Investigating structure and function of the dopaminergic midbrain: with a special focus on the human VTA

Trutti, A.C.

Chapter 7

Exploring neural signatures of the reward prediction error in the human brain using 7 Tesla fMRI

This chapter is in preparation for publication.

Abstract

Predictions driven by reward-related information shape behaviour and learning processes. Understanding the neural underpinnings of the reward prediction error (RPE) is critical in unravelling the complexities of reward processing within the human brain. A large body of work has demonstrated the crucial role of midbrain dopamine neurons in reward-related processes like the RPE. Despite extensive evidence in non-human species, studying RPE in humans faces challenges due to the inaccessibility of midbrain neurons. This study employs ultra-high field 7 Tesla functional MRI (fMRI), and a scanning protocol tailored for subcortical imaging to investigate RPE-related neural activity in the human subcortex. This study utilises various masks of subcortical nuclei, including individualised masks developed by Bazin et al. (2020) and a model-based fMRI approach to understand better the spatial organisation and functional specialisations of the subcortex, with a particular emphasis on the midbrain. The fMRI analysis, spanning whole-brain and region-of-interest (ROI) specific levels, uncovers significant BOLD responses associated with RPE in ventral and dorsal regions of the striatum, amygdala, hippocampus, and cortical areas. Our results demonstrate robust striatal activation across analyses consistent with the literature, while dopaminergic midbrain activation remains elusive. The absence of findings in the midbrain ROIs may be attributed to challenges in functional imaging of the midbrain, even with high-resolution fMRI, and precisely delineating ROIs in the midbrain, especially in dopamine-specific populations.

7.1 Introduction

Survival in the natural world hinges not only on an organism's ability to react to immediate stimuli but also on its capacity to predict and respond to future events. Predictions afford organisms the crucial time needed to prepare behavioural responses, enhancing their ability to make optimal choices. This anticipatory capability becomes especially critical when organisms must navigate choices that may lead to the acquisition of valuable resources. A fundamental aspect of predictions is the concept of reward, ultimately guiding animals toward beneficial outcomes. Rewards function as positive reinforcers, influencing learning processes and sustaining well-established appetitive behaviours. These learning processes are formalised within the reinforcement learning (RL) framework (Sutton & Barto, 1998), which defines the reward prediction error (RPE) as the difference between expectations and actual rewards. The RPE drives learning through integrating past experiences into new expectations. Decades of research in non-human species have shown that midbrain dopamine neurons are pivotal in predicting and encoding reward-related information. Besides responding to rewarding events, dopamine neurons appear to calculate the RPE through changes in their signalling (Bayer & Glimcher, 2005; Enomoto et al., 2011; Mirenowicz & Schultz, 1994; Montague et al., 1996; Schultz et al., 1993, 1997). This RPE signalling is reflected in an increase in the firing of dopaminergic neurons when the received reward is larger than the predicted reward (positive RPE). At the same time, a pause of spontaneous firing of the dopaminergic neurons indicates that the received reward is less than the predicted reward (negative RPE). Along these lines, no change in the firing of the midbrain dopamine neurons suggests a perfect prediction of the perceived reward. Hence, the dopaminergic midbrain neurons and their projection targets have emerged as focal points for investigating the encoding and processing of RPE.

Functional involvement of midbrain dopamine neurons is not limited to rewarding events but in fact, those neurons show activation to diverse occasions, such as aversive (Matsumoto & Hikosaka, 2009; Watabe-Uchida et al., 2012, 2017) and noxious events (Brischoux et al., 2009), salience (Matsumoto & Hikosaka, 2009) and movement (Bogacz, 2020; da Silva et al., 2018; Howe & Dombeck, 2016). For instance, a study in nonhuman primates provided intriguing insights into the organisation of dopamine neurons and delineated two subpopulations of dopamine neurons —saliency-coding and value-coding (Matsumoto & Hikosaka, 2009). While several neurons demonstrated increased firing in response to reward-predicting events and inhibition in response to aversive events (value-coding), some neurons responded with increased firing to both rewarding and aversive events (saliency-

coding). Importantly, these neurons exhibited differential spatial localisation within the midbrain dopamine nuclei, with saliency-coding neurons predominantly located in the dorsolateral part, mainly substantia nigra pars compacta (SNc), and reward-value-coding neurons primarily found in the more ventromedial part linked to the medial SNc and VTA. Other studies suggest a robust association between midbrain dopamine neurons of the VTA and the RPE (Bayer & Glimcher, 2005; Cohen et al., 2012; Schultz et al., 1997; Tian et al., 2016), while it is unclear to what extent the dopamine neurons of the SNc are also engaged (Howe & Dombeck, 2016; Matsumoto & Hikosaka, 2009). In contrast, SNc dopamine neurons demonstrate movement-related activity (Howe & Dombeck, 2016; Schultz et al., 1983).

Despite ample evidence from animal studies for the encoding of the RPE in the midbrain's dopamine neurons, evidence from human studies is still sparse (D'Ardenne et al., 2008; Miletić, 2023; Pauli et al., 2015), as invasive cell-recording techniques are not available in investigations of the human brain, rendering it necessary to study reward-learning processes in feedback-based computer experiments. This absence of empirical support can be explained by the challenges associated with imaging the dopaminergic neurons. For instance, the anatomical location of the dopaminergic midbrain neurons, which are located deep in the brain and thus far away from the magnetic resonance imaging (MRI) coils, along with physiological noise from adjacent vascular and ventricular systems, result in low signal- and BOLD contrast-to-noise ratio (Barry et al., 2013; de Hollander et al., 2017; Trutti et al., 2019) and an increased risk of signal contamination from neighbouring structures (de Hollander et al., 2015). The SN demonstrates high levels of iron (Chen et al., 1989; Haake et al., 2005; Miletić et al., 2022b), while the VTA exhibits a high heterogeneity in tissue properties as it covers diverse neural populations (Morales & Margolis, 2017; Root et al., 2016; Trutti et al., 2019, 2021), rendering customised structural and functional MRI scanning protocols essential. All these factors, in combination with the relatively small volume of both VTA and SN, demand high-field strength, such as 7 Tesla (T), with small voxel sizes and increased signal-to-noise ratios when aiming to localise associated functions, such as the RPE, with increased spatial specificity in humans (de Hollander et al., 2017; van der Zwaag et al., 2015).

