Expectation (and attention) in visual cognition

Christopher Summerfield¹ and Tobias Egner²,³

¹Department of Experimental Psychology, University of Oxford, South Parks Road, Oxford, OX1 3UD, UK
²Department of Psychology and Neuroscience, Duke University, LSRC, Box 90999, Durham, NC 27708, USA
³Center for Cognitive Neuroscience, Duke University, LSRC, Box 90999, Durham, NC 27708, USA

Visual cognition is limited by computational capacity, because the brain can process only a fraction of the visual sensorium in detail, and by the inherent ambiguity of the information entering the visual system. Two mechanisms mitigate these burdens: attention prioritizes stimulus processing on the basis of motivational relevance, and expectations constrain visual interpretation on the basis of prior likelihood. Of the two, attention has been extensively investigated while expectation has been relatively neglected. Here, we review recent work that has begun to delineate a neurobiology of visual expectation, and contrast the findings with those of the attention literature, to explore how these two central influences on visual perception overlap, differ and interact.

Introduction

‘Everyone’, William James famously wrote, ‘knows what attention is’ [1]. A similar claim might be made for expectation: we have a rich and immediate introspective sense of what it means to expect a forthcoming stimulus, or for expectation to be violated (‘Surprise!’). By comparison with attention, however, psychologists and neuroscientists have traditionally dedicated relatively little effort to exploring the brain mechanisms by which expectations about the forthcoming sensory world are generated, represented and implemented. On the contrary, the effects of perceptual expectation often tend to be conflated with those of attention in the empirical literature. For instance, many experimental protocols geared at gauging the effects of attention deploy probabilistic cues that manipulate subjects’ expectations concerning the likely location or feature of a forthcoming target stimulus [2]. However, the behavioural and neural consequences of such probabilistic cueing are typically attributed exclusively to the effects of attention, even though this procedure also engenders perceptual expectations, as well as their fulfilment or violation. The purpose of this review, therefore, is twofold. First, we aim to highlight and synthesize recent empirical and computational work that is fostering an emergent neurobiology of expectation (with a focus on the visual domain). Second, in parallel to reviewing characteristics of visual expectation, we will emphasize how these might differ from, overlap with, or interact with, the much better-characterized processes underlying visual attention. We begin by consider-

ing the purpose and behavioural consequences of visual expectation.

Expectation in visual cognition

Expectations are brain states that reflect prior information about what is possible or probable in the forthcoming sensory environment. For example, on entering a familiar room, we have prior knowledge about the likely configuration of furniture, paintings or plants, gleaned over multiple previous encounters. This type of foreknowledge can be employed to reduce the computational burden of visual perception in two ways. First, expectations can guide the acquisition of visual information; aspects of the environment that are constant (and thus conform to previously formed expectations) do not have to be processed in depth repeatedly. Second, expectation facilitates the interpretation of visual input. Information arriving at the retina is inherently ambiguous: a single pattern of retinal stimulation can be caused by myriad different stimulus configurations, and a single object can elicit many distinct patterns of retinal stimulation, depending on viewpoint, lighting conditions, or the interposition of occluding surfaces. To overcome these ambiguities, the interpretation of a given percept (‘Is this a breadbox or a mailbox?’) can be guided by contextual probabilities (‘Am I in the kitchen or on the street?’). Accordingly, much classical behavioral research (reviewed in [3]) has demonstrated that objects placed within a congruent context (for example, a lamp in a bedroom) are detected faster and more reliably than objects placed in an incongruent context (for example, a lamp on the beach). Similar results have been found in more controlled psychophysical experiments using rudimentary stimuli: for instance, detection of a Gabor patch (sinusoidal gratings) is facilitated by colinearly oriented flanking patches [4]. Despite ongoing controversy [5], these effects do not appear to be attributable to a change in response bias alone [6]. Benefits to visual recognition also accrue from preceding items that have been jointly or conditionally associated with a stimulus or array [7], from the statistics of natural image information [8], or even the mere recent presentation of related information [9]. The perceptual processes underlying visual detection and recognition, thus, have evolved to benefit from probabilistic information about what is likely to occur, via associative learning, priming, or contextual processing of scene information.

