

Chapter 11

Predictive Coding in Sensory Cortex

Peter Kok and Floris P. de Lange

Abstract In recent years, predictive coding has become an increasingly influential model of how the brain processes sensory information. Predictive coding theories state that the brain is constantly trying to predict the inputs it receives, and each region in the cortical sensory hierarchy represents both these predictions and the mismatch between predictions and input (prediction error). In this chapter, we review the extant empirical evidence for this theory, as well as discuss recent theoretical advances. We find that predictive coding provides a good explanation for many phenomena observed in perception, and generates testable hypotheses. Furthermore, we suggest possible avenues for further empirical testing and for broadening the perspective of the role predictive coding may play in cognition.

11.1 Introduction

In recent years, predictive coding has become an increasingly influential model of how the brain processes sensory information [1–3]. This model challenges the traditional view of sensory cortex as a unidirectional hierarchical system that passively receives sensory signals and extracts increasingly complex features as one progresses up the hierarchy. Instead, predictive coding theories state that the brain is constantly trying to predict the inputs it receives, and each region in the sensory hierarchy represents both these predictions and the mismatch between predictions and input (prediction error). Moreover, regions in the sensory hierarchy continually interact, informing each other about what they expect the other region is observing, and how this expectation matches their input.

In this chapter, we will review recent theoretical and empirical advances in the field of predictive coding. First, we will outline the motivations behind predictive coding, and discuss its principal features (§ 2). Then, we will review empirical evidence from cognitive neuroscience for several basic tenets of predictive coding (§ 3), and discuss one of them—the implementation of attention in predictive coding—in more detail

P. Kok (✉) · F. P. de Lange
Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen,
6500 HB, Nijmegen, The Netherlands
e-mail: p.kok@donders.ru.nl

(§ 4). We will end with a discussion of the limitations of current empirical foundations of predictive coding, and suggest future directions to strengthen these foundations and extend the perspective on predictive coding to other cognitive domains (§ 5).

11.2 Predictive Coding

11.2.1 *Perception as Inference*

One of the major motivations behind the development of predictive coding models of sensory processing has been the observation that perception is not solely determined by the input to our eyes, but is strongly influenced by our expectations. Over a century ago, Helmholtz [4] cast perception as a process of unconscious inference, wherein perception is determined by both sensory inputs and our prior experience with the world. For example, many perceptual illusions can be explained as the result of our prior knowledge of the world influencing perceptual inference [5, 6]. When we are presented with four ‘Pac-Man’ figures arranged in a certain way, we perceive an illusory square (Kanizsa square; Fig. 11.1a). Presumably, the brain infers that the most likely cause for such an input, given its prior experience of the world, is a white square overlaying four black circles. Note that occlusion is a ubiquitous feature of the visual world, and inferring the presence of whole objects despite this is key to successful perceptual inference. Furthermore, priors provided by the larger context can help disambiguate local details. For example, the same figure can be perceived as the letter ‘B’ or the number ‘13’, depending on the surrounding figures (‘A’ and ‘C’ or ‘12’ and ‘14’, respectively; Fig. 11.1b). Finally, prior knowledge can improve perception by ‘explaining away’ predictable features (e.g., stripy leaves), leaving unexpected (and potentially vitally important) features to stand out (a tiger!) (Fig. 11.1c). That is, without prior knowledge of the world the image in Fig. 11.1c might look like a mass of incoherent lines, while recognising the majority of the lines as parts of plants allows ‘subtracting’ them from the image. Any features that cannot be explained away as part of the plants (the stripes on the tiger’s back and head) will be all the more salient. In recent years, the idea of perception as inference has enjoyed a revival, benefitting from converging ideas from computer vision research and neuroscience [3, 7–10].

11.2.2 *Coding Scheme*

One model of sensory processing that describes perception as fundamentally inferential is predictive coding [1–3, 11]. In this model (Fig. 11.2a), each cortical sensory region contains two functionally distinct sub-populations of neurons. Prediction (P) units represent the hypothesis that best explains the input the region receives, while prediction error (PE) units represent that part of the input that is not explained by the current hypothesis, i.e. the mismatch between input and prediction. Connected



Fig. 11.1 Examples of perceptual inference. **a** Kanizsa *square*: four ‘Pac-Man’ figures or a *white square* overlaying *black circles*? **b** Context resolves ambiguity: is the figure in the centre the letter ‘*B*’ or the number ‘*13*’? **c** Prior knowledge improves processing of noisy sensory inputs: ‘explaining away’ the leaves makes the tiger stand out more

regions in the cortical hierarchy interact recurrently in a joint effort to find the world model that best explains the sensory inputs in the P units, and thereby reduce the activity of the PE units. This interaction takes place as follows: (1) The PE in one region serves as input to the next region in the cortical hierarchy, triggering that region to select a hypothesis that better matches its input. Note that the representational content of PEs (as opposed to an unspecific “surprise” signal) allows for selection of specific hypotheses in the higher order region. (2) The (newly) selected higher order hypothesis is subsequently sent back as a prediction to the lower order region, where it is compared to the current lower level hypothesis, and (3) the mismatch is represented as the (new) prediction error. The above describes one cycle of hypothesis testing in the predictive coding framework. This is an iterative process, culminating in a state in which PE units are all silenced and an accurate representation of the current sensory world is represented by activity in the relevant P units.

Since top-down predictions suppress expected sensory input (i.e., reduce prediction error), expected stimuli lead to relatively little neuronal firing. Such a coding scheme has several advantages. First, it is metabolically efficient. Second, it makes unexpected (and potentially highly relevant) stimuli more salient: if you ‘explain away’ the striped leaves, the crouching tiger stands out even more (Fig. 11.1c). In fact, it has been proposed that saliency might be equated to the strength of the prediction error [12, 13]. Third, while expected stimuli result in reduced firing in the PE units, the stimulus *representation in the P units* is enhanced [14]. A valid prediction leads to the proper hypothesis being selected prior to sensory input, and since this hypothesis quickly silences these sensory inputs (prediction error) when they arrive, alternative hypotheses are not given a chance to compete (Fig. 11.2b). (Note that this pre-selection does not prevent potentially relevant unexpected inputs from being processed, since such inputs will lead to a large PE, attracting attention and triggering selection of alternative hypotheses, see Fig. 11.2c). In other words, pre-selection of the valid hypothesis makes the stimulus representation more unequivocal, or sharper. Such use of prior expectations helps us make sense of the ambiguous and noisy sensory inputs we receive in everyday life [15]. For this aspect of perceptual inference, the hierarchical nature of predictive coding is crucial [9, 16]. Inference on fine scale

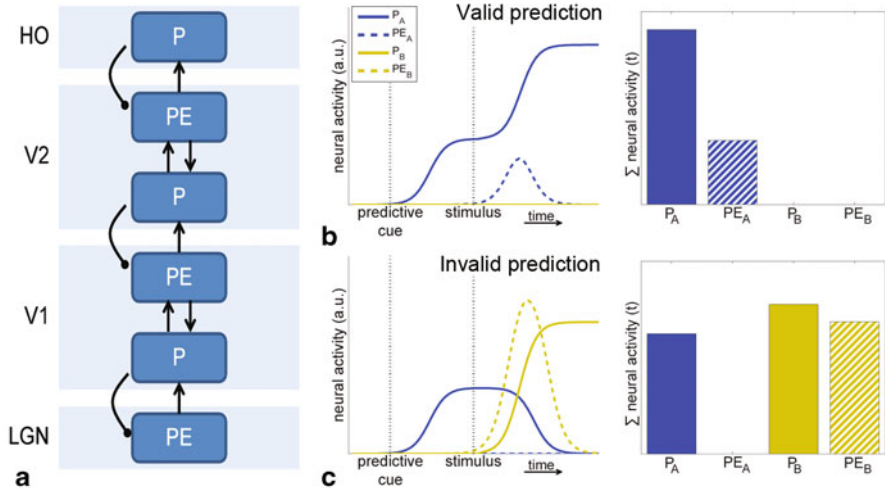


Fig. 11.2 Predictive coding. **a** Schematic predictive coding architecture, with PE units providing excitatory feedforward input, and P units providing inhibitory feedback. **b** Hypothesised neural activity in two populations of P and PE units, each representing a different hypothesis ('A' and 'B', respectively). Here, stimulus A is predicted, and subsequently presented (valid prediction). *Left* panel illustrates schematic timecourse of activity, *right* panel provides integral over time (i.e., proxy of BOLD amplitude). **c** Here, stimulus A is predicted, but B is presented. Activity is higher overall (particularly in PE units), but less unequivocal (in P units)

low level features (black and white stripes, in a seemingly random arrangement) benefits from high level representations (a tiger in a bush). In turn, high level representations can be refined by the high resolution information present in lower order visual areas, e.g., progressing from a coarse representation ('a face') to one reflecting the identity and emotional expression of that face [17].

