NeuroImage 89 (2014) 57–69

Contents lists available at ScienceDirect

NeuroImage

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

Neural systems for choice and valuation with counterfactual learning signals

M.J. Tobia [a,*], R. Guo [b,c], U. Schwarze [a], W. Boehmer [b], J. Gläscher [a], B. Finckh [a], A. Marschner [a], C. Büchel [a], K. Obermayer [b,c], T. Sommer [a]

[a] Department of Systems Neuroscience, University Medical Center Hamburg-Eppendorf, Germany
[b] Department of Software Engineering and Theoretical Computer Science, School IV Electrical Engineering and Computer Science, Technische Universität Berlin, Germany
[c] Bernstein Center for Computational Neuroscience Berlin, Germany

ARTICLE INFO

Article history:
Accepted 28 November 2013
Available online 7 December 2013

ABSTRACT

The purpose of this experiment was to test a computational model of reinforcement learning with and without fictive prediction error (FPE) signals to investigate how counterfactual consequences contribute to acquired representations of action-specific expected value, and to determine the functional neuroanatomy and neuromodulator systems that are involved. 80 male participants underwent dietary depletion of either tryptophan or tyrosine/phenylalanine to manipulate serotonin (5HT) and dopamine (DA), respectively. They completed 80 rounds (240 trials) of a strategic sequential investment task that required accepting interim losses in order to access a lucrative state and maximize long-term gains, while being scanned. We extended the standard Q-learning model by incorporating both counterfactual gains and losses into separate error signals. The FPE model explained the participants' data significantly better than a model that did not include counterfactual learning signals. Expected value from the FPE model was significantly correlated with BOLD signal change in the ventromedial prefrontal cortex (vmPFC) and posterior orbitofrontal cortex (OFC), whereas expected value from the standard model did not predict changes in neural activity. The depletion procedure revealed significantly different neural responses to expected value in the vmPFC, caudate, and dopaminergic midbrain in the vicinity of the substantia nigra (SN). Differences in neural activity were not evident in the standard Q-learning computational model. These findings demonstrate that FPE signals are an important component of valuation for decision making, and that the neural representation of expected value incorporates cortical and subcortical structures via interactions among serotonergic and dopaminergic modulator systems.

Introduction

Computational models of reward-based learning assert that learning occurs when expectancies are violated (Sutton and Barto, 1981). Representations of expected value mediate decision making and are shaped by experience, as in Q-learning (Watkins and Dayan, 1992). It is typically the product of both observed probability and magnitude of gains and losses (Knutson et al., 2005), and is incrementally updated by a reward prediction error (PE), the difference between expected and experienced consequences. Counterfactual consequences, the gains and losses associated with alternative actions that were not executed, affect subsequent choices in a variety of ways (Boorman et al., 2009, 2013; Buchel et al., 2011; Coricelli et al., 2005; Li and Daw, 2011; Lohrenz et al., 2007; Nicolle et al., 2010, 2011), presumably by influencing the computation and representation of expected value via a fictive error signal processed in the ventral and dorsal striatum (Montague et al., 2006; Lohrenz et al., 2007).

A cortical–subcortical system, including the dorsal anterior cingulate cortex (dACC), orbitofrontal cortex (OFC) and ventromedial prefrontal cortex (vmPFC), as well as the striatum and mid-brain structures, is implicated in reward-based learning from expected values and reward PE (Doya, 2008; O'Doherty, 2004). Importantly, the vmPFC activates acquired representations of expected value that correspond to the selected action or stimulus during choice (Gläscher et al., 2009; Jocham et al., 2011), and reward PEs modulate activity in the ventral and dorsal striatum (O'Doherty, 2004; Haruno & Kawato, 2006; Schonberg et al., 2010). Moreover, both dopamine (DA) and serotonin (5HT) are important for the neural computations of this reward-based learning system (Jocham et al., 2011; Montague et al., 2006; Pessiglione et al., 2006; Schonberg et al., 2010; Seymour et al., 2012; Tanaka et al., 2007), although they are thought to play distinct, yet cooperative or even conflicting roles (Boureau and Dayan, 2010; Cools et al., 2011; Rogers, 2011).

