How do expectations shape perception?

Floris P. de Lange[1,*], Micha Heilbron[1,*], Peter Kok[2,*]

[1] Donders Institute for Brain, Cognition and Behaviour, Radboud University, Kapittelweg 29, 6525 EN Nijmegen, the Netherlands

[2] Department of Psychology, Yale University, 2 Hillhouse Avenue, New Haven, CT 06520, USA

[*] All authors contributed equally

Correspondence: floris.delange@donders.ru.nl (F.P. de Lange).

Keywords

Prediction, Perception, Sensory Processing, Bayesian Inference, Predictive Coding, Perceptual Inference

Abstract

Perception and perceptual decision-making are strongly facilitated by prior knowledge about the probabilistic structure of the world. While the computational benefits of using prior expectation in perception are clear, there are myriad ways in which this computation can be realized. Here, we review recent advances in our understanding of the neural sources and targets of expectations in perception. Furthermore, we discuss Bayesian theories of perception that prescribe how an agent should integrate prior knowledge and sensory information, and investigate how current and future empirical data can inform and constrain computational frameworks that implement such probabilistic integration in perception.

.

Expectation in perception

Humans, like many other species, are 'anticipatory systems' [1]. They construct predictive models of themselves and their environment, allowing them to quickly and robustly make sense of incoming data. In line with this notion, the brain has been described as a 'prediction machine' [2] that attempts to match incoming sensory inputs with top-down expectations. Although the concept of the predictive brain is not new, dating back at least to Helmholtz [3], the neural implementation of such a predictive architecture has remained rather elusive. The last decade has witnessed a renewed interest in how neural circuits may capitalize on prior expectations to facilitate neural computations [4] in various domains, ranging from perception and action to high-level reasoning and language. In this article, we will review the general principles that emerge from recent empirical work on how expectations modify the neural computations that underlie sensory processing and perception. We will discuss how expectations change perception, the neural sources and targets of expectation, and we will examine how these empirical data inform and constrain theoretical computational models of perception, such as predictive coding [5]. Finally, we will highlight recent advances from computational psychiatry [6,7] that show how neurodevelopmental disorders like schizophrenia and autism may be understood as resulting from aberrant weighting of expectations [8,9]. We hope that this Review will stimulate more research into what we believe is a central computation of the brain: the prediction of incoming input.

Perceptual consequences of expectation

We live in a highly predictable world, in which most objects remain stable and things change only slowly over time. This allows us to build internal models that can predict upcoming input on the basis of past and present input. Such expectations may prepare sensory cortex for processing, thereby increasing perceptual sensitivity for expected stimulus features. But what are the consequences of expectation for perception?

There are several ways in which expectation can influence perception, as demonstrated by various concrete examples of such influences (see Box 1). When sensory input is weak, noisy or ambiguous, expectation can bias perception: changing not only *how well*, but even *what* is perceived. For instance, the direction in which a cloud of dots is seen to be moving can be biased by an implicit expectation of the most likely direction when the contrast of the dots [10] or the coherence of motion direction in the cloud is low [11]. Similarly, when input is ambiguous, expectation can strongly modulate what people perceive. For instance, after learning to associate a particular set of colored spectacles with either leftward or rightward moving dots, participants were more likely to perceive fully ambiguously moving dots as moving in the direction that was associated with the glasses they wore [12]. Biasing effects of expectation can also occur more indirectly. Moving line segments behind an aperture are consistent with either slow movement perpendicular to the lines, or fast movement oblique to the lines [13]. A recent study showed that changing participants' expectations by exposing them extensively to slow or fast moving lines made them more likely perceive to the perpendicular or oblique movement, respectively [14].

However, when stimuli are unambiguous, or when expectations are highly inaccurate (seeing a face when expecting a house) expectations are less likely to bias the contents of perception. Under such circumstances, the effects of expectations are more subtle and often restricted to the efficiency of perception, resulting in stimuli being detected more rapidly [15–17] and accurately [18] when validly predicted, compared to when not.

In general, then, it can be said that the relative impact of expectations versus sensory input on perception depends on their relative reliability (i.e. 'precision') [19]. Observers rely most strongly on prior knowledge when expectations are reliable and stimuli ambiguous, but rely most strongly on the input when expectations are weak and stimuli reliable. This form of uncertainty weighting fits naturally with computational models that cast perception as Bayesian inference (see 'Expectation in computational models of perception').

Where do expectations come from?

The brain can predict future input by learning about and exploiting statistical regularities in its inputs [20] – but how does it achieve this? Since such regularities come in different shapes and forms (see Box 2), the neural mechanisms likely depend on the type of regularity.

Arguably the simplest regularity in our sensory input is that certain features appear more often, and are thus generally more likely than others. For instance, cardinally oriented (i.e., horizontal and vertical) lines are more prevalent than oblique ones [21] (Fig. 1A), shadows are more likely to appear underneath objects than above them since light usually comes from above [22,23], and objects in the periphery of our visual field often move away from the center of our gaze (centrifugally) due to the optic flow resulting from forward locomotion [24]. Such regularities, which are the result of stable physical features of the world we live in - and thus relatively constant throughout an individual's lifetime - allow the organism to form prior expectations. Due to their static nature, these prior expectations (or priors) are likely to be learnt over relatively long time scales, leading them to become encoded in the tuning properties of our sensory cortices [25]. The overrepresentation of cardinal orientations, for instance, is mirrored by an overrepresentation of early visual neurons that are tuned for cardinal relative to oblique orientations, as well as narrower tuning curves for these orientations, resulting in more precise representations [21,26] (Fig. 1A). Similarly, the bias towards centrifugal motion in our visual periphery has a neural counterpart in terms of an overrepresentation of neurons in motion sensitive area MT tuned to centrifugal motion [24]. Intriguingly, recent research has shown that such long term expectations can nevertheless be modulated by recent experiences [14,20,27], suggesting interactions between long term and dynamic, context-dependent priors that may be encoded in higher-order brain regions. It should be noted that such modulations do not occur in all circumstances [28], and sometimes require extensive and specific cross-modal training [27].

In addition to the simple frequency distribution of sensory inputs, expectations can also arise from conditional probabilities. For instance, the spatial context can constrain which objects are more or less likely to appear: driving down a street makes it more likely that

4

other moving objects are cars rather than sharks, and vice versa if you're scuba diving. In this example, a higher level representation of the scene constrains the lower level elements making up that scene. More generally, it can be said that there is a hierarchical structure to the visual world – with oriented lines and curves making up simple shapes, shapes making up objects, and objects being arranged into scenes – and this visual hierarchy is mirrored by the hierarchical organization of visual cortex [29]. Alongside bottom-up (i.e., low-to-high, or simple-to-complex) signal flow in the visual cortical hierarchy, there is also a wealth of top-down connectivity, which may allow higher-level representations (e.g., a street) to communicate which lower-level features (e.g., a car) to expect [30,31]. Such top-down expectations can occur at all levels of the cortical hierarchy. For example, neural representations of objects are enhanced by the presence of scene context, and this enhancement is correlated with the engagement of scene-selective cortical regions [32] (Fig. 1B). Similarly, simple feature representations in early visual cortex (V1/V2) are modulated by object context from LOC [33–35] and motion context from V5/MT [36,37].

Most of the expectations discussed above occur within one sensory modality (i.e.., vision). However, statistical regularities can also exist between sensory inputs from different modalities, such as when we hear the barking of our dog and have a very specific expectation of what we'll see next. Learning and exploiting such cross-modal associations necessarily requires the involvement of higher-order brain regions that receive inputs from – and send feedback to – multiple sensory modalities. When multimodal associations are largely stable over our lifetime, such as the relationship between lip movements and speech sounds, predictive associations can be encoded in multisensory regions such as in the superior temporal sulcus [38,39]. Other proposals of long term cross-modal predictions include a common prior for different forms of physical magnitude (e.g., distance, rotation angle, time) [40,41], potentially located in posterior parietal cortex [42], and integration of predicted and actual interoceptive signals in insula cortex [43,44]. In short, given that predictions play such a fundamental role in sensory processing, expectations likely exist at all levels of the cortical hierarchy.

