Neuropsychologia 66 (2015) 75–87

Contents lists available at ScienceDirect

Neuropsychologia

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

Processing of action- but not stimulus-related prediction errors differs between active and observational feedback learning

Stefan Kobza [a,*], Christian Bellebaum [b]

[a] Institute of Cognitive Neuroscience, Department of Neuropsychology, Faculty of Psychology, Ruhr University Bochum, Universitätsstraße 150, 44780 Bochum, Germany
[b] Institute for Experimental Psychology, Heinrich Heine University Düsseldorf, Universitätsstraße 1, 40225 Düsseldorf, Germany

ARTICLE INFO

Article history:
Received 18 March 2014
Received in revised form
19 September 2014
Accepted 27 October 2014
Available online 7 November 2014

Keywords:
Reward
Learning
Observation
Striatum
Prediction error
Cerebellum

ABSTRACT

Learning of stimulus–response–outcome associations is driven by outcome prediction errors (PEs). Previous studies have shown larger PE-dependent activity in the striatum for learning from own as compared to observed actions and the following outcomes despite comparable learning rates. We hypothesised that this finding relates primarily to a stronger integration of action and outcome information in active learners. Using functional magnetic resonance imaging, we investigated brain activations related to action-dependent PEs, reflecting the deviation between action values and obtained outcomes, and action-independent PEs, reflecting the deviation between subjective values of response-preceding cues and obtained outcomes. To this end, 16 active and 15 observational learners engaged in a probabilistic learning card-guessing paradigm. On each trial, active learners saw one out of five cues and pressed either a left or right response button to receive feedback (monetary win or loss). Each observational learner observed exactly those cues, responses and outcomes of one active learner. Learning performance was assessed in active test trials without feedback and did not differ between groups. For both types of PEs, activations were found in the globus pallidus, putamen, cerebellum, and insula in active learners. However, only for action-dependent PEs, activations in these structures and the anterior cingulate were increased in active relative to observational learners. Thus, PE-related activity in the reward system is not generally enhanced in active relative to observational learning but only for action-dependent PEs. For the cerebellum, additional activations were found across groups for cue-related uncertainty, thereby emphasising the cerebellum's role in stimulus–outcome learning.

1. Introduction

One of the most important principles of evolution is the selection of traits (of a species or an individual) which fit best in the current environmental conditions. Accordingly, organisms seek to select behaviour which fits best in the current situation, that is, behaviour which is followed by the most positive outcome. In order to maintain a high proportion of positive outcomes in an ever-changing environment, the ability to adapt behaviour is crucial. This adaptive process is reflected in an increase and decrease of the probability of behaviour followed by positive and negative outcomes, respectively, suggesting a preceding learning process: for example, the extent to which an outcome is worse than predicted may serve as a 'teaching signal' to select an alternative action in the future. Conversely, an outcome better than predicted may elicit a signal which reinforces repetition of the preceding action. Animal and human studies have revealed neural correlates of these so-called prediction errors (PEs): Dopamine (DA) neurons in the monkey fire at higher frequency following unexpected reward, whereas the firing rate drops below baseline when expected reward is omitted (Schultz, 1997, 1998a, 1998b; Schultz et al., 1997). A similar pattern was found via microelectrode recordings in Parkinson's Disease (PD) patients during deep brain surgery (Zaghloul et al., 2009). Further evidence in humans stems from functional magnetic resonance imaging (fMRI) studies showing outcome-related activations also in brain regions receiving projections from midbrain DA neurons (Haber and Fudge, 1997), most prominently the basal ganglia (BG; Delgado, 2007; Pagnoni et al., 2002) and the medial prefrontal cortex (mPFC) (O'Doherty et al., 2001; Rolls et al., 2008), particularly the anterior cingulate cortex (ACC; Holroyd et al., 2004; for a review, see Knutson and Cooper, 2005). These structures constitute the so-called reward system. Outcome-related activations have, however, also been found in other structures such as the insula (Clark et al., 2009; Delgado et al., 2000) and the hippocampus (Dickerson et al., 2011). Furthermore, a study on patients with cerebellar lesions suggests that the

