Biological Psychology 125 (2017) 45–57

Contents lists available at ScienceDirect

Biological Psychology

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

Temporal and spatial localization of prediction-error signals in the visual brain

Patrick Johnston[a,b,*], Jonathan Robinson[a,b], Athanasios Kokkinakis[b], Samuel Ridgeway[b], Michael Simpson[b], Sam Johnson[b], Jordy Kaufman[c], Andrew W. Young[b]

[a] Institute of Health and Biomedical Innovation, Queensland University of Technology, Australia
[b] York Neuroimaging Centre and Department of Psychology, University of York, UK
[c] Brain & Psychological Sciences Research Centre, Swinburne University of Technology, Burwood Road, Hawthorn, VIC 3123, Australia

ARTICLE INFO

Article history:
Received 17 October 2016
Received in revised form 17 February 2017
Accepted 17 February 2017
Available online 28 February 2017

Keywords:
Predictive coding
Visual perception
EEG
MEG
N170
M170
Perceptual prediction

ABSTRACT

It has been suggested that the brain pre-empts changes in the environment through generating predictions, although real-time electrophysiological evidence of prediction violations in the domain of visual perception remain elusive. In a series of experiments we showed participants sequences of images that followed a predictable implied sequence or whose final image violated the implied sequence. Through careful design we were able to use the same final image transitions across predictable and unpredictable conditions, ensuring that any differences in neural responses were due only to preceding context and not to the images themselves. EEG and MEG recordings showed that early (N170) and mid-latency (N300) visual evoked potentials were robustly modulated by images that violated the implied sequence across a range of types of image change (expression deformations, rigid-rotations and visual field location). This modulation occurred irrespective of stimulus object category. Although the stimuli were static images, MEG source reconstruction of the early latency signal (N/M170) localized expectancy violation signals to brain areas associated with motion perception. Our findings suggest that the N/M170 can index mismatches between predicted and actual visual inputs in a system that predicts trajectories based on ongoing context. More generally we suggest that the N/M170 may reflect a "family" of brain signals generated across widespread regions of the visual brain indexing the resolution of top-down influences and incoming sensory data. This has important implications for understanding the N/M170 and investigating how the brain represents context to generate perceptual predictions.

1. Introduction

It has long been recognized that top-down influences play a role in perception. An influential refinement of this idea is that rather than passively registering sensory data, the brain is hypothesized to actively generate and test predictions about its likely sensory input on a moment-by-moment basis (Gregory, 1980). Models of perceptual prediction therefore focus upon the need for mechanisms that attempt to minimize prediction error within reciprocally interconnected hierarchical networks (Friston & Kiebel, 2009; Panichello, Cheung, & Bar, 2013; Summerfield & de Lange, 2014). Behaviorally, there is growing support for the existence of such mechanisms. For instance, the phenomenon of *representational momentum* suggests the existence of dynamically evolving representations that model object trajectories (Hubbard, 2005), including biological motion trajectories (Kaufman & Johnston, 2014). Such findings suggest the possibility of identifying brain activity indices reflecting error-checking mechanisms at early stages of visual perception. We propose that the evoked brain response known as the N/M170 may provide such an index.

First reported by Bentin, Allison, Puce, Perez, and McCarthy (1996), the N170 Event Related Potential (ERP) has proven to be a robust and highly replicable index of early visual cognition (Johnston, Molyneux, & Young, 2015). Recorded maximally at occipito-temporal electrodes, the N170 is a negative inflection of the ERP occurring ~150–200 ms following stimulus onset. A corresponding positivity recorded with similar latency at central electrode sites – the Vertex Positive Potential (VPP) – is believed to reflect the same generators (Johnston, Stojanov, Devir, & Schall, 2005; Joyce & Rossion, 2005). There has been much focus on the N170 (M170 in magnetoencephalography/MEG; Halgren, Raij,

* Corresponding author at: School of Psychology & Counseling, Queensland University of Technology, Kelvin Grove, Victoria Park Road, QLD 4059, Australia.
E-mail address: patrick.johnston@qut.edu.au (P. Johnston).

