Cerebral Cortex, April 2020;30: 2005–2018

doi: 10.1093/cercor/bhz218
Advance Access Publication Date: 11 November 2019
Original Article

ORIGINAL ARTICLE

Learning in Visual Regions as Support for the Bias in Future Value-Driven Choice

Sara Jahfari[1,2], Jan Theeuwes[3] and Tomas Knapen[1,3]

[1]Spinoza Centre for Neuroimaging, Royal Netherlands Academy of Arts and Sciences (KNAW), Amsterdam, The Netherlands, [2]Department of Psychology, University of Amsterdam, 1018 WB Amsterdam, The Netherlands and [3]Department of Applied and Experimental Psychology, Vrije Universiteit, 1081 HV Amsterdam, The Netherlands

Address correspondence to Sara Jahfari, Spinoza Centre for Neuroimaging, Meibergdreef 75, 1105 BK Amsterdam, The Netherlands. Email: sara.jahfari@gmail.com

Abstract

Reinforcement learning can bias decision-making toward the option with the highest expected outcome. Cognitive learning theories associate this bias with the constant tracking of stimulus values and the evaluation of choice outcomes in the striatum and prefrontal cortex. Decisions however first require processing of sensory input, and to date, we know far less about the interplay between learning and perception. This functional magnetic resonance imaging study ($N = 43$) relates visual blood oxygen level–dependent (BOLD) responses to value beliefs during choice and signed prediction errors after outcomes. To understand these relationships, which co-occurred in the striatum, we sought relevance by evaluating the prediction of future value-based decisions in a separate transfer phase where learning was already established. We decoded choice outcomes with a 70% accuracy with a supervised machine learning algorithm that was given trial-by-trial BOLD from visual regions alongside more traditional motor, prefrontal, and striatal regions. Importantly, this decoding of future value-driven choice outcomes again highlighted an important role for visual activity. These results raise the intriguing possibility that the tracking of value in visual cortex is supportive for the striatal bias toward the more valued option in future choice.

Key words: Bayesian hierarchical modeling, decoding, perceptual learning, random forest machine learning, reinforcement learning

Introduction

In decision-making, our value beliefs bias future choices. This bias is shaped by the outcomes of similar decisions made in the past where the action, or stimulus chosen, becomes associated with a positive or negative outcome ("value beliefs"). The evaluation of value after an outcome, or the comparison of value in decisions, is traditionally associated with activity in the prefrontal cortex and striatum (O'Doherty et al. 2004, 2017; Daw et al. 2006; Kahnt et al. 2009; Hare et al. 2011; Jocham et al. 2011; Klein et al. 2017).

To underset the bias in action selection, midbrain dopamine neurons are thought to send a teaching signal toward the striatum and prefrontal cortex after an outcome (Montague et al. 1996; Schultz et al. 1997; Tobler et al. 2005). In the striatum, future actions are facilitated by bursts in dopamine after positive outcomes or discouraged by dopamine dips after negative outcomes. The dorsal and ventral parts of the striatum are known to receive differential, but also overlapping, inputs from midbrain neurons (O'Doherty et al. 2004; Atallah et al. 2007). Ventral and dorsal striatum have also been ascribed a differential role during learning by reinforcement learning

theories. Here, the ventral parts of the striatum are involved with the prediction of future outcomes through the processing of prediction errors, whereas the dorsal striatum uses the same information to maintain action values as a way to bias future actions toward the most favored option (Joel et al. 2002; Kahnt et al. 2009; Collins and Frank 2014). Intriguingly, however, before many of these value-based computations can take place, stimuli first have to be parsed from the natural world, an environment where most reward-predicting events are perceptually complex. This suggests that sensory processing might be an important integral part of optimized value-based decision-making.

Here, we investigate whether choice outcomes can modulate the early sensory processing of perceptually complex stimuli to help bias future decisions. Recent neurophysiological studies find visually responsive neurons in the tail of the caudate nucleus, which is part of the dorsal striatum (Kim and Hikosaka 2013; Hikosaka et al. 2014). These neurons encode and differentiate stable reward values of visual objects to facilitate eye movements toward the most valued target, while at the same time inhibiting a movement toward the lesser valued object (Kim et al. 2017). Critically, differential modulations are also observed in the primary visual cortex where stronger cortical responses are seen for objects with higher values (Serences 2008; Serences and Saproo 2010), which is consistent with the response of visual neurons in the caudate. As visual cortex is densely connected to the striatum (Fernandez-Ruiz et al. 2001; Kravitz et al. 2013), prioritized visual processing of high-value stimuli could aid the integration of information regarding the most-valued choice in the striatum (Lim et al. 2011, 2013; Jahfari et al. 2015; Jahfari and Theeuwes 2017). To understand these visual–striatal interactions, we focus on a more detailed parsing of the underlying computations.

Specifically, we explored two questions by reanalyzing functional magnetic resonance imaging (fMRI) data from a probabilistic reinforcement learning task using faces as visual stimuli (Jahfari et al. 2018; Fig. 1*a*). First, we focus on the interplay between learning and visual activity in the fusiform face area (FFA) and occipital cortex (OC). Here, with the use of a Bayesian hierarchical reinforcement learning model (Fig. 1*b*), we outline how trial-by-trial estimates of action values (Q-value) and reward prediction errors (RPEs) relate to the blood oxygen level–dependent (BOLD) response of visual regions and the striatum (O'Doherty et al. 2007; Daw 2011) (Fig. 1*c*). Second, we analyze data from a follow-up transfer phase, where the learning of value was already established. In our analysis, the importance of visual brain activity in the prediction, or decoding, of future value-based decisions is evaluated by using a supervised Random Forest (RF) machine learning algorithm (Breiman 2001, 2004). Specifically, transfer phase single-trial BOLD estimates from anatomically defined visual, prefrontal, and subcortical regions are combined by RF to predict, or decode, choice outcomes in a separate validation set. We focus on classification accuracy and the relative importance of each brain region in the correct classification of future value-based decisions.

Materials and Methods

To understand how value learning relates to the activity pattern in perceptual regions, we reanalyzed the behavioral and fMRI recordings of a recent study (Jahfari et al. 2018). In this study, BOLD signals were recorded while participants performed a reinforcement learning task using male or female faces and a stop-signal task (which was discussed in Jahfari et al. (2018)). The FFA was localized using a separate experimental run.

Participants

Forty-nine young adults (25 male; mean age = 22 years; range 19–29 years) participated in the study. All participants had normal or corrected-to-normal vision and provided written consent before the scanning session, in accordance with the Declaration of Helsinki. The ethics committee of the University of Amsterdam approved the experiment, and all procedures were in accordance with relevant laws and institutional guidelines. In total, six participants were excluded from all analyses due to movement (2), incomplete sessions (3), or misunderstanding of task instructions (1). In total, data from 43 participants were analyzed.

Reinforcement Learning Task

Full details of the reinforcement learning task are provided in Jahfari et al. (2018). In brief, the task consisted of two phases (Fig. 1*a*). In the first learning phase, three male or female face pairs (AB, CD, and EF) were presented in a random order, and participants learned to select the most optimal face (A, C, E) in each pair solely through probabilistic feedback ("correct": happy smiley, "incorrect": sad smiley). Choosing face A lead to "correct" on 80% of the trials, whereas a choice for face B only lead to the feedback "correct" for 20% of the trials. Other ratios for "correct" were 70:30 (CD) and 60:40 (EF). Participants were not informed about the complementary relationship in pairs. All trials started with a jitter interval where only a white fixation cross was presented and had a duration of 0, 500, 1000, or 1500 ms to obtain an interpolated temporal resolution of 500 ms. Two faces were then shown left and right of the fixation cross and remained on screen up to response or trial end (4000 ms). If a response was given on time, a white box surrounding the chosen face was then shown (300 ms) and followed (interval 0–450 ms) by feedback (500 ms). Omissions were followed by the text "miss" (2000 ms). The transfer phase contained the three face pairs from the learning phase and 12 novel combinations, in which participants had to select which item they thought had been more rewarding during learning. Transfer phase trials were identical to the learning phase, with the exception that no feedback was provided. All trials had a fixed duration of 4000 ms, where in addition to the jitter used at the beginning of each trial, null trials (4000 ms) were randomly interspersed across the learning (60 trials; 20%) and transfer (72 trials; 20%) phase. Each face was presented equally often on the left or right side, and choices were indicated with the right-hand index (left) or middle (right) finger. Before the magnetic resonance imaging (MRI) session, participants performed a complete learning phase to familiarize with the task (300 trials with different faces). In the MRI scanner, participants performed two learning blocks of 150 trials each (300 trials total; equal numbers of AB, CD, and EF) and three transfer phase blocks of 120 trials each (360 total; 24 presentations of each pair). All stimuli were presented on a black projection screen that was viewed via a mirror system attached to the MRI head coil.