* Corresponding author at: Department of Systems Neuroscience, Bldg W34, University Medical Center Hamburg-Eppendorf, Martinistrasse 52, D-20246 Hamburg, Germany. Fax: +49 40 7410 59955.
E-mail address: mtobia@uke.de (M.J. Tobia).

1053-8119/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2013.11.051

Numerous studies have examined the effects of fictive error signals on subsequent choices, as well as the shared neural substrates for processing reward PE and fictive error signals using a variety of paradigms (Sommer et al., 2009). But less attention has been given to elucidating how fictive error signals shape the expected values that mediate choice, especially during strategic sequential choices for which optimal performance requires accepting interim losses in order to maximize long-term gains. In addition, the roles of DA and 5HT in acquiring and representing expected value from counterfactual learning signals have not been examined. For these reasons, the goal of this experiment was to investigate the functional–neuroanatomical and neuromodulatory systems involved in representing expected value during reward-based learning with valuation processing that incorporates a fictive prediction error (FPE) signal. This experiment was designed to address three issues: 1) whether or not FPE signals improve computations of expected value during intertemporal choice beyond the contribution from standard reward PE signals, 2) whether or not representations of expected value that incorporate FPE signals are supported by activation of the vmPFC and/or subcortical regions, and 3) to localize the involvement of DA and 5HT in processing and representing expected value computed with FPE signals.

A counterfactual loss (i.e., an amount of reward that was not acquired) occurs on winning trials as a *missed opportunity* for which an alternative action would have returned a greater reward, and is associated with subjectively experienced regret (Camille et al., 2004; Coricelli et al., 2005). Counterfactual losses promote choice repetition (Boorman et al., 2013; Nicolle et al., 2010, 2011) as well as choices that spontaneously deviate from an established preference (Boorman et al., 2009). They can be used to optimize investment magnitude (Lohrenz et al., 2007) and choice strategy (Li and Daw, 2011), and they lead to increased subsequent risk taking (Brassen et al., 2012; Buchel et al., 2011; Coricelli et al., 2005). According to Lohrenz et al. (2007), a counterfactual loss, which may only occur on winning trials, can be used as a fictive error signal (referred to as $f+$ in their study). To investigate whether the fictive error signal stemming from a counterfactual loss contributes to valuation processing we developed a computational model that utilizes the fictive error signal to produce an FPE. This FPE differs from the *fictive error signal* studied by Lohrenz et al. in that it is computed using the temporal difference between *expected values* and counterfactual consequences, rather than simply the difference between an obtained and unobtained outcome. We use FPE+ to refer to the FPE signal on a winning trial because it uses the fictive error signal associated with a counterfactual loss (referred to as $f+$ by Lohrenz and colleagues), rather than the counterfactual outcome itself, to compute the prediction error.

A counterfactual gain (i.e., an amount of punishment that was not suffered) occurs on losing trials as a *reduced cost* for which an alternative action would have cost more, and is associated with subjectively experienced relief (Coricelli et al., 2005; Nicolle et al., 2010). We use FPE− to refer to the FPE signal on a losing trial, which is the counterfactual gain due to this reduced cost. Counterfactual gains reportedly lead to differential changes in subsequent choices and cognitive performance aspects (i.e., speeded response times) of decision making (Fujiwara et al., 2009; Lohrenz et al., 2007), although the precise nature of these effects is not well elucidated in the literature. For example, Lohrenz et al. (2007) included the fictive error stemming from the counterfactual gain (referred to as $f-$ in their study) in their analysis of fictive learning signals, but found that it did not significantly predict subsequent choice behavior. Their study showed that fictive error signals from counterfactual gains and losses have dissociable effects on learning and choice behavior.

