Cerebral Cortex August 2014;24:2009–2021
doi:10.1093/cercor/bht049
Advance Access publication March 8, 2013

Neural Evidence for Adaptive Strategy Selection in Value-Based Decision-Making

Sebastian Gluth[1], Jörg Rieskamp[2] and Christian Büchel[1]

[1]Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Hamburg D-20246, Germany and
[2]Department of Psychology, University of Basel, Basel CH-4055, Switzerland

Address correspondence to S. Gluth, Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Martinistr. 52, D-20246 Hamburg, Germany. Email: sgluth@uke.de

In everyday life, humans often encounter complex environments in which multiple sources of information can influence their decisions. We propose that in such situations, people select and apply different strategies representing different cognitive models of the decision problem. Learning advances by evaluating the success of using a strategy and eventually by switching between strategies. To test our strategy selection model, we investigated how humans solve a dynamic learning task with complex auditory and visual information, and assessed the underlying neural mechanisms with functional magnetic resonance imaging. Using the model, we were able to capture participants' choices and to successfully attribute expected values and reward prediction errors to activations in the dopaminoceptive system (e.g., ventral striatum [VS]) as well as decision conflict to signals in the anterior cingulate cortex. The model outperformed an alternative approach that did not update decision strategies, but the relevance of information itself. Activation of sensory areas depended on whether the selected strategy made use of the respective source of information. Selection of a strategy also determined how value-related information influenced effective connectivity between sensory systems and the VS. Our results suggest that humans can structure their search for and use of relevant information by adaptively selecting between decision strategies.

Keywords: model-based fMRI, reinforcement learning, reward, ventral striatum

Introduction

In the past decade, the understanding of the neural mechanisms that guide human learning and decision-making has been strongly promoted by the combination of brain imaging and computational modeling of cognition and behavior (Daw and Doya 2006; O'Doherty et al. 2007; Rangel et al. 2008; Glimcher et al. 2009). The rationale is that mathematical models provide researchers with precise trial-by-trial predictions of internal states of the cognitive system, which can be regressed against functional magnetic resonance imaging (fMRI) data to track their neural correlates (O'Doherty et al. 2007). Initially, fMRI studies applied traditional reinforcement learning (RL) models, such as the Rescorla–Wagner model (Rescorla and Wagner 1972) or temporal difference (TD) learning models (Sutton and Barto 1998), to unravel the neural mechanisms of elementary learning processes, such as classical and instrumental conditioning (O'Doherty et al. 2003, 2004; Gläscher and Büchel 2005). O'Doherty et al. (2003), for instance, were able to associate reward expectations and reward prediction errors (RPEs) with changes in the blood oxygen level-dependent (BOLD) fMRI signal in the ventral striatum (VS) and the orbitofrontal cortex using a TD model.

However, recent work suggests that in more complex decision-making scenarios, human learning behavior can be better described by more sophisticated, so-called "model-based" learning approaches (Hampton et al. 2006; Behrens et al. 2007; Gläscher et al. 2010; Daw et al. 2011). These learning models assume humans to create an internal representation of the decision environment, which allows speeding up the learning process and facilitating the adaptation to changing environments (Sutton and Barto 1998; Daw et al. 2005; Gershman and Niv 2010). Critically, the advantage of using an internal model to guide decisions depends on whether this model sufficiently matches the environment. Therefore, the decision maker has to learn about the adequacy of different internal models and to select an appropriate one for making accurate inferences. In this vein, a "model" is represented by a decision strategy. The strategy prescribes which sources of information should be taken into account and how this information should be processed to arrive at a final decision (Hutchinson and Gigerenzer 2005). Moreover, by being equipped with a repertoire of strategies (i.e., multiple models of the environment), the decision maker can learn to select the strategy that represents the environment best. In the decision sciences, the concept of a "repertoire" of decision strategies is very prominent (e.g., Tversky and Kahneman 1974; Payne et al. 1988, 1993; Gigerenzer et al. 1999).

Rieskamp and Otto (2006) introduced a computational model of adaptive strategy selection, the strategy selection learning (SSL) theory, and demonstrated its capability to account for human choice behavior and learning in complex decision-making situations (see also Rieskamp 2006, 2008; Mata et al. 2007, 2010; Rieskamp and Hoffrage 2008). SSL assumes that people possess a repertoire of decision strategies that have been proved successful in the past. To solve a new decision problem, a particular strategy is selected based on the subjective expectation that using this strategy is suitable in the current scenario. Learning takes place by updating the strategy's expectation on the basis of received feedback, which may result in adopting a different strategy for future choices. SSL represents an RL approach, because the learning process is solely based on feedback on the outcome of a decision and does not require a teaching signal about the optimal decision (unlike supervised learning).

The purpose of the present study was to test whether SSL can describe human learning and decision-making behavior and unravel its neural correlates in a complex scenario with multiple sources of information. Furthermore, we were interested in how the selection between strategies influences the attention to and use of sensory information that might or might not be relevant to the decision (Khader et al. 2011). To investigate this, we designed a learning task (Fig. 1), in which participants decided between buying or rejecting fictitious

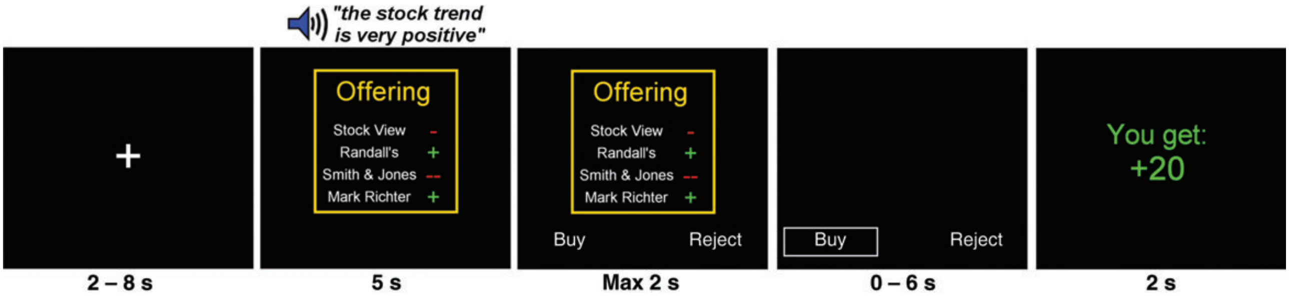

Figure 1. Task design. In each trial, participants were offered a stock of unknown value. In the decision phase, a single auditory and 4 visual cues provided information about the stock. Subsequently, participants had 2 s to respond (buy or reject the offering), and feedback was given after a variable delay to separate BOLD signals for decision and outcome phases.

stock offerings of unknown value. They had to infer the attractiveness of each stock based on numerous cues (i.e., stock ratings and stock trend) presented in 2 different sensory domains (i.e., visual and auditory). Participants were not told to what extent these cues were informative, but that they could learn about their relevance by considering the feedback. In fact, we manipulated cue validities by 2 consecutive environmental conditions: In the compensatory environment, all cues contributed equally to the determination of the stock values, whereas in the noncompensatory environment, only the auditory cue was relevant.

We hypothesized that the complexity of the paradigm (a large amount of information of unknown relevance, short deliberation time, and unspecific feedback) and the disjunction of cues into separate sensory domains (visual, auditory) would drive our participants to use particular strategies that simplify the decisions and the learning process. Reward-based learning and decision-making parameters have been repeatedly associated with the dopaminergic system of the primate brain (Schultz et al. 1997; O'Doherty et al. 2003; Bayer and Glimcher 2005; Knutson et al. 2005; Hampton et al. 2006; Pessiglione et al. 2006; Kable and Glimcher 2007), and this connection appears to apply to model-based RL as well (Hampton et al. 2006; Daw et al. 2011; Wunderlich et al. 2012). Further evidence for a central role of this neuronal circuitry comes from lesion studies (Damasio 1994; Fellows 2011). Thus, we expected to find neural correlates of subjective expected value (EV) and RPE as derived from SSL in the VS and the ventromedial prefrontal cortex (vmPFC). We also tested whether decision conflict (again, quantified on the basis of SSL) is represented in the anterior cingulate cortex (ACC; Botvinick et al. 2004; Pochon et al. 2008; Pine et al. 2009; Venkatraman et al. 2009). To strengthen the evidence for an adaptive strategy selection process, we compared the behavioral and neuronal fit of SSL with an alternative learning approach that is based on updating the relevance of each cue separately, the additive linear model (ALM; Gluck and Bower 1988; Rieskamp 2006; Kelley and Busemeyer 2008). Finally, we hypothesized that, due to the irrelevance of visual information in the noncompensatory environment, participants would learn to adopt a strategy that solely relies on the auditory cue. This should result in a reduced fMRI BOLD signal in the visual system as an effect of shifted attention (O'Craven et al. 1997; Büchel et al. 1998; Shomstein and Yantis 2004). Using dynamic causal modeling (DCM; Friston et al. 2003), we then looked at the modulation of effective connectivity between sensory and reward structures by (sensory-specific) value information and tested whether the strength of this modulation depended on the selected strategy.

Materials and Methods

Participants

Participants were 24 right-handed healthy subjects (mean age = 26.5 years, SD = 2.4, range = 22–32 years; 10 females) with normal or corrected-to-normal vision. All the participants gave written informed consent approved by a local ethics committee. The subjects were reimbursed for participation (10 Euro per hour). They could earn additional money by winning points during the task: The sum of collected points was converted into Euro at the ratio of 100:1 (e.g., 720 points = 7.20 Euro).