However, cortical connections are known to modulate slowly, requiring many exposures and a relatively long time to learn new associations, whereas some expectations need to be learnt very rapidly (Figure IC) or are highly context-dependent, requiring high flexibility (Figure IB). For instance, imagine the task of learning the sound associated with the ice cream van. How does the brain learn to quickly exploit such complex associations? In such instances, brain regions outside of sensory cortex are likely required to quickly learn and store such cross-modal associations. One brain region that possesses all the necessary connectivity and computational machinery for this feat is the hippocampus. In fact, given its bidirectional connectivity with all sensory modalities [45], the hippocampus has been considered the apex of the cortical sensory hierarchy [46]. Furthermore, the hippocampus is known to be involved in rapidly learning associations between arbitrary stimuli [47–49], and, once learned, to retrieve associated items from memory upon presentation of a partial cue, and reinstate these items in sensory cortex [50,51]. Traditionally, these functions have mostly been considered in the context of episodic memory, but these same functions make the hippocampus ideally suited to subserve sensory prediction. In fact, recent studies in humans have shown that the hippocampus can encode predictive relationships between arbitrary stimuli [49,52], and the retrieval of such associations in hippocampus has been linked to the facilitations of predicted stimuli in visual cortex [53] (Fig. 1C). In addition to the hippocampus, there may be a role for other higher-order brain regions specialized in acquiring complex associations, such as in the prefrontal cortex [54–57], particularly when complex, semantic associations are involved.

Finally, sensory expectations can also be formed on the basis of one's own actions. It has long been suspected that the motor system sends 'corollary discharge' to sensory regions to compensate for the expected sensory consequences of motor commands, and there is now extensive evidence for movement-related modulations of sensory neurons, part of which has been traced back to motor and cerebellar cortex [58–60]. Some of these modulations seem to reflect internal predictive models similar to those proposed to underlie other perceptual expectations [58,61,62]. However, the goal of action-based expectations seems fundamentally different: namely to inhibit, rather than potentiate, the representation (and thereby perception) of expected sensory inputs. This

6

is for example apparent in the case of mormyrid electric fish, where inhibition prevents the fish from sensing its own electric discharges [63]. This example also shows that this type of expectation may be highly conserved and independent of the mammalian neocortex that the earlier discussed perceptual expectations arguably rely on.

How do expectations modulate sensory processing?

Expectations, alongside other cognitive factors such as attention and reward, strongly modulate the responsiveness of even the earliest sensory regions [64]. Overall, stimuli that are expected evoke a *reduced* neural response. This pattern has been observed when stimuli are validly predicted by a preceding, arbitrarily paired, stimulus [65–69], when stimuli comply with a higher-order pattern such as a shape or scene [33,34], or when stimuli are the predictable result of an animal's own actions [70]. This phenomenon, termed *expectation suppression* [64,68,71], is present even when low-level adaption is controlled for and has been consistently found across modalities such as vision [65–67,72–74] and audition [68,69,75,76]. Moreover, expectation suppression is observed not only in non-invasive fMRI BOLD [65,72,73] and EEG/MEG signals [68,69,77], but also in spike rates in non-human primates [66,67,74,76] and rodents [70,75,78].

Given the pervasiveness and robustness of this effect, several authors have suggested that expectation suppression reflects a general principle of cortical processing [4,5,64,79]. However, its precise explanation remains debated. Under one account, expected stimuli induce weaker responses because the brain filters out the expected components of sensory inputs. Response strength should then be a function of surprise and expectations hence *dampen* responses in neurons tuned for the expected stimulus (Fig. 2). Computationally, the 'dampening' account casts expectation suppression as a form of redundancy reduction, and the suppression of sensory signals by high-level expectations as analogous to the 'explaining away' of hypotheses in Bayesian networks [4,80]. Alternatively, the reduction of the overall amplitude may reflect *sharpening* of the underlying representation. In this view, neurons encoding not the expected but the

7

unexpected features are suppressed, resulting in a sharper, more selective population response with a lower overall amplitude (Fig. 2). Computationally, the 'sharpening' account of expectation is similar to the notion of neuronal resonance [79,81,82] in the sense that high-level feedback selectively enhances the representation of expected sensory signals, by inhibiting inputs that are inconsistent with feedback signals more strongly than consistent input. A similar physiological mechanism has also been proposed to underlie adaptation [71,83–85] and attention [86].

Several studies tried to arbitrate between these two accounts by investigating how expectations affect the pattern, rather than only the mean, of population activity. For example, it was found that while gratings whose orientation was validly predicted by a cue evoked *weaker* BOLD-responses in V1, the stimulus orientation could be decoded with higher accuracy from the pattern of activity [65]. The authors interpreted this as evidence for a sharpened representation. Further supporting this notion, they found a negative relation between the amount of suppression and stimulus preference: expectation suppression was weakest in voxels that 'preferred' the expected (and presented) orientation. Similarly, a recent study found that expectations weakened response amplitude in macaque IT but improved decoding performance[87]. However, other studies applying similar techniques found evidence in line with dampening. Such studies showed for instance that expectation suppression in macaque IT was associated with decreased decoding performance [66]; that BOLD responses in pSTS to expected words were weaker and contained *less* information, as quantified by RSA [88]; and that there was a *positive* relationship between expectation suppression and stimulus preference in the Macaque face patch system [74] and Macaque IT [67].

Since these studies used different measuring techniques and targeted different cortical regions, the discrepancies are difficult to interpret. Clearly, more work is needed to establish whether sharpening or dampening can best account for expectation suppression. Possibly, both play a role: some theories accommodate both mechanisms, for instance by proposing that dampening and sharpening take place in separate neural populations encoding respectively the errors and predictions [4]. Sometimes, expectations even *increase* the sensory response [89,90]. One explanation for this

effect is that in these experiments expectation ("What is likely?") and attention ("What is relevant?") covaried [64]. Indeed, studies that independently manipulated the two showed that attention can counteract [77,91] or even reverse [92] the suppressive effects of expectation. However, it is unclear if this explanation can account for all paradoxical effects of expectation [93–95] and more work is needed to further elucidate its interplay with attention.

Aside from *how* expectations affect responses, another relevant question is *when* they do so. One possibility is that expectations modulate responses after the initial 'feedforward sweep'. This is in line with some theoretical models [5] and with observations that putatively predictive phenomena such as extra-classical receptive field effects [5] and expectation suppression [68,96] often occur after the earliest components (>100ms) of the response. Another possibility is that expectations have an anticipatory effect and sensitize representations even *before* stimulus onset. Empirical support for such forward-looking perceptual predictions comes from electrophysiological reports of stimulus-specific baseline shifts before stimulus onset [67,97], from time-locked responses to unexpectedly *omitted* stimuli [69,98], and from predictive pre-activation of stimulus-specific activity patterns [99]. We should note, however, that these possibilities are not mutually exclusive and may well depend on the type of expectation at play. For instance, the precise temporal predictions required for anticipatory effects could be more readily available in the auditory domain, where regularities often manifest as patterns unfolding over time, than in the visual domain, where regularities are often marked by their spatial structure.

Expectation in computational models of perception
While it is clear that neural responses are heavily influenced by prior expectations, the computational role of these modulations is not yet fully understood. In general, expectations figure prominently in computational theories that cast perception as a process of probabilistic inference. As the information conveyed by our senses is both noisy and ambiguous, perception has often been conceptualized as a process of probabilistic inference, in which the system decides on the most probable causes of our

9

sensory inputs, based on the sensory data and prior expectations [3,100]. Bayesian probability theory provides a principled way of making such inferences, dictating that agents should form and update their beliefs on the basis of not only the evidence provided by the senses but also the prior probability of the various hypotheses about what is currently present in the world (i.e., expectations) [101]. Importantly, Bayesian inference is a *normative* theory, prescribing the best possible inference a system can make given the data. Interestingly, human observers often approach such an 'ideal observer' [102,but see 103], inspiring the idea that the human perceptual system may implement (approximate) Bayesian inference [104].

