cortex xxx (2016) 1–16

Available online at www.sciencedirect.com

ScienceDirect

Journal homepage: www.elsevier.com/locate/cortex

Special issue: Review

Repetition suppression and its contextual determinants in predictive coding

*Ryszard Auksztulewicz** and *Karl Friston*

Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, London, United Kingdom

ARTICLE INFO

Article history:
Received 8 June 2015
Reviewed 27 August 2015
Revised 7 September 2015
Accepted 11 November 2015
Published online xxx

Keywords:
Repetition suppression
Predictive coding
Mismatch negativity
Perceptual inference
Perceptual learning

ABSTRACT

This paper presents a review of theoretical and empirical work on repetition suppression in the context of predictive coding. Predictive coding is a neurobiologically plausible scheme explaining how biological systems might perform perceptual inference and learning. From this perspective, repetition suppression is a manifestation of minimising prediction error through adaptive changes in predictions about the content and precision of sensory inputs. Simulations of artificial neural hierarchies provide a principled way of understanding how repetition suppression – at different time scales – can be explained in terms of inference and learning implemented under predictive coding. This formulation of repetition suppression is supported by results of numerous empirical studies of repetition suppression and its contextual determinants.

1. Introduction

The effect of stimulus repetition on neural responses is one of the most studied phenomena in neuroscience. Typically, repeated stimuli evoke neural activity with amplitudes smaller than responses to novel stimuli. Although repetition suppression is often portrayed as an expression of relatively simple mechanisms, such as neural fatigue (Grill-Spector, Henson, & Martin, 2006), its dependence on statistical regularities in the environment and other contextual factors casts repetition suppression as a consequence of sensory predictions (e.g., Summerfield, Trittschuh, Monti, Mesulam, & Egner, 2008). The predictive coding framework provides a principled explanation of repetition effects in terms of perceptual inference and learning, mediated by changes in synaptic efficacy (Friston, 2005). Adaptive changes in coupling of neuronal populations within areas and connectivity between areas are means of optimising a neuronal (generative) model of the external world to provide more accurate and precise predictions about sensory inputs. Thus, repetition suppression can be understood in terms of 'explaining away' sensory prediction errors.

In the following, we will review modelling and experimental work on repetition suppression in the setting of predictive coding. First, we introduce the predictive coding

* *Corresponding author*. Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, 12 Queen Square, London WC1N 3BG, United Kingdom.

E-mail addresses: R.Auksztulewicz@ucl.ac.uk (R. Auksztulewicz), K.Friston@ucl.ac.uk (K. Friston).

http://dx.doi.org/10.1016/j.cortex.2015.11.024

framework and portray neuronal message passing in terms of descending predictions, ascending prediction errors, and modulatory precision. We will then show how the predictive coding scheme can be mapped onto a canonical cortical microcircuit. The subsequent section will focus on explaining the dynamics of repetition suppression using simulations and computational modelling of empirical data. This will be followed by a review of empirical studies on repetition suppression and its context sensitivity, with a special focus on the crucial role of predictions and precision in modulating the effects of stimulus repetition.

2. Predictive coding

In order to maintain their integrity (e.g., homoeostasis), biological systems have to minimise the excursions or entropy of their interoceptive and exteroceptive states. Since entropy is the average of *surprise* (also known as surprisal or self-information) over time, biological systems should continually minimise their surprise about sensory states. Mathematically, this is equivalent to maximising the Bayesian evidence for their model of sensory inputs, also known as Bayesian filtering. Predictive coding (Friston, 2005; Mumford, 1992; Rao & Ballard, 1999) is a popular, neurobiologically plausible Bayesian filtering scheme that decomposes the optimisation of the agent's (neuronal) model of the world into two tractable components; namely (1) changes in expectations about the sensory inputs and (2) the computation of prediction errors that are needed to change expectations.

Minimising surprise – or maximising model evidence – lies at the heart of the free energy principle, where free energy provides a proxy for surprise that, under simplifying assumptions, can be reduced to prediction error (Friston, 2010). This means one can understand the process of perception as the resolution of prediction errors, by changing top-down predictions about the causes of sensory input (Fig. 1). Intuitively, the predictions descending along the processing (e.g., cortical) hierarchy are compared against sampled sensory inputs in sensory cortex (or expectations as intermediate hierarchical levels). The ensuing prediction errors are then passed up the hierarchy to optimise expectations and subsequent predictions. If the ascending input matches the descending prediction, the prediction error will be low – as exemplified by repetition suppression. If the predictions are inconsistent with the incoming input, the prediction error will be high – as illustrated by mismatch negativity (Garrido, Kilner, Stephan, & Friston, 2009). In the following, the notion of perception under predictive coding will be unpacked in the context of repetition suppression.