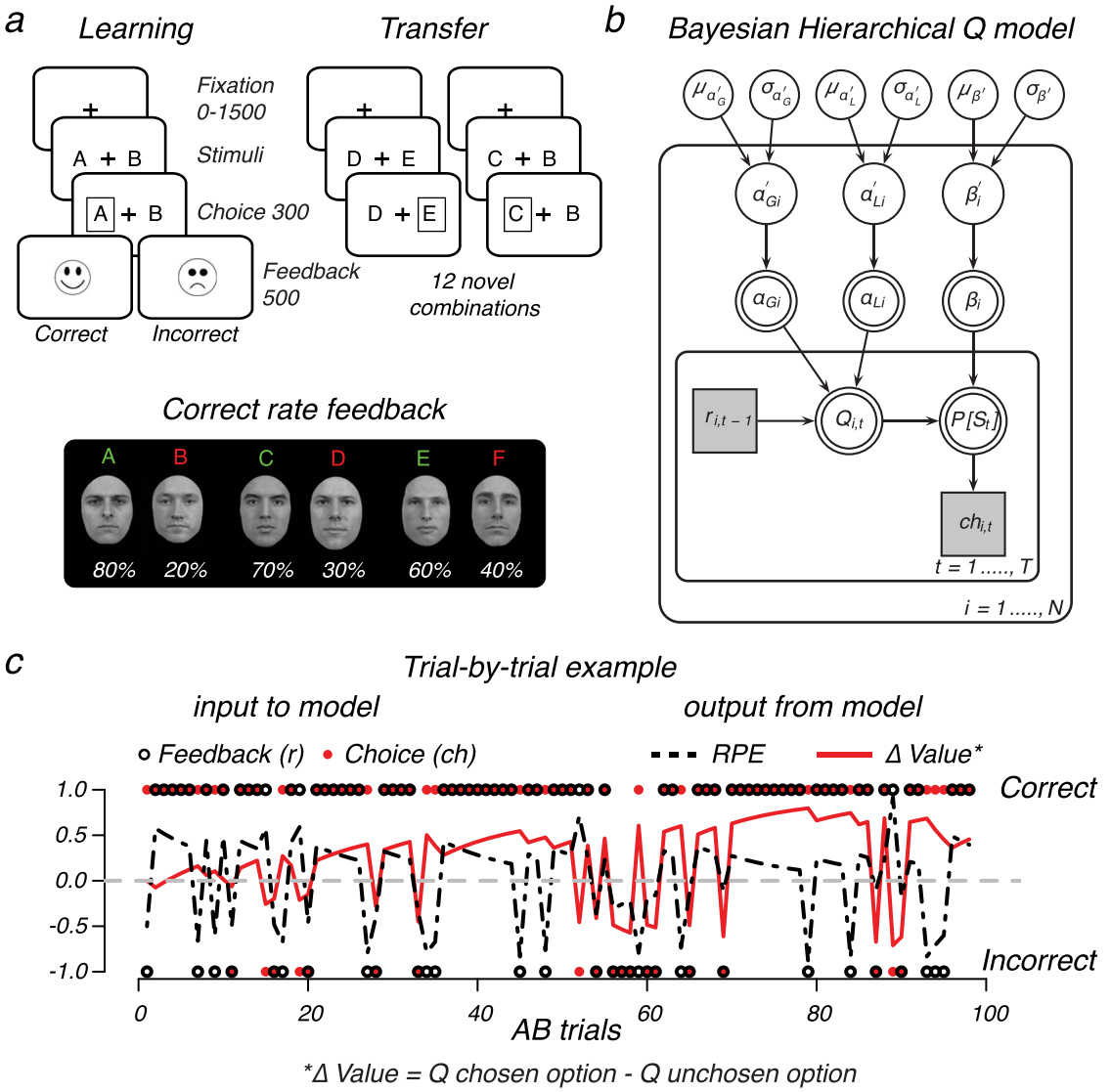

Figure 1. Design and Model. (*a*) Reinforcement learning task using faces. During learning, two faces were presented on each trial, and participants learned to select the optimal face identity (A, C, E) through probabilistic feedback (% of correct is shown beneath each stimulus). The learning phase contained three face pairs (AB, CD, and ED) for which feedback was given. In a follow-up transfer phase, these faces were rearranged into 12 novel combinations to assess learning. These trials were identical to learning trials, with the exception of feedback. (*b*) Graphical Q-learning model with hierarchical Bayesian parameter estimation. The model consists of an outer subject ($i = 1, \ldots, N$) and an inner trial plane ($t = 1, \ldots, T$). Nodes represent variables of interest. Arrows are used to indicate dependencies between variables. Double borders indicate deterministic variables. Continuous variables are denoted with circular nodes and discrete with square nodes. Observed variables are shaded in gray (see Materials and Methods for details about the fitting procedure). (*c*) Illustration of the observed trial-by-trial input (i.e., the choice made and feedback received) and output (i.e., Q for the chosen and unchosen stimulus, ΔValue, and RPE) of the model given the estimated variability in learning rates from either positive (α_{Gi}) or negative (α_{Li}) feedback and the tendency to exploit β_i higher values.

Reinforcement Learning Model

Trial-by-trial updating in value beliefs about the face selected in the learning phase and RPEs (signed expectancy violations) were estimated with a variant of the computational Q-learning algorithm (Watkins and Dayan 1992; Frank et al. 2007; Daw 2011) that is frequently used with this reinforcement learning task and contains two separate learning rate parameters for positive (α_{gain}) and negative (α_{loss}) RPEs (Frank et al. 2007; Kahnt et al. 2009; Niv et al. 2012; Jahfari and Theeuwes 2017; Jahfari et al. 2018). Q-learning assumes participants to maintain reward expectations for each of the six (A-to-F) stimuli presented during the learning phase. The expected value (Q) for selecting a stimulus *i* (could be A-to-F) upon the next presentation is then updated as follows:

$$Q_i(t+1) = Q_i(t) + \begin{cases} \alpha_{gain}\left[r_i(t) - Q_i(t)\right], & \text{if } r = 1 \\ \alpha_{loss}\left[r_i(t) - Q_i(t)\right], & \text{if } r = 0 \end{cases}$$

Where $0 \leq \alpha_{gain}$ or $\alpha_{loss} \leq 1$ represents learning rates, *t* is trial number, and $r = 1$ (positive feedback) or $r = 0$ (negative feedback). The probability of selecting one response over the

other (i.e., A over B) is computed as follows:

$$P_A(t) = \frac{\exp(\beta * Q_t(A))}{\exp(\beta * Q_t(B)) + \exp(\beta * Q_t(A))}$$

With $0 \leq \beta \leq 100$ known as the inverse temperature.

Bayesian Hierarchical Estimation Procedure

To fit this Q-learning algorithm with two learning rate parameters, we used Bayesian hierarchical estimation procedure. The full estimation procedure is explained by Jahfari et al. (2018). To summarize, this implementation assumes that probit-transformed model parameters for each participant are drawn from a group-level normal distribution characterized by group-level mean and standard deviation (SD) parameters: $z \sim N(\mu_z, \sigma_z)$. A normal prior was assigned to group-level means $\mu_z \sim N(0,1)$ and a uniform prior to the group-level SD $\sigma_z \sim U(1, 1.5)$. Model fits were implemented in Stan, where multiple chains were generated to ensure convergence.

Image Acquisition

The fMRI data for the Reinforcement learning task were acquired in a single scanning session with two learning and three transfer phase runs on a 3T scanner (Philips Achieva TX) using a 32-channel head coil. Each scanning run contained 340 functional T2*-weighted echo-planar images for the learning phase and 290 T2*-weighted echo-planar images for the transfer phase (repetition time [TR] = 2000 ms; echo time [TE] = 27.63 ms; flip angle [FA] = 76.1°; 3 mm slice thickness; 0.3 mm slice spacing; field of view [FOV] = 240 × 121.8 × 240; 80 × 80 matrix; 37 slices, ascending slice order). After a short break of 10 min with no scanning, data collection was continued with a three-dimensional T1 scan for registration purposes (TR = 8.5080 ms; TE = 3.95 ms; FA = 8°; 1 mm slice thickness; 0 mm slice spacing; FOV = 240 × 220 × 188) and the fMRI data collection using a stop signal task (described by Jahfari et al. (2018)) and a localizer task with faces, houses, objects, and scrambled scenes to identify FFA responsive regions on an individual level (317 T2*-weighted echo-planar images; TR = 1500 ms; TE = 27.6 ms; FA = 70°; 2.5 mm slice thickness; 0.25 mm slice spacing; FOV = 240 × 79.5 × 240; 96 × 96 matrix; 29 slices, ascending slice order). Here, participants viewed a series of houses, faces, objects as well as phase-scrambled scenes. To sustain attention during functional localization, subjects pressed a button when an image was directly repeated (12.5% likelihood).

fMRI Analysis Learning Phase

The interplay between learning and perceptual activity was examined by evaluating how trial-by-trial computations of value beliefs and RPEs relate to BOLD responses in the OC and FFA. To compare perceptual responses with the more traditional literature, we first show how value beliefs and RPEs relate to the activity pattern of the dorsal (i.e., caudate or putamen) or ventral (i.e., accumbens) parts of the striatum. Region of interest (ROI) templates were defined using anatomical atlases available in FMRIB Software Library (FSL) or the localizer task for FFA. For this purpose, the localizer scans were preprocessed using motion correction, slice-time correction, and prewhitening (Woolrich et al. 2001). For each subject, a general linear model (GLM) was fitted with the following explanatory variables (EVs): for FFA, faces > (houses and objects), for parahippocampal place area, houses > (faces and objects), and for lateral occipital complex, intact scenes > scrambled scenes. Higher level analysis was performed using FLAME Stage 1 and Stage 2 with automatic outlier detection (Beckmann et al. 2003). For the whole-brain analysis Z (Gaussianized T/F), statistic images were thresholded using clusters determined by $z > 2.3$ and $P < .05$ (GRFT) to define a group-level binary FFA region. Templates used for the caudate [center of gravity (cog): (−) 13, 10, 10], putamen [cog: (−) 25, 1, 1], and nucleus accumbens [cog: (−)19, 12, −7] were based on binary masks. Because participants were asked to differentiate faces, for each participant, we multiplied the binary templates of OC [cog: 1, −83, 5] and FFA [cog: 23, −48, −18] with the individual t-stats from the localizer task contrast faces > (houses and objects). All anatomical masks and the localizer group-level FFA mask can be downloaded from github (see acknowledgments).

Deconvolution Analysis Learning Phase

To more precisely examine the time course of activation in the striatal and perceptual regions, we performed finite impulse response (FIR) estimation on the BOLD signals. After motion correction, temporal filtering (third-order Savitzky–Golay filter with window of 120 s), and percent signal change conversion, data from each region were averaged across voxels while weighting voxels according to ROI probability masks and upsampled from 0.5 to 3 Hz. This allows the FIR fitting procedure to capitalize on the random timings (relative to TR onset) of the stimulus presentation and feedback events in the experiment. Separate response time courses were simultaneously estimated triggered on two separate events: stimulus onset and feedback onset. FIR time courses for all trial types were estimated simultaneously using a penalized (ridge) least-squares fit, as implemented in the FIRDeconvolution package (Knapen and Gee 2016), and the appropriate penalization parameter was estimated using cross-validation. For stimulus onset events (i.e., onset presentation of face pairs), response time courses were fit separately for the AB, CD, and EF pairs, while also estimating the time courses of signal covariation with chosen and unchosen value for these pairs. For these events, our analysis corrected for the duration of the decision process. For the feedback events, the co-variation response time course with signed and unsigned prediction errors was estimated. These signal response time courses were analyzed using across-subjects GLMs at each time-point using the statsmodels package (Seabold and Perktold 2010). The α value for the contributions of Q or RPE was set to 0.0125 (i.e., a Bonferroni corrected value of 0.05 given the interval of interest between 0 and 8 s).