Experimental Design and Task

Each trial began with the decision phase (5 s) during which a frame appeared on the screen enclosing the heading "offering" and the names of 4 rating companies together with their ratings (Fig. 1). The order of the companies from top to bottom was fixed. Ratings could be a "+ +," "+," "–," or a "– –." Simultaneously, a female voice announced "the current stock trend" via MR-compatible headphones (auditory stimulation lasted approximately 2.2 s). In analogy to the ratings, the stock trend could be "very positive," "slightly positive," "slightly negative," or "very negative." The participants were told that stocks and rating companies were fictitious and that they would not benefit from any expertise in the field of stock markets. Following the decision phase, labels for "buy" and "reject" appeared below the frame on the left and right sides in a randomized order so that participants would not know in advance which option would occur on which side. They had 2 s to press the respective button (left/right) after appearance of the labels and were asked to already establish the decision itself during the decision phase (missing responses were rare; <4 per subject). As soon as a response was given, the selection was framed and the offer disappeared. The time between response and outcome phases was jittered (0–6 s plus the rest of the 2 s for responding) to separate BOLD signals, indicative of deciding and evaluating feedback. During the outcome phase (2 s), feedback was provided visually ("you get: [no. points won]" colored green for positive values, red for negative values, and white for 0 points). Trials were separated by an inter-trial interval of 2–8 s, showing a white cross. The experiment (160 trials) lasted approximately 43 min. Before the experiment, the participants were instructed on the paradigm: They were told that they would play a stock-buying game in which they could use several pieces of information to decide whether to buy or to reject a stock offer in every trial. They were introduced to the different cues, the potential stock trends and ratings, and the general workflow of a trial. They were also told that taking cues into account could help them improving their decisions. However, they were not informed about the cues' relevance, the existence (and switching) of 2 environments, or that stock values are a weighted linear combination of cue values (see below). Furthermore, the expression "decision strategy" was strictly avoided in the instructions. Following instructions, the participants were trained on the stimulus material. Training trials did not contain meaningful cues or feedback (i.e., they were set to "unknown") to prevent subjects from establishing strategies before the actual experiment. Stimulus presentation was realized using the Presentation Software package (Neurobehavioral Systems).

Reward Contingencies

The relationship between ratings and stock values was manipulated by 2 consecutive environmental conditions. In the compensatory environment, all cues contributed equally to the determination of the stock values; in the noncompensatory environment, only the stock trend (i.e., the auditory cue) predicted its value. Gaussian noise was added to these payoff functions, making the task probabilistic. Each environment consisted of 80 trials, and the order of environments was counterbalanced across participants. In the compensatory environment, the value of a stock (V_t) at trial t was a function of all cues (a_t = auditory cue; $v_{1,t}$ to $v_{4,t}$ = visual cues) plus Gaussian noise ε with $\mu = 0$ and $\sigma = 7.5$:

$$V_t = 10a_t + 10v_{1,t} + 10v_{2,t} + 10v_{3,t} + 10v_{4,t} + \varepsilon, \qquad (1)$$

where $a_t = 2$, 1, −1, or −2 for "very positive," "slightly positive," "slightly negative," or "very negative" stock trends and $v_{1,t}$ to $v_{4,t} = 2$, 1, −1, or −2 for the ratings "+ +," "+," "−," or "− −," respectively. In the noncompensatory environment, the value of a stock was only a function of the auditory cue plus noise:

$$V_t = 10a_t + \varepsilon. \qquad (2)$$

The first cue (as any other cue) can adopt 4 different values (2, 1, −1, and −2), but the sum of all the 5 cues can adopt 21 different values (from 10 to −10). Hence, the range of possible Vs is less restricted in the compensatory environment. To overcome this imbalance, we restricted possible stock offers to those with values of 20, 10, −10, or −20 plus noise for both environments (this restriction induced a small negative correlation between cue values of ∼ −0.2). From the remaining pool of 430 stocks, 80 stocks were randomly selected for each participant with the further restriction that the 2 strategies ALL and AUD (see below) made different suggestions to buy or to reject the stocks in exactly 50% of the trials (this was done to compare the SSL model predictions with participants' choices, see Results; without this restriction, the strategies would have made different suggestions in ∼41% of the trials). The 80 stocks were offered in both environments, but in differently randomized orders. Thus, stimulus material was kept identical between environments and offers only differed in terms of reward contingencies. The noise ε was added in a discrete manner, such that stock values were always a multiple of 5. For instance, if $a_t = 1$ in the noncompensatory environment, the value V_t could be −10, −5, 0, 5, 10, 15, 20, 25, or 30 points with a probability of 0.01, 0.04, 0.10, 0.21, 0.28, 0.21, 0.10, 0.04, or 0.01, respectively.

SSL Model

We modeled participants' behavior using a variant of the original SSL model (Rieskamp and Otto 2006). SSL assumes that people have a set of strategies from which they select. For the sake of simplicity, we restricted the set to those 2 strategies that provide an accurate representation of the 2 environments: 1) ALL is a multiple-cue strategy that sums over all cue values treating them as equally important; this strategy is equivalent to the established decision strategy "Equal weight" (Dawes 1979; Payne et al. 1988), also known as "Tallying" (Todd and Gigerenzer 2007). 2) AUD is a single-cue strategy that focuses exclusively on the auditory cue; this strategy is similar to the established lexicographic decision strategy "Take-the-best" (Gigerenzer and Goldstein 1996). Assuming the existence of only these 2 strategies is certainly a simplification of the actual repertoire of human decision strategies. However, for the present decision problem, this simplification appears justifiable as the 2 strategies provide an accurate model of the 2 environments. Furthermore, the 2 strategies have successfully described behavior in many inference problems in the past (e.g., Payne et al. 1988; Gigerenzer et al. 1999; Rieskamp 2006; Rieskamp and Otto 2006) and they do not represent arbitrary strategies applicable to the decision problem. However, we also tested an alternative SSL model that included a third strategy (see below).

Each strategy i has its expectancy $Q(i)_t$, representing the participant's degree of belief that using the strategy in the current context is appropriate (or in other words that it represents an accurate model of the environment). The probability of selecting a strategy at trial t is a function of its expectancy and the expectancies of all other strategies:

$$P(i)_t = \frac{Q(i)_t}{\sum_{j=1}^{J} Q(j)_t}, \qquad (3)$$

where J = number of strategies = 2 (note that in our case $P(\text{AUD})_t = 1 - P(\text{ALL})_t$). At $t = 1$, we assume that $Q(\text{ALL})_{t=1} = Q(\text{AUD})_{t=1} = 1$, so that $P(\text{ALL})_{t=1} = P(\text{AUD})_{t=1} = 0.5$. Because a strategy is selected with a certain probability in a particular trial, we simulated which strategy was actually selected randomly according to the specified probabilities (i.e., the simulation selected strategy i at trial t with probability $p(i)_t$). According to the 2 strategies, each stock offer can be assigned a strategy value (SV) representing the values of the cues:

$$\text{SV}(\text{ALL})_t = a_t + v_{1,t} + v_{2,t} + v_{3,t} + v_{4,t}, \qquad (4)$$

$$\text{SV}(\text{AUD})_t = a_t, \qquad (5)$$

where the strategy value of ALL depends on all cues and the strategy value of AUD depends only on the auditory cue. To determine the EV of the stock offer, the strategy value of the selected strategy i needs to be multiplied by the subjective value of that strategy, that is:

$$\text{EV}(\text{buy}|i)_t = \text{SV}(i)_t \times Q(i)_t. \qquad (6)$$

Due to learning, the Q-value for the appropriate strategy in an environment should converge to a value of 10, so that the EV of stock offers will approximate the actual reward structure of the environments (i.e., eq. 6 will represent the actual EV of a stock according to eqs 1 or 2, respectively). Likewise, the Q-values of the inappropriate strategy should converge to 0, since using this strategy would lead to an average payoff of 0. The probability of buying the current offer is determined by the softmax choice rule (Sutton and Barto 1998) comparing the EVs of buying the stock versus rejecting it (equivalent to receiving 0):

$$P(\text{buy})_t = \frac{1}{1 + e^{-\gamma \times [\text{EV}(\text{buy}|i)_t - \text{EV}(\text{reject})_t]}}, \qquad (7)$$

where γ is the sensitivity parameter (or inverse "temperature") estimated for each subject. The RPE is defined as the outcome (R) minus the EV (Sutton and Barto 1998):

$$\text{RPE}_t = I_t \times [R_t - \text{EV}(\text{buy}|i)_t], \qquad (8)$$

where I_t is an indicator function indicating whether the participant has actually bought ($I_t = 1$) or rejected ($I_t = 0$) the offer at trial t (this is introduced to set RPE to 0 in the case of rejecting the offer). Finally, SSL updates the Q-value of the selected strategy i for the next trial as follows:

$$Q(i)_{t+1} = Q(i)_t + \alpha \times \text{RPE}_t \times \text{SV}(i)_t, \qquad (9)$$

where α is the learning rate estimated for each participant. Note that strategy values need to be included so that the Q-values are updated according to the predictions of the strategy (e.g., if a person chooses the option opposite to a strategy's prediction, the RPE should influence the corresponding Q-value inversely).

SSL Variations

In addition to the described SSL model, we tested 3 variants of SSL. For the first alternative, the repertoire of strategies was extended by a third strategy VIS, which takes only visual cues into account. This strategy appears reasonable, because participants might have contrasted auditory against visual information. The strategy value for VIS in every trial is:

$$\mathrm{SV}\,(\mathrm{VIS})_t = v_{1,t} + v_{2,t} + v_{3,t} + v_{4,t}. \tag{10}$$

Note that although adding more strategies to SSL does not increase the number of free parameters, the model's complexity can increase because the model can potentially account for more choice patterns (Pitt and Myung 2002). The second SSL variation does not assume that only one strategy is taken at a time, but that the EV is a combination of all strategy values weighted by their probabilities (see, for instance, Wunderlich et al. 2011). The EV for this "probabilistic" SSL version thus becomes:

$$\mathrm{EV}\,(\mathrm{buy})_t = \sum_{j=1}^{J} \mathrm{SV}(j)_t \times Q(j)_t \times P(j)_t. \tag{11}$$

In accordance, the updating rule is changed such that all J strategies are updated in every trial weighted by the respective strategy selection probabilities:

$$Q(j)_{t+1} = Q(j)_t + \alpha \times \mathrm{RPE}_t \times \mathrm{SV}(j)_t \times P(j)_t. \tag{12}$$

The third variation combines these changes, that is, it is a probabilistic SSL model with 3 strategies.