The ability of the brain to infer the causes of its sensations rests upon the presence of statistical structure or contingencies in the environment. These contingencies can be embodied within a generative model describing the hierarchical and dynamic statistics of the external world:

$$D\tilde{x}^{(i)} = f^{(i)}\left(\tilde{x}^{(i)}, \tilde{\upsilon}^{(i)}\right) + \tilde{\omega}_x^{(i)} \tag{1}$$

$$\tilde{\upsilon}^{(i-1)} = g^{(i)}\left(\tilde{x}^{(i)}, \tilde{\upsilon}^{(i)}\right) + \tilde{\omega}_\upsilon^{(i)}. \tag{2}$$

In the equations above, υ denote causes representing (hidden) causes (e.g., the bark of a dog), while x denote states of the world mediating the influence of that cause on sensory signals (e.g., the acoustic consequences of a dog barking). Because these dynamics follow stereotyped trajectories over time, they endow the model with memory. In equations above, tilde is used to augment the variables with their generalised coordinates of motion, i.e.,

$$\tilde{x} = \left[x, x', x'', \ldots\right]. \tag{3}$$

Eq. (1) describes the motion of states $x^{(i)}$ (at i-th hierarchical level) as a nonlinear function f of causes and states themselves. Here D is a block-matrix derivative operator, with identity matrices on its first leading-diagonal. Eq. (2) describes the motion of causes at a hierarchically lower level i-1 as a nonlinear function g of hidden causes and states at the level above. Random fluctuations in hidden causes and states are denoted by $\tilde{\omega}_\upsilon^{(i)}$ and $\tilde{\omega}_x^{(i)}$ respectively.

Since the brain does not have direct access to the causes and states in the external world, it can only infer the most likely values under its generative model: mathematically, these values are expectations. In other words, the generative model maps from causes to sensory consequences, while perception solves the (usually very difficult) inverse problem which is to map from sensations to their underlying causes. An inversion of hierarchical dynamic models can be cast in terms of a hierarchical message passing scheme also known as predictive coding:

$$\dot{\tilde{\mu}}_\upsilon^{(i)} = D\tilde{\mu}_\upsilon^{(i)} - \partial_{\tilde{\upsilon}}\tilde{\varepsilon}^{(i)}\xi^{(i)} - \xi_\upsilon^{(i+1)} \tag{4}$$

$$\dot{\tilde{\mu}}_x^{(i)} = D\tilde{\mu}_x^{(i)} - \partial_{\tilde{x}}\tilde{\varepsilon}^{(i)}\xi^{(i)} \tag{5}$$

$$\xi_\upsilon^{(i)} = \Pi_\upsilon^{(i)}\tilde{\varepsilon}_\upsilon^{(i)} = \Pi_\upsilon^{(i)}\left(\tilde{\mu}_\upsilon^{(i-1)} - g^{(i)}\left(\tilde{\mu}_x^{(i)}, \tilde{\mu}_\upsilon^{(i)}\right)\right. \tag{6}$$

$$\xi_x^{(i)} = \Pi_x^{(i)}\tilde{\varepsilon}_x^{(i)} = \Pi_x^{(i)}\left(D\tilde{\mu}_x^{(i)} - f^{(i)}\left(\tilde{\mu}_x^{(i)}, \tilde{\mu}_\upsilon^{(i)}\right)\right) \tag{7}$$

This message passing suggests two distinct populations of neurons: one encoding the trajectory of the expectations (conditional means) of hidden causes $\dot{\tilde{\mu}}_\upsilon^{(i)}$ and states $\dot{\tilde{\mu}}_x^{(i)}$, which we can therefore label state-units, and one encoding the prediction errors $\tilde{\varepsilon}_\upsilon^{(i)}$ and $\tilde{\varepsilon}_x^{(i)}$ weighted by their respective precisions $\Pi_\upsilon^{(i)}$ and $\Pi_x^{(i)}$, which we can label error-units. These precisions are the inverse amplitude of the random fluctuations above, so that when the fluctuations are small, prediction errors become precise and are amplified. To simplify notation, $\partial_{\tilde{x}}$ and $\partial_{\tilde{\upsilon}}$ are used to denote a partial derivative with respect to hidden states and causes respectively. Temporal derivatives, e.g., $\partial_t x$, are denoted by a dot $\dot{x}$.

These equations may look complicated but are formally simple and quite revealing in terms of which (neuronal) units talk to each other. In brief, the equations suggest that error-units receive messages from populations in the same hierarchical level and the level above, while state-units are driven by error-units in the same level and the level below. The prediction errors from the same level $\xi^{(i+1)}$ and the level below $\xi^{(i)}$ provide lateral and bottom-up messages driving the conditional expectations $\tilde{\mu}^{(i)}$ towards better lateral and top-down