There are several ways in which Bayesian inference can be implemented in a neural architecture [105]. Many frameworks posit that feedback from higher-order areas provides contextual priors [4,79,106–108]. While it may seem natural to implement priors with a feedback process, it should be noted that Bayesian inference can be performed in a purely feedforward manner [109,110]. An advantage of recurrent architectures however is that they can perform *hierarchical* inference, in which the output (or 'posterior') from a higher level serves as an input (or 'prior') to a lower level. This allows the system to dynamically 'home in' on a globally coherent interpretation: a faint edge for instance could turn out to be a critical boundary only after the whole object or the entire scene is interpreted. Such hierarchical Bayesian inference can proceed concurrently across multiple areas, as successive areas in the cortical hierarchy constrain one another's inference in small loops rapidly and continuously as the interpretation evolves [79].

A computational theory that gained traction over the last decade is predictive coding. Predictive coding posits that the brain constructs an internal model of the world, encoding the possible causes of sensory inputs as parameters of a generative model [4,5]. Sensory inputs are compared to 'virtual' inputs that would have been generated by the expected cause. At each processing stage, only the difference between the input and the prediction (the *prediction error*) is propagated to higher regions for further processing, and perception is achieved by minimizing this error so as to match the predictions to the input. According to some more recent formulations of predictive

coding [111,111,112], the propagated error is by weighted by the sensory reliability or *precision*. When sensory precision is low, prediction errors are down-weighted and observers will rely more on predictions. Conversely, when it is high prediction errors are up-weighted and observers will rely more strongly on their input. Using this gain control of prediction errors, these models incorporate a specific version of a Bayesian formulation of attention (as reviewed in [113]).

A common point of confusion when describing predictive coding as a model of cortical function is the misconception that *only* prediction errors are encoded, as the predictable part of the input is subtracted out by the prediction. While there are example systems where the predictable part of the signal is indeed removed (e.g., in the retina, in order to increase the dynamic range [114]), predictive coding models of cortex always contain separate neural populations representing both the current best guess (prediction) as well as the error associated with the guess (prediction error). Another common misconception is that predictive coding is a computational model of the cognitive process of prediction (or 'forecasting'). While it is certainly possible to model such a process within a predictive coding architecture (e.g., by changing the baseline activity of a specific prediction in anticipation of sensory input; see [115,116]), predictive coding itself is a general theory of how the brain can efficiently encode information, and does not have the specific aim of explaining the cognitive process of prediction. In fact, most seminal models of predictive coding in cortex [4,5] do not include forward predictions: in these models, predictions are the result of an initial bottom-up analysis and are only formed *after* the first wave of feedforward activity. Such 'predictions' are thus *hypotheses* about the current sensory input, rather than *forecasts* of what is coming next (see Box 2).

Different versions of predictive coding have been developed [117], differing primarily in how the error is computed (by subtraction or division) and how prediction and prediction error neurons are connected [118]. At the same time, there are several other computational theories of perceptual inference that share the computational goal of optimal inference under uncertainty, such as pattern theory [119], adaptive resonance theory [106], particle filtering [79], free energy and active inference [120], and sampling-

based probabilistic inference [121,122]. The general motif of these theories is the notion that perceptual inference involves a top-down generative component that predicts and constrains the processing of bottom-up input over time [101].

To what extent are theoretical models like predictive coding supported by empirical evidence? While an exhaustive overview is beyond the scope of this Review (but see [2,123,124], we will highlight a few important observations. Firstly, extra-classical receptive field effects, both in space [5,125] and time [126], are readily explained by predictive coding as originating from predictive feedback. In line with this, removal of feedback signals within the visual system abolishes these extraclassical effects [127]. Secondly, the neural response to identical sensory input is markedly reduced when the input is predicted [67,74], putatively due to a reduced response of the error units to predictable input, as well as an improved selection of units encoding the relevant causes of the input. These two mechanisms, which have been referred to as 'shutting up' and 'stop gossiping' [4,104]; see also previous paragraph on dampening vs sharpening) are both accommodated within predictive coding, and operate on distinct populations (error and prediction units respectively).

Note that this also highlights a potential problem: Although many findings are consistent with frameworks like predictive coding, this partly derives from the relatively unspecific nature of some empirical findings. For instance, the fact that expectations tend to attenuate neural responses has been interpreted as a reduction in prediction error [4,5], but it could instead reflect a reinforcement of the expected interpretation at the expense of others [79], and therefore fits both the predictive coding framework [4] and other frameworks (e.g., [79]). Therefore, more direct tests of these computational frameworks will require experiments that target the distinguishing assumptions and predictions of individual models. For instance, while (most of) the aforementioned computational models posit the embodiment of a generative model in feedback connections, the models differ on the role of feedforward activity, with predictive coding proposing it signals the mismatch between the model and the data, while other theories [79,118] suggest that both feedforward and feedback signals contain (lower level vs. higher level) beliefs/hypotheses. One way to possibly arbitrate between these proposals is by

analysis of activity profiles at laminar resolution (See Box 3), which allows to distinguish feedforward and feedback activity within the cortical column. Also, temporally resolved signals may help to identify specific signals related to prediction and prediction error: for example, only prediction signals, but not prediction errors, can precede sensory input [87,128]. Furthermore, prediction error signals should rapidly decay as inference is formed, while confirmed prediction signals should remain stable [129]. As such, lamina-resolved data with high temporal and spatial resolution could have the potential to constrain and validate current computational theories of perceptual inference. Finally, apart from more precise recording techniques, future empirical work can benefit from more precise analyses. While much has been learned using simple approaches that manipulate variables of interest implicitly on a condition basis (e.g. by contrasting 'expected' versus 'unexpected' stimuli) this approach is less likely viable for subtler distinctions (such as between prediction error and precision-weighted prediction error). One way to make more precise comparisons is to use model-based analyses that use computational models of perceptual inference or learning (e.g. [130]) to derive trial-by-trial quantitative estimates of variables of interest, and compare those to the data [87,128,131].

Alterations of expectation in psychopathology

Expectations may not only be important to understand how the human brain implements perceptual inference, but also for understanding various psychopathological conditions. In particular, neurodevelopmental disorders like autism spectrum disorder [132,133] and schizophrenia [134,135] have been linked to an atypical integration of prior and incoming information, with autism even being cast as a 'disorder of prediction' [136]. While both conditions are linked to abberant expectations, they are proposed to fall at opposite sides of a spectrum. On the one hand, the false percepts (hallucinations) and beliefs (delusions) characteristic of schizophrenia are proposed to be caused by the misperception of inner states due to overly strong expectations [9], something that can also occur in healthy individuals [9,137]. Perceptual atypicalities in autism, on the other hand, are suggested to reflect impaired top-down guidance by expectations, resulting in

better processing of local details but impoverished global form perception [138], and possibly hypersensitivity to incoming (unexpected) stimuli. In recent years, these computational ideas are starting to be put to test in the burgeoning field of computational psychiatry [6,7]. For example, using a model of hierarchical Bayesian inference [139], researchers could link autism to a propensity to overestimate the volatility (i.e. changebility and thereby unpredictability) of the environment [8], and link schizophrenia to an overly strong reliance on perceptual priors [9]. More generally, these results underscore the integrative explanatory potential for computational models to reach a more complete understanding of brain function and its alterations [6].

Concluding Remarks and Future Perspectives

In this article we have discussed how the brain capitalizes on prior knowledge to facilitate the neural computations underlying sensory processing. While there are many different forms of prior knowledge, which can be neurally implemented in distinct ways, there appears to be a common currency in terms of their modulatory effect on target regions involved in the processing of sensory data. An interesting question for future research could be how all the prediction signals from different sources are ultimately combined during perceptual inference. One possibility is that all regions send their predictions to a shared 'blackboard' that resides in the primary visual cortex, facilitating the combining of different priors ([140,141]. Alternatively, the site of integration may depend on the level of detail afforded by the prediction [29,142].