In a slightly different take on hierarchical inference, Lee and Mumford [9] proposed a model wherein hypotheses at one level reinforce consistent hypotheses at the lower level. In their approach, multiple hypotheses are kept alive at each level of the cortical hierarchy, and excitatory feedback helps the most likely lower level hypothesis to win the competition. In other words, excitatory feedback collapses the lower level hypothesis space and thereby reduces the overall level of neuronal activity. Strictly taken, this is not a predictive coding model (there is no explicit error representation), but it shares many of its key features (hierarchical perceptual inference) as well as empirical predictions (valid top-down hypotheses lead to reduced activity but improved representations).

11.2.3 Neural Implementation

Several different proposals for the neural architecture underlying predictive coding have been made [1, 3, 18, 19]. All these accounts hypothesise the existence of

separate sub-populations of P and PE neurons, and suggest that these neurons reside in different cortical layers. A major difference lies in the type of information cortical areas exchange: in classical predictive coding schemes [1–3] PE units are the source of feedforward and the target of feedback connections, while in Spratling’s PC/BC model [18] errors are processed intracortically, and P units are reciprocally connected between regions. These schemes result in different predictions regarding the location of the sub-populations of P and PE units, based on known interlaminar and intercortical connectivity patterns [19]. Feedforward connections mainly arise from layers 2/3 and send input to layer 4 of the next higher-order region in the hierarchy, while feedback is sent from layers 5/6 to the agranular layers of the lower-order region [20–22]. Therefore, if feedforward connections carry prediction errors, PE units would be expected to reside in layers 2/3, while feedback-sending P units would reside in layers 5/6 [23]. In the PC/BC model, separate populations of P units would reside in layers 2/3 and 5/6, sending forward and backward predictions, respectively, while PE units reside in layer 4, which does not have interregional outputs. Note that such a separation of forward and backward messages seems necessary for hierarchical inference [9, 19], since these messages need to be tailored to higher-order (larger, more complex receptive fields) and lower-order (smaller, simpler receptive fields) regions, respectively. While these schemes differ in terms of the details of neural implementation, they are computationally equivalent [18].

11.3 Empirical Evidence for Predictive Coding

The recurrent interaction between prediction and prediction error units that characterises predictive coding models leads to several hypotheses that can be empirically tested. For example, since perception reflects an integration of top-down expectations and bottom-up sensory input, the same sensory input should lead to different responses depending on the strength and validity of the expectation. Specifically, the amplitude of the stimulus-evoked response should be lower, the more expected the input is (i.e., the less prediction error there is). Also, top-down expectations may activate hypotheses in sensory regions in the absence of sensory input. Further empirical validation of predictive coding may come from assessing its neural substrate: are there separate units coding predictions and prediction errors? We will discuss each of these points in turn, reviewing evidence from neuroimaging, electrophysiology, and physiology.

11.3.1 *Encoding of Surprise in the Brain*

One of the most robust modulations of sensory responses is repetition suppression (RS): when a stimulus is repeated, the neural response to the second stimulus is reduced compared to the first. This effect holds across a range of modalities, stimulus

properties, and brain areas, and has been considered the result of stimulus-induced neural adaptation [24]. However, if prediction is indeed a fundamental feature of sensory processing, RS may (partly) reflect the fact that the initial presentation of the stimulus induces an expectation of that same stimulus reappearing [25]. To test this hypothesis, Summerfield et al. [26] used functional magnetic resonance imaging (fMRI) to compare the neural response to stimulus repetitions and alternations, in two different contexts. In one context, a face stimulus was likely to be repeated, while in the other it was likely to be followed by a different face. These researchers showed that when stimulus repetitions were likely (i.e., expected), repeated stimuli led to a strongly reduced neural response compared to alternations (strong RS). When repetitions were unlikely however, the RS effect was strongly reduced, suggesting that RS at least partly reflects predictability. Since this study used fMRI to investigate neural activity, the time course of the RS effect (and its modulation by predictability) could not be resolved. Therefore, it was unclear whether predictability had an immediate suppressive effect on the expected sensory signal (prediction error suppression), or whether surprising events (alternations in one context, repetitions in the other) resulted in a reorienting of attention, with the reported effects of predictability reflecting later attentional modulations. In an effort to distinguish between these possibilities, Todorovic et al. [27] used magneto-encephalography (MEG) to investigate the time course of the effects of predictability on RS in auditory cortex. They found that predictability affected early stimulus-evoked components in auditory cortex, from 100 ms post-stimulus onwards (Fig. 11.3a; see also 28, for similar findings in monkey inferotemporal cortex using visual stimuli). Such early modulations are not in line with a late attention effect, but rather suggest predictive suppression of sensory signals. Furthermore, in a follow-up study, Todorovic and De Lange [29] reported dissociable time courses for the effects of repetition (i.e., stimulus-induced adaptation) and predictability, suggesting that prediction has suppressive effects independent of those of bottom-up adaptation. These and other studies [30–32] clearly show that prediction suppresses expected sensory signals.

Other studies have investigated violations of more high-level sensory predictions. One example is apparent motion: Static visual stimuli presented successively at separate spatial locations induce the illusory perception of motion between these locations. Areas of the primary visual cortex (V1) that correspond retinotopically to visual stimulation along the trajectory of illusory motion, but that are not directly stimulated by the static stimuli, have been shown to be active during perception of apparent motion [33]. Presumably, this is caused by higher level motion sensitive areas with larger receptive fields (i.e., MT/V5) inferring a moving stimulus and sending predictions of this inferred stimulus back to the corresponding locations in V1 [34–36]. Interestingly, the study by Ahmed et al. [36] was performed in anaesthetised ferrets, suggesting that this predictive feedback is not dependent on attention, but rather reflects automatic processes inherent to sensory processing in the visual cortical hierarchy. In a recent study, Alink et al. [37] reasoned that if these feedback signals indeed reflect predictions, they should affect the processing of stimuli presented along the apparent motion trajectory. Specifically, stimuli presented in temporal alignment with the inferred motion path should evoke less prediction error

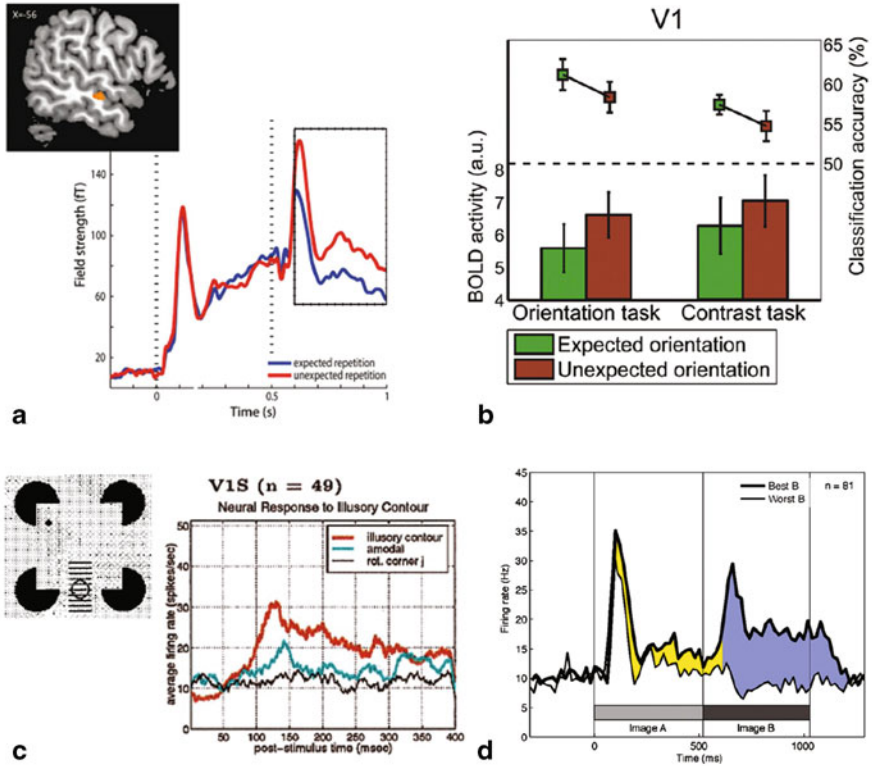


Fig. 11.3 Empirical evidence for predictive coding. **a** Unexpected stimulus repetitions evoke more activity in auditory cortex than expected repetitions. Reprinted from [27] with permission from the authors. **b** Grating stimuli with an expected orientation evoke less activity in V1 (green and red bars), but this activity contains more orientation information (green and red squares). This effect is independent of feature attention: it holds both when orientation is task relevant (leftmost bars) and when it is task irrelevant (rightmost bars). Reprinted from [14] with permission from the authors. **c** Illusory contours evoke activity in V1 cells with a receptive field on the contour, presumably as a result of feedback from higher order regions (Reprinted from [59]). **d** Predictive activity in macaque temporal cortex. After paired-association learning, neurons fire more strongly when the first stimulus of a pair (A) predicts that the second stimulus (B) will be their preferred stimulus ('Best B', thick line), than when stimulus A predicts a non-preferred stimulus B ('Worst B', thin line). Increased firing to preferred stimuli is present not only after stimulus presentation (blue shading), but already before stimulus presentation (yellow shading). Reprinted from [28]