RF Classification

To specify the relevance of perceptual regions in the resolve of future value-driven choices, a RF classifier was used (Breiman 2001, 2004). The RF classifier relies on an ensemble of decision trees as base learners, where the final prediction (e.g., for a given trial is the choice going to be correct/optimal? or incorrect/-suboptimal? given past learning) is obtained by a majority vote that combines the prediction of all decision trees. To achieve controlled variation, each decision tree is trained on a random subset of the variables (i.e., ROI chosen) and a bootstrapped sample of data points (i.e., trials or rows of the matrix in Fig. 2c). In the construction of each tree, about one-third of all trials is left out—termed as the "out-of-bag" sample—and later used to see how well each tree preforms on unseen data in the training

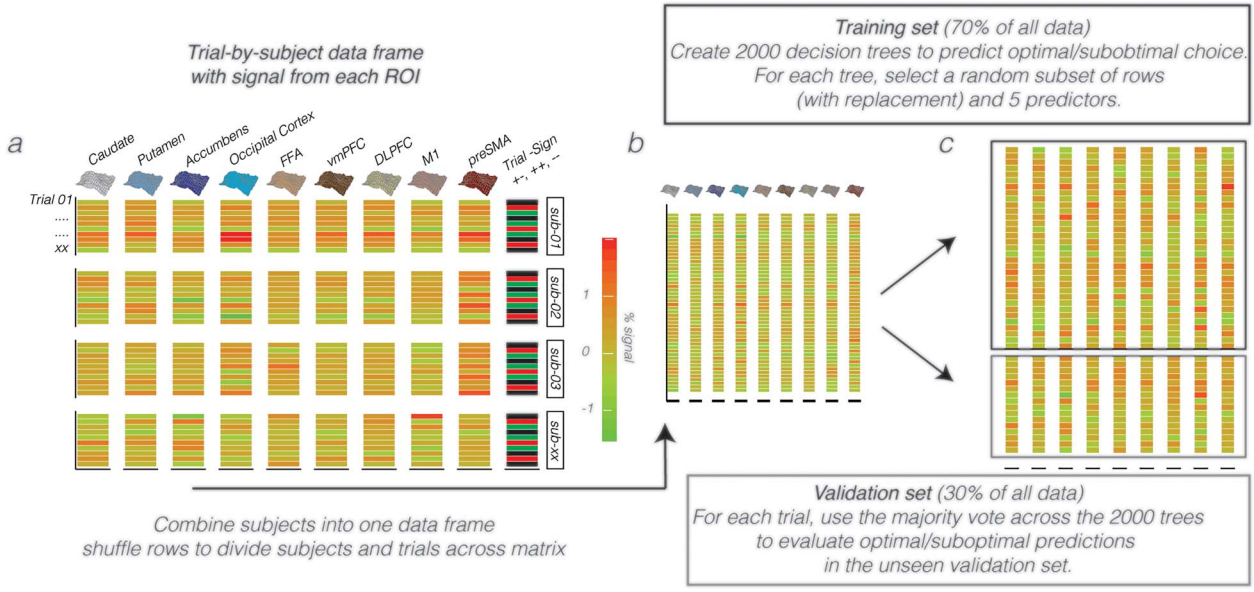

Figure 2. Random Forest input and data-structure. (*a*) Trial-by-subject data matrix with the % signal change drawn for each choice trial in the transfer phase (rows) from 9 apriori defined ROI (columns). In addition to the ROI data, the matrix contained a column with the identity of participants (sub-01, etc.) and Trial Sign, which specified a choice between two positives (+/+; AC, AE, CE), negatives (−/−, BD, BF, DF), or between a negative and positive option (+/−, e.g., AD, CF, etc.) given the feedback scheme in the learning phase. (*b*) The individual subject data frames were then combined into one matrix, in which the rows were subsequently shuffled to randomly distribute trials and subjects across the rows. (*c*) This matrix was then divided into a training set (two-thirds of the data) for the creation of 2000 decision trees of which the majority vote on each trial is then used to evaluate the predictive accuracy of optimal/suboptimal choices in a separate validation set (one-third of the data).

set. Because in RF each tree is built from a different sample of the original data, each observation is "out-of-bag" (OOB) for some of the trees. As such, each OOB sample is offered to all trees where the sample was not used for construction, and the average vote across those trees is taken as the classification outcome. The proportion of times that the classification outcome is not equal to the actual choice is averaged over all cases and represents the RF OOB error estimate. In other words, the generalized error for predictions is calculated by aggregating the prediction for every OOB sample across all trees. In the Results section, the OOB errors obtained from RF during training were well matched with the classification accuracy seen for the validation set given only the "good learners" (OOB = 30%, RF error validation set = 31%) or all participants (OOB = 33%, RF error validation set = 35%). An important feature of the RF classification method is the ease to measure the relative importance of each variable (i.e., region) in the overall predictive performance. That is, it allows for the ranking of all regions evaluated in the prediction of future value-based decisions.

ROI Selection and RF Procedure

This study used the "Breiman and Cutler's Random Forests for Classification and Regression" package in R, termed randomForest (randomForest_4.6–14). RF evaluations relied on the fMRI data recorded during the transfer phase, in a set of 9 ROIs. These ROIs included all templates from the learning phase (i.e., caudate, putamen, accumbens, OC, and FFA), as well as the ventromedial prefrontal cortex (vmPFC), dorsolateral prefrontal cortex (DLPFC), pre-supplementary motor area (preSMA), and the primary motor cortex (M1). The selection of these additional anatomical templates was inspired by our previous analysis of these data with those templates focusing on networks (Pircalabelu et al. 2015; Schmittmann et al. 2015; Jahfari et al. 2018). Specifically, the DLPFC template was obtained from an earlier study, linking especially the posterior part to action execution (Cieslik et al. 2012). The preSMA, vmPFC, and M1 mask were created from cortical atlases available in FSL. Please notice that we used the same anatomical ROIs for both the model-based deconvolution analysis (Figs 4 and 5) and the decoding analysis (Figs 2 and 6). From each ROI, a single parameter estimate (averaged normalized β estimate across voxels in each ROI) was obtained per trial, per subject. All preprocessing steps to obtain single-trial images are described by Jahfari et al. (2018). Single-trial activity estimates were used as input variables in RF to predict choice outcomes (optimal/suboptimal) in the transfer phase. Here, participants choose the best/optimal option based on values learned during the learning phase. We defined optimal choices as correct (i.e., when participants choose the option with the higher value) and suboptimal choices as incorrect. Misses were excluded from RF evaluations.

By design, the transfer phase contained 360 trials including 15 different pairs (12 novel), where each pair was presented 24 times with the higher value presented left in 12 of the 24 presentations and on the right for the other half. With so many subtle value differences across the options presented and only one BOLD estimate per trial/region, the prediction of future choices is underpowered (Fig. 2*a*). Therefore, assuming that all participants come from the same population, a fixed effects approach was taken for evaluations with RF. Here, the trial * region activity matrices for all participants were combined into one big data matrix (Fig. 2*b*) and subsequently shuffled across the rows, so that both participants and trials were rearranged in a random order across rows. Besides the single-trial BOLD estimates from the 9 ROIs, this shuffled matrix contained two additional columns, which specified subject_id (to which subject

does each trial belong) and Trial Sign—that is, is the choice between the two faces about two positive (+/+; AC, AE, CE), negative (−/−; BD, BF, DF), or a positive–negative (+/−; e.g., AD, CF etc.) associations given the task manipulation during learning. Subject_id was included to control for different BOLD fluctuations across participants, whereas Trial Sign was added because both BOLD and choice patterns differ across these options (please see Jahfari et al. (2018)). The shuffled fixed-effect matrix was divided into a separate training (two-thirds of whole matrix) and validation (one-third) set to be used for RF evaluations (Fig. 2c). Based on our previous connectivity work with these data (Jahfari et al. 2018), we were aware that many of our single-trial BOLD response were correlated across time, which potentially results from shared learning effects (Supplementary Fig. 4). With RF, the problem of correlated features is minimized for predictions with variable selection—that is, the random selection of a set of regions to use for each tree. With more variables selected, we get better splits in each tree but also highly correlated decision trees across the forest, which in essence diminishes the forest effect. To find the best balance, this study optimized the number of variables to select with a tuning function using the OOB error estimate. Learning was based on the training set, using 2000 trees with the number of variables (regions) used by each tree optimized with the tuneRF function in R, and accordingly set to 5. For the construction of each tree, about one-third of all trials is left out—termed as the OOB sample—and later used to see how well each tree preforms on unseen data. The generalized error for predictions is calculated by aggregating the prediction for every OOB sample across all trees. Besides this OOB approximation, we evaluated the predictive accuracy of the whole RF on the separate unseen validation set. Additionally, we reasoned that RF predictions can be driven by alternative BOLD fluctuations related to e.g., the buildup of a motor response, the ease of face distinctions, or other to us unknown functional fluctuations. Therefore, prior to the evaluation of region importance (or ranking), we preformed two control analysis ensuring that RF predictions are sensitive to the consistency of past learning and the representation of ΔValue. These are the evaluations comparing "good" with "all" learners, as well as the relationship between ΔValue and RF uncertainty. In addition, while potential confounds of colinearity on the RF ranking cannot be excluded, we tried to minimize this with the use of permutation importance. Here, by using the OOB samples, the importance of each variable (region) is computed as the difference between the model's baseline accuracy and the drop in overall accuracy caused by permuting that column (region). While being slower, permutation importance is described as more robust in comparison with the default (gini) importance computation where only the uncertainty of predictions is evaluated (with no checks on accuracy fluctuations after region permutation). The single-trial data used as input, the RF evaluation codes, and ROI templates can all be downloaded from the github link provided in acknowledgments.

Results

Model and Behavior

As shown in Figure 1a, in the reinforcement learning task, participants learned to select among choices with different probabilities of reinforcement (i.e., AB 80:20, CD 70:30, and EF 60:40). A subsequent transfer phase, where feedback was omitted, required participants to select the optimal option among novel pair combinations of the faces that were used during the learning phase (Fig. 1a). In the learning phase, subjects reliably learned to choose the most optimal face option in all pairs. For each pair, the probability of choosing the better option was above chance ($P < 0.001$), and the effect of learning decreased from AB (80:20) and CD (70:30) to the most uncertain EF (60:40) pair, $F\left(2, 84\right) = 13.74, P < 0.0001$. At the end of learning, value beliefs differentiating the optimal (A, C, E) from the suboptimal (B, D, F) action were very distinct for the AB and CD face pairs but decreased with uncertainty, $F\left(2, 84\right) = 39.70, P < 0.0001$ (Fig. 3a). Value beliefs were estimated using the individual subject parameters of the Q-learning model that best captured the observed data (Fig. 3b–e, reproduced from Jahfari et al. (2018) to show performance).

