

Journal Club

Editor's Note: These short, critical reviews of recent papers in the *Journal*, written exclusively by graduate students or postdoctoral fellows, are intended to summarize the important findings of the paper and provide additional insight and commentary. For more information on the format and purpose of the Journal Club, please see http://www.jneurosci.org/misc/ifa_features.shtml.

Predictive Coding and the Neural Response to Predictable Stimuli

Lee de-Wit, Bart Machilsen, and Tom Putzeys

Laboratory of Experimental Psychology, University of Leuven (K.U. Leuven), B-3000 Leuven, Belgium
Review of Alink et al. and den Ouden et al.

Introduction

Many studies have characterized the brain as an efficient coding system (Olshausen and Field, 1996). A common claim in this “efficient coding” approach is that the brain has evolved to code sensory information in an efficient way by using information-processing strategies optimized to the statistics of the perceptual environment. An influential theoretical and computational framework, closely related to efficient coding, is predictive coding. This framework provides a novel means of understanding the interactions between distinct brain areas. Predictive coding posits that the brain actively predicts upcoming sensory input rather than passively registering it. Predictive coding is efficient in the sense that the brain does not need to maintain multiple versions of the same information at different levels of the processing hierarchy. Indeed, activation in early sensory areas no longer represents sensory information per se, but only that part of the input that has not been successfully predicted by higher level areas. The activity in lower level areas can therefore be considered an “error signal”

that updates the predictions at higher areas and guides learning.

One articulate example of predictive coding is the model of Rao and Ballard (1999). In this model, predictions generated at higher levels are used to “explain away” lower level representations that are compatible with the higher level interpretation. The explaining away component of predictive coding has a remarkable consequence: predictability reduces activity in early areas through feedback from higher level areas. This might seem counter intuitive. Indeed, due to the ambiguity of sensory input and the limited reliability of neural responses, one might expect feedback to enhance early information that is in agreement with higher level representations. However, two recent studies published in *The Journal of Neuroscience* seem to provide fMRI evidence consistent with this notion of explaining away.

Focusing on the primary visual cortex, Alink et al. (2010) observed a lower blood oxygenation level-dependent (BOLD) signal for predictable compared with unpredictable stimuli. These authors exploited an apparent motion illusion, in which the alternation between a stationary bar at the top and at the bottom of the screen leads to the illusory perception of a moving bar between the two physically presented stimuli. Alink et al. (2010) measured the neural response to an additional bar at one location along the path of the illusory moving bar. The timing of this additional bar was manipulated such that

it was presented at a time point that was either consistent with the motion illusion (and thus predictable) or at an inconsistent (unpredictable) time point. The lower BOLD response for the predictable stimulus (in contrast to the unpredictable stimulus) seems to be in line with the explaining away notion of predictive coding: predictable representations at lower levels are actively suppressed by feedback from higher levels.

The lower BOLD response for the predictable versus unpredictable bar was observed in a region of V1 that retinotopically corresponded to the location at which this bar appeared. Because this region did not directly respond to the “inducing” bars, it seems logical to conclude that predictability effects at this location must reflect feedback from higher levels with larger receptive fields. Alink et al. (2010) acknowledge, however, that BOLD signal changes cannot be taken as direct evidence for feedback effects, and they highlight that it is not possible to rule out a role of lateral interactions within V1.

The increased neural response to the unpredictable stimulus could be thought to reflect an increased allocation of attention to a surprising stimulus. In a separate psychophysical experiment however, Alink et al. (2010) found that detectability rates were in fact higher for the predictable stimulus. This behavioral result rules out any straightforward explanation of the changes in fMRI response in terms of an increased allocation of attention to the unpredictable stimulus.

Received May 3, 2010; revised May 28, 2010; accepted June 2, 2010.

This work was supported by a Methusalem grant (METH/08/02) from the Flemish Government. T.P. is research assistant of the Fund for Scientific Research-Flanders (FWO-Vlaanderen). We thank Jonas Kubilius, Hans Op de Beeck, and Johan Wagemans for valuable feedback on this manuscript.

Correspondence should be addressed to Lee de-Wit, Laboratory of Experimental Psychology, Tiensestraat 102 bus 3711, B-3000 Leuven, Belgium. E-mail: lee.dewit@psy.kuleuven.be.

DOI:10.1523/JNEUROSCI.2248-10.2010

Copyright © 2010 the authors 0270-6474/10/308702-02\$15.00/0

It is worth considering how Alink et al. (2010) operationalized predictability, because in their experiment the unpredictable stimulus was, in fact, always presented at the same location and at a constant delay. Thus, although the timing of the unpredictable stimulus was inconsistent with the illusory motion (and therefore unpredictable with respect to the brain's long term assumptions about moving stimuli), it could still be considered predictable in the context of the experiment. If predictive coding is to provide a flexible means of learning about the statistics of one's environment, one might expect the brain to be sensitive to such short-term (within experiment) contingencies.