than stimuli that are not temporally aligned. Indeed, Alink et al. [37] found that a visual stimulus presented along the apparent motion path in temporal alignment with the apparent motion evoked a reduced activation in V1, compared to when it was not temporally aligned. Presumably, such a non-aligned stimulus violates top-down predictions and therefore causes a larger prediction error.

Predictions operate not only within, but also between sensory modalities. An everyday example of this is the perception of audiovisual speech. In natural speech,

visual inputs (i.e., mouth movements) precede auditory input by 100–300 ms [38]. This allows the brain to make predictions about the auditory speech signals before their arrival. Indeed, presence of visual speech signals improves speech perception [39], and speeds up cortical processing of auditory speech [40, 41]. Furthermore, when visual and auditory signals mismatch, this can result in distorted or merged percepts [42]. For example, simultaneous presentation of auditory ‘ba’ and visual ‘ga’ signals is perceived as ‘da’. If visual speech signals indeed result in predictions being sent to auditory sensory areas, mismatch of visual and auditory signals should lead to increased PE in auditory cortex, compared to matching visual and auditory signals. This is indeed what was found by Arnal et al. [41, 43], who used both fMRI and MEG to characterise the response to audiovisual mismatches. Their results show an increased response to incongruent audiovisual stimuli in the superior temporal sulcus—a multisensory region—as well as increased gamma activity in the auditory cortex [43]. Both these responses scaled with the amount of predictive information contained by the visual stimulus: the more informative the visual stimulus was regarding the syllable being pronounced, the stronger the PE when the subsequent auditory stimulus did not match.

Studies on audiovisual speech exploit predictions learned over a lifetime. Recent studies have also shown effects of predictions across modalities when these predictions are learned over the course of the experiment [14, 44, 45]. For example, Den Ouden et al. [44] presented auditory cues that predicted with 80 % likelihood that a visual stimulus would appear. When a visual stimulus was preceded by such a cue, the activity it evoked in V1 was reduced compared to when it was not preceded by a predictive cue. Remarkably, the *omission* of a predicted visual stimulus also evoked more activity in V1 than the omission of a non-predicted stimulus. In this study, both the auditory and visual stimuli were completely irrelevant to participants’ task. These results demonstrate that predictions are learned rapidly, even when irrelevant to the task at hand, and affect sensory responses at the earliest stages of cortical processing. In line with this, we [14] found that auditory cues that predicted the features of a visual stimulus led to reduced activity in V1. Specifically, when the pitch of a preceding auditory tone correctly predicted the orientation of a subsequent grating stimulus, the response to this grating in V1 was reduced, compared to when the prediction was invalid. Furthermore, this study investigated not only the amplitude of the neural response evoked by the stimuli, but also used multivariate pattern analysis (MVPA) methods to probe the amount of information contained in the neural signal. Interestingly, we found that a valid orientation prediction led to a *decrease* in the amplitude of the neural signal in V1, but to an *increase* in the amount of information about the grating orientation in the signal (Fig. 11.3b). This is exactly the pattern of results that is predicted by predictive coding theories of perception (cf. Fig. 11.2b, c): valid predictions lead to selection of the proper hypothesis prior to sensory input, allowing this hypothesis to quickly suppress the sensory signal when it comes in (prediction error suppression), thereby preventing activation of alternative hypotheses (representational sharpening). These results suggest that the population level neural signals measured in humans (with fMRI or EEG/MEG) are a mixture of prediction and prediction error signals [46].

While all these studies exemplify suppressive effects of sensory predictions on the basis of learned contingencies between events in the external world, potentially the largest source of sensory prediction is derived internally, from our motor system. Namely, any movement gives rise to largely predictable sensory input, which according to the predictive coding framework should therefore be suppressed. Some of the clearest demonstrations of this phenomenon so far have come from fish [47]. Many fish are equipped with electroreceptors, allowing them to detect nearby objects (e.g., other fish) through changes in the electric field around them. However, these fishes' own movements (and in some species, self-generated electric currents) also cause disturbances in the electric field around them, such that detecting non-self objects would benefit from suppressing such self-generated signals. Indeed, several types of predictive signals (arising from corollary discharge, proprioception, and higher level electrosensory regions) have been shown to evoke *negative images* of the predicted sensory input in the electrosensory organs of these fish [47]. When the predicted inputs arrive, they are cancelled out by these negative images, enhancing sensitivity to non-self generated signals. These predictive signals have been shown to be highly plastic—when paired with an artificially generated stimulus they adapt within minutes—and highly precise in terms of timing, amplitude, and spatial location. Similar predictive suppression mechanisms have been observed in humans [48–52].

As noted above, a crucial feature of predictive coding is its hierarchical nature: valid high-level hypotheses can enhance representations through reducing prediction error in lower-order regions. Murray and colleagues [53, 54] have shown that when stimuli have lower level features that can be grouped into a higher order shape there is increased activity in shape-selective area LOC, but decreased activity in V1, compared to stimuli for which no such grouping takes place. The researchers ensured that the stimuli were matched for low-level features, precluding an explanation in terms of physical differences between the conditions. Presumably, the inferences of high level areas are subtracted from the incoming sensory signals in lower order areas, leading to reduced activity in V1 whenever such a high level hypothesis is generated.

11.3.2 Encoding of Predictions in the Brain

In a predictive coding framework of perception, prior expectations may be hypothesised to activate representations of predicted stimuli prior to sensory input [55]. One way to test this is to probe activity in sensory cortex when a stimulus is predicted, but no bottom-up input is subsequently provided. In line with this, recent studies have shown increased responses to unexpectedly omitted stimuli in early sensory cortex [44, 56], as early as 100 ms after the stimulus was predicted to appear [27, 30]. Recently, we used multivariate methods to probe the representational content of such omission responses [57]. In this study, we presented subjects with auditory cues (high or low pitch) that predicted the orientation of an upcoming grating stimulus (clockwise or anticlockwise). In 25 % of trials, the grating stimulus was omitted. In these trials, only a prediction-inducing auditory tone was presented. Interestingly,

the pattern of activity evoked in V1 on these omission trials was similar to the pattern evoked by the predicted stimulus (e.g., a clockwise grating). In other words, neural activity in V1 evoked solely by predictions, in the absence of visual input, contained information about the grating orientation that was predicted.

Further evidence for representation-specific signals in early visual cortex in the absence of input comes from a study that presented subjects with naturalistic images of which one quadrant was occluded [58]. These authors used multivariate methods to show that the non-stimulated region of V1 (i.e., retinotopically corresponding to the occluded quadrant) contained information about the presented naturalistic image. Control analyses showed that this could not be explained by the spreading of activity (through lateral connections) from stimulated regions within V1. In a hierarchical inference framework, these results would be taken to reflect predictive feedback from a higher-order region containing a representation of the naturalistic scene as a whole (e.g., the whole car), informing lower-order areas which fine spatial features to expect (e.g., the angle of the tail light).

Single neuron recordings in monkeys have also provided empirical support for neuronal responses to absent but predicted input. Lee and Nguyen [59] presented monkeys with Kanizsa figures (Fig. 11.1a) in such a way that the illusory contours were positioned in the receptive fields of the V1 and V2 neurons they recorded from. Interestingly, both V1 and V2 neurons responded to these illusory contours (Fig. 11.3c). Moreover, V2 neurons responded consistently earlier in time, suggesting a feedback mechanism from V2 to V1. Similarly to the apparent motion example discussed earlier, this can be understood to result from inferences about the presence of a white square occluding the black circles within higher-order visual regions with receptive fields that encompass the whole figure. These inferences are subsequently sent as top-down predictions to those lower-order neurons that are expected to detect the (illusory) sides of the square [60]. It should be noted, however, that some previous studies have observed illusory contour responses in V2, but not in V1 [61]. The presence of predictive feedback to V1 may depend on such factors as stimulus size, attention, and experience with the stimulus.