BOLD Is Modulated by Reliable Value Differences between Faces in Striatal and Visual Regions

For each pair of faces presented during the learning phase (AB, CD, and EF), we asked how the BOLD signal time course in striatal and visual regions relates to trial-by-trial value beliefs about the two faces presented as a choice. First, as a reference, we focused on the activity pattern of three striatal regions. Results showed BOLD responses in dorsal (caudate, putamen) but not ventral (accumbens) striatum to be differentially modulated by the estimated value beliefs of the chosen face (Q_{chosen}) in comparison with value beliefs about the face that was not chosen ($Q_{unchosen}$). Thus, BOLD responses in the dorsal striatum were modulated more strongly by value beliefs about the chosen stimulus (Q_{chosen}; Fig. 4a bottom row). Critically, this differential modulation was only observed with the presentation of AB faces where value differences were most distinct because of the reliable feedback scheme. Next, we evaluated the relationship between value and BOLD in the FFA and OC. Again, only with the presentation of the AB face option, trial-by-trial BOLD fluctuations were differentially modulated by values of the chosen versus unchosen face option (Fig. 4b, bottom row). These evaluations highlight how the BOLD response in striatal and perceptual regions is especially sensitive to values of the (to-be) chosen stimulus when belief representations are stable and distinct.

RPEs in Striatal and Visual Regions

Our findings so far described relationships between BOLD and value time-locked to the moment of stimulus presentation—that is, when a choice is requested. Learning occurs when an outcome is different from what was expected. We therefore next focused on modulations of the BOLD response when participants received feedback. Learning modulations were explored by asking how trial-by-trial BOLD responses in perceptual and striatal regions relate to either signed (outcome was better or worse than expected) or unsigned (magnitude of expected violation) RPEs (Fouragnan et al. 2018). Consistent with the literature, BOLD responses in all striatal regions were modulated by signed RPEs, with larger responses after positive RPEs or smaller responses after negative RPEs (Fig. 5a, bottom row). Activity in the accumbens (ventral striatum) was additionally tied to unsigned RPEs in the tail of the BOLD time course, with larger violations (either positive or negative) tied to smaller dips. Consistently, estimated BOLD responses in both visual regions were modulated by the signed RPE and once more mirrored the striatal modulations with stronger positive

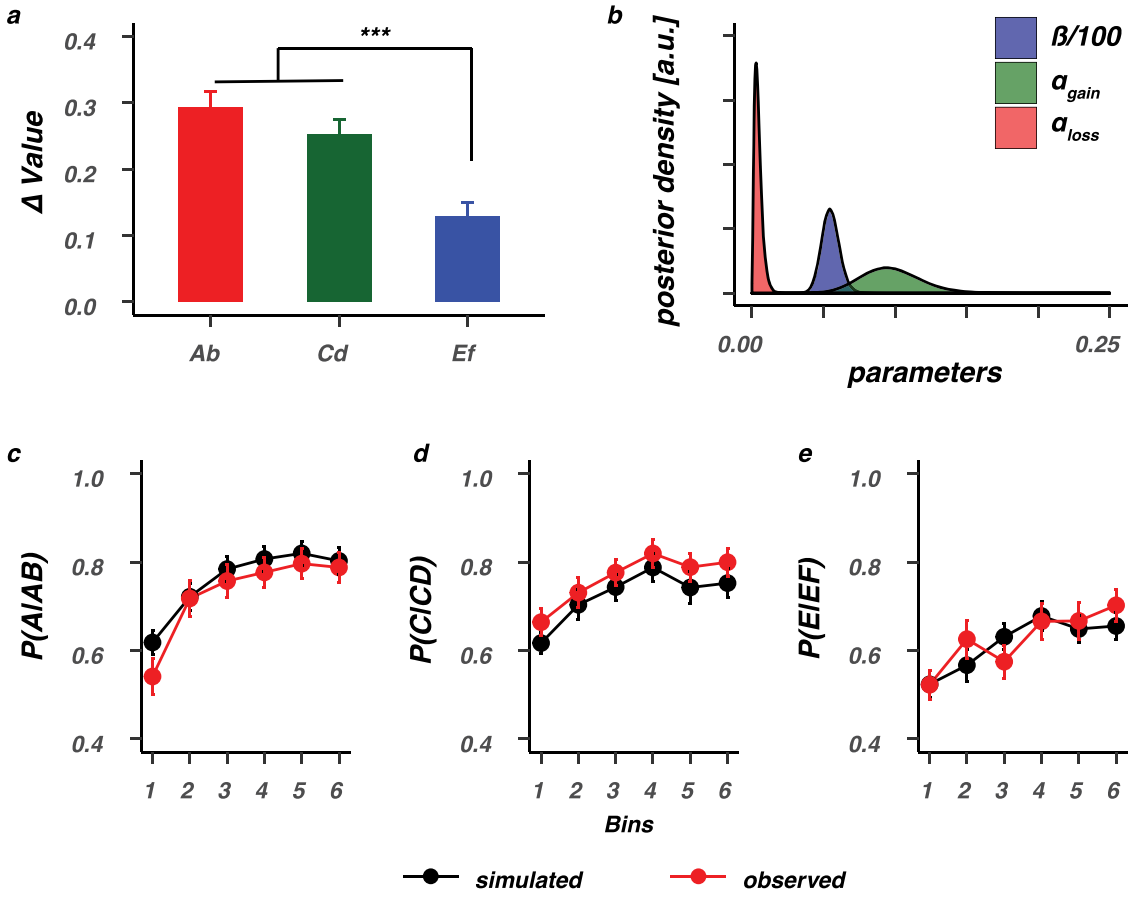

Figure 3. Value differentiation and model performance. (*a*) Value differentiation (ΔValue) for the selection of the optimal (A, C, E) stimuli over the suboptimal (B, D, F) stimuli decreased as a function of feedback reliability and was smallest for the most uncertain EF stimuli. ***$P < 0.0001$, Bonferroni corrected. (*b*) Group-level posteriors for all Q-learning parameters. The bottom row shows model performance, where data were simulated with the estimated individual subject parameters and evaluated against the observed data for the AB (*c*), CD (*d*), or EF (*e*) pairs. Bins contain ±16 trials. Error bars represent standard error of the mean.

RPEs eliciting stronger BOLD responses (Fig. 5*b* bottom row). FFA BOLD responses were additionally modulated by unsigned RPEs. However, in contrast to the relationship found between unsigned RPEs and the accumbens, the FFA modulation was positive and co-occurred with the modulation of the signed RPE. That is, bigger violations and more positive outcomes each elicited a stronger response in the FFA.

Can Past Learning in Visual Regions Support the Prediction of Future Value-Based Decisions?

Stable value representations and RPEs both modulated the activity of visual and striatal regions. These modulations in the striatum are described to bias future actions toward the most favored option (the dorsal striatum) or to predict future reward outcomes (the ventral striatum). To better understand the value and RPE modulations observed in visual regions, we next assessed the importance of these visual regions alongside the striatum in the correct classification (decoding) of future value-driven choice outcomes. Here, activity of prefrontal regions was added to the importance evaluation based on our previous work with these data in the transfer phase (Jahfari et al. 2018) (please see Supplementary Figs 1 and 2 for the evaluation of these regions during learning).

In the transfer phase, participants had to make a value-driven choice based on what was learned before, that is, during the learning phase. To specify the relevance of visual regions in the resolve of value-driven choice outcomes, in the transfer phase, a RF classifier was used (Breiman 2001, 2004; please see Fig. 2*a*–*c* for the procedure). The RF classifier was trained to predict the participant's choice, on each trial, given trial-by-trial BOLD estimates from striatal, prefrontal, and visual regions. The RF classifier relies on an ensemble of decision trees as base learners, where the prediction of each trial outcome is obtained by a majority vote that combines the prediction of all decision trees (Fig. 6*a*). To achieve controlled variation, each decision tree is trained on a random subset of the variables (i.e., subset of columns shown in Fig. 2*a*) and a bootstrapped sample of data points (i.e., trials). Importantly, we ensured that the forest was not simply learning the proportion of optimal choices in the transfer phase by training all models on balanced draws from the training set with equal numbers of optimal and suboptimal choices.

Evaluation of all participants resulted in a classification accuracy of 65% Area Under the Curve (AUC = 0.75) using the trial-by-trial BOLD estimates from the ROIs and increased to 70% with the evaluation of the good learners (AUC = 0.76; $N = 34$, criteria: accuracy > 60% across all three learning pairs). Hence, in 65 (all participants) or 70 (good learners) out of 100 trials, the forest correctly classified whether participants would pick the option with the highest value (optimal choice) or not (suboptimal choice) in the

