Acta Psychologica 169 (2016) 1–10

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

On the joys of perceiving: Affect as feedback for perceptual predictions

Andrey Chetverikov [a,b,c,*], Árni Kristjánsson [a]

[a] Laboratory for Visual Perception and Visuomotor Control, Faculty of Psychology, School of Health Sciences, University of Iceland, Reykjavik, Iceland
[b] Department of Psychology, Saint Petersburg State University, St. Petersburg, Russia
[c] Cognitive Research Lab, Russian Academy of National Economy and Public Administration, Moscow, Russia

ARTICLE INFO

Article history:
Received 21 August 2015
Received in revised form 6 May 2016
Accepted 9 May 2016
Available online 16 May 2016

Keywords:
Affect
Predictive coding
Conflict
Uncertainty
Expectations
Errors

ABSTRACT

How we perceive, attend to, or remember the stimuli in our environment depends on our preferences for them. Here we argue that this dependence is reciprocal: pleasures and displeasures are heavily dependent on cognitive processing, namely, on our ability to predict the world correctly. We propose that prediction errors, inversely weighted with prior probabilities of predictions, yield subjective experiences of positive or negative affect. In this way, we link affect to predictions within a predictive coding framework. We discuss how three key factors – uncertainty, expectations, and conflict – influence prediction accuracy and show how they shape our affective response. We demonstrate that predictable stimuli are, in general, preferred to unpredictable ones, though too much predictability may decrease this liking effect. Furthermore, the account successfully overcomes the “dark-room” problem, explaining why we do not avoid stimulation to minimize prediction error. We further discuss the implications of our approach for art perception and the utility of affect as feedback for predictions within a prediction-testing architecture of cognition.

1. Introduction

Humans continually make predictions about the environment. As early in perceptual processing as in the retina, neurons make predictions based on temporal and spatial regularities (Gollisch & Meister, 2010; Hosoya, Baccus, & Meister, 2005). Recently, a powerful inference-based framework has emerged suggesting that brain activity can be described as prediction error minimization (Clark, 2013; Friston, 2009, 2012; Hohwy, 2012). According to this *predictive coding* approach, the brain uses hierarchical Bayesian inference to build a representation of the world. Conscious experience has been described as the “best hypothesis” (Hohwy, Roepstorff, & Friston, 2008), or the model that makes the most accurate predictions about the environment. However, discrepancies between predictions and outcomes are no less important. Prediction errors signify changes in the external world or in our internal states and a need to modify our predictions. We have suggested that affect serves as feedback on our predictions, reflecting their accuracy and regulating them so that confirmed predictions are more likely to be used again (Chetverikov, 2014; Chetverikov & Kristjansson, 2015). Furthermore, if predictions are confirmed (low prediction error), feedback is weighted with inverse prior probabilities of predictions, so that more probable predictions receive less positive feedback. In other words, confirmation of more probable predictions yields less positive feedback than confirmed less-probable predictions. Notably, within this framework there is no need to invoke additional concepts, such as values or rewards, to explain the relationship between affect and predictions. Affect represents a distinct dimension in experience: in addition to our “best hypothesis” about the world, people experience a feeling of how good this hypothesis actually is. The literature describing affect from this perspective has largely been limited to the perception of art (Salimpoor, Zald, Zatorre, Dagher, & Mcintosh, 2014; van de Cruys & Wagemans, 2011). We fill this gap by providing a more general perspective within a predictive coding framework.

2. Affect as universal currency for predictions

The utility of affect as weighted prediction error lies in its ability to provide a common currency for different predictions and drive behavior out of homeostasis. Human cognition is prone to errors, leading to the problem of verification in perception. How can observers distinguish hallucinations or illusory experiences from what is actually real in the world? A recurrent idea is that even if perception does not completely correspond to the world, researchers should try to understand the mechanisms that make our picture of the world more or less realistic. Instead of looking for a single source of protection from the fragility of perception the goal would be instead to look for numerous “dirty tricks” that our cognitive system utilizes to reach the best possible result (Ramachandran, 1990).