Some important similarities and distinctions can be drawn between these characteristics of visual expectation.
and those of visual attention. Expectation facilitates visual perception by constraining interpretation on the basis of prior likelihood, whereas attention alleviates computational burden by prioritizing processing of that subset of visual information deemed to be of the highest relevance to the organism’s goals [10]. Thus, attention is directed endogenously (‘top-down’) to favour processing of features, objects, or spatial locations that are associated with a task goal, reward, or potential danger [2,11], and attention is attracted exogenously (‘bottom-up’) towards stimuli of high physical salience, such as sudden motion [12]. Critically, the motivational considerations that guide attention are in principle orthogonal to the perceptual regularities that guide visual expectations - although in practice,
expectations and attention often coincide and interact. For example, learned perceptual context can guide attention towards a likely location of a behaviourally relevant item [13] (even unconsciously [14]), and the violation of a perceptual expectation might be employed as a signal for attracting attention towards a potentially significant event [15]. Conversely, attention facilitates the formation of stimulus–stimulus associations [16], and a modicum of attentional resources might in fact be a prerequisite for the acquisition of contextual knowledge [17,18]. Regardless of these conceptual distinctions, it is particularly noteworthy that the behavioural effects of expectation and attention are superficially identical; both attended and expected stimuli are detected and recognized more readily than unattended or unexpected. The exact mechanisms underlying this facilitation, however, remain subject to ongoing debate (Box 1). Given the similarity of their facilitative effects on object recognition, the question arises as to whether expectation and attention might share neural mechanisms, discussed next.

Neural sources of expectation

A natural way of understanding the role of expectation in visual cognition is as a ‘top-down’ process, in which prior information, perhaps stored in higher brain regions, feeds back to modulate basic perceptual processing in sensory cortices. This conceptualization has close contact with that of top-down attentional bias. However, while the major sources of top-down attentional bias have long been located principally (although not exclusively) in the dorsolateral prefrontal and posterior parietal cortices [19], only recently have attempts been made to identify the sources of expectation during object recognition. Intriguingly, the emerging evidence suggests that the neural origins of contextual biasing signals lie more medially in the posterior cortices as well as more ventrally in the frontal lobe than those of attentional bias. For example, parahippocampal and retrosplenial zones seem to encode information about visual scenes that might prompt object recognition [20] and these regions are activated by objects that have strong relative to weak contextual associations [21]. Pioneering work by Bar and colleagues has emphasized the role that the orbitofrontal cortex (OFC) might play in ‘top-down’ object recognition [22,23], arguing that during scene perception, a rapid extraction of low-level, low spatial frequency information might be fed to the OFC, where associative information allows the formation of a predictive ‘initial guess’ about what objects or elements are likely in the scene [24]. This information is then fed back to the extrastriate visual cortices to bias recognition towards contextually congruent objects. This account draws support from a variety of literatures, including the reports of very early, scene- and object-related potentials over the frontopolar cortices [25], models of the OFC as a heteromodal nexus for stimulus–stimulus associations [26], and the long-standing implication of frontopolar regions in retrieval of information from long-term memory [27]. Empirical support has also been provided by brain imaging studies using functional connectivity techniques, revealing a unidirectional flow of information from the medial OFC to the visual cortices during categorization of degraded face images [28] (Figure 1) and perception of low-frequency objects [29].

Neural effects of expectation in visual cortex

A large electrophysiological literature has documented that top-down attention acts by enhancing the responsivity of sensory neurons that code for a motivationally relevant feature, object or location [11], leading to an increase in the neural activity elicited by an attended relative to an unattended stimulus, although the precise nature of this modulation remains controversial (Box 1). Given that attention and expectation have similar facilitatory effect on visual object recognition, one might anticipate that expected (relative to unexpected) stimuli would also be associated with enhanced sensory responses. However - strikingly - the opposite is in fact typically the case: expected stimuli tend to elicit reduced visual responses, relative to their