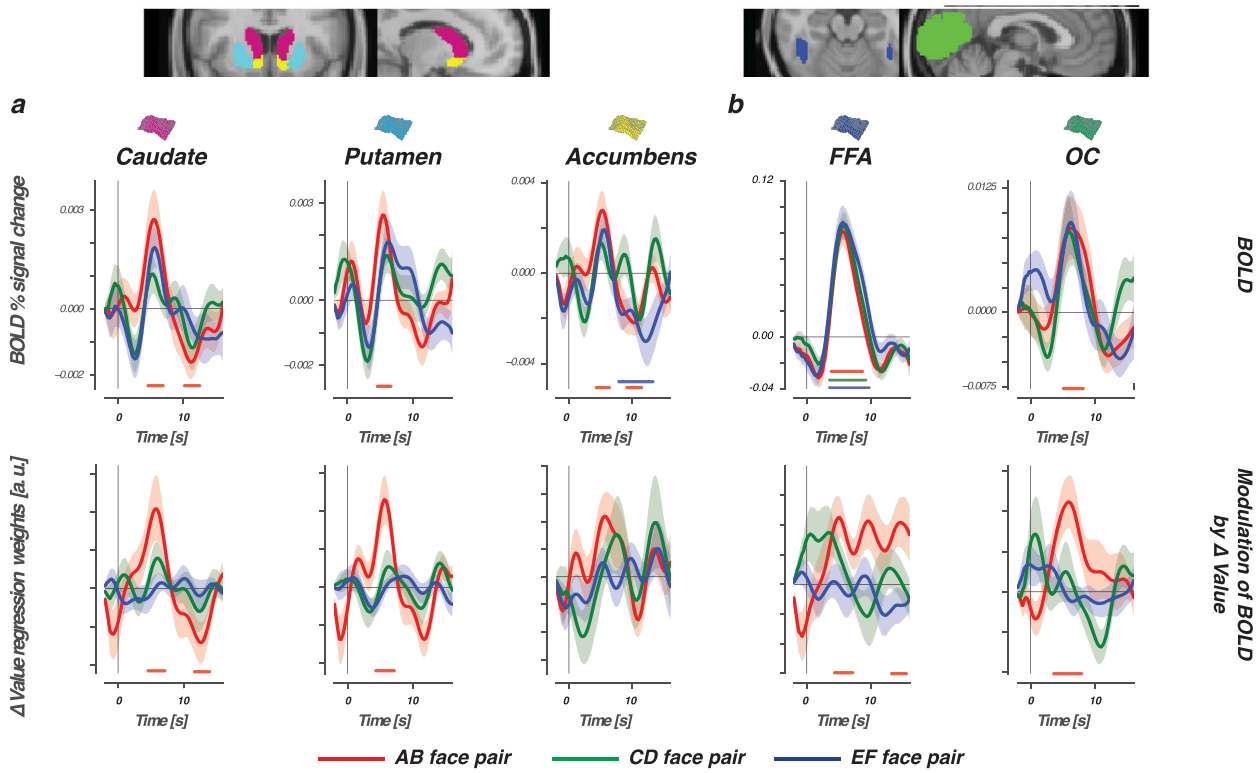

Figure 4. BOLD and the modulation of ΔValue in the learning phase. Top row shows the BOLD signal time course, time-locked to presentations of AB (80:20, red lines), CD (70:30, green lines), and EF (60:40, blue lines) face pairs, for three striatal regions (*a*) and two perceptual regions (*b*). Bottom row displays differential modulation by value (ΔValue = modulation Q_{chosen}—modulation $Q_{unchosen}$). Horizontal lines show the interval in which modulation was significantly stronger for Q_{chosen}. With the presentation of AB faces, BOLD responses in the dorsal striatum (caudate and putamen) and visual regions (FFA and OC) were modulated more by values of the chosen stimulus when compared to values of the unchosen stimulus. Differential AB value modulation was not significant in the ventral striatum (i.e., accumbens). Nor did we observe any differential value modulations with the presentation of the more uncertain CD and EF pairs. Confidence intervals were estimated using bootstrap analysis across participants ($n = 1000$), where the shaded region represents the standard error of the mean across participants (bootstrapped 68% confidence interval).

validation set. RF predictions were substantially lower when labels of the validation set were randomly shuffled (accuracy: all participants = 52%; good learners = 56%).

The improvement of accuracy with the evaluation of only the good learners is remarkable because the classifier was given less data to learn the correct labeling (fewer subjects/trials) and implied that the 2000 decision trees were picking up information related to the consistency of past learning. Further support for this important observation was found by asking how the uncertainty of each prediction (defined as the proportion of agreement in the predicted outcome among the 2000 trees for each trial) relates to the difference in value beliefs (ΔValue) about the two options presented on each trial (computed using the end $Q_{beliefs}$ of participants at the end of learning about face A-to-F; Fig. 6c, left side). As plotted in Figure 6c on the right side, the uncertainty in predicting that a trial choice outcome is optimal—defined as the proportion of disagreement among the 2000 decision trees—decreased with larger belief differences in the assigned values (please see Supplementary Fig. 3 for the evaluation of all participants).

Besides providing insights into how BOLD responses in the transfer phase contribute to predict value-driven choice outcomes (i.e., whether participants would choose the option with the highest value given past learning), the RF algorithm additionally outputs a hierarchy, thereby ranking the contribution of each region in the achieved classification accuracy. Figure 6d shows the ranking of all ROIs for good learners where the model had the highest predictive accuracy. First, regions in the dorsal striatum were most important, which aligned well with both the literature and the BOLD modulations we found by ΔValue and RPE during the learning phase. These regions were next followed by the preSMA. Evaluation of this region during the learning phase showed no modulations by ΔValue or RPE on BOLD (Supplementary Figs 1 and 2). Nevertheless, this region is typically associated with choice difficulty/conflict and might be essential in the resolve of a choice when value differences are small. Remarkably, the third region in this hierarchy was the FFA. In a task where participants pick the most valued face based on past learning, this ranking of the FFA just above the vmPFC implies that the ΔValue and RPE modulations of BOLD observed during learning could function to strengthen the recognition of valuable features. With the evaluation of all participants—including some who were less good in learning—the ranking of both the FFA and vmPFC was much lower (please see Supplementary Fig. 3b), which might be caused by more noise across the group in learning.

Further insights into the role of perceptual regions came from the separate evaluation of RF for only the easiest (with ΔValue between the two choice options being large) or hardest (with small ΔValue) choices (Supplementary Fig. 6). Results showed that when ΔValue is large, that is, the choice is easy, RF predictions are best served by BOLD fluctuations in both dorsal

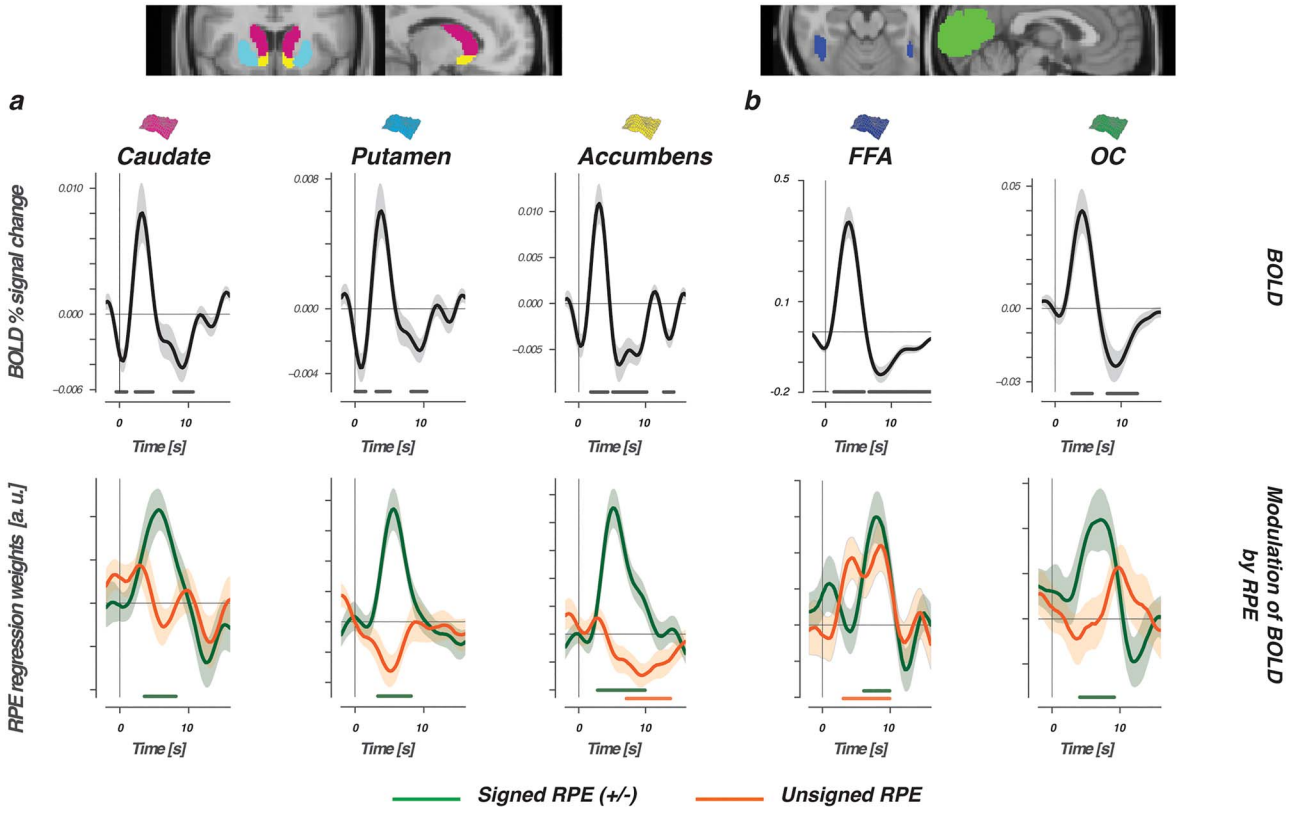

Figure 5. RPEs modulate BOLD in striatal and visual regions. The top row shows the FIR-estimated BOLD signal time course, which was time-locked to the presentation of choice feedback and evaluated for three striatal regions (*a*) and two perceptual regions (*b*). Bottom row displays modulations of the estimated BOLD time course by signed (green lines) or unsigned (orange lines) RPEs. The horizontal lines represent the interval in which signed or unsigned RPEs contributed significantly to the modulation of BOLD in the multiple regression. Note that both variables were always evaluated simultaneously in one GLM.

and ventral striatum, followed by vmPFC, the preSMA, and M1. With easy choices, regions involved with evidence accumulation (DLPFC) or perceptual processing (FFA and OC) rank last. More specifically, the processing of BOLD from OC even has a negative effect on RF accuracy, which means that running RF without OC will improve decoding. At the same time, with the evaluation of the most difficult choices—where participants decide between two very close in value positive (e.g., A or C) or negative (e.g., B or D) faces—we instead find perceptual regions to rank in the top. With difficult choices, where ΔValue is very small, the caudate is followed by the FFA and OC in serving RF predictions. We will return to the interpretation of these different rankings in the discussion.

Finally, we focused on two sets of control analysis. First, we evaluated RF accuracy and ranking with an additional random variable that was sampled from $\mathcal{N}(0,1)$, and unrelated to the BOLD activity of any region, or ΔValue. Here, the added random control region ranks last with negative importance, meaning that removing it improves model performance with 0.5% (good learners) or 0.3% (all learners) points (right side Fig. 6*d*, or Supplementary Fig. 3). Second, RF performance was evaluated with the removal of perceptual, striatal, or frontal regions. Despite the positive ranking of each region shown in Figure 6*d* (or Supplementary Fig. 3b), RF decoding was not affected by the removal of just one or two regions (Supplementary Fig. 5). However, accuracy is reduced when striatal (putamen, caudate, and accumbens), frontal (vmPFC, M1, DLPFC, and preSMA), or perceptual (FFA and OC) regions are evaluated in isolation. These alternative evaluations show not only that RF works best when trial-by-trial BOLD across multiple "learning" brain regions is combined but also that neither of the regions in isolation is crucial for the accuracy of predictions. Moreover, these control checks highlight that when a variable is unrelated to learning, or single-trial BOLD, ranking drops to last (as is to be expected) with counterproductive effects on RF accuracy.

