Bayesian inference. *Nat Neurosci* **14,** 1061–1066 (2011).

44. Li, B., Peterson, M. R. & Freeman, R. D. Oblique Effect: A Neural Basis in the Visual Cortex. *J. Neurophysiol.* **90,** 204–217 (2003).

45. Furmanski, C. S. & Engel, S. A. An oblique effect in human primary visual cortex. *Nature* **3,** 535–536 (2000).

46. Seriès, P. & Seitz, A. R. Learning what to expect (in visual perception). *Front. Hum. Neurosci.* **7,** 1–14 (2013).

47. Geisler, W. S., Perry, J. S., Super, B. J. & Gallogly, D. P. Edge co-occurrence in natural images predicts contour grouping performance. *Vision Res.* **41,** 711–724 (2001).

48. Hess, R. F., May, K. A. & Dumoulin, S. O. in *Oxford Handbook of Perceptual Organization* (ed. Wagemans, J.) 189–206 (Oxford University Press, 2015).

49. Geisler, W. S. & Perry, J. S. Contour statistics in natural images: Grouping across occlusions. *Vis Neurosci* **26,** 109–121 (2009).

50. Bosking, W. H., Zhang, Y., Schofield, B. & Fitzpatrick, D. Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *J. Neurosci.* **17,** 2112–2127 (1997).

51. Gerard-Mercier, F., Carelli, P. V., Pananceau, M., Troncoso, X. G. & Frégnac, Y. Synaptic Correlates of Low-Level Perception in V1. *J. Neurosci.* **36,** 3925–3942 (2016).

52. Fowlkes, C. C., Martin, D. R. & Malik, J. Local figure-ground cues are valid for natural images. *J. Vis.* **7,** 1–9 (2007).

53. Wagemans, J. *et al.* A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychol. Bull.* **138,** 1172–1217 (2012).

54. Peterson, M. A. & Gibson, B. S. Must figure-ground organization precede object recognition? An assumption in peril. *Psychol Sci* **5,** 253–259 (1994).

55. Cacciamani, L., Scalf, P. E. & Peterson, M. A. Neural evidence for competition-mediated suppression in the perception of a single object. *Cortex* **72,** 124–139 (2015).

56. Felleman, D. J. & Van Essen, D. C. Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cereb. Cortex* **1,** 1–47 (1991).

57. Markov, N. T. *et al.* Anatomy of hierarchy: Feedforward and feedback pathways in macaque visual cortex. *Journal of Comparative Neurology* **522,** 225–259 (2014).

58. Self, M. W., van Kerkoerle, T., Supèr, H. & Roelfsema, P. R. Distinct Roles of the Cortical Layers of Area V1 in Figure-Ground Segregation. *Curr. Biol.* **23,** 2121–2129 (2013).

59. Zhaoping, L. Border ownership from intracortical interactions in visual area V2. *Neuron* **47,** 143–153 (2005).

60. Kogo, N., Strecha, C., Van Gool, L. & Wagemans, J. Surface Construction by a 2-D Differentiation-Integration Process: A Neurocomputational Model for Perceived Border Ownership, Depth, and Lightness in Kanizsa Figures. *Psychol. Rev.* **117,** 406–439 (2010).

61. Le, R., Witthoft, N., Ben-Shachar, M. & Wandell, B. The field of view available to the ventral occipito-temporal reading circuitry. *J. Vis.* **17,** –6 (2017).

62. Kaiser, D., Quek, G. L., Cichy, R. M. & Peelen, M. V. Object Vision in a Structured World. *Trends Cog Sci* (2019).

63. Silson, E. H., Groen, I. I. A., Kravitz, D. J. & Baker, C. I. Evaluating the correspondence between face-, scene-, and object-selectivity and retinotopic organization within lateral occipitotemporal cortex. *J. Vis.* **16,** 14–14 (2016).

64. Shettleworth, S. J. *Cognition, Evolution, and Behavior*. (Oxford University Press,

| | **Formatted:** Font: Italic |

2009). n *Learning and memory* 189–201 (2017).

