NeuroImage 56 (2011) 709–715

Contents lists available at ScienceDirect

NeuroImage

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

Decoding different roles for vmPFC and dlPFC in multi-attribute decision making

Thorsten Kahnt [a,b], Jakob Heinzle [a], Soyoung Q. Park [b,c], John-Dylan Haynes [a,b,d,*]

[a] Bernstein Center for Computational Neuroscience, Charité-Universitätsmedizin Berlin, Germany
[b] Berlin School of Mind and Brain, Humboldt Universität zu Berlin, Germany
[c] Department of Education and Psychology, Freie Universität, Berlin, Germany
[d] Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

ARTICLE INFO

Article history:
Received 18 March 2010
Revised 30 April 2010
Accepted 20 May 2010
Available online 25 May 2010

Keywords:
Multi-attribute decision making
Expected value
Functional magnetic resonance imaging (fMRI)
Multivariate decoding
Ventromedial prefrontal cortex (vmPFC)
Dorsolateral prefrontal cortex (dlPFC)

ABSTRACT

In everyday life, successful decision making requires precise representations of expected values. However, for most behavioral options more than one attribute can be relevant in order to predict the expected reward. Thus, to make good or even optimal choices the reward predictions of multiple attributes need to be integrated into a combined expected value. Importantly, the individual attributes of such multi-attribute objects can agree or disagree in their reward prediction. Here we address where the brain encodes the combined reward prediction (averaged across attributes) and where it encodes the variability of the value predictions of the individual attributes. We acquired fMRI data while subjects performed a task in which they had to integrate reward predictions from multiple attributes into a combined value. Using time-resolved pattern recognition techniques (support vector regression) we find that (1) the combined value is encoded in distributed fMRI patterns in the ventromedial prefrontal cortex (vmPFC) and that (2) the variability of value predictions of the individual attributes is encoded in the dorsolateral prefrontal cortex (dlPFC). The combined value could be used to guide choices, whereas the variability of the value predictions of individual attributes indicates an ambiguity that results in an increased difficulty of the value-integration. These results demonstrate that the different features defining multi-attribute objects are encoded in non-overlapping brain regions and therefore suggest different roles for vmPFC and dlPFC in multi-attribute decision making.

Introduction

Successful decision making requires precise anticipatory representations of the reward values that can be obtained from choosing specific behavioral options. However, in everyday life most decision alternatives consist of multiple reward-related attributes. For instance, different attributes of a fruit–size, shape, color and surface texture–signal parts of its nutritional value. To make an optimal choice i.e. to pick the fruit with the highest expected value, the reward predictions of all attributes need to be integrated into a combined value. To describe such decision processes, Multi-Attribute Utility Theory (MAUT) was developed by behavioral decision researchers in the 1970s (Slovic et al., 1977; von Winterfeldt and Fischer, 1975).

Neuroscience has mainly focused on decisions regarding single-attribute options (Daw et al., 2006; Glascher et al., 2009; Hampton et al., 2006; Kim et al., 2006; O'Doherty et al., 2003b). Studies on multiple attributes have typically directly investigated the trade-off between two attributes such as taste vs. health (Hare et al., 2009), amount of money vs. delay (Kable and Glimcher, 2007) and pleasure of acquisition vs. price (Knutson et al., 2007). Furthermore, studies on decisions between real-life objects (comprising multiple attributes) typically did not address their multiple-attribute character explicitly (Chib et al., 2009; FitzGerald et al., 2009; Hare et al., 2009; Knutson et al., 2007; Plassmann et al., 2007). One study aimed to identify brain regions involved in experimentally controlled multi-attribute decisions (Zysset et al., 2006). In this study, however, only the attribute-wise similarity between alternatives i.e. the difficulty of the decision was examined. Taken together, although single- and two-attribute decisions have been studied, no study has moved beyond two attributes and comprehensively investigated how such multi-attribute objects are represented in the brain.

