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, saliency

Abstract

The expectation-suppression effect – reduced stimulus-evoked responses to expected stimuli – is widely considered to be an empirical hallmark of reduced prediction errors in the framework 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 reduced responses to predictable stimuli can also be explained by a reduced saliency-driven allocation of attention. 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; Walsh et al., 2020). 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 or from more selective (or 'sharpened') neural representation of predictable stimuli (Friston, 2005; Kok et al., 2012; Richter et al., 2018; Walsh & McGovern, 2018). In this perspective, we want to discuss the alternative interpretation that expectation suppression results from reduced attention to predictable stimuli (Aitchison & Lengyel, 2017). This would be in line with the view that surprising stimuli are more salient and more likely to attract attention (Desimone & Duncan, 1995; Itti & Baldi, 2009).

Two alternative explanations for expectation suppression

A prominent view suggests 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 was that better detection of predictable stimuli paired with reduced stimulus evoked responses is difficult to reconcile with the positive neural gain mechanism associated with attention (Alink et al., 2010; Treue & Trujillo, 1999).

This independency between attention and expectation has been challenged by a recent study by Richter and de Lange (2019). They showed that diverting attention away from a predictable image completely abolishes the expectation-suppression effect in the visual cortex. Specifically, 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 is interpreted in light of a ***predictive-coding based explanation*** (PBE) by proposing that attending an object boosts prediction error coding. This interpretation is in line with the predictive coding framework, in which attention increases the gain of prediction error units (Feldman & Friston, 2010; Friston, 2009; Spratling, 2008). In addition, Richter and de Lange suggest that the generation of predictions depends on attention. Note that, in contrast to this attention-driven boost of error coding, attention has also been reported to suppress prediction error coding (Sohoglu & Chait, 2016).

In this perspective, we want to discuss an alternative ***attention-based explanation*** (ABE) of expectation suppression which attributes reduced responses to predictable stimuli to a reduced allocation of attentional resources. This alternative is motivated by the notion that attention results from a competition between stimulus representations for the limited information processing capacity of the brain (Desimone & Duncan, 1995). This competition is biased towards behaviourally relevant stimuli, e.g., via enhanced responses to task-relevant or salient stimuli. According to this view, saliency is a stimulus characteristic that automatically results from deviations from both its physical context and from 'higher-level' context based on natural image statistics and experience (Desimone & Duncan, 1995, Kanan et al., 2009). Importantly, the saliency of unpredictable stimuli depends on context: a stimulus is more salient when it is the deviant stimulus in an oddball paradigm as compared to when it is part of a

random stimulus sequence because only the former would result into a contextual violation. Accordingly, ABE proposes that enhanced responses to unexpected stimuli in sensory brain areas reflect a greater saliency-driven allocation of attentional resources towards these stimuli (Figure 1). Note that, in contrast to the PBE, the ABE does not strictly require prediction-error coding as the expectation suppression effect is fully explained by reduced stimulus saliency (but see Spratling, 2008). Finally, we would like to point out that the ABE does not preclude the possibility that attention can be drawn towards predictable events under certain circumstances (e.g., when predictable stimuli are task-relevant, Alamia & Zénon, 2016).

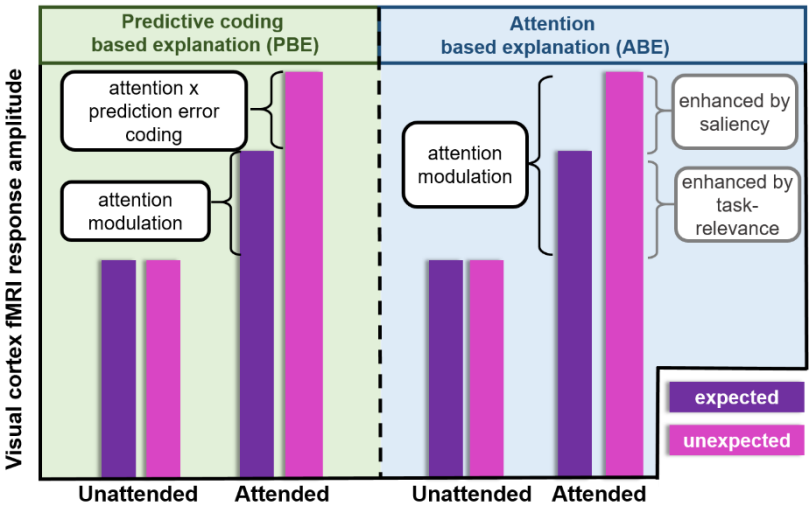

Figure 1 Predictive-coding based explanation (PBE, left) attention increases the responsiveness of neurons sensitive to the discrepancy of the sensory input and the expected stimulus. **Attention-based explanation** (ABE, right) explains expectation suppression as resulting from a reduced stimulus saliency.