65. Dunlap, A. S. & Stephens, D. W. Experimental evolution of prepared learning. *Proc. Natl. Acad. Sci. USA* **111,** 11750–11755 (2014).

66. White, L. E. & Fitzpatrick, D. Vision and Cortical Map Development. *Neuron* **56,** 327–338 (2007).

67. Blakemore, C. & Cooper, G. F. Development of the brain depends on the visual environment. *Nature* **228,** 477–478 (1970).

68. Gandhi, T., Kalia, A., Ganesh, S. & Sinha, P. Immediate susceptibility to visual illusions after sight onset. *Curr. Biol.* **25,** R358–R359 (2015).

69. Hershberger, W. Attached-shadow orientation perceived as depth by chickens reared in an environment illuminated from below. *J Comp Physiol Psych* **78,** 407–411 (1970).

70. Svensson, L., Grant, P. J., Mullarney, K. & Zetterström, D. *Bird Guide*. (Harper Collins Publishers, 2001).

71. Bar, M. Visual objects in context. *Nat. Rev. Neurosci.* **5,** 617–629 (2004).

72. Bar, M. & Ullman, S. Spatial context in recognition. *Perception* **25,** 343–352 (1996).

73. Brandman, T. & Peelen, M. V. Interaction between Scene and Object Processing Revealed by Human fMRI and MEG Decoding. *J. Neurosci.* **37,** 7700–7710 (2017).

74. Sillito, A. M., Cudeiro, J. & Jones, H. E. Always returning: feedback and sensory processing in visual cortex and thalamus. *Trends Neurosci.* **29,** 307–316 (2006).

75. Jones, H. E. *et al.* Figure-ground modulation in awake primate thalamus. *Proc. Natl. Acad. Sci. USA* **112,** 7085–7090 (2015).

76. Lupyan, G. Objective effects of knowledge on visual perception. *J. Exp. Psychol. Hum. Percept. Perform.* **43,** 794–806 (2017).

77. Neri, P. Semantic control of feature extraction from natural scenes. *J. Neurosci.* **34,** 2374–2388 (2014).

78. Christensen, J. H., Bex, P. J. & Fiser, J. Prior implicit knowledge shapes human threshold for orientation noise. *J. Vis.* **15,** 24–24 (2015).

79. Hsieh, P. J., Vul, E. & Kanwisher, N. Recognition Alters the Spatial Pattern of fMRI Activation in Early Retinotopic Cortex. *J. Neurophysiol.* **103,** 1501–1507 (2010).

80. Flounders, M. W., Gonzalez-Garcia, C., Hardstone, R. & He, B. J. Neural dynamics of visual ambiguity resolution by perceptual prior. *eLife* **8,** 1–25 (2019).

81. Wyart, V., Nobre, A. C. & Summerfield, C. Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. *Proc. Natl. Acad. Sci. USA* **109,** 3593–3598 (2012).

82. Griffin, J. D. & Fletcher, P. C. Predictive Processing, Source Monitoring, and Psychosis. *Annu. Rev. Clin. Psychol.* **13,** 265–289 (2017).

83. Neri, P., Luu, J. Y. & Levi, D. M. Meaningful interactions can enhance visual discrimination of human agents. *Nat Neurosci* **9,** 1186–1192 (2006).

84. Moore, J. W., Teufel, C., Subramaniam, N., Davis, G. & Fletcher, P. C. Attribution of Intentional Causation Influences the Perception of Observed Movements: Behavioral Evidence and Neural Correlates. *Front. Psychol.* **4,** 1–11 (2013).

85. Teufel, C. *et al.* Social Cognition Modulates the Sensory Coding of Observed Gaze Direction. *Curr. Biol.* **19,** 1274–1277 (2009).

86. Liepelt, R., Cramon, von, D. Y. & Brass, M. What is matched in direct matching? Intention attribution modulates motor priming. *J. Exp. Psychol. Hum. Percept. Perform.* **34,** 578–591 (2008).

87.   Teufel, C. *et al.* What is social about social perception research? *Front. Integr. Neurosci.* **6,** (2012).

88.   Solomon, S. G. & Kohn, A. Moving Sensory Adaptation beyond Suppressive Effects in Single Neurons. *Curr. Biol.* **24,** R1012–R1022 (2014).

