NeuroImage 83 (2013) 329–334

Contents lists available at ScienceDirect

NeuroImage

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

Characterising reward outcome signals in sensory cortex☆

Thomas H.B. FitzGerald *, Karl J. Friston, Raymond J. Dolan

Wellcome Trust Centre for Neuroimaging, London WC1N 3BG, UK

ARTICLE INFO

Article history:
Accepted 20 June 2013
Available online 27 June 2013

Keywords:
Credit assignment
Reward
Value learning
fMRI adaptation

ABSTRACT

Reward outcome signalling in the sensory cortex is held as important for linking stimuli to their consequences and for modulating perceptual learning in response to incentives. Evidence for reward outcome signalling has been found in sensory regions including the visual, auditory and somatosensory cortices across a range of different paradigms, but it is unknown whether the population of neurons signalling rewarding outcomes are the same as those processing predictive stimuli. We addressed this question using a multivariate analysis of high-resolution functional magnetic resonance imaging (fMRI), in a task where subjects were engaged in instrumental learning with visual predictive cues and auditory signalled reward feedback. We found evidence that outcome signals in sensory regions localise to the same areas involved in stimulus processing. These outcome signals are non-specific and we show that the neuronal populations involved in stimulus representation are not their exclusive target, in keeping with theoretical models of value learning. Thus, our results reveal one likely mechanism through which rewarding outcomes are linked to predictive sensory stimuli, a link that may be key for both reward and perceptual learning.

Introduction

Successful reward learning requires that an organism processes information about appetitive and aversive states, as well as assign causal responsibility for such states to antecedent events, that usually take the form of sensory stimuli. Whilst the first problem has been the subject of considerable neuroscientific study, the second, 'credit assignment' problem has been little explored in humans. One way in which the brain might perform a credit assignment is to direct a 'teaching signal', based on rewarding outcomes, to regions involved in stimulus processing (Friston et al., 1994; Roelfsema et al., 2010). Recently, several studies report evidence consistent with this, showing that rewarding feedback is associated with activity in sensory areas associated with stimulus processing, even in the absence of concurrent stimulation in that modality (Brosch et al., 2011; Pleger et al., 2008, 2009; Weil et al., 2010).

What is less clear is how populations of cells in sensory regions, targeted by an outcome signal, relate to those involved in stimulus representation. Supervised learning schemes, such as error back-propagation (Rumelhart et al., 1986), require generation of error signals tailored for each unit (Roelfsema and van Ooyen, 2005). This predicts a specific reactivation by reward feedback of units involved in stimulus representation. Such schemes are efficient but are considered to lack biological plausibility (Crick, 1989). Value learning models, by contrast, use a non-specific error signal that only modifies eligible connections; namely, those mediating the valuable outcome (Friston et al., 1994; Sutton and Barto, 1998) (visual stimuli within the last five seconds, for example). In a neurophysiological context, this predicts a non-specific input to sensory neurons to enable an associative increase in the strength of synaptic connections between recently active cells (Bailey et al., 2000; Calabresi et al., 2007; Izhikevich, 2007; Roelfsema et al., 2010). Critically, these two possibilities make different predictions about the relationship between spatial patterns of activity reflecting stimulus and outcome processing in stimulus-processing regions of the sensory cortex.

We tested predictions from these frameworks using fMRI adaptation. In brief, when two stimuli that activate the same neurons are presented in close temporal contiguity, the second stimulus produces a reduced BOLD response compared with an equivalent stimulus that does not activate the same population (Grill-Spector et al., 2006; Sawamura et al., 2006). Although the precise electrophysiological correlates remain unclear (Grill-Spector et al., 2006), this methodology has been used to probe stimulus representations across a range of distinct domains (Fang et al., 2007; Sawamura et al., 2006; Winston et al., 2004). We were interested in comparing neuronal responses to stimuli and their reward outcomes. To do this, we adopted a relatively new approach, based on spatial correlations within a region (Kriegeskorte and Bandettini, 2007; Kriegeskorte et al., 2008). We reasoned that if reward signals selectively reactivate sensory neurons involved in representing a preceding stimulus, then activations induced by the stimulus and reward should co-localise and their

---

☆ This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

* Corresponding author at: Wellcome Trust Centre for Neuroimaging, 12 Queen Square, London WC1N 3BG, UK.

E-mail address: thomas.fitzgerald@ucl.ac.uk (T.H.B. FitzGerald).

1053-8119/$ – see front matter © 2013 The Authors. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2013.06.061

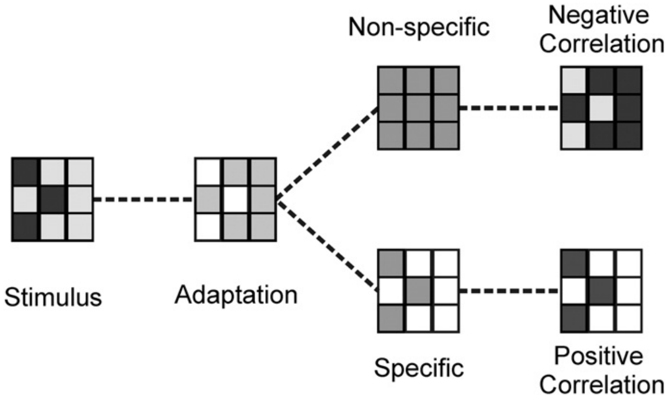

Fig. 1. Cartoon illustrating the effects of non-specific and generalised reward feedback on the spatial correlations between stimulus and outcome activity. (Darker colours indicate greater activity/responsiveness) Stimulus processing produces an adaptation effect, manifest in a decreased responsiveness which is greatest in voxels with the strongest response to the stimulus. If reward feedback signals are non-specific, this leads to a negative spatial correlation between stimulus and outcome activity. If reward feedback signals are specific to those neurons involved in stimulus representation, there will be a positive spatial correlation between stimulus and outcome activity.

patterns, over voxels, should be positively correlated. Conversely, if reward outcomes activate sensory neurons in a non-specific fashion, recently-activated stimulus-specific populations should show adaptation and be less responsive to reward signals, resulting in the activation patterns due to stimulus and reward being negatively correlated (Fig. 1). Crucially, this negative correlation should occur in the context of an overall positive response to reward, distinguishing them from simple reward-induced deactivations.