It should be noted that while the above studies clearly demonstrate representation-specific activity in the absence of sensory input, they cannot strictly distinguish between prediction (P unit) and prediction error (PE unit) activity. Since PE reflects the mismatch between the prior expectation (a specific stimulus) and the input (an empty screen), unexpected omissions could conceivably cause activity in PE units. Evidence of true ‘predictive’ activity would require demonstrating representation-specific activity *prior* to sensory input. Such evidence is provided by paired-association studies in the anterior temporal cortex of macaques [28, 62, 63]. In these studies, monkeys were exposed to pairs of stimuli that were sequentially presented. Learning which stimuli form pairs allowed the monkeys to predict the identity of the second member of a pair upon presentation of the first member. Indeed, after learning, neurons that respond strongly to the second member of a pair already start firing upon presentation of the first member of the pair, i.e., as soon as the identity of the second member can be predicted (Fig. 11.3d). This predictive firing increases until the second stimulus appears, and is higher when monkeys correctly identify

the upcoming stimulus [62]. Furthermore, Erickson and Desimone [63] found that in sessions in which monkeys showed behavioural evidence of association learning, neural activity during the delay period between the two stimuli correlated less strongly with the response to the first stimulus, and more with the response to the second (upcoming) stimulus. In other words, delay activity became more predictive. Meyer and Olson [28] showed that when such a prediction is violated, that is, when the first stimulus is followed by one it has not previously been paired with, neural activity to the second stimulus is increased, suggesting a prediction error response. Similar findings of pair association in the medial temporal lobe have been reported in humans using fMRI [64].

11.3.3 Integration of Predictions and Inputs

In predictive coding theories, P and PE units do not function independently of each other. Rather, the error response in the PE units influences the hypotheses selected in the P units, and vice versa. Therefore, the final hypothesis the brain settles on (the posterior) reflects an integration of prior expectations and sensory input. In other words, if you expect *A*, and get bottom-up input *B*, your percept (posterior) should be somewhere in between. The relative weights of prior and input depend on their precisions: when the input is strong and unequivocal, the prior will have little effect, but if the input is ambiguous, the posterior will be largely determined by the prior [65, 66]. The integration of prior and input has been demonstrated convincingly in behaviour [10, 67, 68], yet neural evidence has been mostly lacking. The main question is whether there is already integration of bottom-up inputs and top-down expectations in sensory regions, as predicted by predictive coding theories, or whether integration takes place in downstream association areas that are proposed to be involved in perceptual decision-making, such as parietal and prefrontal cortex [69]. In line with the former, Nienborg and Cumming [70] have shown that sensory responses in macaque early visual cortex (V2) are dynamic, shifting from the representation of the bottom-up stimulus to the perceptual choice (posterior) within seconds. In a recent study in humans, Serences and Boynton [71] presented participants with ambiguous (0% coherent) moving dot stimuli and forced them to report whether the dots moved toward the top right or bottom left of the screen. Multivariate analysis methods revealed that motion sensitive area MT+ contained information about the (arbitrarily) chosen direction of motion. These studies suggest that early visual areas represent the posterior rather than just the bottom-up input. While the source of the arbitrary choice (i.e., the prior) is unknown and undetermined in the study by Serences and Boynton [71], future studies may test the integration of prior and bottom-up stimulus more directly by explicitly manipulating the prior and probing its effects on stimulus representations in visual cortex.

11.3.4 *Separate Units Coding Predictions and Prediction Errors*

Although many findings of prediction and prediction error effects in cortex have been reported, there is, somewhat surprisingly, a conspicuous lack of direct evidence for separate populations of neurons encoding predictions (P units) and errors (PE units) [72]. However, some conjectures can be made.

Miller and Desimone [73] recorded from single neurons in IT cortex while monkeys performed a delayed match-to-sample task. In this task, monkeys were presented with a sample stimulus, and had to respond when any of the following test stimuli matched the sample. Roughly half of the IT cells recorded showed differential responses to stimuli that matched, compared to stimuli that did not match the sample. Of these, 62 % were *suppressed* by test stimuli that matched the sample, while 35 % showed an *enhanced* response. These effects were present right from the onset of visual responses in IT, about 80–90 ms after stimulus presentation. Only 3 % of cells showed mixed effects, i.e., suppression by some stimuli and enhancement by others, leading the authors to argue that the two classes of cells appear to be distinct. The behaviour of these two classes of cells is reminiscent of PE (suppressed response to matches) and P (enhanced response to matches) units, respectively, though effects of stimulus predictability were not explicitly tested in this study. Woloszyn and Sheinberg [74] also provided evidence for two functionally distinct sub-populations in IT. They found that the maximum response and stimulus-selectivity of excitatory cells were increased for familiar compared to novel stimuli (potentially reflecting enhanced representation in P units), while inhibitory interneurons responded more strongly to *novel* stimuli than to familiar ones (potentially reflecting a larger PE response).

Arguments for a separate population of prediction error neurons have also been inspired by so-called extra-classical receptive field effects in early visual cortex [1]. Certain neurons fire less when a stimulus extends beyond their receptive field [75]. Furthermore, such suppressive surround effects are stronger when the surround is a (predictable) continuation of the centre stimulus, e.g., a continuous line segment or a grating with an iso-oriented surround, compared to when the surround is non-continuous (e.g., a cross-oriented grating) [76–78]. A predictive coding framework can readily explain such responses; a large, continuous stimulus is represented well by a P unit in a higher-order area (e.g., V2), which then sends a prediction to the relevant lower-order (e.g., V1) error neurons, suppressing their response [1]. Indeed, extra-classical receptive field effects have been shown to (partly) depend on feedback from higher-order areas [79, 80]. Hupé et al. [80] showed that feedback from area MT leads to surround suppression in V1, as well as increased responses to stimuli confined to the classical receptive field. In other words, when feedback can successfully predict the lower-order response its effect is inhibitory, but when it cannot it is excitatory.

One (somewhat counterintuitive) feature of the classical predictive coding scheme is that P units in one region send inhibitory feedback to the PE units one step down in the cortical hierarchy. However, in the cortex, interregional feedback connections are

predominantly excitatory [but see [19, 80, 81]]. It is possible that feedback may indirectly target inhibitory interneurons, achieving a net inhibition, as has been observed in surround suppression [79, 80]. Furthermore, there are alternative implementations of predictive coding that do not rely on inhibitory intercortical feedback. In the work of Spratling [18, 83], for example, excitatory feedback is sent from P units in one region to P units in the region below it in the cortical hierarchy. In other words, feedback directly reinforces lower order hypotheses that are consistent with the higher order hypothesis. Here, error suppression is an intracortical phenomenon, consistent with intraregional ‘back projections’ (i.e., from infragranular to supragranular and granular layers) targeting predominantly inhibitory interneurons [84, 85].

In sum, while there is no direct unequivocal evidence for the existence of separate populations of P and PE units, there is suggestive evidence that different layers within each cortical column may implement these distinct computational roles.

11.4 Predictive Coding and Attention

Traditionally, theories of attention and predictive coding have been seen as diametrically opposed [72, 86]. While predictive coding posits that neural responses to expected stimuli should be suppressed, many studies have reported *increased* neural responses to stimuli appearing at expected locations [87, 88]. This increase in activity has been attributed to spatial attention. In fact, studies of visuospatial attention have traditionally used probabilistic cues that predict the upcoming target location as a means of manipulating attention [89, 90]. However, belying this apparent tension between theories of attention and prediction, attention fits quite naturally into a predictive coding framework that takes the relative precision of predictions and sensory input into account [91, 92]. In what follows, we will outline an account of attention in predictive coding, and review empirical evidence for this theory.

11.4.1 Attention and Precision

In the real world, the reliability of sensory signals is changeable: keeping track of a ball is a lot easier in the light of day than at dusk. Perceptual inference must take the precision (inverse variance) of sensory signals (i.e., prediction errors) into account [2, 91, 93]. It is imperative to know whether sensory signals fail to match our prior expectations because they contain information that disproves our current model of the world (e.g., we see and hear a giant luminescent hound), or because the sensory signals are simply too noisy (we hear a dog howl but see only mist). While the former should lead us to update our beliefs (a demon hound!), the latter should not. Specifically, PEs should be weighted by their precision (i.e., reliability), leading to less weight being attributed to less reliable sensory information. In terms of hierarchical inference, perception should be dominated by sensory signals when

their precision is high, and by top-down expectations when their precision is low, e.g., when sensory signals are ambiguous [65, 66].