89.   Webster, M. A. Visual adaptation. *Ann. Rev. Vis. Sci.* **1,** 547–567 (2015).

90.   Grill-Spector, K., Henson, R. & Martin, A. Repetition and the brain: neural models of stimulus-specific effects. *Trends Cog Sci* **10,** 14–23 (2006).

91.   Vogels, R. Sources of adaptation of inferior temporal cortical responses. *Cortex* **80,** 185–195 (2016).

92.   Stocker, A. A. & Simoncelli, E. P. Sensory adaptation within a Bayesian framework for perception. *Adv Neural Inf Process Syst* **18,** 1–8 (2006).

93.   Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M.-M. & Egner, T. Neural repetition suppression reflects fulfilled perceptual expectations. *Nat Neurosci* **11,** 1004–1006 (2008).

94.   Vinken, K., de Beeck, H. P. O. & Vogels, R. Face Repetition Probability Does Not Affect Repetition Suppression in Macaque Inferotemporal Cortex. *J. Neurosci.* **38,** 7492–7504 (2018).

95.   Chopin, A. & Mamassian, P. Predictive Properties of Visual Adaptation. *Curr. Biol.* **22,** 622–626 (2012).

96.   Wacongne, C., Changeux, J.-P. & Dehaene, S. A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. *J. Neurosci.* **32,** 3665–3678 (2012).

97.   Todorovic, A., van Ede, F., Maris, E. & de Lange, F. P. Prior Expectation Mediates Neural Adaptation to Repeated Sounds in the Auditory Cortex: An MEG Study. *J. Neurosci.* **31,** 9118–9123 (2011).

98.   Wacongne, C. *et al.* Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proc. Natl. Acad. Sci. USA* **108,** 20754–20759 (2011).

99.   Schwiedrzik, C. M. & Freiwald, W. A. High-Level Prediction Signals in a Low-Level Area of the Macaque Face-Processing Hierarchy. *Neuron* **96,** 89–97 (2017).

100.   Parras, G. G. *et al.* Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. *Nature Communications* **8,** 2148 (2017).

101.   Ewbank, M. P. *et al.* Changes in 'Top-Down' Connectivity Underlie Repetition Suppression in the Ventral Visual Pathway. *J. Neurosci.* **31,** 5635–5642 (2011).

102.   Ewbank, M. P., Henson, R. N., Rowe, J. B., Stoyanova, R. S. & Calder, A. J. Different Neural Mechanisms within Occipitotemporal Cortex Underlie Repetition Suppression across Same and Different-Size Faces. *Cereb. Cortex* **23,** 1073–1084 (2013).

103.   de Lange, F. P., Heilbron, M. & Kok, P. How Do Expectations Shape Perception? *Trends Cog Sci* **22,** 764–779 (2018).

104.   Brown, H. & Friston, K. J. Free-Energy and Illusions: The Cornsweet Effect. *Front. Psychol.* **3,** (2012).

105.   Corlett, P. R. *et al.* Hallucinations and Strong Priors. *Trends Cog Sci* **23,** 114–127 (2019).

106.   Sterzer, P. *et al.* The Predictive Coding Account of Psychosis. *Biol. Psychiatry* **84,** 634–643 (2018).

107.   Adams, R. A., Stephan, K. E., Brown, H. R., Frith, C. D. & Friston, K. J. The computational anatomy of psychosis. *Front. Integr. Neurosci.* **4,** 1–26 (2013).

108.   Notredame, C.-E., Denève, S. & Jardri, R. What visual illusions teach us about

schizophrenia. *Front. Integr. Neurosci.* **8,** 1–16 (2014).

109. Cornsweet, T. N. *Visual Perception*. (HBJ Publishers, 1970).

110. Purves, D., Shimpi, A. & Lotto, R. B. An empirical explanation of the cornsweet effect. *J. Neurosci.* **19,** 8542–8551 (1999).

111. Dakin, S. C. & Bex, P. J. Natural image statistics mediate brightness filling in. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270,** 2341–2348 (2003).

