Title:

Can expectation suppression be explained by reduced attention to predictable stimuli?

Authors:

Arjen Alink & Helen Blank, University Medical Center Hamburg-Eppendorf

Keywords:

predictive coding, expectation suppression, attention, neuroimaging

Abstract

The expectation-suppression effect – reduced stimulus-evoked responses to expected stimuli – is widely considered to be an empirical hallmark of predictive coding. Here we challenge this notion by proposing that this phenomenon can also be explained by a reduced attention effect. Specifically, we argue that predictable stimuli attract less attention because they are less behaviourally relevant. To resolve whether expectation suppression is best explained by attention or predictive coding, additional research is needed to determine whether attention effects precede the encoding of expectation violations (or vice versa) and to reveal how expectations change neural representations of stimulus features.

Introduction

Over the past decade, predictive coding theories have gained significant traction with high-profile findings across a range of cognitive domains (Friston, 2005; Rao & Ballard, 1999; Summerfield & de Lange, 2014). Central to this theory lies the idea that the computation of prediction errors forms a principal computational mechanism in the brain. One phenomenon that has been frequently reported and cited as empirical support for such theories is the expectation-suppression phenomenon – reduced stimulus-evoked responses to expected stimuli (Alink et al., 2010; Blank & Davis, 2016; Richter et al., 2018; Todorovic & de Lange, 2012; Walsh & McGovern, 2018). The exact neural mechanism via which predictive coding could give rise to expectation suppression, however, remains unclear. For example, a debate is still ongoing on whether it results from dampening (or explaining away) of neural representations of predictable sensory input for from more selective (or 'sharpened') neural representation of predictable stimuli (Friston, 2005; Richter et al., 2018; Walsh & McGovern, 2018).

Two alternative explanations for expectation-suppression

Until recently, the widespread consensus was that expectation suppression occurs irrespectively of whether stimuli are attended (den Ouden et al., 2009; Kok et al., 2012; Summerfield & Egner, 2009, 2016). One argument for this notion has been that enhanced perceptual performance associated with reduced responses to predictable stimuli is incompatible with an attentional mechanism because attention is thought to

boost perceptual performance via a positive neural gain mechanism (Alink et al., 2010; Feldman & Friston, 2010; Ferguson & Cardin, 2020; Ni et al., 2018; Treue & Trujillo, 1999). A recent study by Richter and de Lange (2019), however, fundamentally challenges this view (Summerfield & Egner, 2009, 2016) by showing that diverting attention away from a predictable image completely abolishes the expectation-suppression effect in the visual cortex. Specifically, they found that conditional probabilities of presented object images no longer had an effect on the amplitude of visual cortex responses if the participants' attention was drawn away from the object images by a central fixation task. Nonetheless, this finding has been interpreted in light of a ***predictive-coding based explanation*** (PBE) by proposing that attending an object boosts the generation of predictions for that object, which increases the responsiveness of neurons sensitive to the discrepancy between sensation and expectation (so called 'prediction error units').

In this perspective, however, we put forward an alternative ***attention-based explanation*** (ABE), which explains expectation suppression as being related to a reduced amount of behavioural relevance of for expected stimuli. This explanation is based on a large body of studies suggesting that behavioural relevance is a fundamental determinant of attention-related neural response increases in the visual cortex (Treue, 2003). This neural mechanism can also explain the expectation-suppression effect because surprising events are in general more behaviourally relevant – which should translate into greater attentional gain for their neural representations. This explanation is consistent with the notion that 'Bayesian surprise' attracts human attention (Itti & Baldi, 2009) – which also applies to stimuli that do not directly relate to an explicit task. From this perspective, enhanced responses to unpredictable stimuli in early sensory brain areas reflect a greater allocation of attentional resources towards unpredictable stimuli (Figure 1).

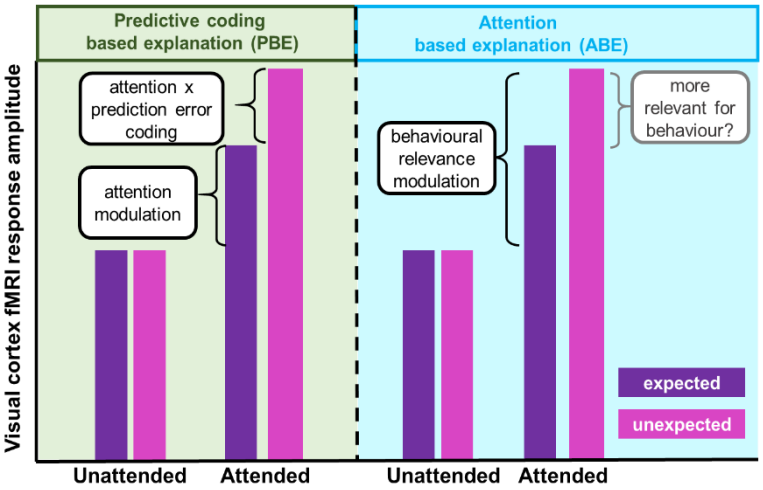

Figure 1 Predictive-coding based explanation (PBE, left) attention increases the responsiveness of neurons sensitive to the discrepancy of the sensory input and the expected stimulus.