ALM Model

We tested the predictions of SSL against an alternative learning model that assumes stock values to be predictable by a weighted linear combination of cues. This model has been shown to account for human behavior in multiple-cue inference task and is known as the ALM (Kelley and Busemeyer 2008) or adaptive network model (Gluck and Bower 1988; Rieskamp 2006). ALM compares the relationship between predictions and outcomes for each cue directly and independently. In more detail, ALM integrates all cues and cue weights into a linear additive function to generate the EV:

$$\mathrm{EV}\,(\mathrm{buy})_t = \sum_{m=1}^{M} w_{m,t} \times c_{m,t}, \tag{13}$$

where $w_{m,t}$ is the cue weight of cue m (M = number of cues = 5) at trial t and $c_{m,t}$, the prediction (stock trend or rating) of cue m at trial t (i.e., $c_{1,t} = a_t$; $c_{2,t} = v_{1,t}$; $c_{3,t} = v_{2,t}$ etc.). The cue weights thus determine the direction and magnitude of the impact of a cue's prediction on the overall EV. Initially, all weights are set to 1, assuming that at the beginning, participants have a weak belief in the positive validities of all cues (we ensured that the setting of initial weights and strategy expectations did not have a substantial effect on our results and conclusions). The softmax choice rule and the calculation of RPEs are equivalent to SSL (see eqs 7 and 8). ALM updates each cue weight independently taking the RPE and cue predictions into account (thus and similar to SSL, updating a cue against its own prediction is prevented) so that,

$$w_{m,t+1} = w_{m,t} + \alpha \times \mathrm{RPE}_t \times c_{m,t}. \tag{14}$$

ALM and SSL share the same number and type of free parameters: A learning rate and a sensitivity parameter.

Model Fitting and Model Comparison

We used maximum likelihood techniques to estimate the parameters of our models for each individual separately. As a goodness-of-fit measure, we calculated the log-likelihood of the data for all trials, given the model

$$\mathrm{LL}_{\mathrm{Model}} = \sum_{t=1}^{N} \ln[f_t(y|\theta)], \tag{15}$$

where N = number of trials = 160. $f_t(y|\theta)$ represents the probability with which the model predicts the choice y of the participant in trial t given the models' parameter set θ. For each participant and model, we estimated the parameters that maximized the likelihood of the data by means of the "fminsearchcon" algorithm as implemented in Matlab (MathWorks). Because the selected strategies for a trial were simulated according to probabilities specified by equation (3), the actual learning process depended considerably on the strategies selected. Therefore, we simulated the learning process for each participant repeatedly for 10 000 simulations per model and used the average model fit for model comparison and the average model variables (such as $\mathrm{EV}[\mathrm{buy}|i]_t$ or $P(\mathrm{ALL})_t$) for the fMRI analysis (see below).

For model comparison, we determined the models' deviances (see Lewandowsky and Farrell 2011):

$$\mathrm{deviance}_{\mathrm{Model}} = -2 \times \mathrm{LL}_{\mathrm{Model}}. \tag{16}$$

We tested each model against the deviance of a Baseline model that predicted all the 160 choices with a pure chance probability of 0.50:

$$\mathrm{deviance}_{\mathrm{Baseline}} = -2 \times N \times \ln(0.5). \tag{17}$$

A log-likelihood ratio test (with a χ^2-distributed test variable with 2 degrees of freedom for the 2 free parameters in the different SSL models and the ALM model, respectively) was used to test whether a particular model predicted choices above chance level for a particular participant (Table 1). SSL and ALM were also tested against each other using the deviances (note that using the Bayesian Information Criterion for comparing SSL with ALM would lead to the same conclusions as SSL and ALM have the same number of free parameters).

fMRI Data Acquisition and Preprocessing

Whole-brain fMRI data were collected on a 3-T Siemens Trio scanner using a 12-channel head coil. Echo-planar T_2*-weighted images were acquired using 40 axial slices and a voxel size of 2 × 2 × 2 mm plus a 1-mm gap between slices (further parameters included: Repetition time 2380 ms, echo time 25 ms, field of view 208 × 208, flip angle 90°). Slice orientation was tilted −30° to the anterior–posterior commissure axis to reduce signal drop in regions of the orbitofrontal cortex (Deichmann et al. 2003). Additionally, a high-resolution T_1-weighted image (voxel size 1 × 1 × 1 mm) was acquired for each subject to improve spatial preprocessing. Preprocessing of fMRI data was performed using SPM8 (Wellcome Trust Center for Neuroimaging, University College London). Preprocessing commenced with slice timing correction to the middle slice of each volume followed by

Table 1
Behavioral model comparison

	$\mathrm{SSL}_{2,\ \mathrm{determ.}}$	$\mathrm{SSL}_{3,\ \mathrm{determ.}}$	$\mathrm{SSL}_{2,\ \mathrm{probab.}}$	$\mathrm{SSL}_{3,\ \mathrm{probab.}}$	ALM
Deviance	118.19 (43.42)	116.02 (41.90)	119.90 (44.05)	119.27 (44.34)	140.83 (34.35)
n (best)	7	8	2	5	2
n (<baseline)	24	24	24	24	22
% correct	82.9 (9.4)	83.0 (9.5)	82.9 (9.4)	82.9 (9.4)	79.1 (8.5)

Note: Values in parentheses represent standard deviations.
SSL: strategy selection learning model (subscripts = the number of strategies and strategy selection rule); ALM: additive linear model; *n*(best): the number of participants for which the respective model performed best; *n*(<baseline): the number of participants for which the respective model performed significantly better than the baseline model; %correct: the average percentage of trials in which the respective model correctly predicted the decision (if buy/reject predictions were assigned according to $P(\mathrm{buy})_t > 0.5/<0.5$).

spatial realignment and unwarping to account for movement artifacts. The individual T_1-weighted image was coregistered to the mean functional image generated during realignment and then segmented into gray matter, white matter, and cerebrospinal fluid. Spatial normalization of functional images to the MNI space was achieved using the normalization parameters from the segmentation. Finally, images were smoothed by a Gaussian kernel of 8-mm full-width at half-maximum.

fMRI Data Analysis

The conventional statistical analysis of fMRI data was based on the general linear model (GLM) approach as implemented in SPM8. The GLM was set up to test the predictions of SSL with respect to the modulation of the fMRI BOLD signal at the time when the decision was made by 1) EV, 2) strategy selection, and 3) decision conflict, and at the time when feedback was provided by RPEs. Subject-specific design matrices were thus generated including an onset vector for estimating the average BOLD response during the decision phase. This onset vector was accompanied by 3 parametric modulators (Büchel et al. 1996) encoding the SSL-based trial-by-trial estimates of 1) the expected value EV(buy | $i)_t$, 2) the probability of selecting strategy ALL $P(\text{ALL})_t$, and 3) the conflict elicited by each decision, which we quantified based on $P(\text{buy})_t$ as

$$\text{conflict} = -|P\,(\text{buy})_t - 0.5|, \tag{18}$$

such that conflict was low when $P(\text{buy})_t$ was close to 0 or 1 and high when $P(\text{buy})_t$ was close to 0.5. Correlations between parametric modulators were very low in general (still, we omitted the automatic, step-wise orthogonalization of parametric modulators in SPM). The GLM further comprised an onset vector for the time point of the button press together with a parametric modulator encoding the specific response (buy or reject) and an onset vector for the outcome phase together with a parametric modulator encoding the reward prediction error, RPE_t (for an illustration of all GLM regressors, see Supplementary Fig. 1). At group level, we used the full factorial design as implemented in SPM8 (controlling for nonsphericity of the error term) to test for effects related to the parametric modulators EV, strategy selection, decision conflict, and RPE. This analysis was repeated for the SSL model with 3 strategies using the sum of probabilities for strategies ALL and VIS (i.e., $P(\text{ALL})_t + P(\text{VIS})_t$) as parametric modulator for strategy selection (since both strategies rely on visual information in contrast to AUD; see Supplementary Figs 2 and 4). The statistical threshold for the imaging results was set to $P < 0.05$, family-wise error (FWE) rate corrected for spherical search volumes (sphere radius: 10 mm) based on previous studies that tested for comparable effects of interest (EV and/or prediction error, decision conflict, visual attention): Center coordinates of spheres were [$x = -3$, $y = 42$, $z = -6$] for vmPFC (Chib et al. 2009), [±14, 10,−10] for VS (O'Doherty et al. 2004), [−6, 24, 38] for ACC (Venkatraman et al. 2009), and [±44, −75, −10] for lateral occipital complex (LOC; Rose et al. 2005). Regions beyond those for which we had a priori hypotheses were reported if they survived a threshold of $P < 0.05$, FWE-corrected for the whole brain. For display purposes, we used a threshold of $P < 0.001$ (uncorrected) with 10 contiguous voxels unless stated otherwise. Activations are depicted on an overlay of the mean structural T_1-weighted image from all participants. Images are presented in neurological convention.

fMRI Model Comparison

In addition to the behavioral model comparison, we tested SSL against ALM on the basis of fMRI data. Since both models make trial-by-trial predictions on EV, decision conflict, and RPE, we tested which of the 2 models better accounts for fMRI signals in the hypothesized brain areas (vmPFC and VS for EV and RPE, respectively; ACC for decision conflict). We used a Bayesian model estimation and selection approach (Penny et al. 2005; Stephan et al. 2009; Rosa et al. 2010), which has already been applied in model-based fMRI research on value-based learning and decision-making (Hare et al. 2011; Wunderlich et al. 2011). The approach comprises 3 steps: First, a conjunction analysis (threshold: $P < 0.001$, uncorrected) is used to determine voxels within the hypothesized brain areas that are predicted by both learning models (i.e., SSL and ALM). Secondly, the GLMs for the 2 learning models are re-run using the Bayesian estimation procedure as described in Penny et al. (2005). This analysis is restricted to the voxels specified by the conjunction analysis. Thirdly, the resulting exceedance probability maps are compared using the random-effects Bayesian model selection (BMS) approach as described in Stephan et al. (2009). We replicated the fMRI model comparison for the SSL model with 3 strategies (Supplementary Fig. 3).