* Corresponding author. Tel.: +49 234 32 23119; fax: +49 234 32 14622.
E-mail addresses: stefan.kobza@ruhr-uni-bochum.de (S. Kobza), christian.bellebaum@hhu.de (C. Bellebaum).

http://dx.doi.org/10.1016/j.neuropsychologia.2014.10.036

cerebellum is also involved in reward-based reversal learning (Thoma et al., 2008), which is in line with anatomical connections between the cerebellar dentate nucleus and the striatum (Hoshi et al., 2005). Consequently, the cerebellum may also play an important role in non-motor (stimulus–outcome) learning.

Importantly, reinforcement-learning is not necessarily restricted to processing of action–outcome related PEs: an action may result in different outcomes depending on the context or, in experimental terms, a preceding informative 'cue'. Notably, one can thus differentiate between two outcome PEs. An action-independent PE reflects the difference between the received outcome and the subjective value (SV) of the cue, which, just as the action value (AV, i.e. subjective value of the chosen action), changes based on outcome history. An action-dependent PE, on the other hand, reflects the difference between the received outcome and the AV. Both PEs appear to be differentially processed in the brain. O'Doherty et al. (2004) showed that the ventral striatum is involved both when outcomes did and did not depend on a preceding action, whereas the dorsal striatum codes action-dependent PEs. It is thus conceivable that especially the dorsal striatum facilitates learning of associations between (own) actions and their consequences. In line with this assumption, Bellebaum et al. (2008) reported disrupted feedback-based reversal learning in BG patients especially when the dorsal striatum was affected.

Stimulus–action–outcome associations can also be learned via observation of another person's actions and the feedback he or she receives. On the one hand, observational learning is characterised by an additional PE which relates to observed actions and which is coded in the dorsolateral PFC (Burke et al., 2010). Furthermore, Monfardini et al. (2013) found activations for observed but not own incorrect outcomes in the posterior medial frontal cortex, the anterior insula, and the posterior superior temporal sulcus. On the other hand, many brain regions are involved in processing of outcome PEs for both active and observational learning, with decreased activity of parts of the 'classical' reward system in observational as compared to active learning (Bellebaum et al., 2010, 2012; Yu and Zhou, 2006). In an fMRI study by Bellebaum et al. (2012), PE-dependent activations in the right putamen were found in both types of learning, with stronger activations in the right anterior caudate nucleus for active learners, suggesting that the processing of PEs is generally reduced in observational learning from feedback. On the other hand, these studies showed that action–outcome associations are learned similarly well in active and observational learning, thereby demonstrating that action–outcome associations can be acquired by observation. We hypothesised, however, that parts of the reward system are dedicated to integrating own (rather than observed) actions with outcomes, and we examined this by differentiating between PEs depending and not depending on the preceding (own or observed) action.

Based on our recent fMRI findings (Bellebaum et al., 2012) and evidence we obtained in PD patients (Kobza et al., 2012), we expected that the BG integrate own actions with outcome information during outcome evaluation in active learning, which would lead to differences between active and observational learning with respect to action-dependent PE processing. We further hypothesised that neural coding of outcome PEs in the reward system is not enhanced in active relative to observational learners if PEs are independent from actions.

For both SVs and AVs, activations have been found in parts of the reward system, such as the orbitofrontal cortex (FitzGerald et al., 2009), the dorsal ACC (Camille et al., 2011), the PFC (Glascher et al., 2009), the supplementary motor cortex (Wunderlich et al., 2009), and the putamen during active learning (FitzGerald et al., 2012). Activations reflecting reward expectation have so far not been investigated in observational learning. Furthermore, activity related to uncertainty has been reported for the amygdala in fMRI studies on aversive conditioning (Buchel et al., 1998; Labar et al., 1998) but also reward learning (Prevost et al., 2011). Uncertainty reflects the extent to which expectations of future reward vary over the course of the task (for the computational definition, see Section 2.4.4): Prior to learning, outcomes are completely unknown, so that uncertainty is at its maximum. Over the course of learning, outcome predictions become more accurate, so that uncertainty decreases. Consequently, uncertainty can be regarded as an inverse indicator of stimulus–outcome learning such as in classical conditioning, which has been shown to depend on the cerebellum (Daum et al., 1993). Thus, the present study also aimed to explore similarities and differences between active and observational learning with respect to the neural representation of SVs, AVs, and uncertainty signals preceding the outcome phase.