Each attribute of a multi-attribute object can have its own predictive information for reward. Importantly, different attributes of the same object can signal *different or even conflicting* reward values. For instance, for one object all attributes could signal an intermediate value, whereas for another object different attributes could signal high and low values. Thus, although both objects have the same combined value (i.e. intermediate) the multi-attribute objects would differ considerably in the *variability* of the rewards predicted by their individual attributes (i.e. low vs. high). Hence, unlike single-attribute objects, different multi-attribute objects can differ not only in their expected value but also in the variability of the rewards predicted by their attributes. In order to understand how decisions are made on the basis of multi-attribute objects we investigated how the

* Corresponding author. Charité-Universitätsmedizin Berlin, Bernstein Center for Computational Neuroscience, Haus 6, Philippstrasse 13, 10115 Berlin.
E-mail address: haynes@bccn-berlin.de (J.-D. Haynes).

1053-8119/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2010.05.058

combined value and the variability of the rewards predicted by the individual attributes are represented in the human brain.

Recently we have shown that expected values can be decoded from distributed fMRI patterns in the ventromedial prefrontal cortex (vmPFC) (Kahnt et al., 2010). This distributed coding is consistent with reports from single-unit recordings showing that different neural populations in value sensitive cortex increase and decrease their firing rate with increasing reward value, respectively (Kennerley et al., 2009; Kobayashi et al., 2010; Morrison and Salzman, 2009; Padoa-Schioppa and Assad, 2006; Schoenbaum et al., 2007). Previous experimental and theoretical work on the human visual system has revealed that applying multivariate pattern analysis (MVPA) techniques to fMRI data is specifically suited to extract information encoded in distributed neural populations (Haynes and Rees, 2005, 2006; Kamitani and Tong, 2005; Norman et al., 2006). Similarly, information about cognitive and decision processes has been shown to be encoded in distributed fMRI patterns in the prefrontal cortex (PFC) (Hampton and O'Doherty, 2007; Haynes et al., 2007; Soon et al., 2008). Thus, it might be expected that distributed fMRI patterns contain more information about the combined value of multi-attribute objects and the variability of rewards predicted by individual attributes than the average fMRI signal. Hence, here we used MVPA techniques (Haynes and Rees, 2006; Norman et al., 2006) to decode information about these two variables.

Materials and methods

Participants

Sixteen right-handed subjects (8 female, mean age = 26.4 ± 1.06 years SEM) participated in the experiment. Subjects had normal or corrected-to-normal vision and gave written informed consent to participate. The study was approved by the local ethics review board of the Charité-Universitätsmedizin Berlin.

Classical conditioning session

In all experiments we used objects that could vary in three visual attributes *shape*, *color* and *coherence of moving dots* with three levels per attribute. In the days prior to scanning (mean 3.19 ± 0.31 SEM) participants performed a *classical conditioning session*, where they learned the association between *single-attribute objects* and different reward values (magnitudes of monetary outcomes). For this behavioral session the objects had only a single feature that varied across three levels, thus resulting in 9 stimuli (3 × shape + 3 × color + 3 × coherence). The three different shapes (diamond, octagon and dodecagon) were presented in white color on black background, the three colors (green, turquoise and blue) were presented in squares on black background and the three coherence levels of moving dots (5%, 35% and 95% coherence) were also presented in white squares on black background (Fig. 1B). The three levels of each attribute were associated with increasing magnitudes of monetary outcomes (0.10 €, 0.20 € and 0.30 €). An example pairing is shown in Fig. 1B; the actual pairings were counter-balanced across subjects and gender. During conditioning, in each trial one single-attribute object (e.g. a green square indicating 0.10 €) was randomly selected and presented for 2000 ms followed by the presentation of its monetary value (1000 ms). Subjects were told that they will receive the money after the experiment. Each stimulus was presented 10 times, resulting in 90 conditioning trials.