Dynamic Causal Modeling

We expected participants to learn the right model representing the compensatory or the noncompensatory environment, that is, they should learn selecting ALL in the compensatory and AUD in the noncompensatory environment. Importantly, in our task, ALL makes use of both auditory and visual information to generate the EV, but AUD solely relies on the auditory cue. We used DCM (Friston et al. 2003) to test whether the selection of a particular strategy is also reflected in the interaction of sensory and reward systems in the brain. Briefly, DCM models the neural dynamics between regions of interest (ROIs) by 3 different sets of parameters: 1) direct inputs of external variables on ROIs, 2) context-independent effective connectivity between ROIs, and 3) context-dependent modulations of this connectivity. In our experiment, we expected the effective connectivity from a sensory area to a reward area to be positively modulated by the value information that is conveyed from the sensory area. For instance, the connectivity between auditory and reward regions should be more positive (negative) if the auditory cue is positive (negative). If, however, the sensory information is not used to generate the value representation (as a consequence of the selected strategy), this context-dependent modulation should not be present. Therefore, we predicted a reduced modulation of the coupling between the visual and reward regions when the strategy AUD was selected.

To test this hypothesis, we first set up a new GLM to separate 1) trials where the strategy ALL was used from those where AUD was used and 2) modulations by value information conveyed by the 2 different sensory systems (auditory and visual). The first separation was realized by splitting trials into ALL and AUD trials based on $P(\text{ALL})_t$ (i.e., split at $P(\text{ALL})_t = 0.5$), for which 2 onset vectors of the decision phase were generated. The second separation was realized by accompanying these onset vectors by 2 parametric modulators coding for the sensory-specific value information:

$$\text{Value}\,(\text{visual})_t = v_{1,t} + v_{2,t} + v_{3,t} + v_{4,t}, \tag{19}$$

$$\text{Value}\,(\text{auditory})_t = a_t. \tag{20}$$

Note that Value(visual)_t and Value(auditory)_t are equivalent to the strategy values for VIS and AUD, respectively (eqs 10 and 5). In the next step, we extracted fMRI BOLD time series from visual, auditory, and reward areas. For anatomical plausibility, we first restricted our search for relevant brain areas: For the primary visual and auditory input systems, we created masks including only the primary visual (V1) and primary auditory cortices (A1), respectively, as defined by an anatomical atlas (Tzourio-Mazoyer et al. 2002); for the reward system, bilateral 10-mm spheres were placed at the VS coordinates as defined above (we selected the VS as reward structure as we obtained the strongest effects related to EV here). Secondly, within each of these ROIs, we identified the peak activations of specific effects (i.e., main effect of decision phase for visual and auditory regions; effect of EV for reward region) separately for each participant. Thirdly, we extracted fMRI time series from 4-mm spheres placed around the individual peak coordinates.

We then defined a set of candidate DCM models to test them against each other (using BMS) to identify the model that provided the Bayesian optimal balance between goodness of fit to the data and model parsimony (Penny et al. 2004; Stephan et al. 2009, 2010). The 8 identified models shared the following features: All the models consisted of 3 ROIs (V1, A1, and VS); V1 and A1 received the external driving input (onset vector of the decision phase) and had directed intrinsic connections toward the VS; these connections were modulated by their sensory-specific value information (i.e., Value(visual) on V1–VS; Value(auditory) on A1–VS). The models differed in terms of 3 variations: 1) whether bidirectional intrinsic connections between V1 and A1 were included or not, 2) whether backward intrinsic connections from VS to V1 and A1 were included or not, and 3) whether Value(visual)/Value(auditory) could also modulate the connections A1–VS/V1–VS or not (see Fig. 6*A* and Supplementary Table 1). We identified the optimal model using the BMS approach (Stephan et al. 2009) (see Fig. 6*B* for the optimal model and Supplementary Fig. 6 for the model comparison). In the final step, we compared the best model's estimated parameters of the modulation of V1–VS and A1–VS connections by Value(visual) and Value(auditory), respectively, for ALL trials against AUD trials (Fig. 6*C*). We replicated the DCM analysis for the SSL model with 3 strategies, comparing AUD trials with ALL and VIS trials together (Supplementary Fig. 5).

Results

Behavioral Results

The participants played 160 rounds of the stock-buying task (Fig. 1) while lying in the MR scanner, 80 trials in the compensatory and 80 trials in the noncompensatory conditions. The order of environments was counterbalanced across participants. To test for learning within each environment, the trials were separated into 16 consecutive blocks of 10 trials, and the average accuracy (i.e., buying good and rejecting poor stocks) per block was introduced into an $8 \times 2 \times 2$ analysis of variance with 8 trial blocks and 2 environments as within-subject factors and 2 environment orders as a between-subjects factor. This analysis revealed a main effect of trial blocks indicating a strong learning effect within each environment ($F_{7,154} = 14.70$, $P < 0.001$).

Figure 2*A* shows that participants increased their performance in the initial 30 trials before reaching a plateau. In the ninth trial block, performance broke down due to the change in the environment, followed by an increase in performance comparable with the first half. A significant interaction of environment × environment order was also found, suggesting a higher average performance in both groups in the first phase of the experiment compared with the second phase ($F_{1,22} = 5.30$, $P = 0.031$). This effect appears to be driven by the severe drop in accuracy in the ninth block. Finally, the interaction of trial block × environment was significant ($F_{7,154} = 2.34$, $P = 0.027$), indicating a somewhat stronger learning effect in the noncompensatory environment. Importantly, the main effects of environment and environment order were not significant, which implies equivalent performance in both environments and for both groups.

Behavioral Model Fit and Comparison

The strategy-based learning model SSL, as it is specified for the current study (see Materials and Methods), assumes participants to select from among 2 strategies: ALL, which sums over all cue values treating them as equally important, and AUD, which focuses exclusively on the auditory cue and does not take the visual cues into account. Thus, ALL and AUD represent 2 models that match the compensatory and the noncompensatory environments, respectively. Therefore, we expected participants to learn to select ALL in the compensatory and AUD in the noncompensatory environments. Note that the 2 strategies made different predictions in 50% of all trials. This allowed us to compare behavior with predictions from the SSL model as follows: We approximated the frequency of selecting a strategy *i* per block as the ratio between the number of trials in which the actual behavior matched exclusively the prediction of strategy *i* and the total number of trials with diverging strategy predictions. As shown in Figure 2*B*, SSL-based strategy selection probabilities closely matched these frequencies in both environments and groups (average correlation per participant was 0.79). Furthermore, by comparing SSL with a Baseline model, we found that SSL predicted behavior significantly better than chance level in all the 24 participants (Table 1).

We also tested a variation of SSL that included a third decision strategy, VIS, which takes only visual information into account. Given the distinction between visual and auditory information, it appears reasonable to assume that participants might have considered this strategy as well. Furthermore, VIS can perform well in the compensatory environment (as it considers 4 of the 5 relevant cues), although it is inferior to ALL. Inspection of the strategy selection probabilities of this 3-strategy SSL version suggests that participants might have alternated between ALL and VIS in the first blocks of the compensatory environment, but then learned to select the more accurate strategy ALL (Fig. 2*C*; note, however, that this suggestion is based on estimated strategy probabilities but not frequencies, which cannot be calculated for the 3-strategy version due to the high overlap of choice predictions for ALL and VIS). In line with this, no participant reported having only considered visual information at the end of the compensatory environment (Supplementary Table 2). When comparing this model with the 2-strategy model by means of their deviances, the 3-strategy version appears to perform slightly (but not significantly) better (Table 1). Note, however, that this model is also more complex, given the inclusion of a third strategy. We further tested 2 probabilistic SSL variations (1 with 2 strategies and 1 with 3 strategies) that assumed the EV to be influenced by all strategies at the same time, weighted by their selection probabilities. These models made very similar predictions compared with the "deterministic" SSL models, but performed slightly (not significantly) worse (Table 1).

Finally, we tested an alternative learning approach (ALM). This model does not assume a strategy selection process, but relies on assigning weights to each cue depending on how well each cue predicted the outcome in the past. These cue weights determine the impact of each cue on the EV of the stock and are updated after feedback. Figure 2*D* illustrates the development of the 5 cue weights during the experiment and it is evident that the model captured the coarse pattern of choice behavior: In the compensatory environment, all weights were approximately equal; in the noncompensatory environment, the auditory cue's weight was much higher than the visual cues' weights. Accordingly, ALM predicted choices above chance level in 22 of 24 participants. When comparing ALM with SSL by means of their deviances, however, the strategy-based learning model proved to be clearly superior (Table 1; all paired *t*-tests between ALM and SSL variants, $P < 0.001$). Furthermore, SSL made significantly better predictions of participants' choices (Table 1; all paired *t*-tests between ALM and SSL variants, $P < 0.001$).