The precision of sensory signals has long been acknowledged to be a matter of importance for models of perceptual inference [2, 93], and recent predictive coding models incorporate it explicitly [91, 92]. In these models, the brain estimates not only PEs themselves, but also their precision. PEs are subsequently weighted by their precision through modulation of synaptic gain of the PE units. One hypothesis suggests that attention is the process whereby the brain optimises precision estimates [91, 92]. By increasing the precision of specific PEs, attention increases the weight these errors carry in perceptual inference. Mechanistically, this is equivalent to proposals of attention increasing synaptic gain (precision) of specific sensory neurons (PE units) [94–96]. Note that in predictive coding models, sensory signals and PE are equivalent, since these errors are the only sensory information that is yet to be explained. Therefore, while casting attention as modulating the precision of PEs may at first glance seem a radically different view of its function, it is in fact fully consistent with contemporary theories of attention. Indeed, in this account, spatial attention is proposed to increase the precision of information coming from a certain region of visual space, similar to the effect of pointing a flashlight in that direction, making spotlight metaphors of attention an intuitive way to think about its functional role [92, 97]. Furthermore, biased competition, a highly influential theory of attention [94], can be shown to emerge from a predictive coding framework in which attention is cast as optimising precision [92].

Crucially, prediction and precision (attention) are not independent. Instead, precision depends on the expected states of the world [92]: expectation of a stimulus on the left side of the visual field leads to expectation of high precision sensory signals at that location. Spatial attention might enhance these sensory signals further by boosting the precision (synaptic gain) at that location. This suggests that attention can be seen as a selective enhancement of sensory data that have high precision (high signal-to-noise ratio) in relation to the brain's current predictions [98]. Mechanistically, this means that attention does not simply boost the synaptic gain of PE units indiscriminately. Rather, it boosts the gain of PE units receiving input from P units (in the same hierarchical level and the level above) that are currently active. Therefore, attending to a region of space where a stimulus is not expected (in this example, to the right) would be relatively ineffective. Put simply, there should be no strong expectation of a high precision sensory signal when a stimulus is not expected to appear.

In sum, recent predictive coding theories propose that prediction is concerned with *what* is being represented, while attention is the process of optimising the *precision* of representations [91]. We will now turn to a discussion of empirical evidence for this proposal.

11.4.2 Empirical Evidence

The account outlined above may reconcile some seemingly contradictory findings in the literature. Generally speaking, it seems that expectation is associated with reduced sensory signals when stimuli are task irrelevant (unattended), but enhanced responses when stimuli are relevant (attended) [99]. For instance, two recent studies found opposite effects of predictive motion trajectories on stimulus processing. Doherty et al. [87] had participants view a red ball moving across the screen in either a regular (predictable) or irregular (unpredictable) trajectory. At some point the ball disappeared behind an occluder, and when it reappeared participants were required to detect a black dot on the surface of the ball as soon as possible. The different types of trajectories meant that the reappearance of the dot could be either predictable (in space and time) or unpredictable. Using EEG, these authors found that predictability enhanced the neural response in early sensory regions. In contrast, Alink et al. [37] found a *reduced* neural response in V1 in response to a stimulus that was congruent with a predictable trajectory (see § 11.3.1 for a more detailed discussion of this study), compared to a stimulus that was incongruent with the trajectory. In their design, subjects did not perform a task that involved the stimuli, instead stimulus presentations were fully irrelevant.

These studies provide a suggestion for a potential interaction between attention and prediction [for a review, see 98]. However, there are also notable differences between these two studies in terms of experimental paradigm (e.g., real vs. apparent motion) and methodology (EEG vs. fMRI). A direct study of the (potential) interaction of prediction and attention has been lacking.

In a recent fMRI study, we [56] manipulated both visuospatial attention (which side of the screen is task-relevant) and visuospatial prediction (on which side of the screen the stimulus is likely to appear). Spatial attention was manipulated on a trial-by-trial basis, by means of a cue that pointed either to the left or to the right. Unlike in typical Posner cueing tasks, in which attention is biased towards one visual hemifield by increasing the probability of the target appearing on that side, in this experiment, the attention manipulation was not probabilistic. The key difference is that in the former, both visual hemifields are potentially task-relevant, with one more likely to be so than the other, while in our study only one hemifield was task-relevant in a particular trial. If a subsequent grating stimulus appeared in the hemifield indicated by the cue, subjects were asked to do an orientation identification task on the grating stimulus. If instead the grating appeared on the other side of the screen, subjects could simply ignore it. Prediction was manipulated in mini-blocks: in each block of eight trials, subjects were told that stimuli were (a) 75 % likely to appear on the left, (b) 75 % likely to appear on the right, (c) 50 % likely to appear on either side (neutral cue). Thereby, attention and prediction were orthogonally manipulated. We reasoned that there were two likely explanations for the seemingly contradictory effects of expectation reported in the literature. One possible explanation is that attention and prediction have opposing main effects, enhancing and suppressing sensory signals, respectively. If the enhancing effect of attention outweighs the suppressive

effect of prediction, this would explain the seemingly enhancing effect of expectation in attentional cueing experiments. Alternatively, attention and prediction might operate synergistically to optimise perceptual inference, with attention boosting the precision (synaptic gain) of PE units that are expected to receive input based on current predictions. Hereby, if attention and prediction are congruent (i.e., a stimulus is expected in the task-relevant hemifield), attention would boost (the precision of) the expected sensory signal. However, if attention and prediction are incongruent (a stimulus is expected in the task-irrelevant hemifield), attention would be relatively ineffective in boosting the sensory signal; there should be no strong expectation that an unpredicted signal is going to be precise. This account would therefore predict an interactive effect of attention and prediction on sensory signals (see Fig. 11.4a).

The data provided support for the latter hypothesis (Fig. 11.4b). When stimuli were task-irrelevant (unattended), predicted stimuli evoked a reduced neural response in V1 compared to unpredicted stimuli. However, when stimuli were task-relevant (attended), this pattern reversed: here, predicted stimuli evoked an enhanced neural response. This interaction is in line with predictive coding models casting attention as optimising precision estimates during perceptual inference [91, 92]. Furthermore, when a stimulus was predicted in the task-relevant hemifield (i.e., there was a strong and precise prediction), we observed an increased response in V1 when this stimulus was omitted (Fig. 11.4c). As discussed in § 11.3.2, this might reflect either the prediction itself, or a prediction error response. In either case, this effect is in line with predictive coding, but hard to reconcile with bottom-up attention accounts of ‘stimulus surprise’, since there was no stimulus to grab attention in these trials (or, rather, a stimulus appeared in the opposite hemifield).

Further support for attention increasing the precision of sensory signals (prediction errors) comes from studies showing that fluctuations in the amplitude of the neural signal in visual areas covary with detection task performance, both pre- [100] and post-stimulus [101]. In other words, activity in these regions is higher when people correctly detect or reject the presence of a stimulus than when they incorrectly report or miss it (although amplitude has also been seen to covary with subjective perception rather than performance accuracy; [102]).

Boosting specific sensory signals results in a gain in signal-to-noise ratio (precision) for those signals. Such a gain could also be achieved by reducing the neural noise in sensory areas. In fact, single cell recordings in macaques have revealed a decrease in neural noise correlations as the result of attention [103, 104]. Furthermore, a recent behavioural study that applied sophisticated signal detection theory analyses showed that whereas prediction increases the baseline activity of stimulus-specific units (*P* units?), attention suppresses internal noise during signal processing [55]. In order to optimally boost the signal-to-noise ratio of selected sensory signals, attention may both increase the amplitude of specific prediction errors, as well as suppress noise fluctuations arising from non-selected sources.