In this paper, we aim to address the challenges of imaging the midbrain and clarify questions surrounding RPE-encoding neurons by employing ultra-high field 7T fMRI. We use a scanning protocol customised for the subcortex to probe the neural activity associated with RPE in the human brain. Using individual masks of the VTA and SN, we seek to address the lingering questions surrounding the spatial

organisation of neurons encoding RPE and shed light on the potential functional distinctions between the VTA and SN. Through this investigation, we aim to enhance our understanding of the functional organisation of the midbrain, the neural basis of the RPE, and its implications for reward-related behaviours in humans.

7.2 Methods

7.2.1 Procedure

37 participants (20 female; mean age 26.65± 5.72; age range 19 – 39 years) took part in the study, approved by the ethical committee at the University of Amsterdam, the Netherlands, and the Regional Committees for Medical and Health Research Ethics, Norway. All participants provided written informed consent and completed MRI screening forms to ensure they were eligible for scanning. The recruitment was conducted at the Norwegian University of Science and Technology. The participants had a corrected-to-normal vision and no history of epilepsy or overt clinical neuropsychiatric disease. Participants completed two blocks of the reversal learning task.

7.2.2 Reversal learning task

The experimental paradigm employed an instrumental learning task, which involved the reversal of previously acquired reward contingencies after roughly 50% of the trials. During the experiment, participants were tasked with making decisions between sets of two abstract choice stimuli, each linked to a predetermined reward probability (see Figure 7.1), within 1.5 s. After each response, feedback was provided as either +0 or +100 points. In every trial, one of the choice options consistently offered a higher expected reward, termed the 'correct' choice. Following each decision, participants received feedback in the form of points. Choice difficulty varied between the two sets, with one set featuring reward probabilities of 80/20% and the other set featuring 70/30% reward probabilities. Each block consisted of 102 trials, totalling 204 trials. Two pairs of stimuli with reward probabilities of 0.8/0.2 and 0.7/0.3 were randomly interleaved in each block. Between trials 15 and 35 (uniformly sampled), the reward probability switched between stimuli for each stimulus pair. Specifically, the stimulus with a pre-reversal reward probability of 0.8/0.7 transitioned to a post-reversal reward probability of 0.2/0.3 and vice versa. Importantly, participants were not informed about these reversals before the experiment commenced.

Each trial lasted 6.9 seconds, equivalent to 5 volumes (TR). The timing of events was varied to decorrelate the design matrix, achieved by pseudo-randomly selecting the duration of each fixation cross from 0.5 s, 1 s, 1.5 s, and 2 s. Furthermore, 13 null trials were incorporated, where the screen stayed empty for 5.53 s.

7.2.3 Cognitive modelling

The reinforcement learning advantage racing diffusion (RL-ARD) model, which has previously demonstrated accurate prediction of response time distributions and learning curves of a similar reversal learning paradigm (Miletić et al., 2021), was fit to the behavioural data. The RL-ARD model belongs to reinforcement learning evidence accumulation models (RL-EAMS), which combines two schools of behavioural modelling: reinforcement learning and evidence accumulation models (see Miletić et al., 2020). In evidence accumulation models (EAM), evidence in favour of a particular decision is modelled as an accumulator and a decision is made once this accumulator reaches an evidence threshold. The RL-ARD shares this common framework with the EAMs.

Precisely, in the RL-ARD, decision-making is modelled as a competition among accumulators, each collecting evidence for a specific choice option by gathering the *advantage* of one choice option relative to another. The speed of evidence accumulation is represented in the drift rate parameter v, which is influenced by a weighted combination of the advantage, the sum of the mean reward expectancies, and an urgency signal (for more information on how v is mathematically conceptualised, see Miletić et al., 2021). The initial accumulator to reach a shared threshold level of evidence a triggers the motor processes that execute the decision. The response time is determined by the time taken to reach the threshold, plus an intercept denoted as $t0$, representing the time required for processes unrelated to evidence accumulation, such as early perceptual encoding and motor response execution. In this version of an RL-EAM model, the reinforcement learning aspect is incorporated through the 'evidence' term, which signifies Q-values. These Q-values denote the participant's internal beliefs about the perceived rewards of each choice option.

The RL-ARD model was fit to the behavioural data utilising Bayesian model estimation, and participant-level posterior distribution was acquired through a combination of differential evolution and Markov-chain Monte Carlo sampling (Ter Braak, 2006; Turner et al., 2013a), employing default setting in the Dynamic models

of Choice (DMC; Heathcote et al., 2019) software package for R (R Core Team, 2017). For a detailed protocol of model-fitting, see Miletić (2023, p. 155). Ultimately, the RL-ARD model featured a total of 6 free parameters: evidence-independent base rate *V0*, weights on the difference and sum of the evidence (w_d and w_s), non-decision time *t0*, learning rate α, and threshold *a*.

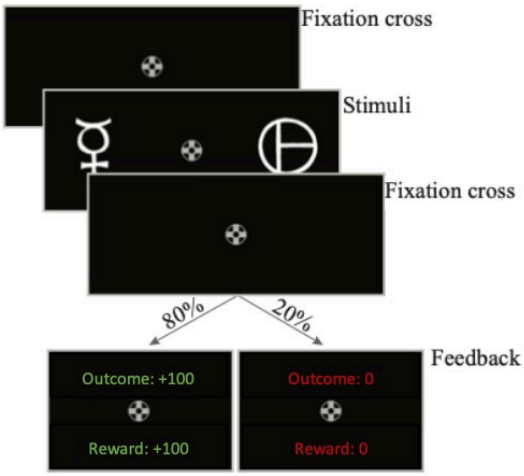

Figure 7.1. In the reversal learning experiment, the trial begins with a fixation cross (0-2 s), then the stimulus is presented for 1.5 seconds, followed by another fixation cross (0-2 s), after which probabilistic feedback is given (1s). The acquisition phase lasted 15–35 trials for each stimulus set (uniformly sampled each block), after which the reward contingencies for each stimulus set reversed. The actual symbols used varied between runs and participants.
Note. Figure reprinted and adapted from '*Modelling structure and function of the human subcortex*', by S. Miletić, 2023, with permission.