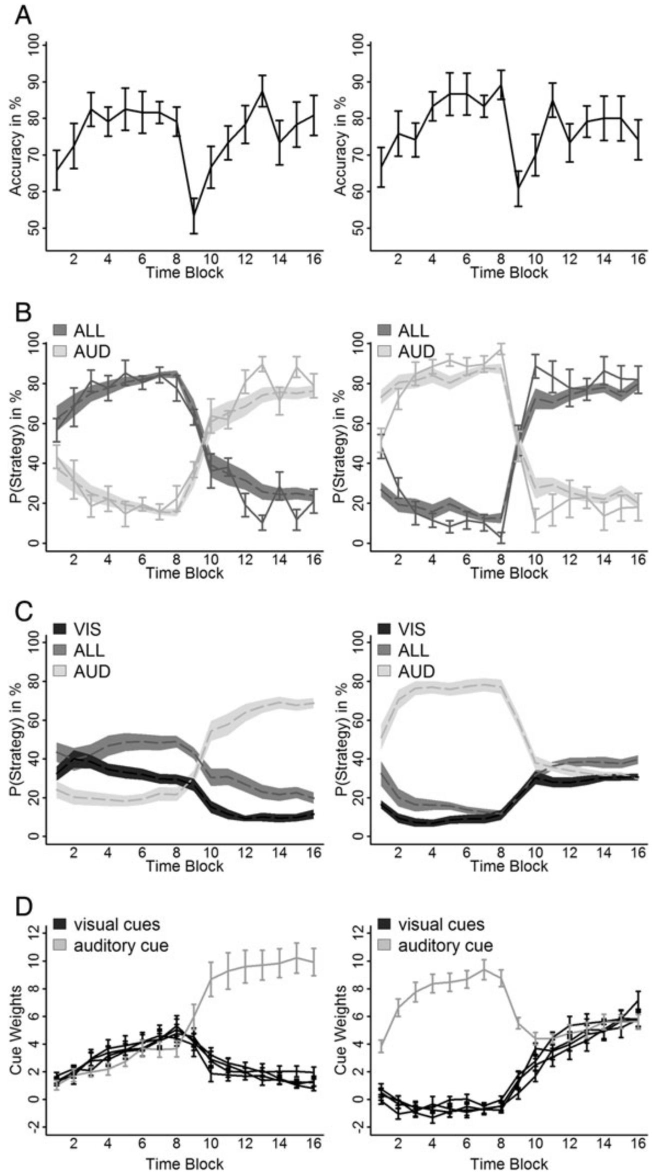

Figure 2. Behavioral results and model fit. All the panels on the left refer to the group that first encountered the compensatory environment; all the panels on the right refer to the group that first encountered the noncompensatory environment. (*A*) Average performance per block (10 consecutive trials). (*B*) Comparison of frequencies (continuous lines with error bars) and SSL-based probabilities (dashed lines with shaded areas) of selecting strategy ALL and AUD. (*C*) Strategy selection probabilities for the 3-strategy SSL model. (*D*) Development of cue weights over time according to the ALM.

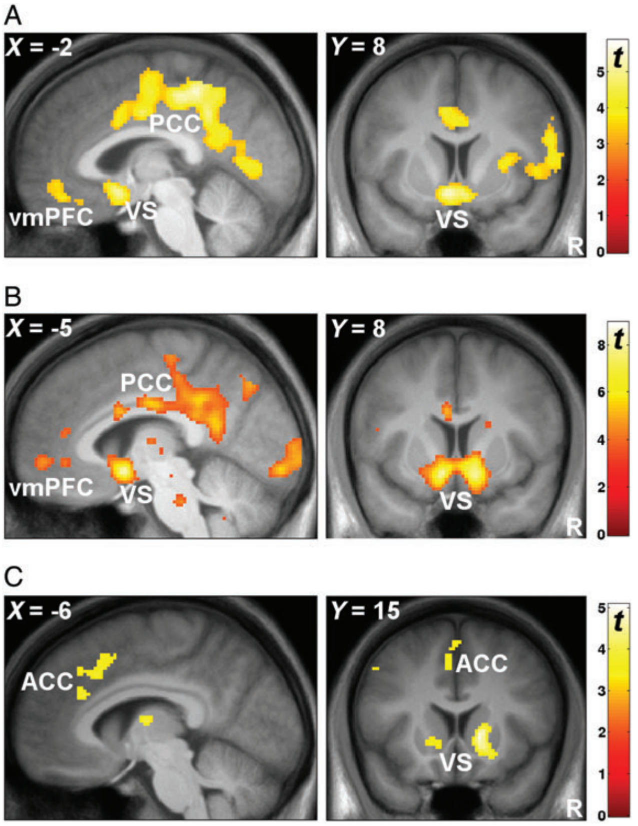

Figure 3. Neural correlates of SSL-based EV, RPE, and decision conflict. (*A*) EV during the decision phase was associated with fMRI signals in VS, vmPFC, and PCC. (*B*) *A* similar brain circuit was activated during the outcome phase as a function of the RPE. (*C*) ACC and VS encoded the decision conflict during the decision phase.

fMRI Results: Expected Value, Prediction Error, and Decision Conflict

SSL makes trial-by-trial predictions for the EV of each stock offering and for the RPE of each reward obtained. We tested whether the neural correlates of these model variables can be linked to the dopaminoceptive reward system of the brain as has been shown for other RL scenarios (O'Doherty et al. 2003, 2004; Hampton et al. 2006; Pessiglione et al. 2008; Gläscher et al. 2009; Daw et al. 2011). For the fMRI analysis, we thus included EV and RPE as parametric modulators of the decision and outcome phases, respectively. Note that we report fMRI results for the original 2-strategy SSL model in the main text, but replicated all results for the 3-strategy version (Supplementary Figs 2–5). We found significant BOLD signals associated with both EV and RPE in the bilateral VS and vmPFC (Fig. 3*A*,*B*; Table 2 provides statistics for all conventional fMRI analyses). The posterior cingulate cortex (PCC) were also activated, which is in line with previous neuroimaging studies (Hampton et al. 2006; Kable and Glimcher 2007; Gläscher et al. 2009; Peters and Büchel 2009).

SSL further allows quantification of the degree of decision conflict based on the model-based probability $P(\text{buy})_t$ that the current stock offer is bought or not: If $P(\text{buy})_t$ is either very high or very low, the decision is comparatively easy, since the stock is either bought or rejected with high probability (i.e., high confidence); if, however, $P(\text{buy})_t$ is at an intermediate level, the decision is comparatively difficult (see eq. 18). We implemented decision conflict as a parametric modulator at the decision phase in our fMRI analysis and obtained significant activation in the ACC (Fig. 3*C*) replicating previous research (Botvinick et al. 2004; Pochon et al. 2008; Pine et al. 2009; Venkatraman et al. 2009). Note that this effect cannot be explained by potential differences in reaction times, since responses were only possible after 5 s of the decision phase (Pochon et al. 2008; Grinband et al. 2011). A second cluster of activation was located in the VS (more dorsal and anterior than the clusters linked to EV and RPE). This additional finding might be related to outcome uncertainty: Decisions are complicated by uncertainty about the value of a stock offer, and this uncertainty has been linked to tonic activity in

Table 2
Peak coordinates and statistics of fMRI analyses

Contrast	Name of region	MNI coordinates in mm			Statistics		
		x	*y*	*z*	*t*-value	*Z*-value	*P*-value
Expected value	Left ventral striatum	−6	8	−12	5.35	4.97	<0.001
	Right ventral striatum	6	8	−12	4.74	4.47	0.001
	vmPFC	−2	42	−14	3.85	3.70	0.010
	Left PCC	12	−22	42	5.68	5.25	0.004
	Right PCC	−10	−50	24	5.86	5.39	0.002
	Precuneus	−6	−38	48	5.53	5.13	0.008
	Angular gyrus	−42	−74	34	5.26	4.91	0.021
Reward prediction error	Left ventral striatum	−10	4	−10	8.93	7.56	<0.001
	Right ventral striatum	12	6	−12	8.08	7.01	<0.001
	vmPFC	−10	40	−6	3.82	3.67	0.011
	PCC	0	−34	28	6.69	6.02	<0.001
	Medial temporal gyrus	58	−40	−12	5.64	5.22	0.005
	Fusiform gyrus	−48	−52	−16	6.62	5.97	<0.001
	Left occipital gyrus	−16	−100	8	6.53	5.90	<0.001
	Right occipital gyrus	26	−96	4	6.11	5.59	0.001
Decision conflict	ACC	−6	22	38	3.77	3.63	0.012
	Left ventral striatum	−10	18	−6	3.55	3.43	0.022
	Right ventral striatum	14	14	−2	4.98	4.68	<0.001
Probability of selecting ADD	Left LOC	−46	−72	−6	3.91	3.75	0.008
	Right LOC	46	−82	−10	4.08	3.90	0.005

Note: Small-volume-corrected regions (ventral striatum, vmPFC, ACC, and LOC) are listed first followed by other regions that survived a threshold of $P < 0.05$ FWE-corrected at the whole brain.

the dopaminoceptive system (Fiorillo et al. 2003; Preuschoff et al. 2006; Bach and Dolan 2012).

fMRI Results: Model Comparison

Next, we compared the learning models, SSL and ALM, in terms of how well they account for the fMRI signals in reward-based (vmPFC and VS) and conflict-related (ACC) brain regions. For reward-based effects, we used a Bayesian model estimation and selection approach (Penny et al. 2005; Stephan et al. 2009) that determines the models' exceedance probabilities for conjointly activated voxels within the vmPFC and VS. The conjunction analysis for SSL- and ALM-based EV and prediction error showed that both models accounted for reward-related fMRI signals in the VS (Fig. 4*A*). The BMS analysis, however, suggested a higher exceedance probability for SSL (78.8%) when compared with the ALM (21.2%) in this region (Fig. 4*B*). For conflict-related effects in the ACC, we could not implement this approach, simply because the ALM-based regressor for decision conflict did not account for fMRI signals in the ACC even at a very liberal threshold of $P < 0.01$ (uncorrected; Fig. 4*C*). Together, these results indicate that SSL better explained the fMRI signals in the proposed brain areas.