Previous studies of counterfactual learning signals have applied variations of the Q-learning (Watkins and Dayan, 1992) model to choice behavior (Chiu et al., 2008; Li and Daw, 2011; Lohrenz et al., 2007), although none have directly incorporated counterfactual consequences into a TD-like error term, an FPE, for valuation. Whereas Lohrenz et al., as well as Chiu et al. (2008), found that a fictive error signal contributed to a change in behavior, they did not examine if an FPE+ or FPE− contributes to valuation in a modified Q-learning model of choice behavior. As suggested by Lohrenz and colleagues, situating the fictive error signal ($f+$ and $f-$) within a machine learning framework, such as Q-learning, could provide additional insight into the contribution of counterfactual consequences to choice behavior. Li and Daw (2011) examined the effects of counterfactual consequences on learning, but they modeled the counterfactual prediction error with a Rescorla–Wagner update rule, which by definition does not take into account future anticipated rewards as does a TD prediction error. They found that their data was more consistent with a policy updating mechanism, as opposed to a system that updates action-specific expected values with counterfactual consequences. Taken together, these studies showed that choice behavior is responsive to counterfactual consequences, but it remains unclear if the various effects of counterfactual consequences on subsequent choices are mediated by a direct effect on action-specific valuation processing.

To investigate the effects of both counterfactual losses and gains on valuation and choice behavior, we designed a strategic sequential investment task (SSIT) that overtly presented the counterfactual outcome on each trial, included action–contingent state transition rules, and modified the Q-learning algorithm to incorporate winning and losing FPEs in a two-stage update process. The computational model incorporated the counterfactual outcome by computing a fictive error signal that is subsequently used to compute an FPE, which then updates action-state pair specific expected values. This allowed observation of which FPE signals (either FPE+ or FPE−, or both) contributed to the action-specific valuation. We expected that incorporating counterfactual learning signals into Q-learning with FPE signals would facilitate model performance, and that expected value signals would modulate the vmPFC during choice. In addition, we aimed to further characterize the functional neuroanatomy representing expected value by localizing the involvement of dopaminergic and serotonergic neuromodulators via an acute amino acid dietary depletion protocol for both DA and 5HT, respectively.

Method

Participants

80 healthy males aged 18–30 years (mean = 24.3, SD = 3.4) participated in the experiment. Participants were screened for mental health disorders during recruitment and provided informed consent. They completed a set of psychological tasks and questionnaires including assessments of risk and loss aversion (Sokol-Hessner et al., 2009), personality traits, and spatial intelligence, over two days in addition to adhering to a restricted diet (details below) for 24 h prior to the experimental session. All 80 participants complied with the diet (see results of the depletion procedure below) and completed the decision making task while being scanned. None were excluded from any analyses. All protocols were approved by the ethics committee of the medical association of Hamburg and carried out in accordance with the Declaration of Helsinki.

Dietary depletion procedure

Participants were randomly assigned in a double-blind placebo-controlled protocol to one of three groups: placebo (n = 30; P), DA-depletion (n = 25; D−), and 5HT-depletion (n = 25; S−); each receiving a different dietary depletion treatment designed to reduce or preclude the metabolism of essential amino acids into various neurotransmitters (Young et al., 1985). All participants received a low protein diet (12 g) provided by the University Medical Center canteen the day before scanning, and fasted overnight (including food, alcohol and caffeine). In the morning of the test day, a baseline blood sample was

drawn from each participant and then they were given a drink containing an array of amino acids. Whereas the mixture for the P group contained the full complement of amino acids, the mixture for the D− group lacked L-tyrosine and L-phenylalanine, and the mixture for the S− group lacked L-tryptophan. As such, metabolic precursors for DA were depleted in the D− group, and the metabolic precursor for 5HT was depleted in the S− group. Following consumption of the beverage, participants had a delay period of 4 h during which they could read or watch DVD videos, and also completed the training exercises. A second blood draw was taken just prior to scanning.

Amino acid mixtures and biochemical measures

Amino acid mixtures were prepared by Meta X Institute for Dietetics (Freiburg, Germany). Three different powdered mixtures were prepared corresponding to the three experimental groups. The powder was stirred into a glass of water (approximately 250 mL) yielding a citrus flavored drink.