112. Anderson, E. J., Dakin, S. C. & Rees, G. Monocular signals in human lateral geniculate nucleus reflect the Craik–Cornsweet–O'Brien effect. *J. Vis.* **9,** 1–18 (2009).

113. Ramachandran, V. S. Perception of shape from shading. *Nature* **331,** 163–166 (1988).

114. Olshausen, B. A. & Field, D. J. Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Res.* **37,** 3311–3325 (1997).

115. Tulver, K., Aru, J., Rutiku, R. & Bachmann, T. Individual differences in the effects of priors on perception: A multi-paradigm approach. *Cognition* **187,** 167–177 (2019).

116. Adams, W. J., Graf, E. W. & Ernst, M. O. Experience can change the 'light-from-above' prior. *Nat Neurosci* **7,** 1057–1058 (2004).

117. Panichello, M. F., Cheung, O. S. & Bar, M. Predictive Feedback and Conscious Visual Experience. *Front. Psychol.* **3,** 620 (2013).

118. Adams, W. J., Kerrigan, I. S. & Graf, E. W. Efficient Visual Recalibration from Either Visual or Haptic Feedback: The Importance of Being Wrong. *J. Neurosci.* **30,** 14745–14749 (2010).

119. Kerrigan, I. S. & Adams, W. J. Learning different light prior distributions for different contexts. *Cognition* **127,** 99–104 (2013).

120. Knill, D. C. Learning Bayesian priors for depth perception. *J. Vis.* **7,** 13–13 (2007).

121. Mamassian, P., Jentzsch, I., Bacon, B. A. & Schweinberger, S. R. Neural correlates of shape from shading. *NeuroReport* **14,** 971–975 (2003).

122. Cuthbert, B. N. & Insel, T. R. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Medicine* **11,** 126 (2013).

123. Xia, C. H. *et al.* Linked dimensions of psychopathology and connectivity in functional brain networks. *Nature Communications* **9,** 3003 (2018).

124. Pollak, T. A. *et al.* Autoimmune psychosis: an international consensus on an approach to the diagnosis and management of psychosis of suspected autoimmune origin. *Lancet Psychiatry* **7,** 93–108 (2020).

125. O'Callaghan, C. *et al.* Visual hallucinations are characterised by impaired sensory evidence accumulation: Insights from hierarchical drift diffusion modelling in Parkinson's disease. *Biol. Psychiatry* **2,** 680-688(2017).

126. Urwyler, P. *et al.* Visual Hallucinations in Eye Disease and Lewy Body Disease. *The American Journal of Geriatric Psychiatry* **24,** 350–358 (2016).

127. Waters, F. & Fernyhough, C. Hallucinations: a systematic review of points of similarity and difference across diagnostic classes. *Schizophrenia Bulletin* **1,** 32–43 (2017).

128. McGrath, J. J. *et al.* Psychotic Experiences in the General Population: A Cross-National Analysis Based on 31 261 Respondents From 18 Countries. *JAMA Psychiatry* **72,** 697–705 (2015).

129. Carhart-Harris, R. L., Friston, K. J. & Barker, E. L. REBUS and the Anarchic Brain: Toward a Unified Model of the Brain Action of Psychedelics. *Pharmacol Rev* **71,**

316–344 (2019).

130. Corlett, P. R., Frith, C. D. & Fletcher, P. C. From drugs to deprivation: a Bayesian framework for understanding models of psychosis. *Psychopharmacology* **206,** 515–530 (2009).

131. Cassidy, C. M. *et al.* A Perceptual Inference Mechanism for Hallucinations Linked to Striatal Dopamine. *Curr. Biol.* **28,** 503–514 (2018).

132. Schmack, K. *et al.* Delusions and the role of beliefs in perceptual inference. *J. Neurosci.* **33,** 13701–13712 (2013).

133. Grzeczkowski, L. *et al.* Is the perception of illusions abnormal in schizophrenia? *Psychiatry Res.* **270,** 929–939 (2018).