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 outlined above, unexpected stimuli can be thought of as more capable of capturing attention due to their saliency. 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 task is argued to completely prevent attention allocation towards expected and unexpected images (Richter & de Lange, 2019).

Can attention and expectation effects be orthogonalized?

An influential view is that it is possible to orthogonalize feature-based attention and expectation effects in experimental designs (Summerfield & Egner, 2016). However, as pointed out by Summerfield and Egner (2009, 2016), many studies on predictive

coding still contain subtle dependencies between them (see also Rungratsameetaweemana & Serences, 2019). For example, the experimental design used by Richter and de Lange (2019) allowed participants to correctly guess the upcoming required response with an accuracy of 70% before seeing predictable test images. This might have reduced the task difficulty for predictable images and made them less task relevant.

Our key hypothesis is that even perfectly orthogonalizing task relevance and stimulus predictability could not preclude the proposed reduced attention effect for predicable stimuli (as outline with the ABE). The reason for this is that the allocation of attention towards a stimulus depends on both task relevance *and* stimulus saliency (as visualized in Figure 1). We argue that it is impossible to experimentally orthogonalize expectation and attention effects because unexpected stimuli are 'per definition' more salient.

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

Over the last decades, theories of predictive coding became more and more popular and expectation suppression is an observation that has frequently been interpreted as supporting such theories. Although we agree that reduced prediction error coding provides an appealing explanation for reduced responses to expected stimuli, we here want to stress that predictive coding should not be accepted as an 'a-priori truth'. Instead, scientist should remain open to considering alternative explanations in order to prevent this theoretical framework from turning into a 'just-so story' that is able to accommodate any empirical finding (Bowers & Davis, 2012; Heilbron & Chait, 2017).

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 results from increased attention allocation towards salient stimuli. As pointed out by Spratling (2008), it is possible that stimulus saliency is computed in a similar manner as prediction errors. However, even if stimulus saliency can be equated to a prediction error coding process, we would like to point out that the ABE and the PBE differ with

regard to the order of attention effects and the encoding of prediction errors/stimulus saliency: The PBE proposes that attention facilitates prediction error coding (Feldman & Friston, 2010; Friston, 2009; Richter & de Lange, 2019; Spratling, 2008) while the ABE suggests that attention effects result from stimulus saliency (Desimone & Duncan, 1995, Kanan et al., 2009). Therefore, the critical difference appears to be that the PBE suggests that attention is *a cause* of the enhanced prediction-error coding while the ABE implies that attention is *an effect* of stimulus saliency.

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 expectation suppression, we have to specify the different hypotheses that these explanations make and test for them empirically. As outlined above, the PBE and the ABE differ with regard to the relative timing of attention allocation and the timing of prediction error coding/stimulus saliency encoding. Therefore, future experiments could discriminate between the ABE and the PBE by using temporally resolved measurements (e.g., using EEG or MEG). For example, by testing the assumption of the PBE (Richter et al, 2019) that attention allocation precedes prediction error coding in the visual cortex (e.g., using an approach like Press et al., 2020).

How else could we determine whether expectation suppression results from predictive coding or from reduced attention 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 (Alink et al., 2010; 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. Since perceptual decision making is typically linked to brain activity in higher-level non-sensory brain areas (e.g., the frontoparietal network, Mulder et al., 2014), enhanced perceptual decision making for predictable stimuli does not necessarily require a facilitation of stimulus representations in sensory brain areas (Rungratsameetaweemana & Serences, 2019).

Could we disentangle the PBE from the ABE by investigating *how* features of expected and unexpected stimuli are encoded in the human brain – e.g., by analysing how expectation suppression affects fMRI stimulus decodability? Attention is known to enhance the responses of neurons tuned to the attended stimulus, leading to greater fMRI decodability (Kamitani & Tong, 2005). Therefore, according to the ABE, unexpected stimuli should be more decodable. In contrast, it is unclear whether according to the PBE unexpected or expected stimuli should be more decodable. If reduced responses to expected stimuli result from 'explaining away' of predicted sensory input (Clark, 2013) unexpected stimuli should be more decodable. However, stimulus predictability has been associated with enhanced fMRI pattern decodability which has led to the proposal that expectations give rise to sharpened neural representations (Kok et al., 2012). Such expectation-induced sharpening would contradict the ABE. At present, however, expectation-driven sharpening remains controversial because enhanced decodability has also been observed for unexpected stimuli (Jiang et al., 2013; Kumar et al., 2017; Richter et al., 2018; Tang et al., 2018) and given the observation that a conjunction of reduced fMRI responses and enhanced decodability can also be explained by a local-scaling effect (Alink et al., 2018). In conclusion, more evidence for sharpened responses to expected stimuli is needed to rule out the ABE.