Attention-based explanation (ABE, right) explains expectation suppression as resulting from a reduced behavioural relevance for expected stimuli.

For illustrative purposes, let us evaluate whether the ABE can accommodate the finding by Richter and de Lange (2019) that expectation-suppression only occurs in the visual cortex when images are attended. As we outlined above, unexpected stimuli can, in general, be thought of as being of greater behavioural relevance – irrespective of explicit tasks. In addition, unexpected stimuli used during this experiment were of greater behavioural relevance with respect to the explicit task. Specifically, the experimental design used allowed participants to correctly guess the upcoming required response with an accuracy of 70% before seeing predictable test images. Consequently, expected images were less likely to require participants to override their response-policy, which reduced their behavioural relevance relative to unpredictable images. Therefore, predictable images should have received less attention which could explain the reduced associated responses in the visual cortex. The reported absence of this expectation-suppression effect for the central fixation task also fits with the ABE as this condition exhausts attentional resources and thereby prevents attention enhancing responses to unpredictable images

Are attention and expectation two independent cognitive processes?

A prominent and influential view in the field is that that attention and expectation affect brain responses via two independent processes - the manipulation of the behavioural relevance of stimulus features (feature-based attention, FBA) and the manipulation of the likelihood of stimulus features (feature-based expectations, FBE) respectively (Summerfield & Egner, 2009, 016). This stance requires experimental designs that completely orthogonalize stimulus predictability and the task-relevance of stimuli. However, as pointed out by Summerfield and Egner (2009, 2016) and illustrated by the

Richer and de Lange study (2019), many studies on predictive coding still contain subtle dependencies between the predictability and the task-relevance of stimuli. A key point of this perspective, however, is that even a fully orthogonal experimental design would not allow one to independently measure effects of FBA and FBA because unexpected stimuli are ‘per definition’ more informative and behaviourally relevant (Itti & Baldi, 2009). Therefore, unexpected stimuli should always attract more attention than expected stimuli, even if they are equally relevant for the ongoing explicit task. For this reason, we suggest that the common assumption of independence of attention and expectation effects in the brain needs to be critically re-evaluated.

Why it is important to differentiate between a predictive coding and an attention-based explanation of repetition suppression?

Over the last decades, theories of predictive coding became more and more popular and a constantly increasing number of findings has been interpreted as being in line with predictive coding (e.g., yearly published articles on this topic has increased from 371 to 2.870 from 2009 to 2019 – google scholar data returned for the search term: "predictive coding" AND brain). Expectation suppression is arguably one of the most frequently used observations in support of predictive coding. We agree that (the lack of) prediction error computation provides a compelling explanation for reduced responses to expected stimuli. However, it is important to not simply accept predictive coding as an ‘a-priori truth’ and instead to constantly consider alternative explanations. Otherwise this would effectively turn this theoretical framework into a ‘just-so story’ able to accommodate any empirical finding (Bowers & Davis, 2012).

In addition, differentiating between the two alternative explanations of expectation suppression is important because they make different predictions about what the increased neural activity for unexpected stimuli signifies: The PBE suggests that this increased activity encodes the discrepancy between the actual and expected input to a brain area, while the ABE suggests that this enhanced activity is driven by an attentional gain for stimuli with greater behavioural relevance. Additionally, the two explanations differ with regard to the temporal order of expectation-violation detection and attention effects in the visual cortex: ABE proposes that expectation-violation detection precedes attentional enhancement of neural responses to surprising stimuli

while PBE suggests that attending a stimulus is prerequisite for stimulus-specific expectation-violation detection (Richter & De Lange, 2019).

Reduced prediction-error coding or a reduced attention effect: How to tell the difference and what is the more parsimonious explanation?

To differentiate between the predictive coding and an attention-based explanation of repetition suppression, we have to specify the different predictions that these explanations make and test them empirically. As outlined above, these two explanations differ with regard to *when* expectation-violation is processed relative to when attention is allocated towards the relevant stimulus. Specifically, according the ABE, expectation violation has to be detected *first* before additional attention resources are directed towards unexpected stimuli. According to the PBE, however, attention towards a relevant stimulus is needed *before* expectation violations can be detected as attention is considered to be a prerequisite for the generation of stimulus-specific expectations (Richter & De Lange, 2019). As a consequence, ABE would predict response enhancements for unexpected stimuli to occur later in time than a typical task-driven attentional gain effect (Serences & Kastner, 2014). The PBE, on the other hand, would be compatible with response enhancements for unexpected stimuli occurring at the same time, or even earlier, than a task-driven attentional gain effect. For this reason, it is important, as also recently discussed by Press and colleagues (Press et al., 2020), to use neuroimaging methods with high temporal resolution, like EEG and MEG, to disentangle the neural mechanisms underlying expectation suppression.