134. Keane, B. P., Cruz, L. N., Paterno, D. & Silverstein, S. M. Self-Reported Visual Perceptual Abnormalities Are Strongly Associated with Core Clinical Features in Psychotic Disorders. *Front. Integr. Neurosci.* **9,** 646 (2018).

135. McGhie, A. & Chapman, J. Disorders of attention and perception in early schizophrenia. *Brit. J. Med. Psychol.* **34,** 103–116 (1961).

136. Geisler, W. S. & Diehl, R. L. Bayesian natural selection and the evolution of perceptual systems. *Phil. Trans. R. Soc. B* **357,** 419–448 (2002).

137. Uexküll, J. *Theoretische Biologie*. (Paetel, 1920).

138. Seth, A. K. in *Open MIND* (eds. Metzinger, T. & Windt, J.) (2015).

139. Srinivasan, M. V., Laughlin, S. B. & Dubs, A. Predictive Coding: A Fresh View of Inhibition in the Retina. *Proc. Royal Soc. B* **216,** 427–459 (1982).

140. Heeger, D. J. Theory of cortical function. *Proc. Natl. Acad. Sci. USA* **114,** 1773–1782 (2017).

141. Blank, H. & Davis, M. H. Prediction Errors but Not Sharpened Signals Simulate Multivoxel fMRI Patterns during Speech Perception. *Plos Biol* **14,** e1002577 (2016).

142. Richter, D., Ekman, M. & de Lange, F. P. Suppressed Sensory Response to Predictable Object Stimuli throughout the Ventral Visual Stream. *J. Neurosci.* **38,** 7452–7461 (2018).

143. Han, B., Mostert, P. & de Lange, F. P. Predictable tones elicit stimulus-specific suppression of evoked activity in auditory cortex. *NeuroImage* **200**, 242-249 (2019).

144. Meijs, E. L., Slagter, H. A., de Lange, F. P. & van Gaal, S. Dynamic Interactions between Top-Down Expectations and Conscious Awareness. *J. Neurosci.* **38,** 2318–2327 (2018).

145. Kok, P., Jehee, J. F. M. & de Lange, F. P. Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron* **75,** 265–270 (2012).

146. Carrasco, M. Visual attention: The past 25 years. *Vision Res.* **51,** 1484–1525 (2011).

147. Desimone, R. & Duncan, J. Neural Mechanisms of Selective Visual-Attention. *Annu. Rev. Neurosci.* **18,** 193–222 (1995).

148. Summerfield, C. & Egner, T. Expectation (and attention) in visual cognition. *Trends Cog Sci* **13,** 403–409 (2009).

149. Jiang, J., Summerfield, C. & Egner, T. Attention Sharpens the Distinction between Expected and Unexpected Percepts in the Visual Brain. *J. Neurosci.* **33,** 18438–18447 (2013).

150. Kok, P., Rahnev, D., Jehee, J. F. M., Lau, H. C. & de Lange, F. P. Attention Reverses the Effect of Prediction in Silencing Sensory Signals. *Cereb. Cortex* **22,** 2197–2206 (2012).

Fig. 1: Possible neural network implementation of predictive coding

a| The basic components of a predictive coding network and the basic equations necessary to perform inference[14]. In the schematic, neural units within a simulated predictive coding network are indicated by circles and facilitatory and inhibitory connections are shown. Two error units and two prediction units are shown, the dashed grey lines indicate potential additional units. The subscripts enumerate the units and their inputs. The activity of prediction units (s) signals the estimate of the current world state. Using this estimate, a generative model of the world (V) generates a prediction about the input (p, equation 1), which is transmitted from prediction units to error units via weighted connections . Differences (prediction errors) between the actual input (i) and the predicted input (p) are indicated by the activity of error units (e, equation 2). The prediction error is then used to recursively adjust the activity of the prediction units so that they better estimate the current world state (equation 3). The parameter μ controls the extent of adjustment in the prediction units; T indicates that the transpose of V is used. While only non-predicted information (the prediction error) is transmitted between error and prediction units, prediction units use this information to hone in on a sharpened representation of the state of the world. In this version of predictive coding, V — which represents the generative model of the world — is often (but not always) thought to be implemented in the synaptic