The mixture for the P group contained: L-alanine (4.1 g), L-arginine (3.7 g), L-aspartic (9.8 g), L-cystine (2.0 g), glycine (2.4 g), L-histidine (2.4 g), L-isoleucine (6.1 g), L-leucine (10.2 g), L-lysine (7.6 g), L-methionine (3.0 g), L-phenylalanine (4.3 g), L-proline (9.3 g), L-serine (5.3 g), L-threonine (4.3 g), L-tryptophan (3.0 g), L-tyrosine (5.3 g), and L-valine (6.8 g).

The mixture for the D- group contained: L-alanine (4.6 g), L-arginine (4.1 g), L-aspartic (10.8 g), L-cystine (2.2 g), glycine (2.7 g), L-histidine (2.7 g), L-isoleucine (6.7 g), L-leucine (11.3 g), L-lysine (8.4 g), L-methionine (3.4 g), L-phenylalanine (0.0 g), L-proline (0.3 g), L-serine (5.8 g), L-threonine (4.8 g), L-tryptophan (3.4 g), L-tyrosine (0.0 g), and L-valine (7.5 g).

The mixture for the S− group contained: L-alanine (4.3 g), L-arginine (3.9 g), L-aspartic (10.1 g), L-cystine (2.1 g), glycine (2.5 g), L-histidine (2.5 g), L-isoleucine (6.3 g), L-leucine (10.5 g), L-lysine (7.8 g), L-methionine (3.1 g), L-phenylalanine (4.5 g), L-proline (9.6 g), L-serine (5.4 g), L-threonine (4.5 g), L-tryptophan (0.0 g), L-tyrosine (5.4 g), and L-valine (7.0 g).

Blood samples were analyzed for plasma free amino acid concentrations. EDTA-blood was collected (10 mL) before ingestion of the amino acid drink and again 5.5 h later to determine the ratio of either tryptophan or tyrosine and phenylalanine amino acids to five large neutral amino acids (LNAA). Blood samples were centrifuged (10 min; 3000 *g*), frozen in liquid nitrogen and stored at −80 °C. Each plasma sample (100 μL) was deproteinized with 5-sulphosalicylic acid (10%, w/v) and centrifuged (10 min; 3000 *g*) after neutralization and adding of the internal standard norleucine. The amino acids in the resulting supernatant were determined using an amino acid analyzer according to standard procedures (Biochrom 30, Laborservice Onken, Gruendau, Germany). A cation exchange chromatography system is coupled with a detection system using post-column derivatization with o-phthaldialdehyde and fluorescent detection (ex: 340 nm; em: 450 nm).

Strategic sequential investment task

The strategic sequential investment task (SSIT) was designed to investigate the potential use of counterfactual consequences as FPE signals during value learning. On each trial, participants decide how much money to invest in a financial market, and then learn about the factual and counterfactual outcomes in succession. The task design included a complex state-space (Fig. 1) comprised of four possible paths (a sequence of three states), each with a different chance of gaining or losing money in the long-run. Each individual state was uniquely identifiable by a different neutral visual background pattern. As illustrated in Fig. 1, the paths leading to states 4 and 6 are associated with long-term gains, with state 4 being the most lucrative.

Participants completed 80 rounds of the SSIT where each round started at state 1, consisted of three decisions and ended in state 4, 5, 6, or 7 (Fig. 1). On each trial, participants choose an amount of money to invest (0, 1, 2 or 3 Euros). Their path through the virtual maze was determined by the magnitude of their investments, rather than the outcome of the trial. Risk averse investments (0–1 Euro; RA) descended throughout the maze leading to a non-lucrative, losing state (e.g., states 5 & 7). Risk seeking investments (2–3 Euro; RS) elevated throughout the maze leading to a lucrative, winning state (e.g., states 4 & 6). In order to identify and follow the optimally lucrative path, participants must make strategic decisions that accept interim losses (at states 1 and 2, for example) in order to gain access to lucrative state 4. As such, decision making based on expected value must take into account anticipated future rewards, rather than only feedback for the current state-action pair.