Discussion

This study provides novel insights into how reinforcements modulate visual activity and specifies its potential in the prediction of future value-driven choice outcomes. First, by focusing on how participants learn, we find BOLD in visual regions to change with trial-by-trial adaptations in value beliefs about the faces presented and then to be subsequently scaled by the signed RPE after feedback. Next, the relevance of these observed value and feedback modulations was sought by exploring the prediction of future value-driven choice outcomes in a follow-up transfer phase where feedback was omitted. Our machine learning algorithm here shows a classification accuracy of 70% for participants who were efficient in learning by combining trial-by-trial BOLD estimates from perceptual, striatal, and prefrontal regions. The evaluation of region importance in these predictions ranked the FFA just after the dorsal striatum and the preSMA, thereby showing an important role for visual regions in

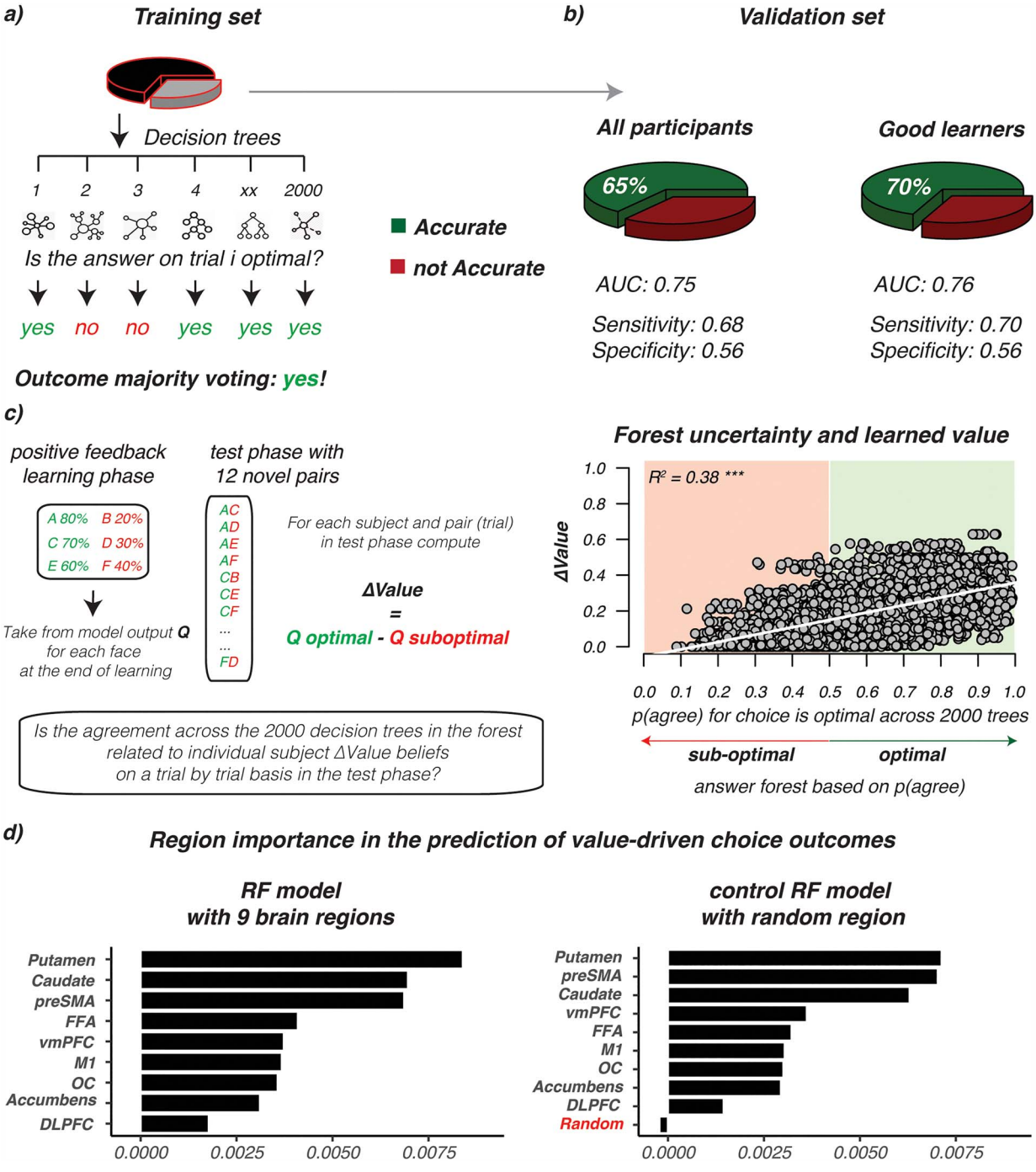

Figure 6. Random Forest performance and importance ranking. Prediction of value-driven choice outcomes in the transfer phase using trial-by-trial BOLD responses from striatal, perceptual, and prefrontal cortex regions. (*a*) Overview of the Random Forest approach where the training set is used to predict choice outcomes for each trial by using the majority vote of 2000 different decision trees. Each tree is built using a different set, or sample, of trials and predictors from the training set. The forest is trained on a training set sampled from all participants ($N = 43$) or only "the good learners" ($N = 34$). (*b*) The classification, or decoding, accuracy (green) given the separate unseen validation sets, for all participants and good learners. (*c*) On the left, overview of the feedback scheme in the learning phase, and the new combination in transfer about which the RF is making a prediction with an illustration of how ΔValue is computed for each trial. ΔValue was computed for each trial in the transfer phase by using the end beliefs (Q) that participants had about each stimulus (A-to-F) at the end of the learning phase. On the right side, plotted relationship between forest uncertainty (i.e., proportion of agreement across 2000 trees) on each prediction/trial (*x*-axis) and ΔValue (*y*-axis) for the model with the highest accuracy (i.e., the good learners). Forest uncertainty is defined as the proportion of trees saying "yes! the choice on this trial was optimal/correct." When this ratio is below 0.5, the forest will predict "no" (suboptimal/wrong choice), otherwise the prediction is "yes! the choice on this trial was optimal/correct" (optimal). R^2=adjusted R^2. Note that the same pattern was found for all participants ($R^2 = 0.41^{***}$, please see Supplementary Fig. 3). (*d*) Ranking of the ROIs in their contribution to the predictive accuracy of the best performing model (i.e., good learners). Left, The original ranking. On the right, we evaluate ranking with all 9 original regions, but now add a control region that was sampled randomly from $\mathcal{N}(0,1)$, and unrelated to the activity of any region, or ΔValue. Notice that the random variable has negative importance in the ranking, meaning that removing it improves model performance with 0.5%.

the prediction of future value-driven choice outcomes in a phase where learning is established.

In a choice between two faces, BOLD responses in both the dorsal striatum and perceptual regions were affected more by values of the chosen face, relative to the unchosen face. Across three levels of uncertainty, we only observed the differential modulation of value on BOLD when belief representations were stable. This specificity aligns with neuronal responses to perceptual stimuli in the caudate tail (Kim et al. 2017), visual cortex (Shuler and Bear 2006; Weil et al. 2010; Cicmil et al. 2015), and imaging work across sensory modalities (Serences 2008; Pleger et al. 2009; Serences and Saproo 2010; Kahnt et al. 2011; Vickery et al. 2011; FitzGerald et al. 2013; Lim et al. 2013; Kaskan et al. 2016), where it fuels theories in which the learning of stable reward expectations can develop to modulate, or sharpen, the representation of sensory information critical for perceptual decision-making (Roelfsema et al. 2010; Kahnt et al. 2011; Cicmil et al. 2015).

After a choice was made, feedback modulations of signed ("valence") and unsigned ("surprise") RPEs (Fouragnan et al. 2018) were evaluated on BOLD responses, by using an orthogonal design where the unsigned and signed RPE compete to explain BOLD variances. Both visual and striatal regions respond to prediction errors (Den Ouden et al. 2012). In the striatum, both valence and surprise are thought to optimize future action selection in the dorsal striatum or the prediction of future rewards in the ventral striatum. In perceptual regions, a mismatch between the expected and received outcome is often explained as surprise where a boost in attention or salience changes the representation of an image without a representation of value per se. We found positive modulatory effects of signed RPEs in all striatal regions, as well as in the FFA and OC. Concurrently, modulations of unsigned RPEs were only observed in the accumbens (ventral striatum) and FFA, where notably the direction of modulation was reversed. We speculate that this contrast arises from the differential role of the regions. In the FFA, specialized and dedicated information processing is essential to quickly recognize valuable face features. Complementary boosts of surprise and valence here could prioritize attention toward the most rewarding face feature to strengthen the reward association in memory or help speed up future recognition (Gottlieb 2012; Gottlieb et al. 2014; Störmer et al. 2014). In the accumbens, boosted effects of positive valence on BOLD were dampened by larger mismatches. Large mismatches in what was expected are rare in stable environments. We therefore reason that in the accumbens the contrast between valence and surprise could function as a scale to refine learning, eventually leading to more reliable predictions of future rewards.

Whereas BOLD in the ventral striatum was shaped by both signed and unsigned RPEs, the dorsal striatum was sensitive to differential value up to a choice and signed RPEs with the presentation of feedback (Kaskan et al. 2016; Lak et al. 2016, 2017; McCoy et al. 2018; Van Slooten et al. 2018). The concurrent modulation of differential value in the primary motor cortex (please see M1 in Supplementary Fig. 1) associates the dorsal striatum with the integration of sensory information (Ding and Gold 2010; Yamamoto et al. 2012; Hikosaka et al. 2013; Kim et al. 2017), where increased visual cortex BOLD responses to faces with the highest value could potentially help bias the outcome of a value-driven choice.