30

weights of the feedback connections between prediction and error units. Over longer time periods, learning can lead to adjustments of this model. **b|** The most influential predictive coding accounts[3,30] assume a hierarchical implementation of the basic components illustrated in part **a**. That is, predictions are generated in units in higher-level cortical areas and are communicated to error units in lower-level cortical areas via feedback connections. For instance, units at level 1 provide feedback to units at level 0 (the input level), while those at level 2 provide feedback for units at level 1. Horizontal connectivity between units at the same hierarchical level ensures that prediction units at different levels of the hierarchy are driven towards mutually consistent predictions by the reciprocal influences of the intervening error units. However, while this predictive coding scheme is currently the most widely used, it is not the only one possible (**Box 2**).

Fig. 2: Context-dependent and context-independent predictions.

Local orientation – the information contained in a small patch of visual space relating to the orientation of, for example, a contour or an edge in that area – provides an illustration of predictions based on context-independent and context-dependent regularities. **a|** Examples of possible natural visual scenes. **b|** Analysis of local orientation across a large number of such images shows that horizontal (0 and 180 degree) and vertical (90 degree) orientations are overrepresented in natural scenes[42] (please note that the images used for this study did not include those shown in part a). The y-axis indicates the probability of a certain orientation being found in a local patch of an image of a natural scene. **c|** Computational and empirical evidence suggests that the early stages of visual processing in humans implicitly exploit this context-independent regularity to constrain bottom-up information processing. Specifically, the tuning properties of orientation-tuned neurons are biased towards more frequently encountered orientations[42,44,45]. Thus, context-independent predictions about inhomogeneities in local orientation are applied to all stimuli fed through the visual system. **d|** The prior used by the human visual system when judging local orientation has been derived from psychophysical data using a Bayesian framework. This prior (dark green line) shows a close correspondence to environmental regularities (light green line). Critically, the simulated, neuronal inhomogeneities shown in c implement a prior that is very similar to that used by human observers **e|** Local orientations also show strong context-dependent regularities. For instance, the recognition of hills in the left image

allows prediction of the trajectory of local orientation along a hill (as indicated by the grey arrows). Conversely, the recognition of a face as a face (right image) and the identification of different parts of the face can be used to predict the likely orientation of, the cheek and jaw line, for example[6] (indicated by the white arrows). These predictions are highly context-specific and differ depending on the type and precise spatial location of the object in the visual scene. **f|** That the visual system relies on such predictions and implements them via top-down modulation of early sensory processes has been shown in a psychophysical study[6]. In this study, observers viewed small 'edge probes' that were embedded within ambiguous stimuli. An example of this type of stimulus is shown in the right panel: please note that this is an illustrative example only and was not one of the stimuli used in the study in REF 6 (which cannot be shown due to copyright). The orientation of the edge probes was determined by the orientation information extracted from contours in an unambiguous version of the same visual scene (left panel). The ambiguous stimuli, within which the probes were embedded, were only perceived as coherent objects after observers received prior object-knowledge by exposure to the umambiguous version of the same visual scene. (before receiving this knowledge they were perceived as meaningless patches)[6]. Prior object-knowledge thus provides control over object representations while sensory evidence remains identical. **g|** The manipulation of object knowledge combined with precise psychophysical measurements demonstrated that, when local oriented edges are embedded within a meaningful object, neurons in the primary visual cortex (V1) dynamically hone their sensitivity to the currently relevant features, leading to a sharpening of low-level detector properties[6]. The plots show the difference in orientation between two edge probes on the horizontal axis, and the performance of observers in a task in which they had to discriminate between the two probes on the vertical axis. Orientation sensitivity is measured as the smallest difference in orientation between two probes that observers are able to discriminate at a pre-specified performance level (red and blue dashed lines). This sensitivity threshold is illustrated by the red and blue double-headed arrows. Observers performed the same task with the same stimuli once before (blue lines) and once after having received prior knowledge allowing manipulation of object perception (red lines). Without prior object knowledge the stimuli were perceived as meaningless patches; however, with knowledge the observers saw objects. The left panel shows performance when edge probes were congruent with the object that the observer perceived. Here, the