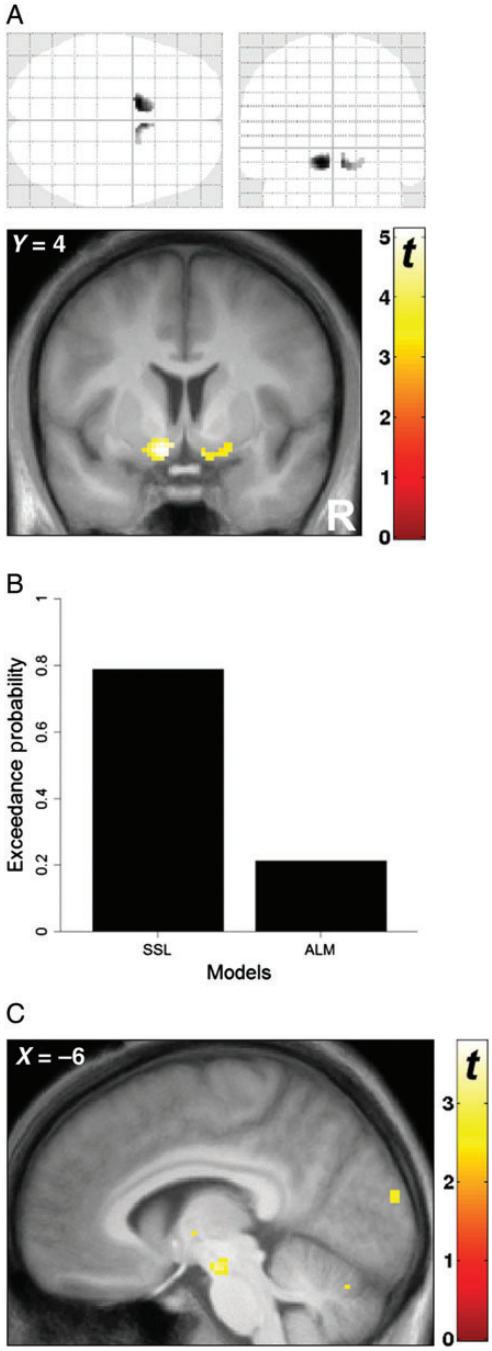

Figure 4. fMRI model comparison. (*A*) The conjunction analysis of SSL- and ALM-based EV and prediction error regressors showed that both models accounted for reward-related fMRI signals in VS. (*B*) A Bayesian model selection suggested, however, that SSL provided a more accurate fit to the data (shown in *A*) than the ALM. (*C*) The ALM-based regressor for decision conflict did not correlate with activity in the ACC even at the very liberal threshold of $P < 0.01$ (uncorrected) shown here.

fMRI Results: Strategy Selection

The behavioral modeling results suggest that participants learned to use a single-cue strategy (AUD) in the noncompensatory environment, in which visual cues were irrelevant for estimating stock values. We propose that using ALL as the decision strategy requires a greater allocation of attention to visual information than using AUD, which should be linked to higher activation in the visual system (O'Craven et al. 1997; Büchel et al. 1998; Shomstein and Yantis 2004). In fact, we do not know which strategy was applied in a single trial, but SSL makes probabilistic predictions for how likely ALL was used in every trial. We implemented this probability, $P(\mathrm{ALL})_t$, in

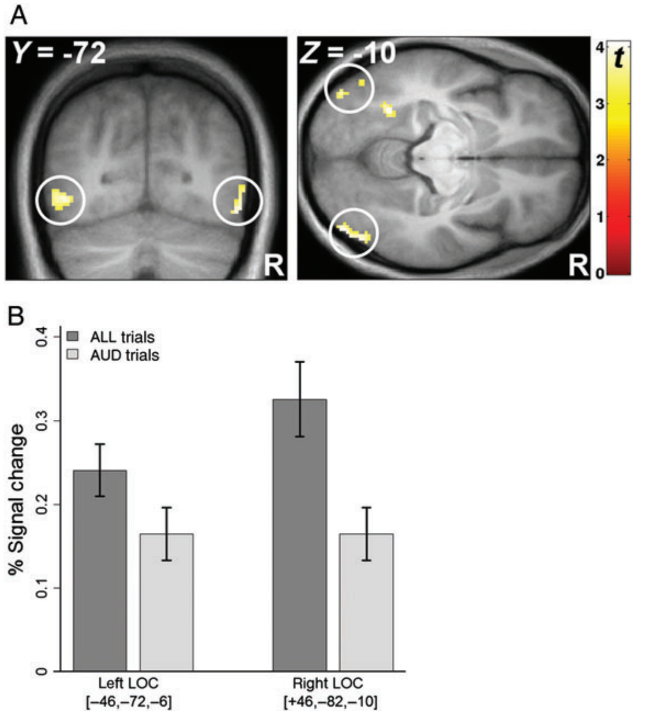

Figure 5. The influence of strategy selection on fMRI signals in the visual system. (*A*) The higher the probability of selecting strategy ALL, the higher the BOLD signal in bilateral LOC. (*B*) BOLD signals extracted from the peak coordinates of the effect in *A* separately for ALL and AUD trials (as defined by a median split at *P*(ALL) = 0.5).

the fMRI model as a parametric modulator of the decision phase to test our hypothesis. As shown in Figure 5*A*, higher probability of selecting ALL was indeed associated with higher BOLD signals in bilateral LOC. Figure 5*B* shows the average percent signal change at the peak coordinates in the left and right LOC separately for ALL and AUD trials (defined by a median split at $P(\text{ALL})_t = 0.5$): Activity in LOC was significantly positive in both trial types (all effects $P < 0.001$), but the strength of the signal was modulated by the selected strategy. We also tested for the reverse contrast (equivalent to $P(\text{AUD})_t$), but did not find any evidence for activation in gray matter structures even at a very liberal threshold of $P < 0.01$ (uncorrected).

DCM Results

The selection of a particular strategy should not only influence the allocation of attention to different sources of information, but also how this information is utilized to generate value expectations. Therefore, we investigated how the selection of strategies influences the neural coupling between sensory and reward systems in the brain. We employed DCM, which among other features allows making inferences on how the intrinsic connectivity between brain regions changes in relationship to experimental manipulations. We expected that the connectivity between cue-specific sensory and reward areas is enhanced when the (sensory-specific) value information is positive: This is because signals from sensory systems should always inform the reward system, but only in the case of positive information, the signal increase in sensory areas (due to stimulus presentation) should be followed by a signal increase in reward areas (due to a positive EV). If the reward system does not make use of the sensory information, however, we hypothesized that this modulation would break down. In our design, such a break down should occur for the coupling of visual and reward areas when strategy AUD is used (as here the EV is solely based on the auditory cue).

To test this prediction, we set up a DCM model that included primary visual (V1) and primary auditory cortex (A1) as sensory and VS as reward brain structure for which time series were extracted (see Materials and Methods for details). V1 and A1 received the driving input and had directed connections to VS. These connections were modulated by value information conveyed by the respective sensory system, that is, the V1–VS connection was modulated by Value(visual) and the A1–VS connection by Value(auditory) (see eqs 19 and 20). Importantly, we analyzed 2 separate sets of models for ALL and AUD trials to compare the estimated DCM parameters of the 2 strategies. Before looking at the DCM model parameters, we identified an optimal model among 8 candidates (Supplementary Table 1) by testing them against each other using BMS (Stephan et al. 2009). Figure 6*A* illustrates which connections were included in all models (continuous lines) and which were variable (dashed lines). Figure 6*B* shows the best model according to BMS (across both trial types as well as for ALL and AUD trials independently; see Supplementary Fig. 6) together with the average connection weights separated for ALL (left panel) and AUD trials (right panel). We used the parameters from the best model to test our hypothesis regarding the difference between ALL and AUD trials in connectivity modulation. We found a strong positive modulation of the V1–VS connection by Value(visual) for ALL trials ($t_{(23)} = 4.82$; $P < 0.001$) that was absent in AUD trials ($P = 0.813$; Fig. 6*C*). The direct comparison of parameters between different trial types was also significant ($t_{(23)} = 4.21$; $P < 0.001$). On the contrary, the A1–VS connection was positively modulated by Value (auditory) for both trial types (ALL trials: $t_{(23)} = 3.11$; $P = 0.005$; AUD trials: $t_{(23)} = 3.27$; $P = 0.003$), and the comparison of parameters did not reveal significant differences ($P = 0.161$). There was only one further connection weight that slightly differed between ALL and AUD trials: The intrinsic connection from V1 to A1 was less negative in ALL trials ($t_{(23)} = 2.12$; $P = 0.045$). DCM model #2, which was very similar to the best model but did not include backwards projections from VS to the sensory areas (Supplementary Table 1), also performed well in terms of BMS (Supplementary Fig. 6) and actually was the preferred model in 9 of the 24 participants. Therefore, we repeated our DCM comparison between ALL and AUD trials using model #2 and again found a difference in the value-related modulation of the V1–VS connection ($P < 0.001$), but not of the A1-VS connection ($P = 0.131$).