We furthermore expect new insights in how the brain combines prior knowledge with input will be obtained from the renewed interest in deep learning algorithms, which has made tremendous progress in successfully implementing object recognition [143] and shares organizational overlap with the ventral visual stream of the brain [144–146]. Finally, technological advances in the measurement of multi-site activity with laminar precision [147,148] may afford the precision required to better understand the information flow within and between cortical modules, thereby constraining neurocomputational theories of perceptual inference. It is our hope that these advances

will ultimately lead us to a more complete understanding of the question how our brain combines the past and present in order to anticipate the future.

Box 1 – Prior knowledge influences perception

Our prior knowledge has a profound impact on the way we perceive the world. This knowledge may be acquired over a lifetime of experience, such as the fact that light usually comes from above, leading us to perceive shapes with shaping at the top (bottom) as concave (convex) (Figure IA).

Knowledge of the current (high-level) context can also impact perception [30,31]. For instance, we clearly perceive a street scene containing a car and a pedestrian in Figure IB. However, in this example, the 'car' and the 'pedestrian' are two identical blurry shapes, differing only in their orientation and location at which they are placed in the scene context [31]. In other words, the shapes themselves contain no information defining them as a car or a person, but it is the context of the street scene that leads us to perceive them as we do.

Prior experience can also modulate perception on very short timescales. Consider Figure IC; if you are unfamiliar with it, it likely appears a meaningless arrangement of black and white blocks. However, once you've been exposed to the grayscale version of the image once, your experience of the black and white image will be immediately and profoundly changed.

Demonstrations of the perceptual effects of expectations are not restricted to vision. In the sensorimotor domain, predictions of the sensory consequences of our motor commands help us distinguish external sensations from the mere consequences of our own movements, leading to the well-known fact that we can't tickle ourselves [149]. In hearing, our brains can 'fill in' corrupted syllables in speech with those that are most likely given the context of the surrounding words and sentences [150]. And just imagine expecting raisin bread but getting a mouthful of olive bread instead – you likely won't even be able to identify what it is you are eating initially.

Box 2 – What's in a prediction?

In cognitive neuroscience, the words 'expectation' and 'prediction' are used for various, seemingly disparate phenomena ranging from low-level biases in orientation judgements [21] to the effects of conscious expectation on speech perception [88]. Such flexible use of terms might seem needlessly inflationary, but becomes sensible when we consider the statistical, rather than mentalistic, definition of these words. Statistically, a *prediction* is an extrapolation from a model to potential observations. Whereas a model is specified via parameters over latent variables (such as weights in a linear regression) a prediction is specified in terms of observable data. Given our model, a prediction thus tells us what observations to expect. The mathematical *expectation* is then simply the predicted long-term average over those observations – or: the mean.

Using this statistical definition is useful because it allows us to see the myriad ways in which the brain can incorporate knowledge of what to expect that do not necessarily involve any "expecting" or "predicting" in the colloquial sense. A clear example is found in Bayesian models of low-level perceptual biases such as those in orientation judgement [21] (Figure 1A). Arguably, these biases reflect how our perceptual system is organized and do not involve neurons "predicting" anything. However, casting them as resulting from an inference process in which the brain incorporates predictions of what to expect allows us to formulate an explanation for *why* these biases arise in the first place – namely, that perception is optimized for the processing of certain (more likely) inputs at the expense of systematic errors for other (less likely) inputs [13,151].

Finally, taking a statistical perspective also helps clarify why predictions are not necessarily the same as forecasts. From a statistical point of view, the essence of prediction is the absence of (sufficient) data. Whether this is because the prediction is about the future or because it is about current but not yet (fully) observed events, is irrelevant. Essential is the extrapolation from a model to new, unobserved data – data not used when estimating the model. In neuroscience, most predictive coding models (see 'Expectation in computational models of perception') use the word 'prediction' only in this statistical sense and do not involve predictions about the future.

Box 3 – Investigating cortical laminae to dissociate top-down from bottom-up

When high-level (e.g., contextual) expectations modulate early sensory responses, it is often assumed that these expectations are encoded at higher cortical regions and modulate earlier regions 'from the top-down'. However, testing this assumption requires directional inference. The traditional way to achieve this is either by perturbing top-down connections, e.g. using cooling or optogenetical (de)activation [152], or by resorting to causal modelling techniques such as DCM [153] and Granger causality [154], which rely on various (contestable) assumptions.

However, cortical anatomy provides us with another method – laminar profiling. Since bottom-up connections originate from superficial layers and terminate in the middle layer, while top-down connections originate from deep layers and avoid the middle layer (Fig IA), it is possible to use the laminar activity profile to infer whether an activity modulation is likely top-down or bottom-up. Critically, it has recently become possible to measure laminar activity profiles non-invasively in humans (Fig IB-C), using ultra-high field-strength fMRI [147,155]. For instance, in a recent study Kok et al. [156] found that an illusory figure induced by the spatial context (i.e., the Kanizsa illusion) specifically activated the deep (but not the middle and superficial) layers of V1 (Fig ID), in line with this activity arising from top-down feedback.

In addition to establishing the directionality of signal flow, laminar profiling of expectation effects can also help distinguish between computational theories of perception [148]. For instance, classical predictive coding theory proposes that neurons in the deep layers represent our current hypotheses, or predictions, about the causes of our sensory inputs, while neurons in superficial layers encode the mismatch between these predictions and actual sensory inputs (i.e., prediction error) [4,5].

Furthermore, there has been great debate on whether or not expectation differs mechanistically from other top-down processes such as attention, working memory, and imagery [157,158]. One possibility is that these processes differ in which layers of visual cortex they target [159] – a hypothesis that laminar fMRI would allow to test in humans.

In sum, laminar profiling, invasively in animals and non-invasively in human fMRI, provides exciting opportunities for studying the neural basis of expectation and of other cognitive influences on perception such as attention and working memory.

Outstanding Questions

- What is the computational role of expectation: suppressing expected (i.e., redundant) signals, improving sensory representations, increasing processing efficiency, or a combination of these? If the latter, how are the various goals combined in the neural circuit?

- How are expectations from different sources combined to jointly modulate sensory processing?

- Does sensory cortex contain explicit prediction error neurons, signaling the mismatch between expectations and sensory inputs?

- What is the relationship between expectation ('What is likely?') and attention ('What is relevant?')? Are these two cognitive processes subserved by the same or different neural mechanisms? And if the latter, how do expectation and attention interact?

- Can we extrapolate from the computational principles discovered in early sensory cortex to the rest of the brain? For instance, do similar neural expectation effects underlie other inferential processes, such as in social cognition, language, and motor control?

References

1   Rosen, R. (2012) Anticipatory systems. In *Anticipatory systems* pp. 313–370, Springer

2   Clark, A. (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204

3   Von Helmholtz, H. (1866) *Handbuch der physiologischen Optik: mit 213 in den Text eingedruckten Holzschnitten und 11 Tafeln*, Voss.