![Figure 1. Frontal contributions to perceptual inference.](image)
unexpected counterparts [30,31], and an extensive literature has documented the corresponding phenomenon in the auditory domain in the form of the ‘mismatch negativity’ event-related potential [32]. In line with these findings, recent neuroimaging studies employing formal measures derived from information theory [33], or reinforcement learning [34], have documented increased visual cortical responses to visual surprise. Unfortunately, an unambiguous characterization of expectation effects at the level of single neurons is currently lacking in the literature (but see the section ‘Is there evidence for error and representation neurons?’ below, for a discussion of suggestive evidence). One possible interpretation of the event-related potential and neuroimaging results is that unexpected stimuli evoke stronger responses than expected ones because they attract attention. However, attentional effects alone are unlikely to account for these data. First, many experiments have demonstrated sensorimotor responses to expectancy violations in a task-irrelevant stimulus stream, while attention was occupied elsewhere [35–37]. Second, the differential response between expected and unexpected stimuli is not solely a function of enhanced activity evoked by the unexpected stimulus, but is substantially driven by accumulative decreases in the response to stimuli that conform to expectation [38,39]. If attention is not responsible for the enhanced neural responses to unexpected relative to expected stimuli, what could be the mediating mechanism?

**Predictive coding models of visual expectation**

Models of visual attention are concerned with explaining the enhancement of neural responses to attended relative to unattended stimuli (Box 1). As described in the previous section, the effects of expectation on visual responses, however, exhibit the exact opposite pattern. A class of models aimed at explaining the effects of prior knowledge on visual perception offers an intriguing explanation for these neuroscientific findings. The formal quantitative [40] and computational [41–43] framework known as ‘predictive coding’ proposes that information flowing forwards through a hierarchy of sensory regions is met at each stage by a top-down ‘prediction’ projected back from the layer immediately above (Figure 2). Neural implementations of this framework assume that there are two interacting classes of sensory neurons at each level of the visual processing stream. **Representational neurons** (R units) encode conditional (context-dependent) expectations about the identity of a stimulus and project their predictions to the next lower level of the processing hierarchy. These units are complemented by **error neurons** (E units) that encode ‘prediction error’ when the expected and observed sensory information diverge, and pass this error message to the R units, where predictions are updated accordingly (these proposed error units represent a sensory equivalent to the well-studied dopaminergic reward prediction error neurons of the midbrain [44]). Predictive coding argues that perceptual inference proceeds as expected and observed information are iteratively reconciled across multiple levels of the visual processing hierarchy, resulting in a progressive reduction in prediction error as the visual system settles on a single perceptual interpretation of the sensory input. Importantly, the proposal of co-existing representational and error neuron populations provides a natural account for the observation that cortical responses scale with sensory surprise: when stimulation conforms to expectations, error neuron activity will be minimal, but when sensory evidence diverges from predictions, error units will respond vigorously.

Indirect support for predictive coding includes the realization that the spatial range over which contextual effects in visual detection (such as facilitation by colinear flankers) have been observed precludes a mechanism that relies exclusively on lateral interactions among neurons in a given cortical stage (for example, via horizontal connections within V1). Rather, these effects - and ‘extra-classical’ receptive field (RF) effects, where the response properties of a neuron are modulated by information falling beyond its classical RF - can only be explained by feedback from...
How can expectation and attention be integrated?

Predictive coding provides a parsimonious account of the neural signature of visual expectations, but how can the effects of top-down attention be accommodated in this scheme? This question provides a major challenge to visual neuroscience, but only few theoretical proposals have been put forward thus far, and, to our knowledge, not a single empirical study has manipulated attention and expectation orthogonally and assessed the resulting effects on visual responses. One proposal has attempted to reconcile top-down mechanisms of expectation and attention by merging predictive coding [41] with standard accounts of attentional biasing under a common computational framework [43]. Under this scheme, visual neurons receive top-down input that facilitates their responding, but rather than competing for processing resources via lateral inhibition as previously proposed [50], these neurons (R units) compete for inputs from a distinct population of neurons [51] that in turn receive feedforward input from the immediately preceding cortical or thalamic level - akin to the error neurons (E units) in the predictive coding scheme (Figure 2). Because top-down information upregulates responding in R units, the model accounts for a range of data describing single neuron responses to attended and unattended features and locations [52]. Critically, however, the iterative interactions between R and E units preserve the main tenet of predictive coding, that predictions are ‘subtracted’ away from observed sensory input via feedback from representation to error neurons [40]. Theoretically, thus, visual ‘surprise’ responses could be accounted for as the responses of these E units to unexpected stimuli. However, it has yet to be reported whether the model can account for data from experiments that manipulated contextual expectations, or expectation and attention simultaneously. It is therefore currently not clear whether (and how) this scheme, which in its current form treats expectation and attention as interchangeable top-down influences on visual neurons, could reproduce the seemingly distinct effects of expectation and attention at the level of neuronal populations discussed in the previous section.