Evidence for a sensitivity to short-term contingencies is in fact provided by another paper recently published in *The Journal of Neuroscience*. den Ouden et al. (2010) found that when one of two tones consistently preceded the visual onset of either a face or a house, activation in brain areas recruited when processing these stimulus types [fusiform face area (FFA) and parahippocampal place area (PPA), respectively] was reduced. Importantly, this reduction occurred even though the association between the tone and stimulus type was artificial and the probability with which a given tone would be followed by a given stimulus changed across the experiment. Furthermore, participants were not explicitly informed when these probability rates would change; they were simply informed that the probabilistic relations between the auditory cues and the target would change over the experiment. Nevertheless, den Ouden et al. (2010) found a negative correlation between the neural response to a particular stimulus type and the predictability of that stimulus type. Thus, den Ouden et al.'s (2010) results seem to corroborate Alink et al.'s (2010) finding of reduced activity levels in sensory areas in the context of predictable stimulus contingencies, but demonstrate this effect for an arbitrary contingency developed over a short time scale.

In addition to this sensory sensitivity to predictability, for faces in FFA and houses in PPA, den Ouden et al. (2010) reported higher levels of activation in the putamen whenever a tone was not followed by the stimulus type it predicted. The putamen therefore seemed to signal a generic (nonstimulus specific) "prediction error," which was elicited for both face and house stimuli. den Ouden et al. (2010) used dynamic causal modeling to

suggest that activation in this area gated the connection between perceptual (FFA/PPA) and motor areas (responsible for generating the participants' response). It appears, therefore, as if the error detection in the putamen ensures that, whenever predictions are not met, the standard perceptual resources used in decoding a particular stimulus have a stronger influence on response selection. This mechanism could play a role in changing connectivity weights that might mediate long-term learning.

Viewed in parallel, Alink et al. (2010) and den Ouden et al. (2010) provide evidence that across different time scales, experimental contexts, and levels of the visual hierarchy, the brain is not only highly sensitive to predictability, but that this predictability can lead to a reduction of activation in sensory areas. This finding is consistent with Rao and Ballard's (1999) hypothesis that predictions in higher level areas are fed back and compared with incoming sensory signals, such that lower areas come to represent the discrepancy, or error signal, and not the stimulus per se. However, it is important to remember that the fMRI BOLD signal is an indirect measure of the underlying neural representation, where each voxel can reflect an average of >100,000 neurons. Indeed, reflecting upon similar reduced-activation levels in early visual areas in the context of perceptual grouping, Murray et al. (2004) highlight that, although a reduction in activation is consistent with an overall reduction in the strength of that representation, it could also be the case that the reduction reflects a sharpened tuning for predictable stimuli that lead to them being represented more efficiently (i.e., a lower average activation with a higher signal-to-noise ratio).

Thus, although the BOLD signal reduction found in sensory areas by Alink et al. (2010) and den Ouden et al. (2010) could well be described in terms of a higher level of representation influencing the activations at a lower level, it is not possible to conclude whether predictable stimuli are explained away at low levels and represented at higher levels, as predicted by the predictive coding framework of Rao and Ballard, or whether the representation of predictable stimuli is adjusted at the lower level itself. The latter possibility is consistent with a recent study (Nienborg and Cumming, 2009) suggesting that efficient decoding mechanisms may be implemented at the level of sensory neurons through top-down changes

in neuronal gain. Distinguishing between these "fine-tuning" and explaining-away explanations at different levels of the visual hierarchy will be an important goal of future research.

In summary, Alink et al. (2010) and den Ouden et al. (2010) highlight different ways in which the predictability of sensory input alters the processing of those stimuli by the brain. Alink et al. (2010) show that even at the first stage of cortical visual information processing, predictable and unpredictable stimuli are processed differently. den Ouden et al. (2010) not only highlight how this effect of predictability can manifest at higher stages of visual information processing but also demonstrate that it can develop for arbitrary contingencies over short time scales. Furthermore, den Ouden et al. (2010) provide evidence for a more generic error-signaling system, contingent on the putamen, that appears to act to functionally increase the connection between sensory and motor responses when our predictions are not met, providing a link between predictive coding and learning. Further experiments however, ideally combining fMRI and single-cell electrophysiology (Tsao et al., 2006), are needed to identify whether the reductions in BOLD signal really do reflect a predictive coding-based explaining away or a fine tuning of sensory representations.

References

- Alink A, Schwiedrzik CM, Kohler A, Singer W, Muckli L (2010) Stimulus predictability reduces responses in primary visual cortex. *J Neurosci* 30:2960–2966.
- den Ouden HEM, Daunizeau J, Roiser J, Friston KJ, Stephan KE (2010) Striatal prediction error modulates cortical coupling. *J Neurosci* 30:3210–3219.
- Murray SO, Schrater P, Kersten D (2004) Perceptual grouping and the interactions between visual cortical areas. *Neural Netw* 17:695–705.
- Nienborg H, Cumming BG (2009) Decision-related activity in sensory neurons reflects more than a neuron's causal effect. *Nature* 459:89–92.
- Olshausen BA, Field DJ (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 381:607–609.
- Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat Neurosci* 2:79–87.
- Tsao DY, Freiwald WA, Tootell RB, Livingstone MS (2006) A cortical region consisting entirely of face-selective cells. *Science* 311:670–674.