In sum, the empirical findings discussed above are in line with predictive coding models incorporating precision estimates, and casting attention as the process of optimising those estimates. This framework resolves the apparent tension between

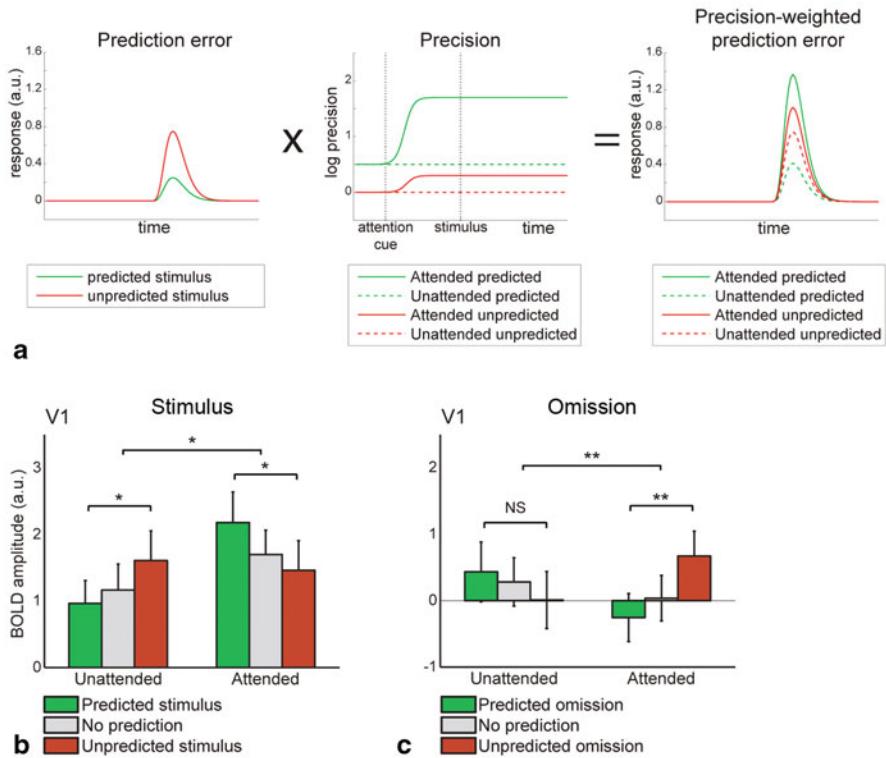


Fig. 11.4 Attention and precision. **a** *Left* panel: The hypothetical prediction error responses to physically identical stimuli, preceded by either a valid (*green*) or invalid (*red*) prediction cue. *Middle* panel: In recent predictive coding models, attention increases the precision (synaptic gain) of prediction errors. This enhancement of precision by attention occurs in relation to current predictions, reflected here by the fact that attention hardly increases precision when no stimulus is predicted to occur. The order of magnitude of the precision values displayed here was based on figures in Feldman and Friston [91], the exact values were chosen arbitrarily, and their evolution over time was simplified. *Right* panel: Prediction errors are weighted by their precision, calculated here as a simple multiplication of prediction error (*left* panel) and precision (*middle* panel). The fact that attention enhances precision *in relation to current predictions* leads to an interactive effect of prediction and attention on the amplitude of the prediction error response. **b** When stimuli are unattended (task irrelevant), predicted stimuli evoke a reduced response in V1 compared to unpredicted stimuli. On the other hand, when stimuli are attended, predicted stimuli evoked a larger response than unpredicted stimuli. This is exactly the interaction between attention and prediction that is hypothesised by recent predictive coding models, see **a**. **c** In visual cortex corresponding to the visual field where no stimulus appeared, i.e. ipsilateral to the stimulus, unpredicted omission of a stimulus in the attended visual field evoked a larger response in V1 than predicted omission of a stimulus. Figures reprinted from [56], with permission from the authors

theories of attention and predictive coding [18], and explains the seemingly contradictory findings in the literature regarding the effects of expectation on neural activity [72, 99].

11.5 Concluding Remarks

In this chapter, we reviewed recent theoretical and empirical advances in the field of predictive coding. Although we have shown that predictive coding makes predictions that can be tested by cognitive neuroscience, and which have been supported by the extant data reasonably well, we would like to stress that more evidence is needed. Particularly, direct evidence for separate sub-populations of P and PE units is lacking. Since these two sub-populations are proposed to co-exist in every (sensory) cortical region, high-resolution methods are required to simultaneously sample neural activity from multiple sites at high spatial resolution. Specifically, given the speculations on different laminar distributions of P and PE units (see § 11.2.3), multicontact laminar electrodes [e.g., [104] or high-resolution laminar fMRI [106] could provide such evidence. So far, there have been no studies using these methods that have focused on the effects of prediction on sensory responses. Under a predictive coding framework, it may be hypothesised that, preceding stimulus onset, expectation would lead to activity in cortical layers containing P units, while after stimulus onset, activity would scale with prediction error in layers dominated by PE units. At the level of single neurons, P and PE units are predicted to be reciprocally connected, with the strength of the excitatory forward connection between individual PE and P units being equal to the strength of the inhibitory backward connection between these same neurons [1, 18]. In V1, it seems conceivable that simple and complex cells [75] could be interpreted as PE and P units, respectively. If this is true, complex cells are expected to inhibit the simple cells that provide them with input. This is a testable hypothesis. In the coming years, studies testing these hypotheses will provide us with much needed answers regarding the possible implementation of predictive coding in the human cortex.

So far, studies of predictive coding have mostly focused on the effects of prediction on the processing of sensory inputs. However, the model also provides a natural explanation for top-down activations of representations in the absence of sensory input. For example, processes like working memory and mental imagery (and even dreaming) might reflect activating part of one's internal model of the (visual) world [2]. These activations would come about through a different flow of information, compared to stimulus-driven activations: whereas the latter would arrive as input into layer 4 and sent onwards to supra- and infragranular layers, the former would bypass layer 4 and directly target agranular layers [107]. Crucially, these opposite flows of information could result in identical representations being activated (in agranular layers). Indeed, recent neuroimaging studies suggest that working memory [108], mental imagery [109, 110], and even dreaming [111] share sensory representations with perception. Such offline activations of the brain's internal model could serve several purposes, such as simulating scenario's not (yet) encountered but consistent with the model (e.g., mental rehearsal), and consolidating synaptic connections between representations within and across different levels of the cortical hierarchy. Speculatively, dreams may subserve both these functions.

Future work might also focus on the link between the neuronal substrate of predictive coding and subjective perception. It seems natural to assume that the contents of perception reflect the current hypothesis represented in the P units across the cortical hierarchy. Might the intensity (e.g., brightness, contrast, duration) of the percept then scale with the prediction error [25]? This account would predict that valid expectations lead to percepts that are ‘sharper’ (improved representation in P units) but less intense (reduced PE), in line with neural effects of expectation in sensory cortex [14]. Indeed, oddball stimuli (that is, unexpected deviants) are perceived as being of longer duration than standards [25, 112, 113]. Also, this account can explain the fact that representations activated by top-down processes such as working memory and imagery are not perceived as vividly as those activated during normal perception; presumably the former bypass PE units and directly activate P units. Furthermore, since attention is proposed to boost the synaptic gain of PE units (see § 4), the increase in perceived contrast observed as a result of attention fits naturally in this framework [114]. Finally, psychosis has been conjectured to involve aberrantly increased prediction errors, and indeed patients report more intense percepts (brighter colours, louder sounds) in early stages of the disease [115]. In fact, it is interesting to note that many positive and negative symptoms of syndromes like schizophrenia [116–118], psychosis [115], and autism [119, 120] can be explained in terms of specific failures of predictive coding mechanisms.

In sum, predictive coding provides a good explanation for many phenomena observed in perception, and generates testable predictions. In this chapter, we have reviewed existing empirical evidence for some of these predictions, as well as outlined possible future directions for further empirical testing and for broadening the perspective of the role predictive coding may play in cognition.

Exercises

1. Does the suppressed response in V1 to predicted stimuli [37, 44, 53] mean that there is less stimulus information in V1 for such stimuli? Why/Why not?
2. In what respect are the neural effects of prediction and attention opposite to each other, and in what respect are they similar?
3. Come up with an experiment that could potentially falsify predictive coding.
4. Given that top-down predictions silence prediction errors in lower-order regions; does predictive coding require inhibitory feedback between cortical regions? Read the paper by Spratling [18] and prepare a 15 min presentation on the differences between the physiological implementation implied by "classical" predictive coding models [1, 3] and that implied by Spratling's PC/BC model.
5. During hallucinations, schizophrenic and psychotic patients perceive things that are not actually there. Autistic patients, on the other hand, sometimes seem to perceive things more precisely or truthfully than non-autistics. How could these symptoms be understood in terms of predictive coding?
6. Read the Corlett et al. [115] paper, and prepare a 30 min presentation on the relationship between predictive coding and psychosis.