7.2.4 MRI data acquisition

Each participant was scanned four times using a 7T Siemens MAGNETOM TERRA scanner (strength = 80 mT/m at 200 T/m/s) equipped with a 32-channel head coil. Anatomical scans were collected in the first session, and the remaining sessions involved different functional scans. The data discussed in this article were collected in one of the functional sessions that included four functional runs in which the participants completed two different tasks. One task was the reversal-learning task (Figure 1), which consisted of two runs of 102 trials each. The other

task was the reference-back task and is not analysed in the present article (see Chapter 6). The anatomical session involved acquiring a multi-echo gradient recalled echo scan (GRE) and an MP2RAGE scan. The GRE scan parameters were as follows: TR = 31.0 ms, TE1 = 2.51 ms, TE2 = 7.22 ms, TE3 = 14.44 ms, TE4 = 23.23 ms, FA = 12°, FOV = 240 × 240 × 168 mm. The MP2RAGE scan parameters were TR = 4300 ms, TE = 1.99 ms, inversions TI1 = 840 ms, TI2 = 3270 ms, flip angle 1 = 5°, flip angle 2 = 6°, FOV = 240 x 240 x 168 mm, and bandwidth (BW) = 250 Hz/Px (Marques et al., 2010). The experimental session comprised four functional echo-planar imaging (EPI) runs with 4 EPI volumes acquired with opposite phase encoding directions for susceptibility distortion correction. The functional data was collected using a single-echo 2D-EPI BOLD sequence with the following parameters: TR = 1380 ms, TE = 14 ms, MB = 2, GRAPPA = 3, voxel size = 1.5 mm isotropic, partial Fourier = 6/8, flip angle = 60°, MS mode = interleaved, FOV = 192 x 192 x 128 mm, matrix size = 128 x 128, BW = 1446 Hz/Px, slices = 82, phase encoding direction = A >> P, and echo spacing = 0.8 ms. Each of the two tasks of the functional session had a total of 2 runs, each lasting 13 minutes and 45 seconds, resulting in a total of 4 runs and 55 minutes. To help co-register the functional scans to the high-resolution data from the anatomical session, another anatomical MP2RAGE scan (1mm) was collected at the end of the functional session. Physiological data, including heart and breathing rate, were recorded for all participants to assess the impact of physiological noise on the fMRI data. An 18-regressor RETROICOR model (Glover et al., 2000) was employed for this purpose. The model comprised a fourth-order phase Fourier expansion of the heart rate signal, a second-order phase expansion of the respiration signal, and a second-order phase Fourier expansion of the interaction between heart rate and respiration (Harvey et al., 2008). Additional regressors accounted for heart rate variability (HRV; Chang et al., 2009) and respiratory volume per time unit (RVT; Birn et al., 2008; Harrison et al., 2021). The PhysIO toolbox (Kasper et al., 2017) implemented in the TAPAS software (Frässle et al., 2021) was employed to estimate the physiological regressors.

7.2.5 MRI data preprocessing

The imaging data was preprocessed using the neuroimaging preprocessing tool *fMRIPrep 20.2.0* (Esteban et al., 2018; Esteban et al., 2019), which is based on *Nipype 1.7.0* (Gorgolewski et al. (2011); Gorgolewski et al. (2018); RRID:SCR_002502). The anatomical data preprocessing involved multiple steps such as intensity and non-uniformity correction (using `N4BiasFieldCorrection`,

Tustison et al., 2010), skull-stripping (using *Nipype's* `antsBrainExtraction.sh`), and tissue segmentation (using *FSL*'s fast, Zhang et al., 2001) of the T1-weighted images. The brain-extracted T1-weighted scans were normalised using volume-based spatial nonlinear registration to standard space (*MI152NLin2009cAsym*; Fonov et al. (2009), RRID:SCR_008796) using `antsRegistration` (*ANTs 2.3.3*). For more information on anatomical data preprocessing, see Miletić (2023). For each of the two functional (BOLD) runs per task per participant, the following preprocessing was performed. A reference volume and its skull-stripped version were generated by aligning and averaging one single-band references (SBRefs). A B0-nonuniformity map (or fieldmap) was estimated based on two echo-planar imaging (EPI) references with opposing phase-encoding directions, with `3dQwarp` (Cox & Hyde, 1997; AFNI 20160207). Based on the estimated susceptibility distortion, a corrected EPI (echo-planar imaging) reference was calculated for a more accurate co-registration with the anatomical reference. The BOLD reference was then co-registered to the T1w reference using `bbregister` (FreeSurfer) which implements boundary-based registration (Greve & Fischl, 2009). Co-registration was configured with six degrees of freedom. Head-motion parameters with respect to the BOLD reference (transformation matrices and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using `mcflirt` (FSL 5.0.9, Jenkinson et al., 2012). BOLD runs were slice-time corrected using `3dTshift` from AFNI 20160207 (Cox & Hyde, 1997; RRID:SCR_005927). A reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. The BOLD time series (including slice-timing correction when applied) were resampled onto their original, native space by applying a single composite transform to correct for head motion and susceptibility distortions. These resampled BOLD time series will be referred to as preprocessed BOLD in the original space or just preprocessed BOLD. Several confounding time series were calculated based on the preprocessed BOLD: framewise displacement (FD), 'DVARS' (the spatial standard deviation of difference images), and three region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions, Power et al., 2014) and Jenkinson (relative root mean square displacement between affines, Jenkinson et al., 2002). FD and DVARS are calculated for each functional run, both using their implementations in *Nipype* (following the definitions by Power et al., 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (*CompCor*, Behzadi et al., 2007). Principal components are estimated after high-pass filtering the preprocessed

BOLD time series (using a discrete cosine filter with 128 s cut-off) for the two *CompCor* variants: temporal (*tCompCor*) and anatomical (*aCompCor*). *tCompCor* components are then calculated from the top 2% variable voxels within the brain mask. For *aCompCor*, three probabilistic masks (CSF, WM and combined CSF + WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. in that instead of eroding the masks by 2 pixels on BOLD space, the *aCompCor* masks are subtracted from a mask of pixels that likely contain a volume fraction of GM. This mask is obtained by dilating a GM mask extracted from the *FreeSurfer's* `aseg` segmentation, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each *CompCor* decomposition, the k components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 per cent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components were excluded. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded by including temporal derivatives and quadratic terms for each (Satterthwaite et al., 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardised DVARS were annotated as motion outliers. All resamplings can be performed with a single interpolation step by composing all the pertinent transformations (i.e., head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using `antsApplyTransforms` (*ANTs*), configured with Lanczos interpolation to minimise the smoothing effects of other kernels (Lanczos, 1964). Non-gridded (surface) resamplings were performed using `mri_vol2surf` (*FreeSurfer*). Many internal operations of fMRIPrep use *Nilearn 0.6.2* (Abraham et al., 2014, RRID:SCR_001362), mostly within the functional processing workflow.