Discussion

We have shown that a computational model of adaptive strategy selection is appropriate to describe the cognitive processes of learning and decision-making in a complex, multiple-cue learning context as well as to reveal the underlying neural mechanisms. Activation in the VS and the vmPFC were correlated with the SSL-based regressors of EV and RPE and ACC activity reflected decision conflict. By providing cues in separate sensory domains and changing the reward contingencies of these cues, we could show that the selection of a

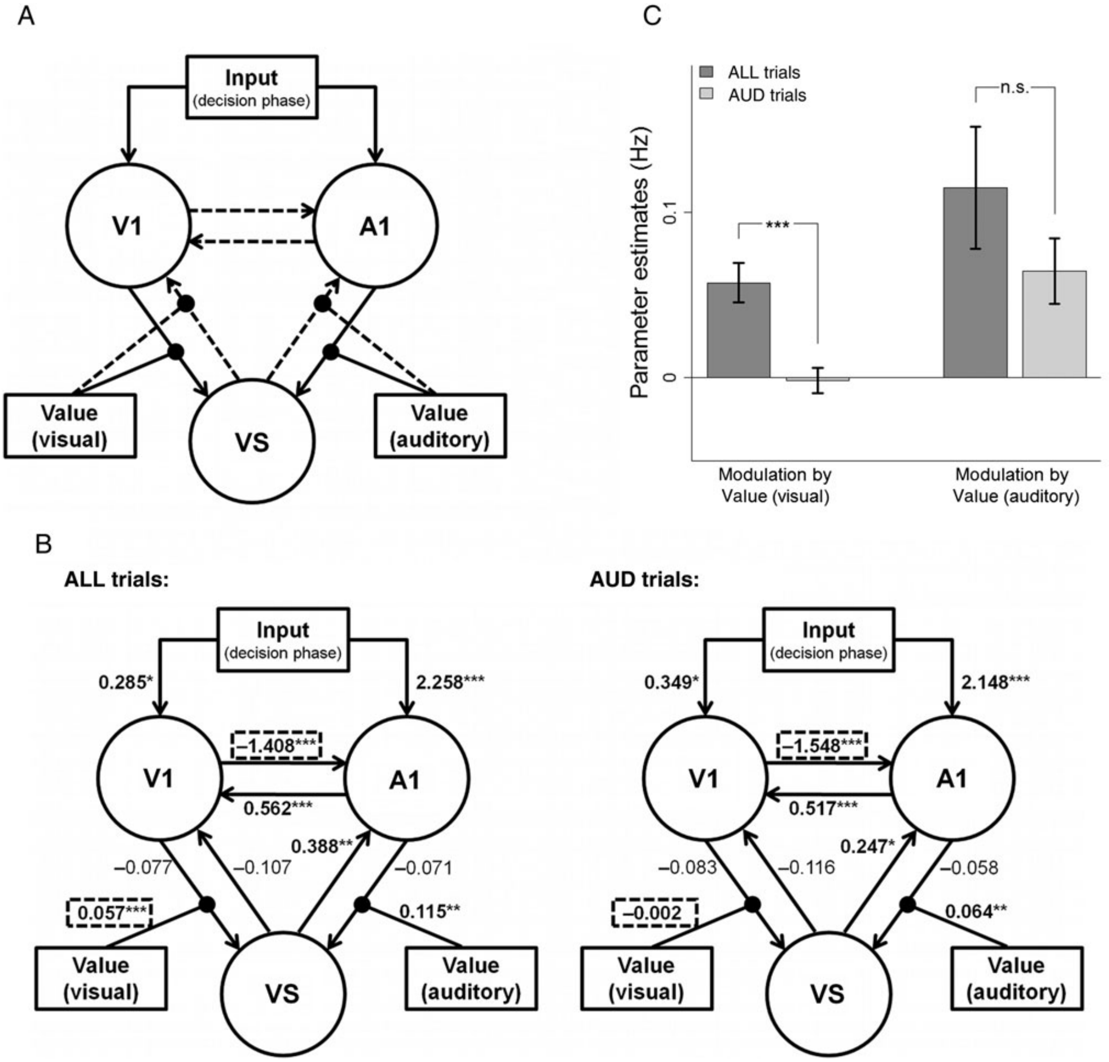

Figure 6. DCM model and results. (*A*) Overview of the variables, regions of interests, and connections used for the DCM analyses. Continuous lines refer to connections included in all 8 models and dashed lines refer to model-specific connections (Supplementary Table 1). (*B*) Illustration of the best performing model with connection weights for ALL (left panel) and AUD trials (right panel). Significant differences between the 2 trial types are indicated by dashed boxes around the respective coefficient. (*C*) Parameter estimates for the modulation of the connection between V1–VS and A1–VS by Value(visual) and Value(auditory), respectively, for the DCM models shown in *B* (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$).

particular strategy influences the reliance on different sources of information for generating value expectations.

In general, models of RL have been successfully applied to account for human learning behavior and to track its neural correlates. Whereas rather simple model-free RL approaches might be sufficient to understand elemental learning phenomena such as classical and instrumental conditioning (O'Doherty et al. 2003, 2004; Gläscher and Büchel 2005; Pessiglione et al. 2008), model-based RL seems to better explain human performance in more complex settings (Hampton et al. 2006; Behrens et al. 2007; Gläscher et al. 2010; Daw et al. 2011). Model-based RL assumes people to acquire a "model" representation of the environment, which allows them to infer the best action. An important question is how the acquisition of an accurate internal model takes place, that is, how people learn to use relevant and to ignore irrelevant information in complex environments. One solution to this task is to assign weights to each piece of information and to adjust these weights according to feedback. An alternative solution is to generate decision strategies that appear adequate for solving the problem. Here, learning takes place by selecting a particular strategy and evaluating its adequacy based on the outcome. We compared these 2 cognitive models (ALM and SSL) and found that SSL described behavioral and neuronal responses better in the context of our paradigm. An explanation for this result could be that ALM requires the decision maker to keep each of the 5 cue predictions in mind until the outcome is revealed in order to update each cue weight separately. In contrast to this cognitively very demanding approach, SSL only requires remembering the prediction of a single (selected) strategy. Hence, a strategy-based learning approach might be particularly useful to predict human decisions when the amount of potentially relevant information is very large. The SSL approach becomes, however, problematic if the contingency rules between cues and outcomes are more difficult than in our task (e.g., if cue validities are graded). Here, the limited set of strategies we assumed would not be sufficient. This raises the questions of how many strategies people possess and how new strategies are acquired. SSL does not make explicit predictions about the number of strategies. The theory is imbedded in the "bounded rationality" research framework, in which the existence of simple decision strategies that reflect adaptations to our decision environments has been proposed repeatedly (Simon 1956; Tversky and Kahneman 1974; Payne et al. 1988; Gigerenzer et al. 1999). The strategies used in the present study are similar to the strategies already tested in this framework and are also obvious, given the dissociation between auditory and visual information in our design. Nevertheless, future studies should directly investigate how many strategies people consider in common decision scenarios and whether strategies are refined or new strategies are acquired if the standard repertoire of strategies fails to provide sufficient results (cf. Scheibehenne et al. 2013).

The probability of using the multiple-cue strategy ALL was associated with the fMRI BOLD signal in bilateral LOC. We assume this to be an effect of shifted attention (O'Craven et al. 1997; Büchel et al. 1998; Shomstein and Yantis 2004): As people learned to use AUD in the noncompensatory environment, their attention was focused on the auditory cue and visual information was disregarded. This finding is consistent with a recent fMRI study, which demonstrated that in memory-based decisions, the reactivation of a specific sensory region depends on the relevance of the sensory-specific cues (Khader et al. 2011). Note, however, that in the study by Khader and colleagues, the use of the decision strategy was instructed, and the relevance of information did not have to be learned via feedback. In addition, our data show that the value-related modulation of the connectivity between sensory and reward areas disappears if the selected decision strategy does not require the respective sensory information to generate value expectations: When participants used strategy AUD (as indicated by SSL), the value information conveyed in the visual cues did not modulate the connectivity between V1 and the VS anymore. On the contrary, auditory information was relevant to the A1–VS coupling across both strategies just as both strategies required taking auditory information into account. These results promote our understanding of how humans extract the relevant from the large amount of available information in the environment to motivate their decisions. The data agree with recent literature toward a critical impact of attention on the neural circuitry of value computation (Hare et al. 2009, 2010; Krajbich et al. 2010; Krajbich and Rangel 2011; Lim et al. 2011).

Beside the model-based RL approaches we considered (SSL and ALM), there are many other learning theories that might account for the effects in our experiment. One attractive alternative could be learning models that are inspired by Bayesian probability theory, as these models would allow formulating and testing hypotheses on the relevance of auditory and visual information (Dayan et al. 2000; Yu and Dayan 2005; Gershman and Niv 2010). If the Bayesian learner thus discovers the irrelevance of visual information in the noncompensatory environment, he/she can save cognitive capacity by drawing attention to the auditory cue.

To conclude, our data suggest that by means of adaptive strategy selection, humans structure their environment when there are multiple sources of information available. Attention is focused on the putatively relevant information as reflected by neural activity in the respective sensory systems. Similarly, effective connectivity between sensory- and reward-related brain structures is positively affected by value as long as the specific sensory information is considered relevant to the decision.

Supplementary Material

Supplementary material can be found at: http://www.cercor.oxfordjournals.org/.

Funding

This work was supported by the Deutsche Forschungsgemeinschaft (GRK 1247/1, TRR 58, and SFB 936 TP A6 to S.G. and C.B.), the Bundesministerium für Bildung und Forschung (01GQ0912 to C.B.), and the Swiss National Science Foundation (100014_126721 to J.R.).

Notes

Conflict of Interest: None declared.

References

Bach DR, Dolan RJ. 2012. Knowing how much you don't know: a neural organization of uncertainty estimates. Nat Rev Neurosci. 13:572–586.

Bayer HM, Glimcher PW. 2005. Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron. 47:129–141.

Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS. 2007. Learning the value of information in an uncertain world. Nat Neurosci. 10:1214–1221.

Botvinick MM, Cohen JD, Carter CS. 2004. Conflict monitoring and anterior cingulate cortex: an update. Trends Cogn Sci. 8:539–546.

Büchel C, Josephs O, Rees G, Turner R, Frith CD, Friston KJ. 1998. The functional anatomy of attention to visual motion. A functional MRI study. Brain. 121(Pt 7):1281–1294.

Büchel C, Wise RJ, Mummery CJ, Poline JB, Friston KJ. 1996. Nonlinear regression in parametric activation studies. Neuroimage. 4:60–66.