difference between the red and the blue double-headed arrows illustrates that observers were more sensitive to the orientation of the edge when the probe was embedded in a meaningful object percept than when they perceived the edge as being embedded in meaningless patches. The panel on the right illustrates the same task except that edges were incongruent with the observer's object percept. In this case, there is no difference in sensitivity before and after having received prior object knowledge. The right hand image in part a is reproduced with permission from REF 17. Parts b-d are adapted, with permission, from REF[42]. Parts e (right hand image) and f are adapted, with permission from REF 17. Panel g is adapted, with permission, from REF[6].

Box 1: Learning from cybernetics

In the field of cybernetics (as in the fields of reinforcement learning and artificial intelligence) it is convenient to divide systems, sets of interacting components that form a unified whole, into agents and their environments. What is considered the system, agent, and environment depends on the specific question. For example, the system might be a fish (agent) responding to the eddies and currents of a stream (environment), or it might be subcomponents of the fish's skeletal apparatus (agent) acting within the overall body (environment). Alternatively, we might think of the system as the stream and all of its lifeforms. Cybernetics is concerned with how agents interact with their surroundings to maintain internal stability in the face of changes in their environment. Early cyberneticists recognised that this premise could also be used to frame investigations into the brain and were mindful that the optimal responses of an agent to environmental disturbances would be proactive rather than reactive, i.e., prediction is required[1]. One of pioneers of this field, W. R. Ashby, and his co-workers produced two axioms highlighting the nature of agents that successfully resist environmental perturbation:

- **Variety in the agent is required to deal with variety in its environment** – Within the field of cybernetics 'variety' refers to the number of states that a system (the environment or the agent) can adopt. If an agent is to control or regulate the impact of its environment in order to maintain its own internal stability, then it must have a repertoire of states at least as great as the number of ways in which the

environmental can affect it. This is the so-called Law of Requisite Variety[35]. Note that this law refers to the agent's capacity to maintain the parameters of its own states within desirable limits by responding to environmental influences, rather than to its capacity to influence directly the environment. Thus, the number of requisite states in the agent must match or exceed the number of ways in which its environment may perturb or influence it, rather than the number of states that this environment can visit.

- ***Good regulators are models of their environment***: Intimately related to the Law of Requisite Variety is the Good Regulator Theorem[1], a mathematical formalisation of the observation that, for an agent to successfully mitigate the impact of environmental states, it must, in some sense, be a model of its environment: that is, it must have a variety of states that map onto, or directly correspond to relevant environmental states.

Early cyberneticists were mainly concerned with performance of simple agents. One celebrated example of this was a series of autonomous 'tortoise' robots that engaged in complex interactions with their environments based on a few simple reactive responses to environmental changes[34]. Cyberneticists quickly recognised that the same principles applied to information processing in the brain[1]: successful responses to environmental influences require an agent's control structures to model and predict the relevant aspects of the environment. This means that we can learn much about the brain's design features by scrutinising its environment.

Box 2: Predictive coding and related computational approaches

The basic idea underlying predictive coding is that the brain capitalises on an internal, or generative, model of the word to actively predict incoming information (Fig. 1a). The deviation of this predicted information from the actual information received is used to drive the inferential process towards a best estimate of the current state of the world[3,13,14,30]. A number of algorithms have been proposed for predictive coding, differing in fundamental aspects such as the form of the generative model, the criteria used to drive inference and the nature of the information being transmitted for further processing[14]. Hypotheses about the neural implementation of predictive coding have also varied: while some models suggest a bottom-up form of predictive coding[30,139], one of the most influential forms of predictive coding[3,13] proposes a hierarchically organised system (Fig. 1b), in which predictions are generated at higher levels to be fed back for comparison to inputs at earlier levels. Following such comparison, the processing of correctly-predicted information is suppressed, whereas prediction errors are passed on to the next level of neural processing for further processing and to drive inference to achieve the overall goal of prediction-error minimisation.