The task was presented to participants in the scanner as an event-related design (Fig. 1, bottom) with 5 stimulus events during each trial. Each trial started with the presentation of a state (indicated by a unique visual background cue) and a randomly initialized response meter to indicate the amount of money to invest on the current trial (i.e., choice phase). Participants could move the indicator bar on the response meter using an MR-compatible mouse according to the value of their desired investment (0–3 Euro). This stimulus remained onscreen for 3000 ms (fixed duration). This was followed by a brief (500 ms, fixed duration) anticipation phase, and then factual (i.e., outcome phase) and counterfactual (i.e., missed outcome phase) outcomes were presented in succession. The outcome presentation (3000–5000 ms, jittered) informed participants of the amount of money that had been gained or lost on that trial, indicated by a stack of coins. The counterfactual outcome presentation (3000–5000 ms, jittered) informed participants of a greater amount of money that could have been won or lost if the maximum investment, i.e. 3 Euros, was selected. This was symbolized by a second stack of coins that highlighted the difference between factual and counterfactual outcomes. Previously, Lohrenz et al. employed a similar task design but did not explicitly present the counterfactual outcome on each trial. Instead, they computed a fictive error signal ($f+$ or $f-$) implicitly, based on the difference between factual obtained reward and what would have been obtained if a maximal investment had been selected on that trial.

Each trial concluded with the presentation of a state transition stimulus event (2700 ms, fixed duration) that highlighted whichever of two possible subsequent states had been selected based on the magnitude of the investment (i.e., transition phase). The two possible state transitions were shown simultaneously at the lower and upper portions of the display in a random fashion. This transition event was substituted for by an additional feedback stimulus after the third trial of each round (the task always returned to state 1 as the first trial of each round), which indicated the total amount of money gained or lost over the previous three decisions (i.e., multi-trial feedback phase).

Two short rounds of practice trials familiarized participants to the task stimuli and the mouse controls for indicating their choice, as well as for making ratings of win expectancies, but did not reveal information about the actual win probabilities or expected values that defined each state and path from the task. The first round of practice trials was self-paced. The second round of practice trials was presented at the same speed as the task would be presented during fMRI scanning. The practice trials did not reveal any information about the contingencies that were active during the experiment. The task was presented to participants in 10 blocks of 3 trials during each of 8 scanning runs (240 total trials). In between each scanning run participants completed visual analog ratings (1–5; anchored at 'never' and 'very often') of the win expectancy associated with each individual state (i.e., "how frequently do you win in this state?").