2. Material and methods

2.1. Subjects

33 healthy, right-handed adult volunteers participated in the study. Two participants were excluded due to data acquisition problems. Out of the remaining 31 subjects (age range 20–34 years), one group of 16 subjects (6 female; mean [M] age=25.1 years; standard deviation [SD]=3.7 years) engaged in an active feedback learning task, whereas the 15 subjects (9 female; M=23.9 years; SD=4.5 years) of a second group learned by observing the choices and following feedback in another person (see Section 2.2 for details of the learning tasks). The mean age did not differ between groups (p=.45). The current IQ as estimated via the Multiple Choice Vocabulary Test (Mehrfachwahl-Wortschatz-Test, MWT, version B; Lehrl, 2005) was also comparable (p=.48) between groups who learned actively (M=116.3; SD=13.9) or by observation (M=119.8; SD=14.0). All participants had normal or corrected-to-normal vision. Apart from standard exclusion criteria applied in fMRI studies – such as artificial cardiac pacemakers, metallic implants, diagnosed or reported claustrophobia – a history of neurological or psychiatric disease and regular medication affecting the central nervous system led to exclusion from the study. Prior to participation, subjects gave written informed consent. The study conforms to the Declaration of Helsinki and received ethical clearance by the Ethics Board of the Faculty of Psychology of the Ruhr University Bochum, Germany.

2.2. The learning tasks

In the present study, two feedback learning tasks were used: one in which subjects learned from own choices and the following outcomes, and one in which subjects learned from choices of another subject and the following outcomes. Both tasks are based on a probabilistic learning card-guessing paradigm introduced by Delgado et al. (2005). As in our previous studies on differences between active and observational learning (Bellebaum et al., 2010; Kobza et al., 2012), we applied a between-subjects design. This design rules out carry-over effects, which may occur in within-subject designs, in which learning from own choices in one phase may influence behaviour and/or neural coding of learning by observation in another phase of the experiment and vice versa.

Recording of participants' responses and timing of stimuli – presented via MRI video goggles (Resonance Technology, Inc.; http://www.mrivideo.com) – was controlled by Presentation Software (Neurobehavioral Systems, Inc.; http://www.neurobs.com).

2.2.1. Active feedback learning task

The subjects of the first group learned from their own choices. Each active learning trial started with the presentation of a fixation cross with a duration of 4000, 8000, and 12000 ms on 58.3%, 29.2%, and 12.5% of trials, respectively. Then one out of five different 'cards' (represented by frames including a circle, triangle, square, star, or hexagon) was presented as cue. Subjects were instructed that on the back of each card a number would be printed. After another fixation cross, subjects were asked to guess whether the number was lower or higher than the number 5, i.e. between 1 and 4 or between 6 and 9, by selecting a downward-directed arrow (presented on the left) or an upward-directed arrow (presented on the right) using the left or right button of a response box via index or middle finger of the right hand, respectively. Following the response, the chosen arrow was surrounded by a red circle. If subjects did not respond within 2000 ms, they were prompted to respond faster. Otherwise, after another presentation of a fixation cross, positive ('+50c' in green characters, indicating a monetary win of 50 cents) or negative ('−50c' in red characters, indicating a monetary loss of 50 cents) feedback was given for a correct or an incorrect guess, respectively (see Fig. 1A for the sequence of events in one learning trial and for the duration of stimulus presentation).