Although suppression or 'dampening' of correctly-predicted information is emphasised in this model of predictive coding recent psychophysical evidence suggests that prediction can serve to make the perceptual representations of predicted events more distinct than those of unpredicted events[6,76-78] (Fig. 2f,g). This sharpening is not necessarily inconsistent with predictive coding theory, which suggests that activity in units that generate predictions rapidly converges on a fine-tuned representation. Nevertheless, other computational approaches that include predictive components place a stronger emphasis on this sharpening[2,32,140] and might prove useful in helping to explain it. Neuroimaging and electrophysiological findings are mixed, with some suggesting that prediction leads to dampening of predicted information[141-143] and others suggest sharpening[79,80,144,145].

Box 3: Attention, expectation, and constraints

Information processing in the brain is limited by the computational capacity of the neural apparatus. Attention is used to mitigate this burden and to ensure that limited resources are flexibly dedicated to where they are most needed, leading to a range of different top-down modulatory effects on early information processing[146]. Prioritisation of information by attentional mechanisms is typically thought to be based on behavioural relevance[146,147]: attention selects parts of the sensory input for further processing depending on an agent's task, goals, and intentions. This situation contrasts with the predictive processes we discuss in the main text. These processes respond to context-dependent statistical regularities in the external environment, rather than to internal motivational and intentional factors. Such processes are often conceptualised in terms of an agent's expectation[24,103,148]. In short, in this conceptualisation, attention modulates information processing in relation to what is relevant, whereas expectation modulates information processing by predicting what is likely to happen.

Many studies in the attention literature have conflated behavioural relevance and likelihood of occurrence[24,103,148]. Consequently, the differences in the behavioural and neural consequences of expectation and attention are not well understood. Furthermore, there is currently no consensus on exactly how the terms 'attention' and 'expectation' should be theoretically demarcated. However, a growing number of studies, including most of those we discuss in this article, employ experimental manipulations to isolate or independently target these processes, showing distinguishable effects of expectation and attention on information processing[6,77,81,149,150]. Thus, the distinction between attention and expectation has not only clear heuristic value but also growing empirical support[24]. We argue that we should further distinguish between the expectation of context-dependent events and constraints that are based on context-independent regularities.

36

Glossary

Agent: In the fields of cybernetics, reinforcement learning, and artificial intelligence an agent is an entity that is capable of acting autonomously to self-regulate in the face of changes in its environment.

Information theory: The mathematical formulation of how information is coded, transmitted, and processed. Informally, information can be thought of as a measure of the reduction of uncertainty. The field of information theory emerged from attempts to solve the problem of how to transfer large datasets within limited capacity systems and has proven useful in thinking about how neural systems deal with a similar problem.

Bayesian decision theory: A theory that describes how decisions are optimised by application of principles from Bayesian probability: that is, by drawing on probability distributions that quantify prior probabilities of events or states. These probabilities are referred to as priors and reflect beliefs about a state before new evidence is taken into account*.*

Perceptual and cognitive inference: The process by which perceptions and beliefs arise from the combination of sensory evidence and information based on prior experience or knowledge. The process of inference may be optimised by using prior knowledge according to Bayes theorem.

Prediction: An estimate of unobserved or missing information on the basis of a model. Within the predictive processing framework, the model is provided by prior knowledge of the world. Note that, prediction can be (but is not necessarily) future-oriented.

Predictive coding: Within neuroscience, a family of algorithms aiming to capture how the brain performs probabilistic inference using the mismatch between predicted and expected magnitude of a signal.

Priors: In Bayesian models of perception, action, and cognition, the term is used as shorthand for 'prior probability distributions', which model the system's information about a world state before current evidence is assessed. Importantly, priors provide information that is the basis of the formation of predictions. It is important to note that the term is agnostic as to how this prior information is implemented, making combined terms, such as 'top-down prior', which implies a specific mechanism, confusing.

Utility: In Bayesian decision theory, a function that determines the value of a possible situation or outcome.

TOC blurb

Prior experience is incorporated into the brain's predictive models of the world, enabling the accurate interpretation of and responses to new sensory information. In this Perspective, Teufel and Fletcher make the case for an important distinction between two forms of prediction that may advance our understanding of brain function.