Is there evidence for error and representation neurons?

This account [43], and related computational models [42] could yet offer a fruitful path to understanding why visual...
regions seem to respond to both attended, and yet also to unexpected, visual stimuli. However, one limitation of these models - and of predictive coding in general - is that to date no single-neuron study has systematically pursued the search for sensory prediction error responses. Nevertheless, indirect evidence exists that both representational error responses might be observed in visual regions. First, some early single-neuron work identified extrastriate visual neurons that responded to orientations that were ‘matches’ or ‘mismatches’ to a cued target orientation [53–55]. These ‘match’ neurons additionally responded to the instruction cue designating the target orientation, even when it was presented in a nonvisual modality [54]. This behaviour seems strikingly like that of the proposed representation (match) and error (mismatch) neurons proposed by predictive coding (see also Figure 3), but contextual expectations were not formally manipulated in these studies. Second, single neurons in the visual cortex have been identified that display robust adaptation effects [56]. Adaptation might reflect neuronal fatigue or sharpening [57], but could also reflect the dampening of surprise responses in error neurons to the second or subsequent presentation of a stimulus [40,58]. Other, more recent studies have emphasized that pyramidal cells and interneurons in V1 might make distinct contributions to attentional biasing (showing enhanced responses to a preferred stimulus and suppressed responses to competitors, respectively [59,60]), but there is as yet no evidence that their responses map onto those of the proposed R and E units. Clearly, additional work is required to understand the local circuit interactions that give rise to visual ‘surprise’ responses alongside more classically described modulations by attention. Important steps towards realizing this goal will require a careful characterization of visual responses under conditions where expectations and attention are orthogonalized.

Moreover, despite its apparent success in explaining a range of neurophysiological phenomena, predictive coding can be criticized on a number of grounds. First, the sheer speed at which object recognition can occur might preclude a slow, iterative resolution of object identity via local circuit interactions in the visual cortex [22,25]. Second, researchers have noted that linear models of recognition based on subtractive, match–mismatch interactions might become catastrophically unstable [61], and a divisive variant of predictive coding, in which observed sensory evidence is normalized by expectations, might be a more plausible alternative [52].

Conclusions
We have reviewed recent advances in understanding a traditionally neglected aspect of top-down biases on visual processing, namely, the use of perceptual expectation to contextually bias the accumulation and interpretation of visual information. This literature suggests that fulfilled expectations are associated with facilitated object recognition but attenuated neural responses; predictive coding models provide an elegant theoretical account for these effects. A central goal of the current review was to emphasize similarities and distinctions between expectation and top-down attention, in particular because these processes might often be conflated in studies of attention. The above survey of their paradoxical behavioural and neural effects reinforces the importance of carefully distinguishing between expectation and attention in experimental design and interpretation: for instance, when attention and expectation toward a target stimulus coincide, both processes will contribute to behavioural benefits in stimulus recognition, but they will exert opposing effects on neural activity in visual regions representing the stimulus. However, the detailed nature of these interactions between expectation and attention is not currently understood, as modelling work on this topic in its infancy, and experimental data are scarce (for some particularly pressing questions, see Box 2). We hope that this review serves to stimulate additional interest in pursuing an improved understanding of this important and exciting issue.

Box 2. Outstanding questions

- How do behavioral and neural effects of expectation and attention relate to each other when these factors are manipulated independently in a single experiment?
- What is the precise mechanism (or mechanisms) by which expectations affect perceptual decision-making (Box 1)?
- Do dedicated ‘prediction error’ neurons exist in sensory cortices?
- Where might expectation-related codes reside in the brain?
- To what degree are neural and behavioral effects that have traditionally been attributed to attention (for example, in probabilistic cueing paradigms) confounded by perceptual expectations?

References

1 James, W. (1890) The Principles of Psychology. Henry Holt