7.2.6 Regions-of-interest

We employed the multi-contrast anatomical subcortical structure parcellation (MASSP) algorithm (Bazin et al., 2020) to generate individualised anatomical masks for 17 subcortical structures. MASSP utilises various contrasts, including quantified susceptibility (QSM) values, longitudinal relaxation rates (R1), and effective transverse relaxation rates (R2*). For more information on the MRI

data processing to run MASSP, see Miletić (2023). The MASSP algorithm delineated 17 subcortical structures: Amygdala (AMG), claustrum (CL), fornix (fx), external and internal segments of the globus pallidus (GPe, GPi), internal capsule (ic), periaqueductal grey (PAG), pedunculopontine nucleus (PPN), red nucleus (RN), substantia nigra (SN), subthalamic nucleus (STN), striatum (STR), thalamus (THA), ventral tegmental area (VTA), and lateral, third, and fourth ventricles (LV, 3V, 4V). For all regions except fx, 3V, and 4V, separate masks were obtained for both hemispheres. We focused solely on grey matter structures, excluding the internal capsule, fornix, and ventricles from the subsequent ROI analyses, resulting in 12 bilaterally considered ROIs.

To explore potential divergent regional specificity in the RPE's neural signature, probabilistic maps of the SNc and two VTA subnuclei, as well as masks of the striatum's subdivision, putamen (Pu), caudate (Ca) and nucleus accumbens (NAcc), were selected from the atlas of Pauli, Nili, and Tyszka (2018). The two VTA masks from the Pauli atlas fall into the region defined as VTA by MASSP (Trutti et al., 2019), a dorsolateral located parabrachial pigmented (PBP) nucleus and the 'VTA', which refers to a small, ventromedially located nucleus (referred to as the VTA nucleus or VTAnc in the following).

7.2.7 fMRI statistical analysis

Initially, a whole-brain analysis was performed to examine brain activation using statistical parametric maps (SPMs) for each contrast at a global level. Subsequently, region of interest (ROI)-wise GLM analyses were carried out to investigate the role of subcortical nuclei in reversal learning. Details about the GLM analysis at both the whole-brain and region-specific levels are provided below. A canonical double gamma hemodynamic response function (HRF) with a temporal derivative was utilised as the basis set for all analysis methods (Glover, 1999).

7.2.7.1 Whole-brain generalised linear models (GLMs)

We fit whole-brain voxel-wise GLMs to the fMRI data, including regressors for each event and model parameters for each voxel. The list of regressors included 'response left', 'response right', 'stimulus', 'value', 'feedback', and 'RPE' along with their temporal derivatives. The initial two regressors are associated with motor responses and are modelled based on the initiation of button presses. The stimulus and feedback effects are both represented by an intercept (capturing the impact of stimulus/feedback presentation) and a parametrically varying regressor (representing

trial-specific changes) in line with Miletić (2023, p. 160). The amplitude of the regressors for 'stimulus', 'stimulus value', 'feedback' and 'RPE' varied parametrically across trials, where each trial's amplitude was based on the fitted RL-ARD model (see Miletić, 2023, p. 160, for a detailed protocol on the model simulations), and subsequently, the regressors were z-scored. In addition to the task-specific regressors, our design matrix incorporated six motion parameters (comprising three translational and three rotational parameters), along with 'DVARS' and framewise displacement estimates obtained during preprocessing. Furthermore, we included physiological regressors derived from *RETROICOR* estimations relating to heart rate and respiration (Harvey et al., 2008; Birn et al., 2008; Harrison et al., 2021). The *physIO toolbox* (Kasper et al., 2017) of the TAPAS software package (Frässle et al., 2021) was used to estimate physiological regressors. For three participants, either both runs or only the second run of physiological data were not collected due to technical issues. In these cases, we substituted the first 20 components obtained through *aCompCor*, as described by Behzadi et al. (2007). Before GLM fitting, a minimal smoothing of the data was performed using SUSAN (Smith & Brady, 1997) with a kernel size of FWHM of 1.5 mm. GLMs for each run were estimated using FSL FEAT (Woolrich et al., 2001), and the two run-level GLMs per participant were subsequently brought together through a fixed-effects analysis. Group-level models were calculated using FSL FLAME1+2 (Woolrich et al., 2004). All group-level SPMs underwent correction for the false discovery rate (FDR) using the Benjamini-Hochberg procedure ($q < 0.05$; Yekutieli & Benjamini, 1999).

7.2.7.2 ROI-wise GLMs

The final analysis, the ROI GLM analysis, was conducted as follows: Mean time series data were extracted from the unsmoothed fMRI data for each subcortical ROI using the masks provided by the individual MASSP parcellation (Bazin et al., 2020) and selected masks of the atlas by Pauli et al. (2018), as described in 2.6. Subsequently, the time series data were transformed into percentage signal change values by dividing each timepoint by the mean signal of the time series, multiplying the result by 100, and then subtracting 100. To streamline the hierarchical GLM fitting process, we initially filtered the time series by employing least square regression on the identical confounds utilised in the whole-brain GLMs. This step aimed to diminish physiological noise and eliminate low-frequency drifts from the signal, contributing to a more focused parameter estimation. Furthermore, the individual runs were concatenated for analysis, and hierarchical GLMs were fit to the mean time series separately for each ROI. For a more comprehensive description of

the hierarchical Bayesian parameter estimation procedure, see Miletić (2023, p. 162-164). Due to a malfunction in the physiological recording system, the data from one participant were excluded. The second run data for two other participants was omitted for the same reason. One participant did not respond to trials in the second run of the session, leading to excluding this run from the ROI analysis. We infer exclusively from positive BOLD responses, given the disagreements surrounding negative BOLD responses (Schridde et al., 2008; Wade, 2002).