How else could we determine whether expectation suppression results from reduced prediction errors or from reduced attentional resources being allocated to expected stimuli? As pointed out above (e.g., Figure 1), univariate responses cannot differentiate between these two possibilities, because PBE and ABE both predict reduced responses to expected stimuli (Blank & Davis, 2016). Also, the observation that stimulus predictability improves perceptual decisions (Kok et al., 2012; Schwiedrzik et al., 2007) does not enable us to rule out PBE or ABE because they both are consistent with perceptual decisions being made based on sensory input *and* prior expectations - which in general leads to improved perceptual decisions for predictable stimuli. Therefore, disentangling of PBE from ABE requires further investigation of *how*

features of expected and unexpected stimuli are encoded differently in the human brain – e.g., by analysing the information content of neural response patterns.

Several studies have done so by investigating how stimulus predictability affects the extent to which fMRI response patterns encode stimulus information (Blank & Davis, 2016; Kok et al., 2012; Kumar et al., 2017). One of these studies (Kok et al., 2012) found that expectation suppression was accompanied by enhanced visual orientation decodability and suggested that the conjunction of these two observations implies sharper stimulus encoding for expected stimuli. Such expectation-driven sharpening would speak against ABE because attention is thought to enhance encoding of stimulus information in the visual cortex (Jehee et al., 2011). Therefore ABE, would predict enhanced stimulus information encoding to occur for unexpected stimuli. At present, however, the existence of sharpened responses for expected stimuli is controversial. First of all, because several studies have found enhanced stimulus information encoding for *unexpected* stimuli (Jiang et al., 2013; Kumar et al., 2017; Richter et al., 2018; Tang et al., 2018) – which would be consistent with ABE. Second, it is not clear if fMRI response reductions associated with enhanced fMRI pattern decodability necessarily implies neural sharpening. For example, it has been recently shown that the conjunction of these two effects can be better explained by a local-scaling effect (Alink et al., 2018) or dampening (Richter et al., 2018). Therefore, ruling out the attention-based account of expectation suppression by demonstrating expectation-induced sharpening would require additional empirical evidence.

ABE and PBE can both explain expectation suppression in V1 in the context of attention. But which account offers the more parsimonious explanation? At first glance, PBE appears to be more parsimonious because prediction-error signalling in V1 can directly explain expectation suppression whereas ABE requires a two-stage process during which attention is allocated towards unexpected stimuli after detection of an expectation violation. However, when taking a closer look at the PBE – as for example outlined by Richter and de Lange (2019) – it becomes clear that it entails two additional assumptions adding to its complexity: PBE requires 1) a population of ‘prediction error units’ in the primary visual cortex and 2) two types of top-down signals – one for attentional modulation *and* one for conveying predictive signals to V1 – while ABE only requires top-down attention signal. Therefore, we argue that the PBE is not necessarily

more parsimonious than the ABE, which explains expectation suppression without the computation of prediction errors.

Concluding remarks and future perspectives

To conclude, with this perspective, we hope to reinvigorate the discussion about how effects of expectation can be differentiated from attention effects - by highlighting that previously reported expectation-suppression effects could also be explained by reduced behavioural-relevance driven attention towards expected stimuli. To distinguish between these two alternative accounts of expectation suppression, we suggest that future studies should assess whether attention effects precede the encoding of expectation violations (or the other way around) and reveal how expectations change the way in which stimulus information is encoded in sensory brain areas. Finally, given the entangled nature of predictive coding and attention, it might be necessary to also explore whether a single overarching (e.g., Bayesian) process could explain both the effects of attention and expectation – without requiring attentional facilitation of prediction error coding nor expectation-violation guidance of attention.

References

Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., & Muckli, L. (2010). Stimulus predictability reduces responses in primary visual cortex. *The Journal of Neuroscience*, *30*(8), 2960–2966. https://doi.org/10.1523/JNEUROSCI.3730-10.2010

Alink, Arjen, Abdulrahman, H., & Henson, R. N. (2018). Forward models demonstrate that repetition suppression is best modelled by local neural scaling. *Nature Communications*, *9*(1), 1–10.