This is a parallel processing approach, where each piece of data is scrupulously analyzed with various tools for identifying stimuli. This parallel analysis could be implemented within an inference-based framework, such as predictive coding (Clark, 2013; Friston, 2009,

---

* Corresponding author at: Department of Psychology, University of Iceland, Saemundargata 2, 101 Reykjavík, Iceland.
E-mail address: andrey@hi.is (A. Chetverikov).

http://dx.doi.org/10.1016/j.actpsy.2016.05.005
0001-6918/

2012; Hohwy, 2012). Bayesian inference combines prior probabilities accumulated from experience (e.g., the probability of seeing a tree in a forest is high) with likelihood (how well actual input corresponds to the prediction of a tree) to determine posterior probabilities (the probability of a tree given the resemblance of sensory input to a tree and that we are in a forest). Predictive coding approaches suggest that cognitive architecture is organized in levels, each receiving predictions from higher levels that send error feedback on discrepancies between prediction and input. This information is, in turn, based on predictions that are then conveyed to lower levels, and so on (see Fig. 1, and below, for discussion of when predictions from differing levels may be in conflict).

Prediction error reflects discrepancy between prediction and input and allows comparison of qualitatively different predictions. For example, when one needs to identify an object, one could predict its identity based on recent experience, the probability of encountering it, context, color, semantic cues, shape, motion cues, and many other sources. It is hard to compare the results of such predictions directly, because they are expressed in different cognitive languages: shape, for example, involves spatial relations that are not necessary for color-based predictions. But prediction errors from differing cognitive levels can be compared, circumventing this problem, informing us which predictions are most accurate even if they are in conflict, for example, if shape analysis predicts a lamppost while context predicts a pedestrian.

Yet, prediction error may not always guide behavior optimally. As put by Clark (2013, p. 13), "staying still inside a darkened room would afford easy and nigh-perfect prediction of our own unfolding neural states" but it is obvious that this neither describes human behavior nor is this behavior adaptive. One way to solve this "dark room" problem is to posit inherent meta-priors that make dark rooms improbable with no possibility for correction of this model (Friston, Thornton, & Clark, 2012). Such meta-priors can be evolutionarily determined or learned through experience because humans are used to constant exposure to external stimulation.

We take a different approach, however, suggesting that behavior is guided by affect, defined as an experience of prediction error weighted with inverse prior probability of prediction. Prediction error is low inside the dark room while prior probabilities are high and low positive affect will therefore drive observers out of it. In the dark room, predictions become more and more accurate, but a continuous iterative weighting process of the inverse prior probabilities reduces positive affect. In contrast to the meta-priors idea we do not suggest that a high level of stimulation is always expected, but simply that low stimulation levels usually do not allow new and accurate predictions. Note that we do not reject the notion of predictions regarding stimulation levels. However, such predictions are not likely to be set in stone. For example, moving from the countryside to a big city or vice versa may lead to a troubled sleep due to changes in the level of audial stimulation. But after some time, expectations change and things return to normal.

Our approach shares characteristics with other accounts linking affect to predictions (Joffily & Coricelli, 2013; Schmidhuber, 2013; Van de Cruys & Wagemans, 2011; Van de Cruys, 2014). Most commonly, affect is linked to an experience of change in prediction errors. When prediction errors increase over time, observers supposedly experience negative affect while reduction of prediction error is associated with positive affect. For example, when observers are able to perceive an image in more detail than before, reduction of prediction error will lead to more positive affect. The affect in such accounts involves a second-order prediction, that is, a prediction regarding predictions. People expect their predictions not simply to be accurate (low error for first-order predictions) but more accurate than previous predictions.