Scanning session

In each trial of the fMRI experiment (Fig. 1A), one level of every single-attribute was combined into a *multi-attribute object* (e.g. shape: octagon indicating 0.20 €, color: blue indicating 0.30 € and coherence: 5% indicating 0.10 €) and presented for 2000 ms. After a variable delay (4000–8000 ms) subjects were asked to rate the combined value of that object on a continuous, circular rating scale (without labeling) using an MRI compatible trackball. Once the rating was made the cursor was blocked and the rating scale stayed presented for a total of 2500 ms (maximum rating time). In the scanning session the objects consisted of *combinations* of all three attribute levels, thus resulting in 27 different multi-attribute objects (3 × 3 × 3 = 27). Each was presented two times in each of the 4 scanning runs. Subjects were informed that they would receive 50% of the combined value (sum of single-attribute values) of each multi-attribute object they evaluated during the experiment. Before scanning, subjects repeated the conditioning procedure described above and practiced on the circular rating scale. Furthermore, they went through one run (54 trials) of the experiment as a practice.

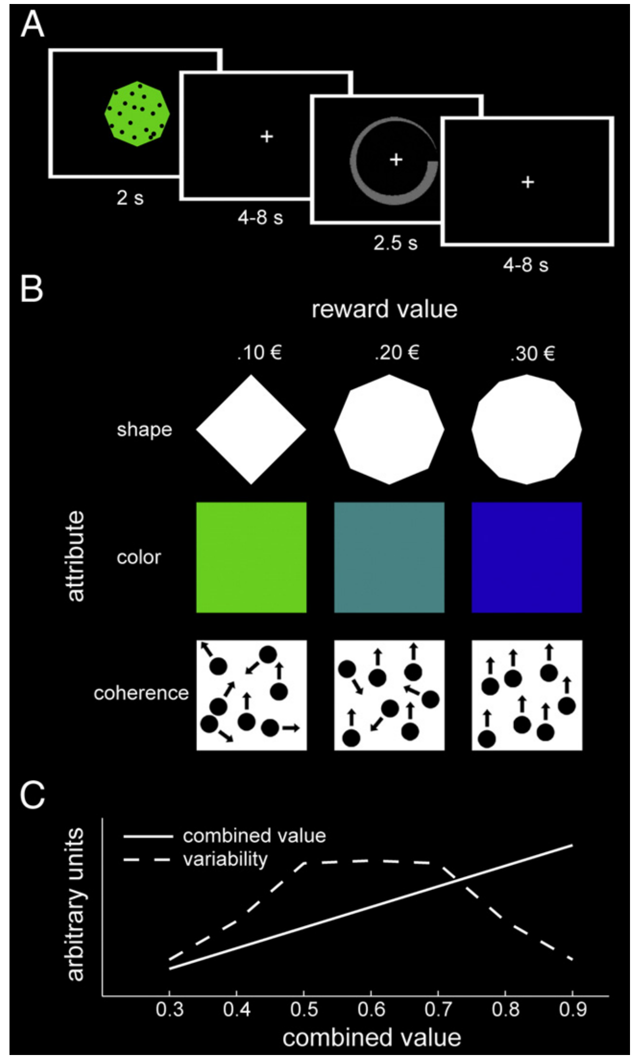

Fig. 1. Experimental design. (A) Example trial of the task used in the scanner. In each trial one multi-attribute object was presented for 2 s. After a variable delay of 4–8 s subjects had to rate the combined value of the object (maximum rating time 2.5 s). Trials were separated by a variable interval of 3.5–7.5 s. To avoid motor preparation and confounds related to the motor component of the rating, a circular rating scale with randomized orientation was used. (B) Example of associations between single attributes and rewards. Different rows indicate the attributes shape, color and coherence (of moving dots) and columns indicate the reward value that was associated with the visual cues in the cells during the classical conditioning procedure. Associations were counter-balanced across subjects and gender. (C) Average relationship between the combined value and the variability of multi-attribute objects. Combined value (solid line) and variability (dashed line) are plotted as a function of combined value.