4   Friston, K. (2005) A theory of cortical responses. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 360, 815–836

5   Rao, R.P. and Ballard, D.H. (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2, 79–87

6   Montague, P.R. *et al.* (2012) Computational psychiatry. *Trends Cogn. Sci.* 16, 72–80

7   Wang, X.-J. and Krystal, J.H. (2014) Computational psychiatry. *Neuron* 84, 638–654

8   Lawson, R.P. *et al.* (2017) Adults with autism overestimate the volatility of the sensory environment. *Nat. Neurosci.* 20, 1293–1299

9   Powers, A.R. *et al.* (2017) Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. *Science* 357, 596–600

10   Chalk, M. *et al.* (2010) Rapidly learned stimulus expectations alter perception of motion. *J. Vis.* 10, 2–2

11   Kok, P. *et al.* (2013) Prior Expectations Bias Sensory Representations in Visual Cortex. *J. Neurosci.* 33, 16275–16284

12   Sterzer, P. *et al.* (2008) Believing is seeing: expectations alter visual awareness. *Curr. Biol.* 18, R697–R698

13   Weiss, Y. *et al.* (2002) Motion illusions as optimal percepts. *Nat. Neurosci.* 5, 598–604

14   Sotiropoulos, G. *et al.* (2011) Changing expectations about speed alters perceived motion direction. *Curr. Biol.* 21, R883–R884

15   Stein, T. and Peelen, M.V. (2015) Content-specific expectations enhance stimulus detectability by increasing perceptual sensitivity. *J. Exp. Psychol. Gen.* 144, 1089–1104

16   Chang, A.Y.-C. *et al.* (2015) Cross-modal prediction changes the timing of conscious access during the motion-induced blindness. *Conscious. Cogn.* 31, 139–147

17   Pinto, Y. *et al.* (2015) Expectations accelerate entry of visual stimuli into awareness. *J. Vis.* 15, 13

18   Wyart, V. *et al.* (2012) Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. *Proc. Natl. Acad. Sci.* 109, 3593–3598

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

20   Seriès, P. and Seitz, A.R. (2013) Learning what to expect (in visual perception). *Front. Hum. Neurosci.* 7, 668

21   Girshick, A.R. *et al.* (2011) Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nat. Neurosci.* 14, 926–932

22   Gerardin, P. *et al.* (2010) Prior knowledge of illumination for 3D perception in the human brain. *Proc. Natl. Acad. Sci. U. S. A.* 107, 16309–16314

23   Sun, J. and Perona, P. (1998) Where is the sun? *Nat. Neurosci.* 1, 183–184

24   Albright, T.D. (1989) Centrifugal directional bias in the middle temporal visual area (MT) of the macaque. *Vis. Neurosci.* 2, 177–188

25   Cloherty, S.L. *et al.* (2016) Sensory experience modifies feature map relationships in visual cortex. *eLife* 5,

26   Li, B. *et al.* (2003) Oblique Effect: A Neural Basis in the Visual Cortex. *J. Neurophysiol.* 90, 204–217

27   Adams, W.J. *et al.* (2004) Experience can change the "light-from-above" prior. *Nat. Neurosci.* 7, nn1312

28   Hershberger, W. (1970) Attached-shadow orientation perceived as depth by chickens reared in an environment illuminated from below. *J. Comp. Physiol. Psychol.* 73, 407–411

29   Hochstein, S. and Ahissar, M. (2002) View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron* 36, 791–804

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

31   Oliva, A. and Torralba, A. (2007) The role of context in object recognition. *Trends Cogn. Sci.* 11, 520–527

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

33   Kok, P. and de Lange, F.P. (2014) Shape perception simultaneously up- and downregulates neural activity in the primary visual cortex. *Curr. Biol. CB* 24, 1531–1535

34   Murray, S.O. *et al.* (2002) Shape perception reduces activity in human primary visual cortex. *Proc. Natl. Acad. Sci.* 99, 15164–15169

35   Smith, F.W. and Muckli, L. (2010) Nonstimulated early visual areas carry information about surrounding context. *Proc. Natl. Acad. Sci.* 107, 20099–20103

36   Ekman, M. *et al.* (2017) Time-compressed preplay of anticipated events in human primary visual cortex. *Nat. Commun.* 8, 15276

37   Muckli, L. *et al.* (2005) Primary Visual Cortex Activity along the Apparent-Motion Trace Reflects Illusory Perception. *PLoS Biol.* 3, e265

38   Arnal, L.H. *et al.* (2009) Dual Neural Routing of Visual Facilitation in Speech Processing. *J. Neurosci.* 29, 13445–13453

39   Beauchamp, M.S. *et al.* (2010) fMRI-guided TMS reveals that the STS is a Cortical Locus of the McGurk Effect. *J. Neurosci. Off. J. Soc. Neurosci.* 30, 2414–2417

40   Petzschner, F.H. and Glasauer, S. (2011) Iterative Bayesian Estimation as an Explanation for Range and Regression Effects: A Study on Human Path Integration. *J. Neurosci.* 31, 17220–17229

41   Petzschner, F.H. *et al.* (2015) A Bayesian perspective on magnitude estimation. *Trends Cogn. Sci.* 19, 285–293

42   Bueti, D. and Walsh, V. (2009) The parietal cortex and the representation of time, space, number and other magnitudes. *Philos. Trans. R. Soc. B Biol. Sci.* 364, 1831–1840

43   Geuter, S. *et al.* (2017) Functional dissociation of stimulus intensity encoding and predictive coding of pain in the insula. *eLife* 6, e24770

44   Seth, A.K. *et al.* (2012) An Interoceptive Predictive Coding Model of Conscious Presence. *Front. Psychol.* 2,

45   Lavenex, P. and Amaral, D.G. (2000) Hippocampal-neocortical interaction: a hierarchy of associativity. *Hippocampus* 10, 420–430

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

47   Cohen, N.J. *et al.* (1997) Memory for Items and Memory for Relations in the Procedural/Declarative Memory Framework. *Memory* 5, 131–178

48   Garvert, M.M. *et al.* (2017) A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *eLife* 6, e17086

49   Schapiro, A.C. *et al.* (2012) Shaping of Object Representations in the Human Medial Temporal Lobe Based on Temporal Regularities. *Curr. Biol.* 22, 1622–1627

50   Bosch, S.E. *et al.* (2014) Reinstatement of Associative Memories in Early Visual Cortex Is Signaled by the Hippocampus. *J. Neurosci.* 34, 7493–7500

51   Gordon, A.M. *et al.* (2014) Cortical Reinstatement Mediates the Relationship Between Content-Specific Encoding Activity and Subsequent Recollection Decisions. *Cereb. Cortex* 24, 3350–3364

52   Davachi, L. and DuBrow, S. (2015) How the hippocampus preserves order: the role of prediction and context. *Trends Cogn. Sci.* 19, 92–99

53   Hindy, N.C. *et al.* (2016) Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nat. Neurosci.* 19, nn.4284

54   Bar, M. (2007) The proactive brain: using analogies and associations to generate predictions. *Trends Cogn. Sci.* 11, 280–289

55   Fletcher, P.C. *et al.* (2001) Responses of human frontal cortex to surprising events are predicted by formal associative learning theory. *Nat. Neurosci.* 4, 1043–1048

56   van Kesteren, M.T.R. *et al.* (2012) How schema and novelty augment memory formation. *Trends Neurosci.* 35, 211–219

57   Wessel, J.R. and Aron, A.R. (2017) On the Globality of Motor Suppression: Unexpected Events and Their Influence on Behavior and Cognition. *Neuron* 93, 259–280

58   Eliades, S.J. and Wang, X. (2008) Neural substrates of vocalization feedback monitoring in primate auditory cortex. *Nature* 453, 1102–1106

59   Petreanu, L. *et al.* (2012) Activity in motor-sensory projections reveals distributed coding in somatosensation. *Nature* 489, 299–303

60   Schneider, D.M. *et al.* (2014) A synaptic and circuit basis for corollary discharge in the auditory cortex. *Nature* 513, 189–194

61   Keller, G.B. *et al.* (2012) Sensorimotor Mismatch Signals in Primary Visual Cortex of the Behaving Mouse. *Neuron* 74, 809–815

62   Keller, G.B. and Hahnloser, R.H.R. (2009) Neural processing of auditory feedback during vocal practice in a songbird. *Nature* 457, nature07467

63   Bell, C.C. (2001) Memory-based expectations in electrosensory systems. *Curr. Opin. Neurobiol.* 11, 481–487

64   Summerfield, C. and De Lange, F.P. (2014) Expectation in perceptual decision making: neural and computational mechanisms. *Nat. Rev. Neurosci.* 15, nrn3838

65   Kok, P. *et al.* (2012) Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron* 75, 265–270

66   Kumar, S. *et al.* (2017) Encoding of Predictable and Unpredictable Stimuli by Inferior Temporal Cortical Neurons. *J. Cogn. Neurosci.* 29, 1445–1454