We explored this line of reasoning with the prediction of value-driven choice outcomes in a follow-up transfer phase after learning. In recent years, machine learning approaches have become increasingly important in neuroscience (Naselaris et al. 2011; Hassabis et al. 2017; Hebart and Baker 2018; Snoek et al. 2019), where the ease of interpretation has often motivated a choice for linear methods above nonlinear methods (Naselaris et al. 2011; Kriegeskorte and Douglas 2018). Despite that nonlinear methods are less constrained, and therefore sometimes able to reach a better classification accuracy by capturing nonarbitrary or unexpected relationships (King et al. 2018). Value-driven choices after a phase of initial learning are influenced by the consistency of past learning, memory updating, and attention. All of these processes are affected by both linear and nonlinear neurotransmitter modulations (Aston-Jones and Cohen 2005; Yu and Dayan 2005; Cools and D'Esposito 2011; Beste et al. 2018). Our RF approach was unconstrained by linearity with classification accuracies well above chance and improved with the evaluation of only the good learners, despite substantial decreases in data given to the algorithm to learn the correct labeling. Critically, we additionally found that the uncertainty of trial-by-trial predictions made by RF is tied to the differentiability of value beliefs—an index that we could compute for the novel pair combination in the transfer phase by using the value (Q) beliefs that participants had about each face at the end of learning. These results showcase how trial-by-trial BOLD fluctuations in striatal, prefrontal, and sensory regions can be combined by machine learning, or decoding, algorithms to reliably predict the outcome of a value-driven choice. In addition, we combine the RF output with the cognitive computational modeling of the learning process to refine the interpretation of the nonlinear decoding predictions. With this combination, we essentially show how the uncertainty of RF predictions is tied to value beliefs acquired with learning in the past.

An important evaluation intended with our machine learning approach was the ranking of regions by their contribution to the predictive (decoding) accuracy in the transfer phase. After the observed modulations of BOLD in the learning phase, this explorative analysis sought the relevance of learning-BOLD relationships in the resolve of future choices. Here, the ranking made by RF first identified signals from the dorsal striatum (putamen and caudate) as most important followed by the preSMA and, then most notably, visual regions. That is, when the quality of learning was high across participants, FFA ranked just above traditional regions such as the vmPFC and the accumbens (O'Doherty et al. 2003, 2017; Hare et al. 2011; Niv et al. 2012; Klein et al. 2017). Notably, FFA was replaced by OC in ranking with the evaluation of all participants (please see Supplementary Fig. 3b). This difference could occur because the quality of learning was more variable across all participants or because RF predictions based on the heterogeneous data from all participants were less accurate. In general, the shift in ranking implies that when learning is less consistent choice outcomes are better predicted by fluctuations in OC—perhaps with the identification of rewarding low-level features. With better or more consistent learning, however, participants should increasingly rely on memory and specialized visual areas. Thus, search for specific face features associated with high value by recruiting the FFA in the visual ventral stream. Consistent with this reasoning, recent neuronal recordings show rapid visual processing of category-specific value cues in the ventral visual stream. These specific value cues are only seen for well-learned reward categories and, critically, precede the processing of value in prefrontal cortex (Sasikumar et al. 2018).

Additionally, in the learning phase, both OC and the FFA were modulated more by values of the (to be) chosen stimulus

when belief representations were stable and distinct—that is, we only observed differential Q-value modulations for the most reliable and easy to learn AB pair. This combined with the RPE modulations found in the same regions suggests an effect of value and learning on perceptual regions that is both specialized (FFA) and global (OC). Note however that this possibility must be studied further with designs that can zoom in on specificity with the separation of different perceptual dimensions (e.g., houses vs. faces). Our transfer phase results imply a differential role for the specialized FFA, and the more low-level general OC, with the comparison of good versus all learners. Tasked with predicting the outcome of future value-driven choices, RF rankings showed a specialized and prominent FFA role for good/efficient learners, whereas OC was more important with the evaluation of all participants (where learning was less consistent or noisier across participants). Recent work on the interplay between learning and attention suggests a bidirectional relationship between learning and attention: we learn what to attend from feedback and, in turn, use selective attention to constrain learning toward relevant value dimensions (Leong et al. 2017; Rusch et al. 2017). In our study, better learning helps a more refined identification of rewarding features in a face, which we interpret as a narrower focus of selective attention in the FFA during learning (Niv et al. 2015). With past learning being more noisy, or less established, extraction of relevant features is less straightforward with attention being more spread to both specialized and global regions. Additionally, we observed both FFA and OC to only rank in the top (just after the caudate) when ΔValue was very small (Supplementary Fig. 6). With easy choices, this effect was reversed where processing of OC BOLD even declined the RF accuracy. This contrast suggests that especially when the options to choose from are just too similar in value (i.e., think of the options A:C or B:D), past learning in perceptual regions could serve the striatum with a selective boost to highlight the most rewarding face features. In contrast, when the distinction is easy and clear-cut, choices depend far more on inputs from the ventral striatum and vmPFC.

We note that although BOLD fluctuations in the preSMA ranked second in the prediction of value-driven choice outcomes, no reliable modulations of BOLD were observed by either differential value or RPEs in the learning phase. The preSMA is densely connected to the dorsal striatum and consistently associated with action-reward learning (Jocham et al. 2016) or choice difficulty (Shenhav et al. 2014). The lack of associations in this study might result from our noisier estimates of the BOLD response that is typical for regions in the prefrontal cortex (Pircalabelu et al. 2015; Bhandari et al. 2018), the anatomical masks selected, or smaller variability across trials in the learning phase (i.e., 3 pairs in learning phase vs. 15 pairs in transfer phase). Nevertheless, the importance indicated by RF, combined with our previous analysis of this transfer phase data (Jahfari et al. 2018), implies an important role for the preSMA in the resolve of value-driven choices in concert with the striatum. More research with optimized sequences to estimate BOLD in PFC is required to clarify the link between learning and transfer.

To summarize, we find an important role for perceptual regions in the prediction of future value-driven choice outcomes, which coincides with the sensitivity of BOLD in visual regions to differential value and signed feedback. With the integration of value and feedback, visual regions could learn to prioritize the processing of high-value features to support optimal response selection via the dorsal striatum in future choices.

Funding

This project was funded by an ABC Talent Grant from the University of Amsterdam to S.J., an ERC grant to J.T. (ERC-2012-AdG-323413), and a NWO-CAS grant to T.K. (012.200.012).

Notes

The code and preprocessed files for behavioral and decoding analyses can be downloaded from: https://github.com/sarajahfari/Pearl3T.git, and fMRI preprocessing and deconvolution analysis code are available at https://github.com/tknapen/pearl_3T. The raw data can be downloaded from openneuro.org in BIDS at: https://openneuro.org/datasets/ds002068. *Conflict of Interest:* None declared.

Author Contributions

S.J. and T.K. developed the questions and analysis plan for the reanalysis, contributed novel methods, and analyzed the data. S.J. wrote the first draft of the manuscript with edits from T.K. J.T. commented on the final draft.

Supplementary Material

Supplementary material is available at *Cerebral Cortex* online.

References

Aston-Jones G, Cohen JD. 2005. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. *Annu Rev Neurosci*. 28:403–450.

Atallah HE, Lopez-Paniagua D, Rudy JW, O'Reilly RC. 2007. Separate neural substrates for skill learning and performance in the ventral and dorsal striatum. *Nat Neurosci*. 10:126–131.

Beckmann CF, Jenkinson M, Smith SM. 2003. General multilevel linear modeling for group analysis in fmri. *Neuroimage*. 20:1052–1063.

Beste C, Adelhöfer N, Gohil K, Passow S, Roessner V, Li S-C. 2018. Dopamine modulates the efficiency of sensory evidence accumulation during perceptual decision making. *Int J Neuropsychopharmacology*. 21:649–655.

Bhandari A, Gagne C, Badre D. 2018. Just above chance: Is it harder to decode information from human prefrontal cortex blood oxygenation level-dependent signals? *J Cogn Neurosci*. 30:1473–1498.

Breiman L. 2001. Random forests. *Mach Learn*. 45:5–32.

Breiman L. 2004. *Consistency for a simple model of random forests*. Statistical Department. University of California at Berkeley. Technical Report, (670).

Cicmil N, Cumming BG, Parker AJ, Krug K. 2015. Reward modulates the effect of visual cortical microstimulation on perceptual decisions. *eLife*. 4:e07832.

Cieslik EC, Zilles K, Caspers S, Roski C, Kellermann TS, Jakobs O, Langner R, Laird AR, Fox PT, Eickhoff SB. 2012. Is there "one" dlpfc in cognitive action control? Evidence for heterogeneity from co-activation-based parcellation. *Cereb Cortex*. 23:2677–2689.

Collins AGE, Frank MJ. 2014. Opponent actor learning (opal): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychol Rev*. 121:337–366.

Cools R, D'Esposito M. 2011. Inverted-u–shaped dopamine actions on human working memory and cognitive control. *Biol Psychiatry*. 69:e113–e125.

Daw ND. 2011. Trial-by-trial data analysis using computational models. In: *Decision making, affect, and learning: attention and performance XXIII*. Vol Vol. 23, pp. 3–38. Delgado MR, Phelps EA, and Robbins TW. Oxford University Press.

Daw ND, O'doherty JP, Dayan P, Seymour B, Dolan RJ. 2006. Cortical substrates for exploratory decisions in humans. *Nature*. 441:876–879.

Den Ouden HEM, Kok P, De Lange FP. 2012. How prediction errors shape perception, attention, and motivation. *Front Psychol*. 3:548.

Ding L, Gold JI. 2010. Caudate encodes multiple computations for perceptual decisions. *J Neurosci*. 30:15747–15759.

Fernandez-Ruiz J, Wang J, Aigner TG, Mishkin M. 2001. Visual habit formation in monkeys with neurotoxic lesions of the ventrocaudal neostriatum. *Proc Natl Acad Sci*. 98:4196–4201.

FitzGerald THB, Friston KJ, Dolan RJ. 2013. Characterising reward outcome signals in sensory cortex. *Neuroimage*. 83:329–334.

Fouragnan E, Retzler C, Philiastides MG. 2018. Separate neural representations of prediction error valence and surprise: evidence from an fMRI meta-analysis. *Human Brain Mapp*. 39:2887–2906.

Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE. 2007. Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proc Natl Acad Sci*. 104:16311–16316.

Gottlieb J. 2012. Attention, learning, and the value of information. *Neuron*. 76:281–295.

Gottlieb J, Hayhoe M, Hikosaka O, Rangel A. 2014. Attention, reward, and information seeking. *J Neurosci*. 34:15497–15504.