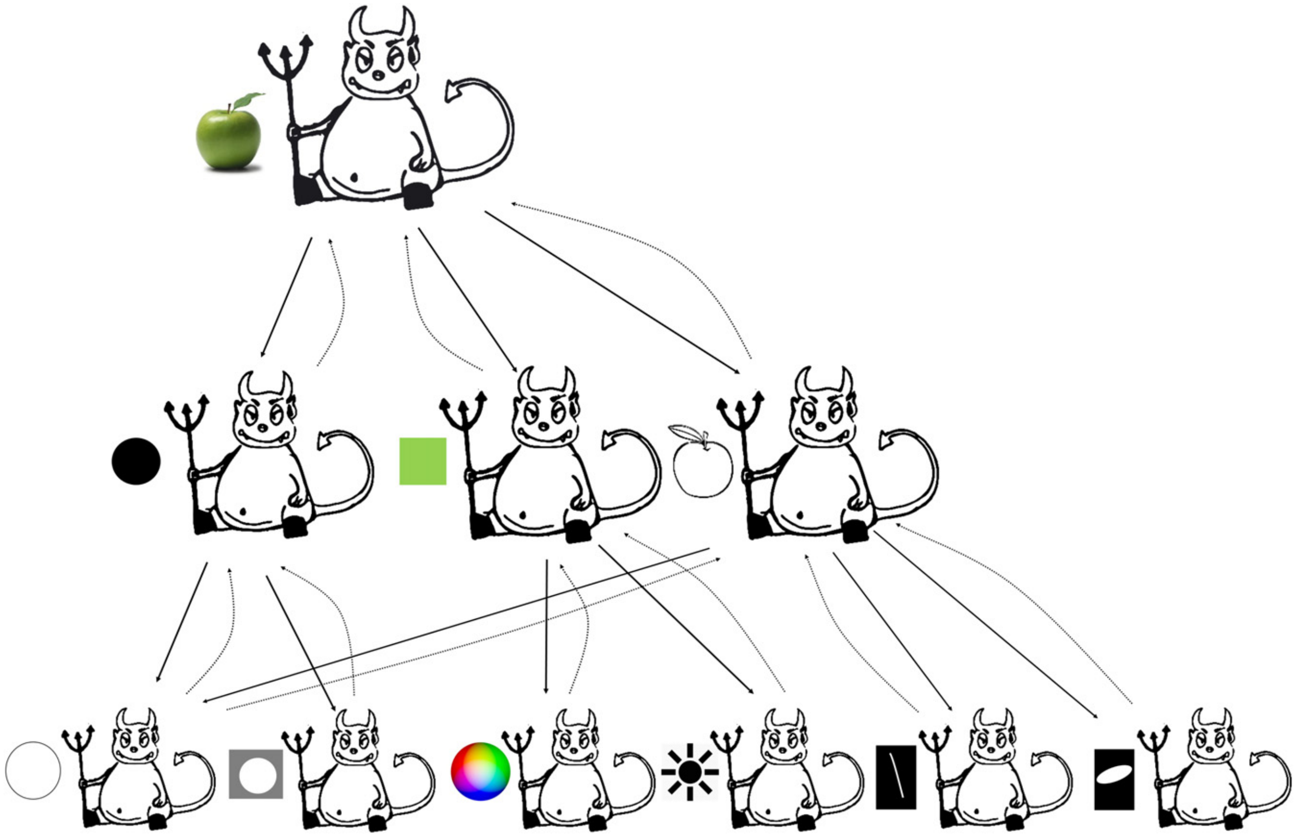

Fig. 1. Schematic representation of a predictive coding approach to perceiving an apple. A hierarchy of "predictive modules" (shown as demons echoing Selfridge's (1959) pandemonium model) are shown, with lower levels representing more granular predictions. In this example, the demons at the top level predict that one sees an apple. The prediction is translated by the second level of demons into predictions of "something circular and filled", "green" and "resembling the contours of an apple". These predictions are in turn split into simpler ones, relating to contours, lines, hue, lightness, etc. Solid arrows denote predictions, dotted arrows - prediction error. Images near the demons show the content of the predictions.