7.3 Results

7.3.1 Behavioural analysis

As the experimental paradigm falls within the reinforcement learning domain, the behavioural data was examined to assess whether participants exhibited learning effects and derive descriptive statistics. Overall, the mean response time (RT) was 0.802 s *(SD* = 0.246), and the mean accuracy of choosing the choice option with a high probability of reward was 0.69 (*SD* = 0.463). RT for correct and incorrect choices was 0.796 s (*SD* = 0.239) and 0.817 (*SD* = 0.259), respectively. See Figure 2D for mean RT across trial bins for correct and incorrect choices. Pre-reversal accuracy was 0.803 (*SD* = 0.398), and post-reversal accuracy was 0.574 (*SD* = 0.495). Participants showed the expected learning effects by means of increased accuracy in the acquisition phase, which decreases after reward contingencies reverse before response accuracy increases again (Figure 7.2 A, C).

7.3.2 fMRI analysis

7.3.2.1 Whole-brain GLMs

Whole-brain GLMs were conducted to explore BOLD responses correlated with the magnitude of the RPE estimated by the RL-ARD model. Several significant BOLD clusters were found to covary with the size of the RPE spanning cortical and subcortical regions (Figure 7.3, Table 7.1). Specifically, the ventral striatum exhibited distinct BOLD clusters in both the left and the right hemisphere (Figures 7.3 and 7.4) in accordance with previous work that highlighted an important role of the ventral striatum, in particular the nucleus accumbens, in RPE-processing (D'Ardenne et al., 2008; Hare et al., 2008; Lefebvre et al., 2017; O'Doherty et al., 2003; Tobler et al., 2006). Other subcortical BOLD clusters associated with the RPE were found in the putamen (dorsal striatum, Figure 4), the (left) amygdala, and the hippocampus. These findings are consistent with the literature (Cooper et al., 2012; Corlett et al., 2022; Lefebvre et al., 2017) and data from another reinforcement

learning paradigm collected in a different session as part of this larger data collection effort (Miletić, 2023, p. 169). No evidence for BOLD clusters in areas of the dopaminergic midbrain was revealed in the whole-brain GLMs.

In the cortex, a large cluster of activation was found in the occipital pole, consistent with previous studies that localised BOLD responses covarying with RPEs in occipital regions (Corlett et al., 2022; Lefebvre et al., 2017; Miletić 2023, p. 169; Rohe et al., 2012). Our data also revealed BOLD cluster in the frontal pole, insular cortex, and (anterior) cingulate cortex, which have previously been suggested to play important roles in RPE processing (Hauser et al., 2017b; Lefebvre et al., 2017; Preuschoff et al., 2008; Rohe et al., 2012). Inconsistent with previous work that implicates the orbitofrontal cortex with RPE processing (O'Doherty et al., 2003; Lefebvre et al., 201; Miletić, 2023, p. 169; Tobler et al., 2006), we only found small clusters of activation in orbitofrontal regions, mostly limited to the left hemisphere. Overall, activation in the frontal pole was limited to a small cluster and predominantly found lateralised. For a complete list of BOLD activation clusters and corresponding peak voxels, see Table 7.1.

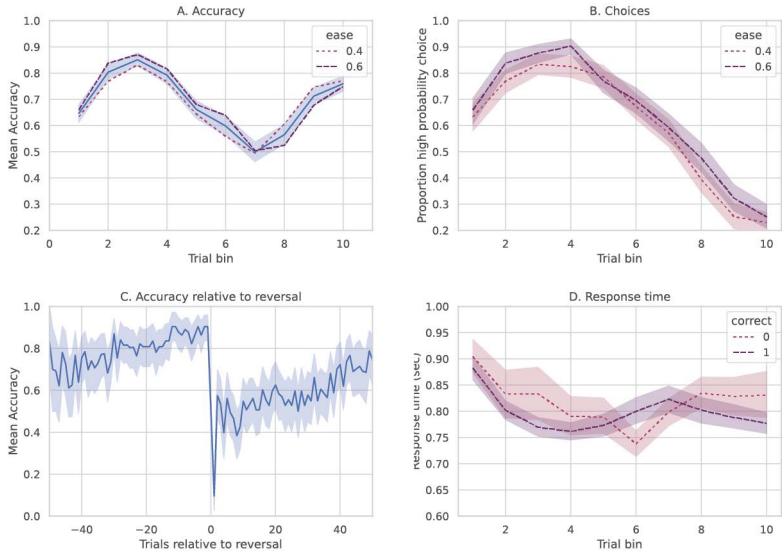

Figure 7.2. Behavioural data illustrating the response behaviour of participants. Mean accuracy across trial bins is shown in panel (A) across stimulus sets (solid line), separately with the stimulus sets (dashed lines). Panel (B) illustrates the probability of choosing the high probability choice for each trial bin. Panel (C) demonstrates choice accuracy relative to the reversal point. And Panel (D) shows the response time for correct and incorrect choices across trial bin. Shaded areas correspond to the 95% confidence interval.

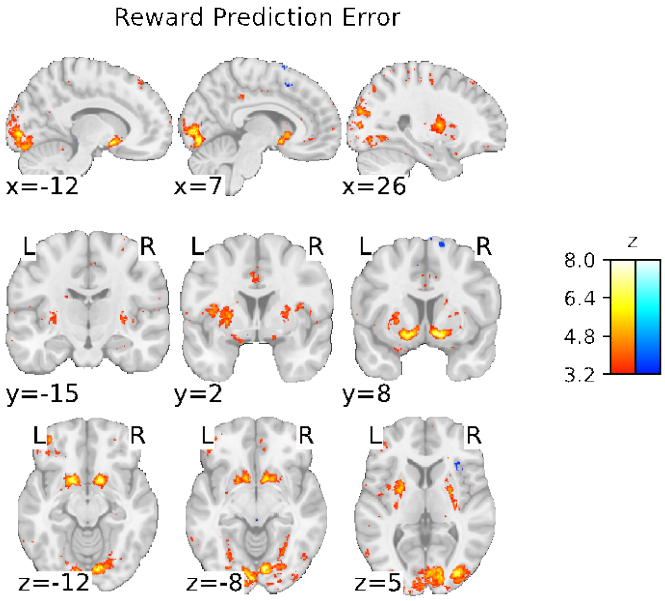

Figure 7.3. Statistical parametric maps of the reward prediction error contrast. Coordinates are in MNI152NLin2009cAsym (1 mm) space. Thresholds were determined with the false discovery rate (z = 3.19).