To test our hypotheses, we analysed high resolution fMRI data (FitzGerald et al., 2012) collected whilst subjects performed an instrumental learning task with visual cues and auditory feedback (Fig. 2). Specifically, we examined spatial correlations within an area of visual cortex responsive to cues and rewarding outcomes.

Materials & methods

Subjects

Twenty six (ten female) right-handed subjects, age range = 19–28 years, all free of psychiatric or neurological disease, participated in the study. The study was approved by the Joint National Hospital for Neurology and Neurosurgery (University College London Hospitals NHS trust) and Institute of Neurology (University College London) Ethics Committee. The subjects were paid according to their performance during the task (from £21.80–£28.80).

Stimuli & task

The subjects performed an instrumental learning task with visual cues and auditory feedback (FitzGerald et al., 2012) (Fig. 2). On each trial of the experiment, the subjects were presented with a visual cue consisting of a box with a coloured pattern, and made either a 'left' or a 'right' response by pressing a button on the corresponding keypad. After 2.5 s, they were played either a higher pitched 'win' sound, or a lower pitched 'no win' sound, each lasting for 1 s. The visual cue disappeared at the end of the sound. There was a variable inter-trial interval of 1–3 s between the trials. The subjects received 10 pence for each win.

Each cue had one of eight contingency types (win probabilities of [0.05 0.30], [0.05 0.55], [0.3 0.55], [0.4 0.9], with either $P(\text{Win}|\text{Chose Right}) > P(\text{Win}|\text{Chose Left})$ or the converse). Over the course of the experiment, each contingency type was repeated three times, using a total of 24 cues. The experiment was separated into blocks of 44 trials. In each block, two cues appeared in pseudo randomised order (we applied the constraint that no cue could be presented on more than three consecutive trials). Cues with identical or mirror image contingencies were never presented together in the same block. The subjects performed 6 blocks in each of two scanning sessions (12 in total). Each cue was presented in only one block. The subjects responded using two fMRI-compatible button boxes, one held in each hand.

Behavioural analysis

Behavioural analysis was performed, as previously described, by fitting a *Q*-learning algorithm, incorporating a softmax decision rule (FitzGerald et al., 2012). *Q*-learning updates the values of individual stimulus action pairs $Q(s,a)$ according to a reward prediction error weighted by a learning rate α (Watkins and Dayan, 1992).

$$Q(s_{t+1},a_{t+1}) = Q(s_t,a_t) + \alpha(R_t - Q(s_t,a_t))$$

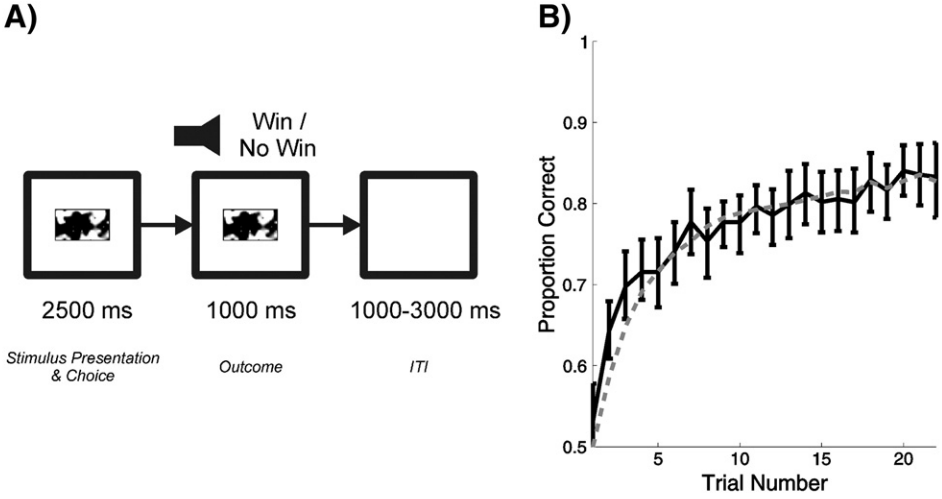

Fig. 2. A: Reward learning task. Subjects were presented with a visual stimulus, and given 2500 ms to make one of two responses, which were rewarded according to fixed probabilities for each stimulus-action pairing. Outcomes were then signalled with two different sounds, which were presented for 1000 ms — followed by a jittered inter-trial interval (1000–3000 ms, uniform distribution). B: Learning curve, averaged across all cues and subjects. Subjects increasingly chose the objectively higher-valued option, indicating that they were able to acquire appropriate responses to the reward contingencies. (Solid line: learning curve based on observed choice behaviour. Dashed line: learning curve based on Q-learning models fitted to individual subject behaviour. Error bars indicate bootstrapped 95% confidence intervals for the observed choice behaviour).