Blank, H., & Davis, M. H. (2016). Prediction Errors but Not Sharpened Signals Simulate Multivoxel fMRI Patterns during Speech Perception. *PLOS Biology*, *14*(11), e1002577. https://doi.org/10.1371/journal.pbio.1002577

Bowers, J. S., & Davis, C. J. (2012). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, *138*(3), 389–414. https://doi.org/10.1037/a0026450

den Ouden, H. E. M., Friston, K. J., Daw, N. D., McIntosh, A. R., & Stephan, K. E. (2009). A Dual Role for Prediction Error in Associative Learning. *Cerebral Cortex*, *19*(5), 1175–1185. https://doi.org/10.1093/cercor/bhn161

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*, 215. https://doi.org/10.3389/fnhum.2010.00215

Ferguson, K. A., & Cardin, J. A. (2020). Mechanisms underlying gain modulation in the cortex. *Nature Reviews Neuroscience*, *21*(2), 80–92. https://doi.org/10.1038/s41583-019-0253-y

Friston, K. (2005). A theory of cortical responses. *Philos Trans R Soc London [Biol]*, *360*(1456), 815–836.

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Visual Attention: Psychophysics, Electrophysiology and Neuroimaging*, *49*(10), 1295–1306. https://doi.org/10.1016/j.visres.2008.09.007

Jehee, J. F. M., Brady, D. K., & Tong, F. (2011). Attention improves encoding of task-relevant features in the human visual cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(22), 8210–8219. PubMed. https://doi.org/10.1523/JNEUROSCI.6153-09.2011

Jiang, J. F., Summerfield, C., & Egner, T. (2013). Attention Sharpens the Distinction between Expected and Unexpected Percepts in the Visual Brain. *Journal of Neuroscience*, *33*(47), 18438–18447. https://doi.org/10.1523/Jneurosci.3308-13-2013

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron*, *75*(2), 265–270. https://doi.org/10.1016/j.neuron.2012.04.034

Kumar, S., Kaposvari, P., & Vogels, R. (2017). Encoding of Predictable and Unpredictable Stimuli by Inferior Temporal Cortical Neurons. *Journal of Cognitive Neuroscience*, *29*(8), 1445–1454. https://doi.org/10.1162/jocn_a_01135

Ni, A. M., Ruff, D. A., Alberts, J. J., Symmonds, J., & Cohen, M. R. (2018). Learning and attention reveal a general relationship between population activity and behavior. *Science*, *359*(6374), 463. https://doi.org/10.1126/science.aao0284

Press, C., Kok, P., & Yon, D. (2020). The Perceptual Prediction Paradox. *Trends in Cognitive Sciences*, *24*(1), 13–24. https://doi.org/10.1016/j.tics.2019.11.003

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci*, *2*(1), 79–87. https://doi.org/10.1038/4580

Richter, D., & De Lange, F. P. (2019). Statistical learning attenuates visual activity only for attended stimuli. *Elife*, *8*.

Richter, D., Ekman, M., & de Lange, F. P. (2018). Suppressed sensory response to predictable object stimuli throughout the ventral visual stream. *The Journal of Neuroscience*. https://doi.org/10.1523/JNEUROSCI.3421-17.2018

Schwiedrzik, C. M., Alink, A., Kohler, A., Singer, W., & Muckli, L. (2007). A spatio-temporal interaction on the apparent motion trace. *Vision Research*, *47*(28), 3424–3433.

Serences, J. T., & Kastner, S. (2014). A multi-level account of selective attention. *The Oxford Handbook of Attention*, 76–104.

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nat Rev Neurosci*, *15*, 745–756. https://doi.org/10.1038/nrn3838

Summerfield, Christopher, & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13*(9), 403–409.

Summerfield, Christopher, & Egner, T. (2016). Feature-based attention and feature-based expectation. *Trends in Cognitive Sciences*, *20*(6), 401–404.

Tang, M. F., Smout, C. A., Arabzadeh, E., & Mattingley, J. B. (2018). Prediction error and repetition suppression have distinct effects on neural representations of visual information. *ELife*, *7*, e33123. https://doi.org/10.7554/eLife.33123

Todorovic, A., & de Lange, F. P. (2012). Repetition Suppression and Expectation Suppression Are Dissociable in Time in Early Auditory Evoked Fields. *Journal of Neuroscience*, *32*(39), 13389–13395. https://doi.org/10.1523/jneurosci.2227-12.2012

Treue, S. (2003). Visual attention: The where, what, how and why of saliency. *Current Opinion in Neurobiology*, *13*(4), 428–432.

Treue, S., & Trujillo, J. C. M. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*(6736), 575–579. https://doi.org/10.1038/21176

Walsh, K. S., & McGovern, D. P. (2018). Expectation Suppression Dampens Sensory Representations of Predicted Stimuli. *The Journal of Neuroscience*, *38*(50), 10592. https://doi.org/10.1523/JNEUROSCI.2133-18.2018