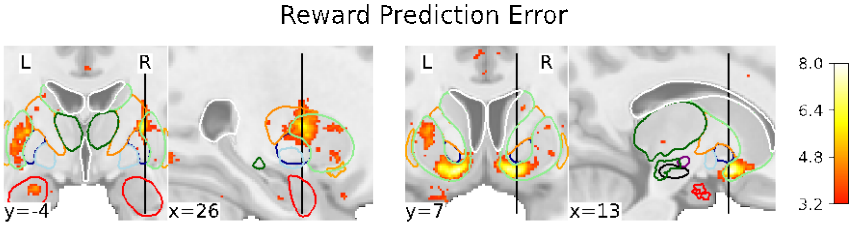

Figure 7.4. Subcortical statistical parametric maps of the reward prediction error contrast with a focus on dorsal striatal (left) and ventral striatal (right) clusters. Coordinates are in MNI152NLin2009cAsym (1 mm) space. Thresholds were determined with the false discovery rate (z = 3.19). Contours illustrate the following regions: Amg (red), Cl (orange), fx (light gray), GPe (dark blue), GPi (light blue), ic (dark orange), PAG (grey), RN (red), SN (black), STN (purple), Str (light green), Tha (dark green), lateral, 3rd and 4th ventricles (white), and VTA (dark green). Masks represent probabilistic maps derived from first, applying the MASSP algorithm to the data from Miletić et al. (2022b) and second, non-linearly warping the obtained masks to MNI152NLin2009cAsym.

Table 7.1. Whole-brain Analysis' List of Peak Activation of Significant Clusters in MNI2009c Coordinates.

			MNI			
	hem	**volume (mm³)**	**x**	**y**	**z**	***Z***
Occipital pole	r	25374	7.00	-87.00	2.00	8.089
Putamen	r	1532	13.00	8.00	-12.00	7.448
Nucleus accumbens	l	1627	-12.00	7.00	-11.00	7.193
Putamen	r	1290	25.00	-3.00	8.00	6.240
Putamen	l	2213	-26.00	-2.00	5.00	5.956
Frontal pole	l	2067	-38.00	47.00	-14.00	5.561
Insular cortex	l	659	-39.00	1.00	8.00	5.464
Fusiform gyrus	l	399	-28.00	-69.00	-4.00	4.969
Precuneus cortex	r	137	15.00	-36.00	46.00	4.848
Inferior parietal lobule	l	198	-45.00	-47.00	52.00	4.772
Parietal operculum cortex	r	406	52.00	-28.00	21.00	4.712
Frontal pole	l	210	-14.00	39.00	53.00	4.708
Lateral occipital cortex	l	521	-50.00	-66.00	46.00	4.689
Post. Cingulate gyrus	r	412	4.00	-35.00	38.00	4.653
Amygdala	l	116	-22.00	-3.00	-19.00	4.606
Frontal orbital cortex	l	111	-23.00	29.00	-16.00	4.589
Superior parietal lobule	r	549	34.00	-43.00	60.00	4.568
Anterior cingulate cortex	l	221	-4.00	3.00	40.00	4.531
Precentral gyrus	r	144	19.00	-27.00	61.00	4.497
Central opercular cortex	r	131	37.00	2.00	13.00	4.403
Lateral occipital cortex	r	100	28.00	-87.00	36.00	4.352
Precentral gyrus	r	116	28.00	-12.00	55.00	4.341
Supramarginal gyrus	l	415	-60.00	-32.00	46.00	4.284
Prim. somatosensory cortex	r	233	44.00	-24.00	41.00	4.244
Superior frontal gyrus	r	183	24.00	27.00	59.00	4.236
Lateral occipital cortex	l	249	-58.00	-66.00	1.00	4.205
Lateral occipital cortex	r	114	25.00	-73.00	47.00	4.156
Lateral occipital cortex	r	131	45.00	-69.00	-5.00	4.065
Occipital pole	r	112	14.00	-97.00	-5.00	4.062
Middle temporal gyrus	l	153	-63.00	-50.00	-4.00	4.018
Precentral gyrus	r	117	35.00	-5.00	59.00	3.915

Note. Table limited to clusters with a minimum volume of 100mm³.

7.3.2.2 ROI-wise GLMs

We conducted two sets of ROI analyses in order to investigate the involvement of subcortical nuclei in RPE processing. For this, we first investigated the subcortex masks provided MASSP (Bazin et al., 2020) that provide a high degree of individual anatomical detail because they were parcellated on the individual brain. Subsequently, we investigated subregions particularly associated with the dopaminergic system in the subcortex by means of exploring the evidence for BOLD responses in subregions of the striatum, SN, and VTA employing the probabilistic atlas by Pauli et al. (2018).

In line with the whole-brain GLM results and previous work (Garrison et al., 2013; Miletić, 2023, p. 170), significant bilateral activation of the striatum was found to covary with the size of the reward prediction error in our ROI-wise GLMs using the individual parcellations derived from MASSP (Figure 7.5, Table 7.2). Contrary to previous work (Garrison et al., 2013; Miletić, 2023, p. 170), ROI-wise GLMs indicated no evidence for the involvement of other subcortical nuclei in reward prediction error processing except for lateralised activation in the GPe (Figure 7.5). Despite the robust association between the dopaminergic neurons of the midbrain and reward prediction error signals, our data indicated no activation in the ROIs of the midbrain (SN and VTA).

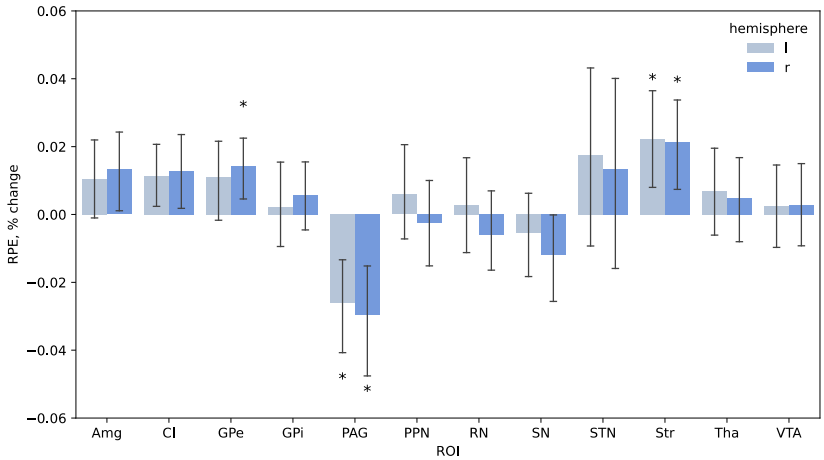

Figure 7.5. Results of the region-of-interest-wise GLMs. Error bars represent the 95% confidence intervals. * Indicate $p < 0.05$.