Chib VS, Rangel A, Shimojo S, O'Doherty JP. 2009. Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. J Neurosci. 29: 12315–12320.

Damasio AR. 1994. Descartes' error: emotion, reason, and the human brain. New York: Putnam.

Daw ND, Doya K. 2006. The computational neurobiology of learning and reward. Curr Opin Neurobiol. 16:199–204.

Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ. 2011. Model-based influences on humans' choices and striatal prediction errors. Neuron. 69:1204–1215.

Daw ND, Niv Y, Dayan P. 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci. 8:1704–1711.

Dawes RM. 1979. The robust beauty of improper linear models in decision making. Am Psychol. 34:571–582.

Dayan P, Kakade S, Montague PR. 2000. Learning and selective attention. Nat Neurosci. 3(Suppl):1218–1223.

Deichmann R, Gottfried JA, Hutton C, Turner R. 2003. Optimized EPI for fMRI studies of the orbitofrontal cortex. Neuroimage. 19:430–441.

Fellows LK. 2011. Orbitofrontal contributions to value-based decision making: evidence from humans with frontal lobe damage. Ann N Y Acad Sci. 1239:51–58.

Fiorillo CD, Tobler PN, Schultz W. 2003. Discrete coding of reward probability and uncertainty by dopamine neurons. Science. 299:1898–1902.

Friston KJ, Harrison L, Penny W. 2003. Dynamic causal modelling. Neuroimage. 19:1273–1302.

Gershman SJ, Niv Y. 2010. Learning latent structure: carving nature at its joints. Curr Opin Neurobiol. 20:251–256.

Gigerenzer G, Goldstein DG. 1996. Reasoning the fast and frugal way: models of bounded rationality. Psychol Rev. 103:650–669.

Gigerenzer G, Todd PM, ABC Research Group. 1999. Simple heuristics that make us smart. New York: Oxford University Press.

Gläscher J, Büchel C. 2005. Formal learning theory dissociates brain regions with different temporal integration. Neuron. 47:295–306.

Gläscher J, Daw N, Dayan P, O'Doherty JP. 2010. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron. 66:585–595.

Gläscher J, Hampton AN, O'Doherty JP. 2009. Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. Cereb Cortex. 19:483–495.

Glimcher PW, Camerer C, Poldrack RA, Fehr E, editors. 2009. Neuroeconomics decision making and the brain. Amsterdam; Boston; Heidelberg: Elsevier/Academic Press.

Gluck MA, Bower GH. 1988. From conditioning to category learning: an adaptive network model. J Exp Psychol Gen. 117:227–247.

Grinband J, Savitskaya J, Wager TD, Teichert T, Ferrera VP, Hirsch J. 2011. The dorsal medial frontal cortex is sensitive to time on task, not response conflict or error likelihood. Neuroimage. 57: 303–311.

Hampton AN, Bossaerts P, O'Doherty JP. 2006. The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. J Neurosci. 26:8360–8367.

Hare TA, Camerer CF, Knoepfle DT, Rangel A. 2010. Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. J Neurosci. 30:583–590.

Hare TA, Camerer CF, Rangel A. 2009. Self-control in decision-making involves modulation of the vmPFC valuation system. Science. 324:646–648.

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 USA. 108:18120–18125.

Hutchinson JMC, Gigerenzer G. 2005. Simple heuristics and rules of thumb: where psychologists and behavioural biologists might meet. Behav Processes. 69:97–124.

Kable JW, Glimcher PW. 2007. The neural correlates of subjective value during intertemporal choice. Nat Neurosci. 10:1625–1633.

Kelley H, Busemeyer J. 2008. A comparison of models for learning how to dynamically integrate multiple cues in order to forecast continuous criteria. J Math Psychol. 52:218–240.

Khader PH, Pachur T, Meier S, Bien S, Jost K, Rösler F. 2011. Memory-based decision-making with heuristics: evidence for a controlled activation of memory representations. J Cogn Neurosci. 23:3540–3554.

Knutson B, Taylor J, Kaufman M, Peterson R, Glover G. 2005. Distributed neural representation of expected value. J Neurosci. 25:4806–4812.

Krajbich I, Armel C, Rangel A. 2010. Visual fixations and the computation and comparison of value in simple choice. Nat Neurosci. 13:1292–1298.

Krajbich I, Rangel A. 2011. Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. Proc Natl Acad Sci USA. 108:13852–13857.

Lewandowsky S, Farrell S. 2011. Computational modeling in cognition: principles and practice. Los Angeles (CA): Sage Publications.

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.

Mata R, Schooler LJ, Rieskamp J. 2007. The aging decision maker: cognitive aging and the adaptive selection of decision strategies. Psychol Aging. 22:796–810.

Mata R, Von Helversen B, Rieskamp J. 2010. Learning to choose: cognitive aging and strategy selection learning in decision making. Psychol Aging. 25:299–309.

O'Craven KM, Rosen BR, Kwong KK, Treisman A, Savoy RL. 1997. Voluntary attention modulates fMRI activity in human MT-MST. Neuron. 18:591–598.

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, Dayan P, Friston K, Critchley H, Dolan RJ. 2003. Temporal difference models and reward-related learning in the human brain. Neuron. 38:329–337.

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.

Payne JW, Bettman JR, Johnson EJ. 1993. The adaptive decision maker. Cambridge, New York (NY): Cambridge University Press.

Payne JW, Bettman JR, Johnson EJ. 1988. Adaptive strategy selection in decision making. J Exp Psychol Learn Mem Cogn. 14:534–552.

Penny WD, Stephan KE, Mechelli A, Friston KJ. 2004. Comparing dynamic causal models. Neuroimage. 22:1157–1172.

Penny WD, Trujillo-Barreto NJ, Friston KJ. 2005. Bayesian fMRI time series analysis with spatial priors. Neuroimage. 24:350–362.

Pessiglione M, Petrovic P, Daunizeau J, Palminteri S, Dolan RJ, Frith CD. 2008. Subliminal instrumental conditioning demonstrated in the human brain. Neuron. 59:561–567.

Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD. 2006. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature. 442:1042–1045.

Peters J, Büchel C. 2009. Overlapping and distinct neural systems code for subjective value during intertemporal and risky decision making. J Neurosci. 29:15727–15734.

Pine A, Seymour B, Roiser JP, Bossaerts P, Friston KJ, Curran HV, Dolan RJ. 2009. Encoding of marginal utility across time in the human brain. J Neurosci. 29:9575–9581.

Pitt MA, Myung IJ. 2002. When a good fit can be bad. Trends Cogn Sci. 6:421–425.

Pochon J-B, Riis J, Sanfey AG, Nystrom LE, Cohen JD. 2008. Functional imaging of decision conflict. J Neurosci. 28:3468–3473.

Preuschoff K, Bossaerts P, Quartz SR. 2006. Neural differentiation of expected reward and risk in human subcortical structures. Neuron. 51:381–390.

Rangel A, Camerer C, Montague PR. 2008. A framework for studying the neurobiology of value-based decision making. Nat Rev Neurosci. 9:545–556.

Rescorla RA, Wagner AR. 1972. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Black AH, Prokasky WF, editors. Classical conditioning II: current research and theory. New York: Appleton-Century-Crofts. p. 64–99.

Rieskamp J. 2008. The importance of learning when making inferences. Judgm Decis Mak. 3:261–277.

Rieskamp J. 2006. Perspectives of probabilistic inferences: reinforcement learning and an adaptive network compared. J Exp Psychol Learn Mem Cogn. 32:1355–1370.

Rieskamp J, Hoffrage U. 2008. Inferences under time pressure: how opportunity costs affect strategy selection. Acta Psychol (Amst). 127:258–276.

Rieskamp J, Otto PE. 2006. SSL: a theory of how people learn to select strategies. J Exp Psychol Gen. 135:207–236.

Rosa MJ, Bestmann S, Harrison L, Penny W. 2010. Bayesian model selection maps for group studies. Neuroimage. 49:217–224.

Rose M, Schmid C, Winzen A, Sommer T, Büchel C. 2005. The functional and temporal characteristics of top-down modulation in visual selection. Cereb Cortex. 15:1290–1298.

Scheibehenne B, Rieskamp J, Wagenmakers E-J. 2013. Testing adaptive toolbox models: a Bayesian hierarchical approach. Psychol Rev. 120:39–64.

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

Shomstein S, Yantis S. 2004. Control of attention shifts between vision and audition in human cortex. J Neurosci. 24:10702–10706.

Simon HA. 1956. Rational choice and the structure of the environment. Psychol Rev. 63:129–138.

Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ. 2009. Bayesian model selection for group studies. Neuroimage. 46:1004–1017.

Stephan KE, Penny WD, Moran RJ, Den Ouden HEM, Daunizeau J, Friston KJ. 2010. Ten simple rules for dynamic causal modeling. Neuroimage. 49:3099–3109.

Sutton RS, Barto AG. 1998. Reinforcement learning: an introduction. Cambridge (MA): MIT Press.

Todd PM, Gigerenzer G. 2007. Environments that make us smart: ecological rationality. Curr Dir Psychol Sci. 16:167–171.

Tversky A, Kahneman D. 1974. Judgment under uncertainty: heuristics and biases. Science. 185:1124–1131.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage. 15:273–289.

Venkatraman V, Rosati AG, Taren AA, Huettel SA. 2009. Resolving response, decision, and strategic control: evidence for a functional topography in dorsomedial prefrontal cortex. J Neurosci. 29:13158–13164.

Wunderlich K, Beierholm UR, Bossaerts P, O'Doherty JP. 2011. The human prefrontal cortex mediates integration of potential causes behind observed outcomes. J Neurophysiol. 106:1558–1569.

Wunderlich K, Smittenaar P, Dolan RJ. 2012. Dopamine enhances model-based over model-free choice behavior. Neuron. 75:418–424.

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