67   Meyer, T. and Olson, C.R. (2011) Statistical learning of visual transitions in monkey inferotemporal cortex. *Proc. Natl. Acad. Sci.* 108, 19401–19406

68   Todorovic, A. and De Lange, F.P. (2012) Repetition Suppression and Expectation Suppression Are Dissociable in Time in Early Auditory Evoked Fields. *J. Neurosci.* 32, 13389–13395

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

70   Rummell, B.P. *et al.* (2016) Attenuation of Responses to Self-Generated Sounds in Auditory Cortical Neurons. *J. Neurosci. Off. J. Soc. Neurosci.* 36, 12010–12026

71   Summerfield, C. *et al.* (2008) Neural repetition suppression reflects fulfilled perceptual expectations. *Nat. Neurosci.* 11, 1004–1006

72   Alink, A. *et al.* (2010) Stimulus Predictability Reduces Responses in Primary Visual Cortex. *J. Neurosci.* 30, 2960–2966

73   Egner, T. *et al.* (2010) Expectation and Surprise Determine Neural Population Responses in the Ventral Visual Stream. *J. Neurosci.* 30, 16601–16608

74   Schwiedrzik, C.M. and Freiwald, W.A. (2017) High-Level Prediction Signals in a Low-Level Area of the Macaque Face-Processing Hierarchy. *Neuron* 96, 89-97.e4

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

76   Rubin, J. *et al.* (2016) The Representation of Prediction Error in Auditory Cortex. *PLoS Comput. Biol.* 12, e1005058

77   Garrido, M.I. *et al.* (2017) Bayesian Mapping Reveals That Attention Boosts Neural Responses to Predicted and Unpredicted Stimuli. *Cereb. Cortex* DOI: 10.1093/cercor/bhx087

78   Vinken, K. *et al.* (2017) Recent Visual Experience Shapes Visual Processing in Rats through Stimulus-Specific Adaptation and Response Enhancement. *Curr. Biol. CB* 27, 914–919

79   Lee, T.S. and Mumford, D. (2003) Hierarchical Bayesian inference in the visual cortex. *J. Opt. Soc. Am. A Opt. Image Sci. Vis.* 20, 1434–1448

80   Murray, S.O. *et al.* (2004) Perceptual grouping and the interactions between visual cortical areas. *Neural Netw.* 17, 695–705

81   Carpenter, G.A. and Grossberg, S. (1987) A Massively Parallel Architecture for a Self-organizing Neural Pattern Recognition Machine. *Comput Vis. Graph Image Process* 37, 54–115

82   Rumelhart, D.E. and McClelland, J.L. (1982) An interactive activation model of context effects in letter perception: Part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychol. Rev.* 89, 60–94

83 Desimone, R. (1996) Neural mechanisms for visual memory and their role in attention. *Proc. Natl. Acad. Sci. U. S. A.* 93, 13494–13499

84 McMahon, D.B.T. and Olson, C.R. (2007) Repetition suppression in monkey inferotemporal cortex: relation to behavioral priming. *J. Neurophysiol.* 97, 3532–3543

85 Wiggs, C.L. and Martin, A. (1998) Properties and mechanisms of perceptual priming. *Curr. Opin. Neurobiol.* 8, 227–233

86 Martinez-Trujillo, J.C. and Treue, S. (2004) Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr. Biol. CB* 14, 744–751

87 Bell, A.H. *et al.* (2016) Encoding of Stimulus Probability in Macaque Inferior Temporal Cortex. *Curr. Biol. CB* 26, 2280–2290

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

89 Doherty, J.R. *et al.* (2005) Synergistic Effect of Combined Temporal and Spatial Expectations on Visual Attention. *J. Neurosci.* 25, 8259–8266

90 Jaramillo, S. and Zador, A.M. (2010) The auditory cortex mediates the perceptual effects of acoustic temporal expectation. *Nat. Neurosci.* 14, nn.2688

91 Jiang, J. *et al.* (2013) Attention Sharpens the Distinction between Expected and Unexpected Percepts in the Visual Brain. *J. Neurosci.* 33, 18438–18447

92 Kok, P. *et al.* (2012) Attention Reverses the Effect of Prediction in Silencing Sensory Signals. *Cereb. Cortex* 22, 2197–2206

93 Barascud, N. *et al.* (2016) Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. *Proc. Natl. Acad. Sci.* 113, E616–E625

94 Southwell, R. *et al.* (2017) Is predictability salient? A study of attentional capture by auditory patterns. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 372,

95 St. John-Saaltink, E. *et al.* (2015) Expectation Suppression in Early Visual Cortex Depends on Task Set. *PLOS ONE* 10, e0131172

96 Chen, I.-W. *et al.* (2015) Specific Early and Late Oddball-Evoked Responses in Excitatory and Inhibitory Neurons of Mouse Auditory Cortex. *J. Neurosci. Off. J. Soc. Neurosci.* 35, 12560–12573

97 Sakai, K. and Miyashita, Y. (1991) Neural Organization for the Long-Term Memory of Paired Associates. *Nat. Lond.* 354, 152–5

98 Hughes, H.C. *et al.* (2001) Responses of human auditory association cortex to the omission of an expected acoustic event. *NeuroImage* 13, 1073–1089

99 Kok, P. *et al.* (2017) Prior expectations induce prestimulus sensory templates. *Proc. Natl. Acad. Sci.* 114, 10473–10478

100 Gregory, R.L. (1980) Perceptions as Hypotheses. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 290, 181–197

101 Yuille, A. and Kersten, D. (2006) Vision as Bayesian inference: analysis by synthesis? *Trends Cogn. Sci.* 10, 301–308

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

103 Laquitaine, S. and Gardner, J.L. (2018) A Switching Observer for Human Perceptual Estimation. *Neuron* DOI: 10.1016/j.neuron.2017.12.011

104 Kersten, D. *et al.* (2004) Object perception as Bayesian inference. *Annu. Rev. Psychol.* 55, 271–304

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

106 Grossberg, S. (2006) Adaptive Resonance Theory. In *Encyclopedia of Cognitive Science* John Wiley & Sons, Ltd

107 Heeger, D.J. (2017) Theory of cortical function. *Proc. Natl. Acad. Sci.* DOI: 10.1073/pnas.1619788114

108 McClelland, J.L. (2013) Integrating probabilistic models of perception and interactive neural networks: a historical and tutorial review. *Front. Psychol.* 4,

109 Beck, J.M. *et al.* (2008) Probabilistic population codes for Bayesian decision making. *Neuron* 60, 1142–52

110 Norris, D. *et al.* (2016) Prediction, Bayesian inference and feedback in speech recognition. *Lang. Cogn. Neurosci.* 31, 4–18

111 Feldman, H. and Friston, K.J. (2010) Attention, Uncertainty, and Free-Energy. *Front. Hum. Neurosci.* 4, 215

112 Bastos, A.M. *et al.* (2012) Canonical Microcircuits for Predictive Coding. *Neuron* 76, 695–711

113 Yu, A.J. (2014) Bayesian Models of Attention. *Oxf. Handb. Atten.* DOI: 10.1093/oxfordhb/9780199675111.013.025

114 Srinivasan, M.V. *et al.* (1982) Predictive coding: a fresh view of inhibition in the retina. *Proc R Soc Lond B Biol Sci* 216, 427–59

115 Kiebel, S.J. *et al.* (2008) A Hierarchy of Time-Scales and the Brain. *PLOS Comput. Biol.* 4, e1000209

116 Wacongne, C. *et al.* (2012) A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. *J. Neurosci.* 32, 3665–3678

117 Spratling, M.W. (2017) A review of predictive coding algorithms. *Brain Cogn.* 112, 92–97

118 Spratling, M.W. (2008) Reconciling predictive coding and biased competition models of cortical function. *Front. Comput. Neurosci.* 2, 4

119 Mumford, D. (1994) *Neuronal architectures for pattern-theoretic problems*, Large-Scale Theories of the Cortex. Cambridge, MA: MIT Press.