The probabilistic atlas by Pauli and colleagues (2018) provides subregions of both the striatum (caudate, putamen, nucleus accumbens) and the dopaminergic midbrain (SNc, VTAnc, PBP), which were subsequently analysed in order to explore differences in the localisation of subcortical RPE-related BOLD signal change in more detail. Consistent with the whole-brain analysis and previous work (Corlett et al., 2022; Lefebvre et al., 2017; Jackson et al., 2018; Morris et al., 2016), bilateral BOLD signal change was observed in both the ventral (nucleus accumbens) and the dorsal striatum (putamen) (Figure 7.6). In accordance with previous work (Miletić 2023, p. 170) and matching the findings from both, the whole-brain GLMs and the first set of ROI-wise GLMs, there was no evidence for RPE-related activation in the subdivisions of the dopaminergic midbrain (Figure 7.6, Table 7.3).

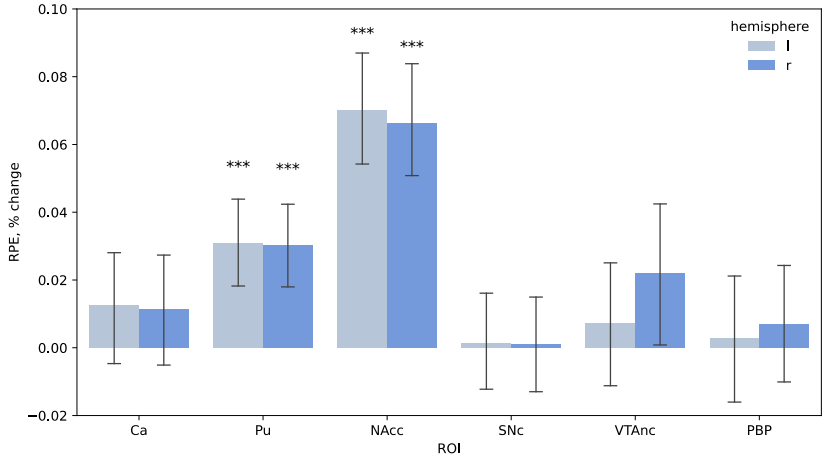

Figure 7.6. Results of the region-of-interest-wise GLMs using the masks provided by Pauli et al., 2018. Error bars represent the 95% confidence intervals. *** indicate $p < 0.001$.

Table 7.2. ROI analysis' mean signal change in the reward prediction error contrast.

ROI	hem	*t*	*p*
Tha	l	1.003	0.527
	r	0.761	0.571
Str	l	3.124	0.020 *
	r	3.063	0.020 *
GPe	l	1.782	0.200
	r	3.133	0.020 *
GPi	l	0.340	0.736
	r	1.121	0.498
STN	l	1.359	0.366
	r	0.944	0.527
Amg	l	1.729	0.202
	r	2.177	0.109
CL	l	2.377	0.092
	r	2.232	0.109
SN	l	-0.846	0.566
	r	-1.877	0.184
VTA	l	0.385	0.736
	r	0.461	0.736
RN	l	0.362	0.736
	r	-0.948	0.527
PPN	l	0.808	0.566
	r	-0.372	0.736
PAG	l	-3.671	0.013 *
	r	-3.550	0.013 *

Note. *p*-value corrected for the false discovery rate using the Benjamini-Hochberg procedure (FDR; $q < 0.05$).

Table 7.3. ROI analysis' mean signal change in the reward prediction error contrast using masks from Pauli et al. (2018).

ROI	hem	*t*	*p*
Ca	l	1.472	0.301
	r	1.382	0.302
Pu	l	4.574	0.000*****
	r	4.789	0.000*****
NAcc	l	8.206	0.000*****
	r	7.746	0.000*****
SNc	l	0.185	0.875
	r	0.158	0.875
VTAnc	l	0.799	0.628
	r	2.019	0.124
PBP	l	0.302	0.875
	r	0.729	0.628

Note. p-value corrected for the false discovery rate using the Benjamini-Hochberg procedure (FDR; $q < 0.05$).

7.4 Discussion

In this model-based fMRI study, we utilised a reversal learning paradigm in combination with 7T fMRI, and a scanning protocol optimised for the subcortex. The aim was to investigate the neural substrates of the reward prediction error (RPE) modelled with the RL-ARD model (Miletić et al., 2021) with a specific focus on the subcortex, particularly the midbrain. Therefore, only subcortical regions-of-interest (ROIs) were included in the ROI analysis that followed the analysis on a whole-brain level.

Consistent with previous work (D'Ardenne et al., 2008; Corlett et al., 2022; Hare et al., 2008; Lefebvre et al., 2017; Miletić 2023; O'Doherty et al., 2003; Preuschoff et al., 2008; Rohe et al., 2012; Tobler et al., 2006), our findings strongly implicate the nucleus accumbens of the ventral striatum in the RPE contrast. We also found activation in the dorsal striatum, particularly the putamen, covarying with the size of the RPE in line with the reversal learning literature (Jackson et al., 2018; Morris et al., 2016). Both whole-brain GLM and ROI-wise GLM demonstrated these striatal increases in activation. Enhanced activity in other subcortical nuclei was only found for the left amygdala and the hippocampus and was limited to whole-brain GLM. ROI-wise GLMs also implicated the left GPe in the RPE contrast, but this finding was not corroborated by the whole-brain GLM. Despite expectations, neither results from the whole-brain GLM nor the ROI-wise GLM showed increased activations in regions of the midbrain in RPE processing. Furthermore, a

particularly large activation cluster was found in the occipital pole. Other relatively large cortical clusters of activation were found in areas of the frontal pole, mainly in the left hemisphere, the (anterior) cingulate, and the insular cortex.