http://dx.doi.org/10.1016/j.biopsycho.2017.02.004
0301-0511/

Marinkovic, Jousmäki, & Hari, 2000) as an index of face-sensitive processes, since the N/M170 is generally larger to faces than to other object categories (Eimer, 2011; Liu, Higuchi, Marantz, & Kanwisher, 2000; Rossion & Jacques, 2008). However, the N/M170 is also robustly elicited by non-face stimuli including objects of expertise (Tanaka & Curran, 2001) visual word-forms (McCandliss, Cohen, & Dehaene, 2003), naked bodies (Hietanen & Nummenmaa, 2011), and conditioned danger signals (Levita, Howsley, Jordan, & Johnston, 2015).

A common assumption has been that the N/M170 predominantly reflects stimulus-driven processes – indeed the dominant view is that the "face N/M170" indexes the structural encoding of faces preceding facial identification (Eimer, 2011). However, the N/M170 may be subject to influences of top-down modulation (Furl, van Rijsbergen, Treves, Friston, & Dolan, 2007; Hietanen & Astikainen, 2013; Johnston, Overell, Kaufman, Robinson, & Young, 2016; Righart & de Gelder, 2006) and visual salience (Hietanen & Nummenmaa, 2011; Levita et al., 2015).

Importantly, the N/M170 is believed to be the first component of the ERP capable of indexing higher-level vision, since earlier components (e.g. P1) are not sensitive to stimulus category (Rossion & Caharel, 2011). This makes it a natural potential candidate for indexing predictive mechanisms since the early stages of higher-level vision are likely to involve the resolution of incoming sensory data with top-down influences on perception. However, the standard visual ERP paradigm involves stimuli being selected at random from a fixed set and presented following a near-blank "fixation" screen. Whilst this has advantages in terms of experimental control, it means that every trial is, in effect, a quasi-independent context-free event. By contrast, in everyday life, many aspects of the visual environment are predictable and discontinuities in our visual input are mostly due to natural external properties such as occlusion or internally generated events such as blinks and saccades. Thus, adherence to the standard visual ERP paradigm may have masked some important aspects of the N/M170. We propose that the N/M170 may, in part, index "visual surprise" to the unpredicted appearance of a potentially important stimulus change.

We tested this idea in a series of three EEG experiments and a fourth experiment using MEG source reconstruction. In each experiment participants viewed a sequence of five successive static images on each trial. The first four images in each sequence created a contextual trajectory of implied movement, such as a regular series of changes in position. The final fifth image in each sequence either conformed to, or violated, the expected trajectory. Experiment 1 used facial expression trajectories, Experiment 2 used rigid-body rotation trajectories (of heads and body images) and Experiment 3 used locational trajectories (for faces and shapes). Experiment 4 was identical to the second experiment but was performed in the MEG scanner. Each experiment was conducted using a new sample of participants. The design of each experiment was such that exactly the same pairs of images were used to create predictable or unpredictable final stimulus transitions, to provide compelling demonstrations that any differences in neural responses were directly due to the preceding context and not to visual properties of the stimuli themselves.

2. Methods

2.1. Experiment 1

We first examined facial expression trajectories. Trials consisted of the presentation of sequences of five static images that followed a consistent direction of change in expression across the first four images, and where the final image either did or did not conform to the established pattern. We predicted that if the N170 can serve as an index of predictive mechanisms, amplitudes to Unpredictable final images would be larger than to Predictable final images.

2.1.1. Participants

There were 20 right-handed participants (10 female; mean age 23.5 years, SD 4.0). All were an opportunity sample from the undergraduate and postgraduate community at the University of York. Participants were offered 8GBP as compensation for their participation. This study was approved by the University of York Psychology Department Ethics Committee.