120 Friston, K. (2010) The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138

121 Fiser, J. *et al.* (2010) Statistically optimal perception and learning: from behavior to neural representations. *Trends Cogn. Sci.* 14, 119–130

122 Orbán, G. *et al.* (2016) Neural Variability and Sampling-Based Probabilistic Representations in the Visual Cortex. *Neuron* 92, 530–543

123 Heilbron, M. and Chait, M. (2017) Great expectations: Is there evidence for predictive coding in auditory cortex? *Neuroscience* DOI: 10.1016/j.neuroscience.2017.07.061

124 Kok, P. and de Lange, F.P. (2015) Predictive coding in sensory cortex. In *An introduction to model-based cognitive neuroscience* pp. 221–244, Springer

125 Spratling, M.W. (2010) Predictive Coding as a Model of Response Properties in Cortical Area V1. *J. Neurosci.* 30, 3531–3543
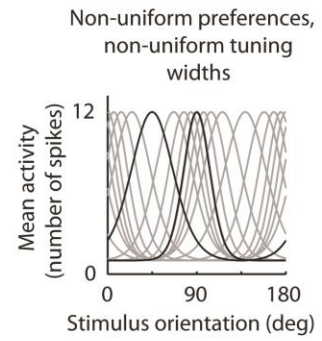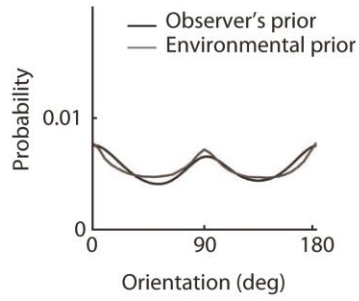
126 Jehee, J.F.M. and Ballard, D.H. (2009) Predictive Feedback Can Account for Biphasic Responses in the Lateral Geniculate Nucleus. *PLoS Comput. Biol.* 5, e1000373

127 Bolz, J. and Gilbert, C.D. (1986) Generation of end-inhibition in the visual cortex via interlaminar connections. *Nature* 320, 362–365

128 Sedley, W. *et al.* (2016) Neural signatures of perceptual inference. *eLife* 5, e11476

129 Maier, A. *et al.* (2011) Infragranular Sources of Sustained Local Field Potential Responses in Macaque Primary Visual Cortex. *J. Neurosci.* 31, 1971–1980

130 Mathys, C. *et al.* (2011) A bayesian foundation for individual learning under uncertainty. *Front. Hum. Neurosci.* 5, 39

131 Iglesias, S. *et al.* (2013) Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. *Neuron* 80, 519–530

132 Pellicano, E. and Burr, D. (2012) When the world becomes 'too real': a Bayesian explanation of autistic perception. *Trends Cogn. Sci.* 16, 504–510

133 Van de Cruys, S. *et al.* (2014) Precise minds in uncertain worlds: predictive coding in autism. *Psychol. Rev.* 121, 649–675

134 Fletcher, P.C. and Frith, C.D. (2009) Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat. Rev. Neurosci.* 10, 48–58

135 Corlett, P.R. *et al.* (2011) Glutamatergic Model Psychoses: Prediction Error, Learning, and Inference. *Neuropsychopharmacology* 36, 294

136 Sinha, P. *et al.* (2014) Autism as a disorder of prediction. *Proc. Natl. Acad. Sci.* 111, 15220–15225

137 Pajani, A. *et al.* (2015) Spontaneous Activity Patterns in Primary Visual Cortex Predispose to Visual Hallucinations. *J. Neurosci.* 35, 12947–12953

138 Happé, F. and Frith, U. (2006) The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *J. Autism Dev. Disord.* 36, 5–25

139 Mathys, C.D. *et al.* (2014) Uncertainty in perception and the Hierarchical Gaussian Filter. *Front. Hum. Neurosci.* 8,

140 Bullier, J. (2001) Integrated model of visual processing. *Brain Res. Brain Res. Rev.* 36, 96–107

141 Roelfsema, P.R. and de Lange, F.P. (2016) Early Visual Cortex as a Multiscale Cognitive Blackboard. *Annu. Rev. Vis. Sci.* 2, 131–151

142 Koster-Hale, J. and Saxe, R. (2013) Theory of mind: a neural prediction problem. *Neuron* 79, 836–848

143 LeCun, Y. *et al.* (2015) Deep learning. *Nature* 521, 436

144 DiCarlo, J.J. *et al.* (2012) How does the brain solve visual object recognition? *Neuron* 73, 415–434

145 Güçlü, U. and Gerven, M.A.J. van (2015) Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. *J. Neurosci.* 35, 10005–10014

146 Kriegeskorte, N. (2015) Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing. *Annu. Rev. Vis. Sci.* 1, 417–446

147 Lawrence, S.J.D. *et al.* (2017) Laminar fMRI: Applications for cognitive neuroscience. *NeuroImage* DOI: 10.1016/j.neuroimage.2017.07.004

148 Stephan, K.E. *et al.* Laminar fMRI and computational theories of brain function. *NeuroImage* DOI: 10.1016/j.neuroimage.2017.11.001

149 Blakemore, S.J. *et al.* (2000) Why can't you tickle yourself? *Neuroreport* 11, R11-16

150 Leonard, M.K. *et al.* (2016) Perceptual restoration of masked speech in human cortex. *Nat. Commun.* 7, 13619

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

152 Hupé, J.M. *et al.* (1998) Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature* 394, 784

153 Stephan, K.E. *et al.* (2010) Ten simple rules for dynamic causal modeling. *NeuroImage* 49, 3099–3109

154 Seth, A.K. *et al.* (2015) Granger Causality Analysis in Neuroscience and Neuroimaging. *J. Neurosci.* 35, 3293–3297

155 Muckli, L. *et al.* (2015) Contextual Feedback to Superficial Layers of V1. *Curr. Biol.* 25, 2690–2695

156 Kok, P. *et al.* (2016) Selective Activation of the Deep Layers of the Human Primary Visual Cortex by Top-Down Feedback. *Curr. Biol.* 26, 371–376

157 Pearson, J. and Westbrook, F. (2015) Phantom perception: voluntary and involuntary nonretinal vision. *Trends Cogn. Sci.* 19, 278–284

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

159 Shipp, S. (2016) Neural Elements for Predictive Coding. *Front. Psychol.* 7,

160 Hsieh, P.-J. *et al.* (2010) Recognition Alters the Spatial Pattern of fMRI Activation in Early Retinotopic Cortex. *J. Neurophysiol.* 103, 1501–1507
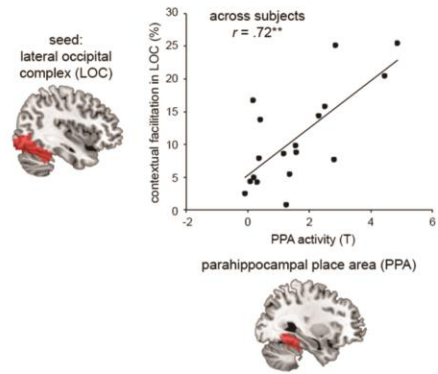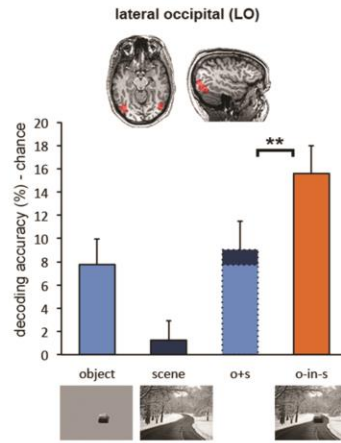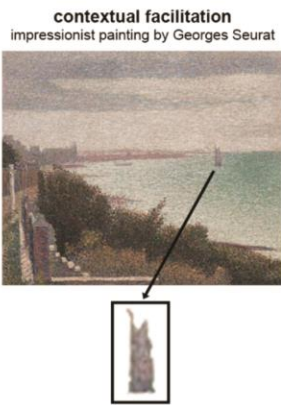
Figure Legends