Since the ABE and the PBE can both explain expectation suppression, one would ideally want to apply Ockham's razor to determine which provides the more parsimonious explanation (Heilbron & Chait, 2017). Such an approach, however, is challenging because attention and predictive coding models are typically not formulated at the same level of description (Marr, 1982). Therefore, resolving whether the ABE or the PBE is more parsimonious requires a reformulation of these explanations at a comparable level of description (e.g., see Spratling, 2008).

Can attention also explain other expectation effects?

Here, we discuss if predictive coding provides an alternative explanation of expectation suppression in comparison to traditional models of attention. Similar concerns may apply for other expectation-induced modulations of sensory responses, such as omission-evoked responses resembling responses to the expected stimulus (Bendixen et al., 2009) or feature-specific pre-stimulus activity induced by expectations (Fiser et

al., 2016; Kok et al., 2017). Although these observations are frequently proposed to result from predictive coding, increased neural activity in the absence of sensory stimulation has also been attributed to directed attention (Kastner et al., 1999).

Predictive coding, more than "New Labels for Old Ideas"?

A current debate around theories of predictive coding is whether they provide new constraints on brain function that have not already been provided by traditional models (Cao, 2020). In this perspective, we highlight that this applies to the expectation-suppression effect by pointing out that a traditional biased-competition model of attention can also explain reduced responses to expected stimuli. The current challenge for predictive coding is to conclusively show that it provides a more accurate description of how expectations suppress neural responses in sensory brain areas than traditional models.

Acknowledgements

We thank Rik Henson, David Richter, and Floris de Lange for thoughtful comments on earlier versions of this commentary.

References

Aitchison, L., & Lengyel, M. (2017). With or without you: Predictive coding and Bayesian inference in the brain. *SI: 46 : Computational Neuroscience (2017)*, *46*(Supplement C), 219–227. https://doi.org/10.1016/j.conb.2017.08.010

Alamia, A., & Zénon, A. (2016). Statistical Regularities Attract Attention when Task-Relevant. *Frontiers in Human Neuroscience*, *10*. https://doi.org/10.3389/fnhum.2016.00042

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, A., 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.

Bendixen, A., Schröger, E., & Winkler, I. (2009). I Heard That Coming: Event-Related Potential Evidence for Stimulus-Driven Prediction in the Auditory System. *The Journal of Neuroscience*, *29*(26), 8447. https://doi.org/10.1523/JNEUROSCI.1493-09.2009

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

Cao, R. (2020). New Labels for Old Ideas: Predictive Processing and the Interpretation of Neural Signals. *Review of Philosophy and Psychology*. https://doi.org/10.1007/s13164-020-00481-x

Clark, A. (2013). Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

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

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annu Rev Neurosci*, 30.

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

Fiser, A., Mahringer, D., Oyibo, H. K., Petersen, A. V., Leinweber, M., & Keller, G. B. (2016). Experience-dependent spatial expectations in mouse visual cortex. *Nature Neuroscience*, *19*(12), 1658–1664. https://doi.org/10.1038/nn.4385

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

Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301.

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

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

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

Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, *8*(5), 679–685. https://doi.org/10.1038/nn1444

Kanan, C., Tong, M. H., Zhang, L., & Cottrell, G. W. (2009). SUN: Top-down saliency using natural statistics. *Visual Cognition*, *17*(6–7), 979–1003. https://doi.org/10.1080/13506280902771138

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the

absence of visual stimulation. *Neuron*, *22*(4), 751–761. https://doi.org/10.1016/S0896-6273(00)80734-5

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

Kok, P., Mostert, P., & de Lange, F. P. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.1705652114

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

Mulder, M. J., van Maanen, L., & Forstmann, B. U. (2014). Perceptual decision neurosciences—A model-based review. *Neuroscience*, *277*, 872–884. https://doi.org/10.1016/j.neuroscience.2014.07.031

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*, e47869. https://doi.org/10.7554/eLife.47869

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

Rungratsameetaweemana, N., & Serences, J. T. (2019). Dissociating the impact of attention and expectation on early sensory processing. *Current Opinion in Psychology*, *29*, 181–186. https://doi.org/10.1016/j.copsyc.2019.03.014

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.

Sohoglu, E., & Chait, M. (2016). Detecting and representing predictable structure during auditory scene analysis. *ELife*, *5*, e19113. https://doi.org/10.7554/eLife.19113

Spratling, M. W. (2008). Reconciling predictive coding and biased competition models of cortical function. *Front Comput Neurosci*, *2*, 4. https://doi.org/10.3389/neuro.10.004.2008

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., & 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

Walsh, K. S., McGovern, D. P., Clark, A., & O'Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences*, *1464*(1), 242–268. https://doi.org/10.1111/nyas.14321