Further Reading

1. For an introduction to the principles of predictive coding and a global perspective of its implications for cognition and action, see the recent review by Andy Clark [121].
2. Friston [3] offers a comprehensive and mathematical description of predictive coding, including a proposal for its neuronal implementation.
3. Summerfield and Egner [72] review the commonalities and differences between theories of predictive coding and attention.
4. In a clearly written and succinct paper, Spratling [18] presents the computational principles of predictive coding and biased competition, and shows that—under certain assumptions—they are equivalent.
5. Lee and Mumford [9] offer a slightly different take on hierarchical inference during perception that shares many of the principles of predictive coding.

References

1. Rao RPN, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 2(1):79–87
2. Mumford D (1992) On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biol Cybern* 66(3):241–251
3. Friston KJ (2005) A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci* 360(1456):815–836
4. Helmholtz H (1867) *Handbuch der physiologischen optik*. L. Voss, Leipzig
5. Weiss Y, Simoncelli EP, Adelson EH (2002) Motion illusions as optimal percepts. *Nat Neurosci* 5(6):598–604
6. Gregory RL (1997) Knowledge in perception and illusion. *Phil Trans R Soc Lond B* 352:1121–1128
7. Knill DC, Richards W (Eds) (1996) *Perception as Bayesian inference*. Cambridge University Press: Cambridge
8. Yuille A, Kersten D (2006) Vision as Bayesian inference: analysis by synthesis? *Trends Cogn Sci* 10(7):301–308
9. Lee TS, Mumford D (2003) Hierarchical Bayesian inference in the visual cortex. *J Opt Soc Am A Opt Image Sci Vis* 20(7):1434–1448
10. Kersten D, Mamassian P, Yuille A (2004) Object perception as Bayesian inference. *Annu Rev Psychol* 55:271–304
11. Den Ouden HEM, Kok P, De Lange FP (2012) How prediction errors shape perception, attention, and motivation. *Front Psychol* 3:548
12. Spratling MW (2012) Predictive coding as a model of the V1 saliency map hypothesis. *Neural Networks* 26:7–28
13. Pearce JM, Hall G (1980) A model for Pavlovian learning: variations on the effectiveness of conditioned but not of unconditioned stimuli. *Psychol Rev* 87(6):532–552
14. Kok P, Jehee JFM, De Lange FP (2012) Less is more: expectation sharpens representations in the primary visual cortex. *Neuron* 75:265–270
15. Bar M (2004) Visual objects in context. *Nat Rev Neurosci* 5(8):617–629
16. Ahissar M, Hochstein S (2004) The reverse hierarchy theory of visual perceptual learning. *Trends Cogn Sci* 8(10):457–464
17. Sugase Y et al (1999) Global and fine information coded by single neurons in the temporal visual cortex. *Nature* 400:869–873

18. Spratling MW (2008) Reconciling predictive coding and biased competition models of cortical function. *Front Comput Neurosci* 2:4
19. Bastos AM et al (2012) Canonical microcircuits for predictive coding. *Neuron* 76:695–711
20. Felleman DJ, Van Essen DC (1991) Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* 1(1):1–47
21. Maunsell JHR, Van Essen DC (1983) The connections of the middle temporal visual area (MT) and their relationship to a cortical hierarchy in the macaque monkey. *J Neurosci* 3(12):2563–2586
22. Rockland K, Pandya D (1979) Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Res* 179:3–20
23. Rao RPN, Sejnowski, TJ (2002) Predictive coding, cortical feedback, and spike-timing dependent plasticity. In: Rao RPN, Olshausen BA, Lewicki MS (eds) *Probabilistic models of the brain: perception and neural function*. MIT Press, Cambridge, pp 297–315
24. Grill-Spector K, Henson R, Martin A (2006) Repetition and the brain: neural models of stimulus-specific effects. *Trends Cogn Sci* 10(1):14–23
25. Pariyadath V, Eagleman D (2007) The effect of predictability on subjective duration. *PLoS ONE* 2(11):e1264
26. Summerfield C et al (2008) Neural repetition suppression reflects fulfilled perceptual expectations. *Nat Neurosci* 11(9):1004–1006
27. Todorovic A et al (2011) Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. *J Neurosci* 31(25):9118–9123
28. Meyer T, Olson CR (2011) Statistical learning of visual transitions in monkey inferotemporal cortex. *Proc Natl Acad Sci U S A* 108(48):19401–19406
29. Todorovic A, De Lange FP (2012) Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *J Neurosci* 32(39):13389–13395
30. 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(51):20754–20759
31. Wacongne C, Changeux JP, Dehaene S (2012) A neuronal model of predictive coding accounting for the mismatch negativity. *J Neurosci* 32(11):3665–3678
32. Schröger E, Wolff C (1996) Mismatch response of the human brain to changes in sound location. *Neuroreport* 7:3005–3008
33. Muckli L et al (2005) Primary visual cortex activity along the apparent-motion trace reflects illusory perception. *PLoS Biol* 3(8):e265
34. Sterzer P, Haynes JD, Rees G (2006) Primary visual cortex activation on the path of apparent motion is mediated by feedback from hMT + /V5. *Neuroimage* 32:1308–1316
35. Wibrals M et al (2009) The timing of feedback to early visual cortex in the perception of long-range apparent motion. *Cereb Cortex* 19:1567–1582
36. Ahmed B et al (2008) Cortical dynamics subserving visual apparent motion. *Cereb Cortex* 18:2796–2810
37. Alink A et al (2010) Stimulus predictability reduces responses in primary visual cortex. *J Neurosci* 30(8):2960–2966
38. Chandrasekaran C et al (2009) The natural statistics of audiovisual speech. *PLoS Comput Biol* 5(7):e1000436
39. Sumbly WH, Pollack I (1954) Visual contribution to speech intelligibility in noise. *J Acoust Soc Am* 26(2):212–215
40. Van Wassenhove V, Grant KW, Poeppel D (2005) Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci U S A* 102(4):1181–1186
41. Arnal LH et al (2009) Dual neural routing of visual facilitation in speech processing. *J Neurosci* 29(43):13445–13453
42. McGurk H, MacDonald J (1976) Hearing lips and seeing voices. *Nature* 264:746–748
43. Arnal LH, Wyart V, Giraud A (2011) Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nat Neurosci* 14(6):797–801
44. Den Ouden HEM et al (2009) A dual role for prediction error in associative learning. *Cereb Cortex* 19:1175–1185

45. Den Ouden HEM et al (2010) Striatal prediction error modulates cortical coupling. *J Neurosci* 30(9):3210–3219
46. Eger T, Monti JM, Summerfield C (2010) Expectation and surprise determine neural population responses in the ventral visual stream. *J Neurosci* 30(49):16601–16608
47. Bell CC (2001) Memory-based expectations in electrosensory systems. *Curr Opin Neurobiol* 11:481–487
48. Blakemore SJ, Wolpert DM, Frith CD (1998) Central cancellation of self-produced tickle sensation. *Nat Neurosci* 1(7):635–640
49. Houde JF et al (2002) Modulation of the auditory cortex during speech: an MEG study. *J Cognitive Neurosci* 14(8):1125–1138
50. Martikainen MH, Kaneko K, Hari R (2005) Suppressed responses to self-triggered sounds in the human auditory cortex. *Cereb Cortex* 15:299–302
51. Shergill SS et al (2013) Modulation of somatosensory processing by action. *Neuroimage* 70:356–362
52. Crapse TB, Sommer MA (2008) Corollary discharge across the animal kingdom. *Nat Rev Neurosci* 9:587–600
53. Murray SO et al (2002) Shape perception reduces activity in human primary visual cortex. *Proc Natl Acad Sci U S A* 99(23):15164–15169
54. Fang F, Kersten D, Murray SO (2008) Perceptual grouping and inverse fMRI activity patterns in human visual cortex. *J Vision* 8(7):2–9
55. Wyart V, Nobre AC, Summerfield C (2012) Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. *Proc Natl Acad Sci U S A* 109(9):3593–3598
56. Kok P et al (2012) Attention reverses the effect of prediction silencing sensory signals. *Cereb Cortex* 22(9):2197–2206
57. Kok P, Failing FM, De Lange FP (2014) Prior expectations evoke stimulus templates in the primary visual cortex. *J Cogn Neurosci* 26(7):1546–1554
58. Smith FW, Muckli L (2010) Nonstimulated early visual areas carry information about surrounding context. *Proc Natl Acad Sci U S A* 107(46):20099–20103
59. Lee TS, Nguyen M (2001) Dynamics of subjective contour formation in the early visual cortex. *Proc Natl Acad Sci U S A* 98(4):1907–1911
60. Kok P, De Lange FP (2014) Shape perception simultaneously up- and downregulates neural activity in the primary visual cortex. *Curr Biol* 24:1531–1535
61. Von Der Heydt R, Peterhans E, Baumgartner G (1984) Illusory contours and cortical neuron responses. *Science* 224(4654):1260–1262
62. Sakai K, Miyashita Y (1991) Neural organization for the long-term memory of paired associates. *Nature* 354:152–155
63. Erickson CA, Desimone R (1999) Responses of macaque perirhinal neurons during and after visual stimulus association learning. *J Neurosci* 19(23):10404–10416
64. Schapiro AC, Kustner LV, Turk-Browne NB (2012) Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Curr Biol* 22(17):1622–1627
65. Sterzer P, Frith C, Petrovic P (2008) Believing is seeing: expectations alter visual awareness. *Curr Biol* 18(16):R697–R698
66. Denison RN, Piazza EA, Silver MA (2011) Predictive context influences perceptual selection during binocular rivalry. *Front Human Neurosci* 5:166
67. Chalk M, Seitz AR, Seriès P (2010) Rapidly learned stimulus expectations alter perception of motion. *J Vision* 10(8):1–18
68. Sotiropoulos G, Seitz AR, Seriès P (2011) Changing expectations about speeds alters perceived motion direction. *Curr Biol* 21(21):R883–R884
69. Gold JI, Shadlen MN (2007) The neural basis of decision making. *Annu Rev Neurosci* 30:535–74
70. Nienborg H, Cumming BG (2009) Decision-related activity in sensory neurons reflects more than a neuron's causal effect. *Nature* 459:89–92
71. Serences JT, Boynton GM (2007) The representation of behavioral choice for motion in human visual cortex. *J Neurosci* 27(47):12893–12899