Hare TA, Schultz W, Camerer CF, O'Doherty JP, Rangel A. 2011. Transformation of stimulus value signals into motor commands during simple choice. *Proc Natl Acad Sci*. 108:18120–18125.

Hassabis D, Kumaran D, Summerfield C, Botvinick M. 2017. Neuroscience-inspired artificial intelligence. *Neuron*. 95:245–258.

Hebart MN, Baker CI. 2018. Deconstructing multivariate decoding for the study of brain function. *Neuroimage*. 180:4–18.

Hikosaka O, Kim HF, Yasuda M, Yamamoto S. 2014. Basal ganglia circuits for reward value–guided behavior. *Ann Rev Neurosci*. 37:289–306.

Hikosaka O, Yamamoto S, Yasuda M, Kim HF. 2013. Why skill matters. *Trends Cogn Sci*. 17:434–441.

Jahfari S, Ridderinkhof KR, Collins AGE, Knapen T, Waldorp LJ, Frank MJ. 2018. Cross-task contributions of frontobasal ganglia circuitry in response inhibition and conflict-induced slowing. *Cereb Cortex*. 29:1969–1983.

Jahfari S, Theeuwes J. 2017. Sensitivity to value-driven attention is predicted by how we learn from value. *Psychon Bull Rev*. 24:408–415.

Jahfari S, Waldorp L, Ridderinkhof KR, Scholte HS. 2015. Visual information shapes the dynamics of corticobasal ganglia pathways during response selection and inhibition. *J Cogn Neurosci*. 27:1344–1359.

Jocham G, Boorman E, Behrens T. 2016. Neuroscience of value-guided choice. The Wiley handbook on the cognitive neuroscience of. *Learning*. 554–591.

Jocham G, Klein TA, Ullsperger M. 2011. Dopamine-mediated reinforcement learning signals in the striatum and ventromedial prefrontal cortex underlie value-based choices. *J Neurosci*. 31:1606–1613.

Joel D, Niv Y, Ruppin E. 2002. Actor–critic models of the basal ganglia: new anatomical and computational perspectives. *Neural Netw*. 15:535–547.

Kahnt T, Heinzle J, Park SQ, Haynes J-D. 2011. Decoding different roles for vmPFC and dlPFC in multi-attribute decision making. *Neuroimage*. 56:709–715.

Kahnt T, Park SQ, Cohen MX, Beck A, Heinz A, Wrase J. 2009. Dorsal striatal–midbrain connectivity in humans predicts how reinforcements are used to guide decisions. *J Cogn Neurosci*. 21:1332–1345.

Kaskan PM, Costa VD, Eaton HP, Zemskova JA, Mitz AR, Leopold DA, Ungerleider LG, Murray EA. 2016. Learned value shapes responses to objects in frontal and ventral stream networks in macaque monkeys. *Cereb Cortex*. 27:2739–2757.

Kim HF, Amita H, Hikosaka O. 2017. Indirect pathway of caudal basal ganglia for rejection of valueless visual objects. *Neuron*. 94:920–930.

Kim HF, Hikosaka O. 2013. Distinct basal ganglia circuits controlling behaviors guided by flexible and stable values. *Neuron*. 79:1001–1010.

King J-R, Gwilliams L, Holdgraf C, Sassenhagen J, Barachant A, Engemann D, Larson E, Gramfort A. 2018. *Encoding and decoding neuronal dynamics: methodological framework to uncover the algorithms of cognition*. hal-01848442

Klein TA, Ullsperger M, Jocham G. 2017. Learning relative values in the striatum induces violations of normative decision making. *Nat Commun*. 8:16033.

Knapen T, Gee J. 2016. *FIRDeconvolution*. 10.5281/zenodo.46216

Kravitz DJ, Saleem KS, Baker CI, Ungerleider LG, Mishkin M. 2013. The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends Cogn Sci*. 17: 26–49.

Kriegeskorte N, Douglas PK. 2018. Interpreting encoding and decoding models. arXiv preprint arXiv:181200278.

Lak A, Nomoto K, Keramati M, Sakagami M, Kepecs A. 2017. Midbrain dopamine neurons signal belief in choice accuracy during a perceptual decision. *Curr Biol*. 27:821–832.

Lak A, Stauffer WR, Schultz W. 2016. Dopamine neurons learn relative chosen value from probabilistic rewards. *eLife*. 5:e18044.

Leong YC, Radulescu A, Daniel R, DeWoskin V, Niv Y. 2017. Dynamic interaction between reinforcement learning and attention in multidimensional environments. *Neuron*. 93:451–463.

Lim S-L, O'Doherty JP, Rangel A. 2011. The decision value computations in the vmPFC and striatum use a relative value code that is guided by visual attention. *J Neurosci*. 31:13214–13223.

Lim S-L, O'Doherty JP, Rangel A. 2013. Stimulus value signals in ventromedial pfc reflect the integration of attribute value signals computed in fusiform gyrus and posterior superior temporal gyrus. *J Neurosci*. 33:8729–8741.

McCoy B, Jahfari S, Engels G, Knapen T, Theeuwes J. 2019. Dopaminergic medication reduces striatal sensitivity to negative outcomes in Parkinson's disease. Brain. 142:3605–3620. 10.1093/brain/awz276.

Montague PR, Dayan P, Sejnowski TJ. 1996. A framework for mesencephalic dopamine systems based on predictive hebbian learning. *J Neurosci*. 16:1936–1947.

Naselaris T, Kay KN, Nishimoto S, Gallant JL. 2011. Encoding and decoding in fMRI. *Neuroimage*. 56:400–410.

Niv Y, Daniel R, Geana A, Gershman SJ, Leong YC, Radulescu A, Wilson RC. 2015. Reinforcement learning in multidimensional environments relies on attention mechanisms. *J Neurosci*. 35:8145–8157.

Niv Y, Edlund JA, Dayan P, O'Doherty JP. 2012. Neural prediction errors reveal a risk-sensitive reinforcement-learning process in the human brain. *J Neurosci*. 32:551–562.

O'Doherty J, Critchley H, Deichmann R, Dolan RJ. 2003. Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. *J Neurosci*. 23:7931–7939.

O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ. 2004. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*. 304:452–454.

O'Doherty JP, Cockburn J, Pauli WM. 2017. Learning, reward, and decision making. *Ann Rev Psychol*. 68:73–100.

O'Doherty JP, Hampton A, Kim H. 2007. Model-based fMRI and its application to reward learning and decision making. *Ann N Y Acad Sci*. 1104:35–53.

Pircalabelu E, Claeskens G, Jahfari S, Waldorp LJ. 2015. A focused information criterion for graphical models in fMRI connectivity with high-dimensional data. *Ann Appl Stat*. 9:2179–2214.

Pleger B, Ruff CC, Blankenburg F, Klöppel S, Driver J, Dolan RJ. 2009. Influence of dopaminergically mediated reward on somatosensory decision-making. *PLoS Biol*. 7:e1000164.

Roelfsema PR, van OA, Watanabe T. 2010. Perceptual learning rules based on reinforcers and attention. *Trends Cogn Sci*. 14:64–71.

Rusch T, Korn CW, Gläscher J. 2017. A two-way street between attention and learning. *Neuron*. 93:256–258.

Sasikumar D, Emeric E, Stuphorn V, Connor CE. 2018. First-pass processing of value cues in the ventral visual pathway. *Curr Biol*. 28:538–548.

Schmittmann VD, Jahfari S, Borsboom D, Savi AO, Waldorp LJ. 2015. Making large-scale networks from fMRI data. *PLoS One*. 10:e0129074.

Schultz W, Dayan P, Montague PR. 1997. A neural substrate of prediction and reward. *Science*. 275:1593–1599.

Seabold S, Perktold J. 2010. Statsmodels: econometric and statistical modeling with python. In: *Proceedings of the 9th python in science conference*, pp. 57–61. https://www.researchgate.net/profile/Josef_Perktold/publication/264891066_Statsmodels_Econometric_and_Statistical_Modeling_with_Python/links/5667ca9308ae34c89a0261a8/Statsmodels-Econometric-and-Statistical-Modeling-with-Python.pdf

Serences JT. 2008. Value-based modulations in human visual cortex. *Neuron*. 60:1169–1181.

Serences JT, Saproo S. 2010. Population response profiles in early visual cortex are biased in favor of more valuable stimuli. *J Neurophysiol*. 104:76–87.

Shenhav A, Straccia MA, Cohen JD, Botvinick MM. 2014. Anterior cingulate engagement in a foraging context reflects choice difficulty, not foraging value. *Nat Neurosci*. 17:1249.

Shuler MG, Bear MF. 2006. Reward timing in the primary visual cortex. *Science*. 311:1606–1609.

Snoek L, Miletić S, Scholte HS. 2019. How to control for confounds in decoding analyses of neuroimaging data. *Neuroimage*. 184:741–760.

Störmer V, Eppinger B, Li S-C. 2014. Reward speeds up and increases consistency of visual selective attention: a lifespan comparison. *Cogn Affect Behav Neurosci*. 14:659–671.

Tobler PN, Fiorillo CD, Schultz W. 2005. Adaptive coding of reward value by dopamine neurons. *Science*. 307:1642–1645.

Van Slooten JC, Jahfari S, Knapen T, Theeuwes J. 2018. How pupil responses track value-based decision-making during and after reinforcement learning. *PLoS Comput Biol*. 14:e1006632.

Vickery TJ, Chun MM, D L. 2011. Ubiquity and specificity of reinforcement signals throughout the human brain. *Neuron*. 72:166–177.

Watkins CJCH, Dayan P. 1992. Q-learning. *Mach Learn*. 8:279–292.

Weil RS, Furl N, Ruff CC, Symmonds M, Flandin G, Dolan RJ, Driver J, Rees G. 2010. Rewarding feedback after correct visual discriminations has both general and specific influences on visual cortex. *J Neurophysiol*. 104:1746–1757.

Woolrich MW, Ripley BD, Brady M, Smith SM. 2001. Temporal autocorrelation in univariate linear modeling of fMRI data. *Neuroimage*. 14:1370–1386.

Yamamoto S, Monosov IE, Yasuda M, Hikosaka O. 2012. What and where information in the caudate tail guides saccades to visual objects. *J Neurosci*. 32:11005–11016.

Yu AJ, Dayan P. 2005. Uncertainty, neuromodulation, and attention. *Neuron*. 46:681–692.