No evidence of RPE activation in the midbrain

Although there is strong evidence that activations in the nucleus accumbens and putamen, located in the ventral and dorsal striatum, respectively, are related to the size of the RPE, our data does not indicate any connection between the midbrain ROIs VTA (incl. VTAnc and PBP) and SN (and SNc), regions that all contain dopaminergic neurons, and the RPE. Since the nucleus accumbens is innervated by the dopaminergic neurons that make up the mesolimbic pathway associated with the VTA, and the dorsal striatum (including the putamen) is connected via the nigrostriatal pathway with the SN dopamine neurons (Fu et al., 2016; Haber, 2003; Halliday et al., 2012), it is very probable that RPE-related activations of the midbrain dopamine neurons drive the activity found in the striatum. The significantly stronger activations of the ventral striatum, particularly the nucleus accumbens, compared to the dorsal striatum, suggest that predominantly mesolimbic neurons of the VTA are activated in response to the reward feedback. However, the projections from the midbrain to the striatum are not as clearly arranged as traditional ideas of the dopaminergic pathways suggest: the ventral striatum receives input from dopamine neurons of the dorsal tier of the midbrain, which extend from the VTA to the dorsal parts of the SNc (Haber, 2003; Haber & Knutson, 2010). As a consequence, neither VTA nor SN cell groups can be exclusively linked to the RPE encoding, and our results do not provide any more detail into the spatial organisation of the midbrain, rendering it unclear where the neural populations that encode the RPE are precisely located in the human midbrain.

Given the strong evidence from animal studies (Bayer & Glimcher, 2005; Enomoto et al., 2011; Mirenowicz & Schultz, 1994; Montague et al., 1996; Schultz et al., 1993, 1997) for the RPE encoding of midbrain dopamine neurons and several human 3T fMRI studies (D'Ardenne et al., 2008; Limbrick-Oldfield et al., 2019; Richter et al., 2020; Rohe et al., 2012; Zhang et al., 2017) for reward-related responses in midbrain neurons, it is puzzling that the current study found no evidence of RPE-related processes in the midbrain, especially since we were using high-resolution 7T fMRI in combination with a subcortex-focussed scanning protocol and individually parcellated masks. These techniques are expected to provide a higher level of detail.

One explanation for the lack of detection of increases in BOLD-signal in the midbrain ROIs is the definition of the masks serving as ROIs in this study, which may not be well-defined enough to detect the activity of the RPE-encoding neurons. Suppose only a few voxels of the individual masks cover the RPE-encoding neurons while others encompass GABA-ergic or glutamatergic neurons that are inactive in the RPE. In that case, the BOLD-signal change might be cancelled out when averaging across all voxels in the applied ROI analysis. The midbrain contains a large diversity of neural populations of which a subset of dopaminergic neurons responds to reward-related events and encodes the RPE. In light of such ROI analyses, dopamine neuron-specific ROIs would be optimal to directly and accurately test hypotheses of reward-based functioning in the human brain. However, dopaminergic neurons are difficult to localise in the living human brain even with high-resolution MRI, rendering the use of tools that help navigate and localise brain structures, such as probabilistic atlas maps, necessary. Probabilistic atlases of brain structures are preferably based on a large number of participants in order to increase the anatomical validity of a map on a group level; however, at the same time, they do not take into account the individual anatomy, which can markedly affect the accurate mapping of smaller brain structures. Therefore, parcellations of individual brains are a preferred tool to study, for instance, variations in BOLD-signal for a specific ROI, as they provide more individual anatomical detail. However, today, there is no algorithm for the automatic parcellation of individual dopaminergic neural populations in the midbrain available. The multi-contrast anatomical subcortical structure parcellation (MASSP) algorithm (Bazin et al., 2020) employed in this study differentiates between the individual SN and VTA. However, the MASSP maps of both structures do not claim to include the dopaminergic populations of each structure exclusively. Given technical limitations that limit identifying individual hubs of specifically dopamine-producing cells in the first place, but also considering the size of these neural populations, even with high-resolution imaging, it is not technically feasible to accurately delineate dopamine-cell groups in the individual brain. It is important to note that this, of course, also holds for probabilistic atlases like the one used in this study (Pauli et al., 2018), which, despite further subdividing SN and VTA, does not claim to be entirely restricted to dopamine neurons either.

Limitations

It is possible that our study did not provide enough statistical power to detect RPE-related BOLD signal change in the midbrain, as the number of trials the

participants performed was relatively low. Hence, it is possible that including more trials would have increased the likelihood of detecting true effects. However, another study from the same data collection effort (which included the same participants, the same high-resolution 7T fMRI scanning protocol customised for the subcortex, and the same individual parcellations MASSP) also used a reinforcement learning paradigm, excluding reversal but including speed versus accuracy cues, and similarly failed to provide evidence for SN or VTA activation in the RPE contrast, despite including more trials (Miletić, 2023).

Along these lines, it remains possible that even our optimised scanning protocol did not provide enough temporal signal-to-noise ratio (tSNR) to pick up on variations in BOLD-signal changes in regions near ventricle and vascular systems and large distance to the receiver coils (Barry et al., 2013; de Hollander et al., 2017; Trutti et al., 2019). After all, compared to the striatum ROIs, smaller and more ventrally located ROIs, like the midbrain ROIs, exhibit relatively smaller tSNR estimates (see Appendix C, Figure C.1). This might partially explain that RPE-related BOLD signal change was limited to the ROIs of the striatum.

Furthermore, it is possible that the firing of dopamine neurons associated with the RPE does not produce a signal change strong enough to be detected by BOLD in general. Previous research on cortical BOLD signal change suggests that the BOLD signal is mainly influenced by the post-synaptic input processing to an area (local field potentials) (Goense & Logothesis, 2006). This would also explain why a strong BOLD response is recorded mainly in the ventral striatum since mesolimbic neurons from the midbrain excite medium spiny neurons (MSN) of the ventral striatum through dopamine release. MSN neurons express high levels of dopamine receptors that render them sensitive towards available dopamine and, this way, facilitate strong striatal responses to local increases in dopamine (Surmeier et al., 2007).

Conclusion

While the nucleus accumbens and putamen showed RPE-related activations in this model-based fMRI study using 7T fMRI and a subcortex-optimized scanning protocol, midbrain regions associated with dopamine neurons did not exhibit such changes. The absence of midbrain findings may be attributed to challenges in precisely defining midbrain regions, especially dopamine-specific populations. Current methods, including probabilistic atlases and individual parcellations, lack the accuracy to delineate dopamine-cell groups in the living human brain. Technical limitations, such as temporal signal-to-noise ratio issues and the influence of post-

synaptic input processing, may contribute to the difficulty in detecting midbrain BOLD-signal changes. Despite these limitations, our results align with another study using a similar protocol, suggesting challenges in detecting midbrain activations related to RPE. The intricate relationship between dopamine neurons and BOLD signals in the midbrain warrants further exploration and methodological advancements for a more nuanced understanding.