72. Summerfield C, Egnér T (2009) Expectation (and attention) in visual cognition. *Trends Cogn Sci* 13(9):403–409
73. Miller EK, Desimone R (1994) Parallel neuronal mechanisms for short-term memory. *Science* 263(5146):520–522
74. Woloszyn L, Sheinberg DL (2012) Effects of long-term visual experience on responses of distinct classes of single units in inferior temporal cortex. *Neuron* 74:193–205
75. Hubel DH, Wiesel TN (1968) Receptive fields and functional architecture of monkey striate cortex. *J Physiol* 195(1):215–243
76. Jones HE, Wang W, Sillito AM (2002) Spatial organization and magnitude of orientation contrast interactions in primate V1. *J Neurophysiol* 88:2796–2808
77. Knierim JJ, Van Essen DC (1992) Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *J Neurophysiol* 67:961–980
78. Sillito AM et al (1995) Visual cortical mechanisms detecting focal orientation discontinuities. *Nature* 378:492–496
79. Angelucci A, Bullier J (2003) Reaching beyond the classical receptive field of V1 neurons: horizontal or feedback axons? *J Physiology-Paris* 97:141–154
80. Hupé JM et al (1998) Cortical feedback improves discrimination between figure and background by V1, V2 and V3 neurons. *Nature* 394:784–787
81. Johnson RR, Burkhalter A (1996) Microcircuitry of forward and feedback connections within rat visual cortex. *J Comp Neurol* 368:383–398
82. Johnson RR, Burkhalter A (1997) A polysynaptic feedback circuit in rat visual cortex. *J Neurosci* 17(18):7129–7140
83. Spratling MW (2008) Predictive coding as a model of biased competition in visual attention. *Vision Res* 48:1391–1408
84. Thomson AM, Bannister AP (2003) Interlaminar connections in the neocortex. *Cereb Cortex* 13:5–14
85. Olsen SR et al (2012) Gain control by layer six in cortical circuits of vision. *Nature* 483:47–52
86. Koch C, Poggio T (1999) Prediction the visual world: silence is golden. *Nat Neurosci* 2(1):9–10
87. Doherty JR et al (2005) Synergistic effect of combined temporal and spatial expectations on visual attention. *J Neurosci* 25(36):8259–8266
88. Chaumon M, Drouet V, Tallon-Baudry C (2008) Unconscious associative memory affects visual processing before 100 ms. *J Vis* 8(3):1–10
89. Posner MI (1980) Orienting of attention. *Q J Exp Psychol* 32(1):3–25
90. Mangun GR, Hillyard SA (1991) Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. *J Exp Psychol* 17(4):1057–1074
91. Friston KJ (2009) The free-energy principle: a rough guide to the brain? *Trends Cogn Sci* 13(7):293–301
92. Feldman H, Friston KJ (2010) Attention, uncertainty, and free-energy. *Front Human Neurosci* 4:215
93. Rao RPN, Ballard DH (1997) Dynamic model of visual recognition predicts neural response properties in the visual cortex. *Neural Comput* 9:721–763
94. Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. *Annu Rev Neurosci* 18:193–222
95. Maunsell JHR, Treue S (2006) Feature-based attention in visual cortex. *Trends Neurosci* 29(6):317–322
96. Reynolds JH, Heeger DJ (2009) The normalization model of attention. *Neuron* 61(2):168–185
97. Posner MI, Snyder CR, Davidson BJ (1980) Attention and the detection of signals. *J Exp Psychol* 109(2):160–174
98. Hohwy J (2012) Attention and conscious perception in the hypothesis testing brain. *Front Psych* 3:96
99. Rauss K, Schwartz S, Pourtois G (2011) Top-down effects on early visual processing in humans: a predictive coding framework. *Neurosci Biobehav Rev* 35:1237–1253

100. Hesselmann G et al (2010) Predictive coding or evidence accumulation? False inference and neuronal fluctuations. *PLoS ONE* 5:e9926
101. Ress D, Backus BT, Heeger DJ (2000) Activity in primary visual cortex predicts performance in a visual detection task. *Nat Neurosci* 3(9):940–945
102. Ress D, Heeger DJ (2003) Neuronal correlates of perception in early visual cortex. *Nat Neurosci* 6(4):414–420
103. Mitchell JF, Sundberg KA, Reynolds JH (2009) Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron* 63:879–888
104. Cohen MR, Maunsell JHR (2009) Attention improves performance primarily by reducing interneuronal correlations. *Nat Neurosci* 12(12):1594–1601
105. Lakatos P et al (2008) Entrainment of neuronal oscillations as a mechanism of attentional selection. *Science* 320(5872):110–113
106. Koopmans PJ, Barth M, Norris DG (2010) Layer-Specific BOLD Activation in Human V1. *Hum Brain Mapp* 31:1297–1304
107. Takeuchi D et al (2011) Reversal of interlaminar signal between sensory and memory processing in monkey temporal cortex. *Science* 331:1443–1447
108. Harrison SA, Tong F (2009) Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458:632–635
109. Lee SH, Kravitz DJ, Baker CI (2012) Disentangling visual imagery and perception of real-world objects. *Neuroimage* 59:4064–4073
110. Albers AM et al (2013) Shared representations for working memory and mental imagery in early visual cortex. *Curr Biol* 23:1427–1431
111. Horikawa T et al (2013) Neural decoding of visual imagery during sleep. *Science* 340:639–642
112. Schindel R, Rowlands J, Arnold DH (2011) The oddball effect: perceived duration and predictive coding. *J Vision* 11(2):17
113. Tse PU et al (2004) Attention and the subjective expansion of time. *Percept Psychophys* 66(7):1171–1189
114. Carrasco M, Ling S, Read S (2004) Attention alters appearance. *Nat Neurosci* 7(3):308–313
115. Corlett PR et al (2011) Glutamatergic model psychoses: prediction error, learning, and inference. *Neuropsychopharmacol* 36:294–315
116. Fletcher PC, Frith CD (2009) Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia. *Nat Rev Neurosci* 10(1):48–58
117. Eagleman DM, Pariyadath V (2009) Is subjective duration a signature of coding efficiency? *Phil Trans R Soc B* 364:1841–1851
118. Blakemore SJ et al (2000) The perception of self-produced sensory stimuli in patients with auditory hallucinations and passivity experiences: evidence for a breakdown in self-monitoring. *Psychol Med* 30:1131–1139
119. Pellicano E, Burr D (2012) When the world becomes ‘too real’: a Bayesian explanation of autistic perception. *Trends Cogn Sci* 16(10):504–510
120. Van de Cruys S et al (2013) Weak priors versus overfitting of predictions in autism: reply to Pellicano and Burr (TICS, 2012). *Iperception* 4:95–97
121. Clark A (2013) Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci* 36:181–253