2.1.2. Stimuli

Stimuli were derived from images of a single male and a single female model from the NIMSTIM set (Tottenham et al., 2009). For these two models, closed mouth happy and neutral expression images were selected, and image-morphing techniques were used to create a sequence of images representing a morph-continuum between the two expressions. Following (Mayes, Pipingas, Silberstein, & Johnston, 2009) 45 fiducial points were used to identify corresponding spatial locations across the two images. These were placed at key locations on the face including the inner and outer canthi of the eyes, the centres of the pupils, multiple locations along the top and bottom of the upper and lower lips, and the face outline. Abrosoft Fantamorph (V 3.0) was then used to generate six intermediate images for each model, leading to a continuum of eight images for each model (2 original expressions, and 6 interpolated morphs between these). An oval frame was placed around each image to remove hair and background.

2.1.3. Procedure

Participants viewed a series of trials that consisted of the presentation of a sequence of five images, in which each image was displayed for 517 ms and then immediately replaced by the next image (0 ms ISI). Stimuli were presented centrally, subtending a visual angle of approximately 3°. Sequences consisted of stepwise images from one of the morph-continua, either commencing with a relatively neutral image (Image 1 or Image 2 from the continuum) that was followed by three progressively more happy images, or commencing with a relatively happy image (Image 7 or Image 8) that was followed by three progressively less happy images. The fifth image in each sequence was then used to create Predictable or Unpredictable experimental conditions. In Predictable sequences the 5th and final image conformed to the trajectory established by the preceding four images (either towards the full happy expression, or towards the neutral expression), whereas in Unpredictable sequences the 5th and final image reversed the direction of the trajectory established by the four preceding images (for example, if the first four images were in increasing morphed steps towards happiness, the final image in an Unpredictable sequence would involve a morphed step back towards neutral).

By using both happy to neutral and neutral to happy sequences in the first 4 images, we were able to match the set of final image transitions across the Predictable and Unpredictable conditions. That is to say each possible penultimate-to-final image transition for a trial in the Predictable condition was matched to an identical penultimate-to-final image transition for an Unpredictable trial whose initial trajectory was in the opposite direction (see Fig. 1). Thus, the set of Predictable trials consisted of image sequences 1-2-3-4-5, 2-3-4-5-6, 8-7-6-5-4 and 7-6-5-4-3, whilst the set of Unpredictable trials of image sequences was 1-2-3-4-3, 2-3-4-5-4, 8-7-6-5-6 and 7-6-5-4-5. This balancing of the fourth and fifth images in each sequence across conditions means that any differences in the ERPs to Predictable versus Unpredictable trials must arise as a consequence of the sequence of images preceding the final (fourth to fifth) image transition (i.e. the context), and cannot be due to any property of the final image transitions themselves (as exactly the same transitions were used in each condition).

It is important to note, that with a frame rate of <2 frames per second (fps) our stimuli are not perceived as fluid motion (which requires a minimum of 12 fps) but as a series of still images. The transitions between images are "jumps". The extent to which these "jumps" may be perceived as "motion" is post hoc since the spatial translation of the stimuli implies that motion must have occurred.

There were equal numbers of Predictable and Unpredictable sequence trials (160 each), presented in a randomized order. Trials were separated by a 1017 ms inter-trial interval during which a central fixation-cross was presented in an otherwise blank screen.

Participants were asked to maintain their gaze on the central fixation point. In order to maintain visual attention, they completed a simple vigilance task. This involved a set of 32 trials that were randomly interleaved with the main experimental trials. They were identical to the experimental trials with the exception that one of the images in the sequence included a small red-dot appearing at a location on the face close to the centre of the screen. This dot could appear (with approximately equal likelihood) on any of the five images constituting the trial. Participants were required to respond via a button press whenever they saw an image containing a red-dot. "Red-dot" trials were coded separately, and were not included in the analysis of ERP data. We set a criterion that any participant who was unable to respond to at least 90% of the "red-dot" trials should be excluded from analysis, on the basis of insufficient attention to the task. Apart from this red-dot monitoring task, the experiment involved passive viewing of the stimuli – participants were never asked anything about whether the sequences were predictable or unpredictable. There were no practice trials.

The task was delivered using Psychopy software (version 1.75) running on an Intel Pentium 4 HT computer, and the visual stimuli were presented on a 23″ TFT LCD