Figure 1. Examples of prior expectations and their neural implementation. **(A)** Cardinal expectations: cardinal (horizontal and vertical) orientations are more prevalent in natural images than oblique ones (left panel). Human observers' perception is biased towards these cardinal orientations, exactly as would be expected from the statistics of natural images (middle panel). This bias could be implemented by a relative overrepresentation of neurons in early visual cortex tuned for cardinal orientations, as well as a sharper tuning of these neurons (right panel). Reprinted from [21]. **(B)** Objects are more easily recognized when presented in their usual context than in isolation (left panel). In the brain, this is reflected by neural representations in object-selective cortex containing more object information for objects presented in the context of a scene (middle panel). This context facilitation in object selective cortex is correlated with the amplitude of neural signals in scene selective cortex, in line with the proposal that the latter may provide predictive feedback signals to object selective cortex (right panel). Reprinted from [32]. **(C)** Expectations can stem from arbitrary regularities between stimuli. For instance, here image A was followed by B when participants pressed the left button, and by C when they pressed the right button (left panel). In the CA and DG subfields of the hippocampus, presentation of a partial sequence (e.g., A-left) leads to the retrieval of the full sequence (A-left-B; middle panel). On trials when the hippocampus correctly retrieves the full sequence, the expected image (e.g., B) is activated in early visual cortex (right panel). Reprinted from [53].
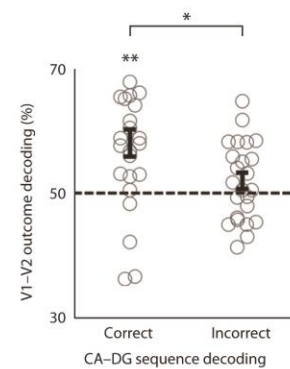
A

Non-uniform preferences, non-uniform tuning widths

- Observer's prior
- Environmental prior

B

contextual facilitation
impressionist painting by Georges Seurat

lateral occipital (LO)

seed: lateral occipital complex (LOC)

across subjects
r = .72**

parahippocampal place area (PPA)

C

Full sequence

Press left   Press right

Cue + action

Press left   Press right

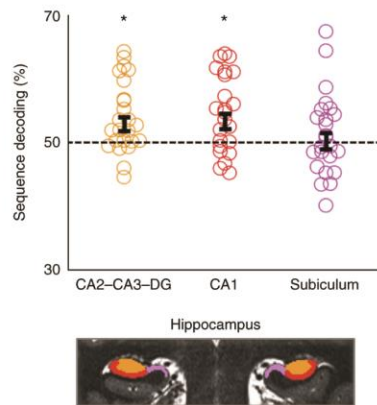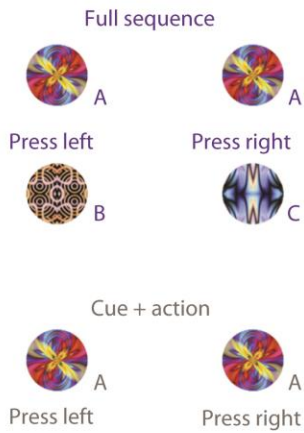Hippocampus

CA–DG sequence decoding

Figure 2. Expectation suppression via sharpening or dampening of neural responses. Stimulus-specific suppression of responses by expectation can be explained by two different mechanisms that make distinct predictions at the neuronal (single or multi-unit) and population level. Under the dampening account, expectation suppression most strongly affects neurons that are tuned to the expected stimulus. At the neuronal level (upper panel), this implies that neurons with the highest selectivity for the expected stimulus are suppressed most strongly. At the population level (lower panel), dampening predicts a decrease in the contrast of activity patterns evoked by stimuli, since the relative suppression of the most selective units of observation (e.g. voxels) reduces differences in activity between units (lower panel, middle). Under the Sharpening account, by contrast, expectation suppression most strongly affects neurons that are *not* tuned to the expected stimulus. At the neural level (upper panel, right), this implies that neurons with the lowest selectivity for the expected stimulus are suppressed most strongly, and that neurons with the highest selectivity may even be enhanced. At the population level (lower panel), Sharpening predicts an increase the contrast of activity patterns evoked by specific stimuli, since the relative suppression of the least selective units should *increase* differences in activity between units, leading to a 'sharpened' representation (lower panel, right). While the above population level patterns are suggestive of sharpening and dampening at the neuronal level, one has to be careful in interpreting population responses with coarse methods like fMRI that aggregate brain activity from large numbers of neurons that span the entire tuning range, as there is not always a direct correspondence between neural-level and voxel-level selectivity patterns. When making inferences about selectivity-dependent suppression at the voxel level, one is ideally guided by a formal and quantitative linking model incorporating all relevant parameters such as tuning, suppression and voxel properties.
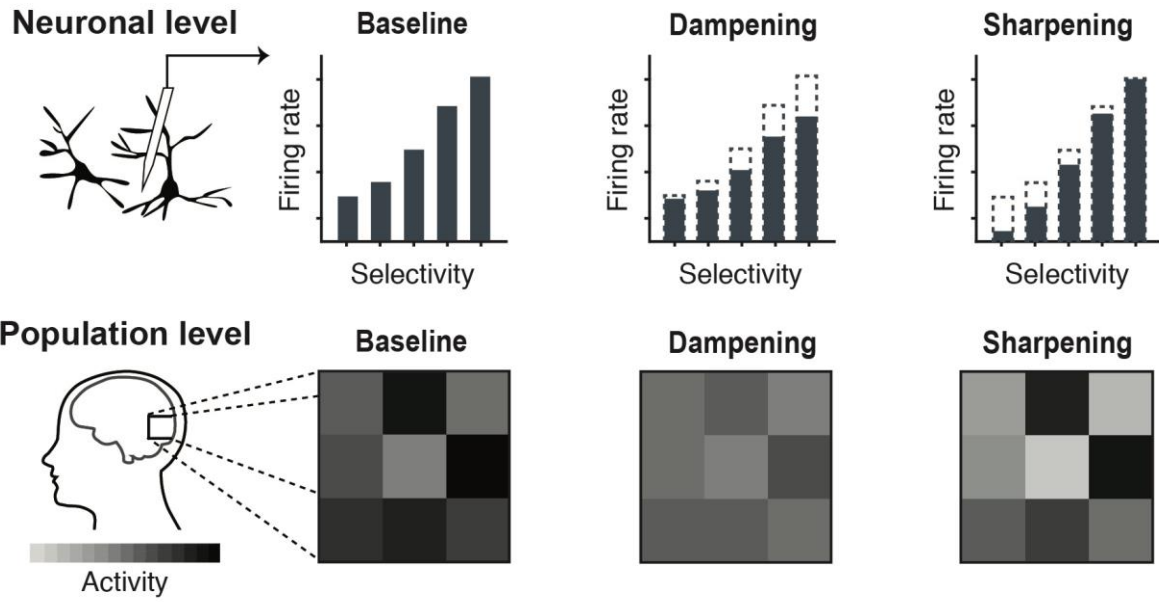
Neuronal level

Baseline · Dampening · Sharpening

Firing rate vs Selectivity

Population level

Baseline · Dampening · Sharpening

Activity

Figure Box 1. Examples of prior knowledge influencing perception. (A) The light-from-above prior. (B) Spatial contextual prior knowledge. Image reproduced from [31]. (C) Prior experience. Images reproduced from [160].

Figure Box 3. Non-invasive imaging of cortical laminae. (A) Inter-regional feedforward (red arrows) and feedback (blue arrows) connections between human LGN, V1 and V2. Reproduced from [147]. **(B)** Example of an fMRI voxel (red square) and its distribution of three gray matter layers. Layer weights can be used as the basis of a regression approach to obtain layer-specific BOLD responses [156]. **(C)** Sagittal slice showing a grid of 0.8 mm isotropic fMRI voxels measured at 7T field strength. Gray-white matter (yellow line) and gray matter-CSF (red line) boundaries are overlaid onto the volume. Image reproduced from [156]. **(D)** BOLD activity in V1 elicited by an illusory figure in V1. Image reproduced from [156].