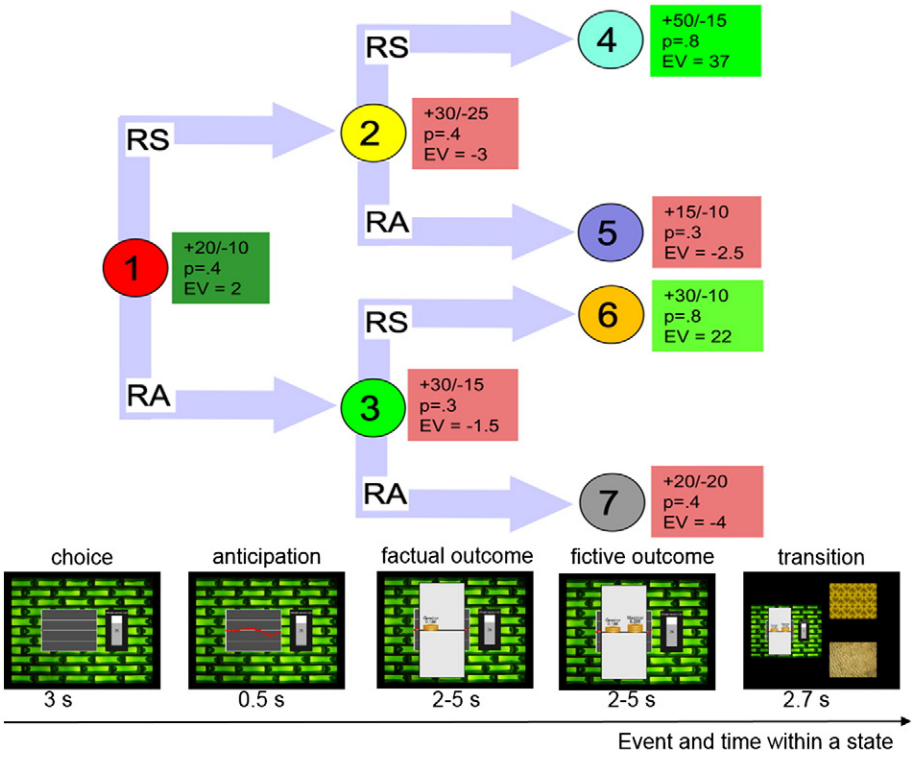

Fig. 1. Task design and presentation. Strategic sequential investment task. The task is based on a complex state space whose underlying structure is not known to the participants (upper part of the figure). The seven states differ with respect to their winning and losing probabilities as well as the mean amount of monetary gains and losses. In each state the underlying outcome is generated by a bi-Gaussian distribution (p as the win probability and $1-p$ as the loss probability). The two numbers on the top in the square next to each state are the mean of the win Gaussian and the loss Gaussian and the expected value (EV) is the mean outcome of the specific state. For example in state 1 (red), $EV = 0.4 \times 20 + 0.6 \times (-10)$. States where the state characteristics are indicated in green squares have positive EVs, i.e. states 1, 4, and 6, whereas states with red squares have negative EVs, i.e. 2, 3, 5, and 7. Each state is associated with a particular neutral background (see lower part of the figure for an example). By this background color, participants can learn over the 240 trials of the experiment to associate each state with an EV. In each trial (lower part of the figure), participants decide how much to invest, i.e. 0, 1, 2, or 3 Euros in the stock market of the current state ('choice'). The amount of the investment is indicated in the bar right to the market chart where the starting amount at the beginning of the choice phase, i.e. 0, 1, 2, or 3 Euros, was random. During the brief anticipation phase participants observe how the market develops. Then they learn in the outcome phase how much they won or lost, which is the product of their investment and the market change. The outcome was presented in numbers but also visualized by a positive or negative stack of coins. In the following fictive outcome phase, subjects learned how much they would have won or lost when they would have invested the maximum of 3 Euros. This phase was included to foster counterfactual comparisons which result in a fictive prediction error, i.e. the difference between the factual and the counterfactual outcome. Participants started each round in state 1 and were then transferred through the state space following a transition rule that was unknown for them. In particular, risk averse (RA) investments of 0 or 1 Euro led to different states than risk seeking (RS) investments. At the end of each trial, the two possible next states were shown to the participant in the transition phase in random vertical order. Then subjects were transferred to the state according to their decision. After 3 trials, the round ended and subjects were informed about the total win or loss of this round, and then transferred back to state 1.

Computational reinforcement Q-learning model

Q-learning is a model-free reinforcement learning technique that learns an action–value function according to the temporal difference (TD) between obtained and expected rewards (Watkins and Dayan, 1992). However, the TD signal only involves factual consequences about the selected action. In order to assess a counterfactual learning process, we modified the standard Q-learning model by incorporating counterfactual consequences into valuation computations in a two-stage update process.

The SSIT consists of seven states (s_1 to s_7), each with four possible actions (a_1, a_2, a_3 and a_4). The goal of both the standard Q-learning model and the FPE model is to learn a state-action value function $Q(s_t,a_t)$ at each trial t, which is defined to be the expected discounted sum of future payoffs obtained by taking action a from state s and following an optimal policy thereafter.

For both the standard model and the computational model modified with FPEs, Q values were initialized to 0 and then updated on each trial with a two-stage update process. In this first stage, Q values are updated with the factual outcome r_t according to a standard temporal difference (TD) learning rule as shown below:

$$Q(s_t,a_t) = Q(s_t,a_t) + \alpha\left[r_t + \gamma \max_{a_{t+1}} Q(s_{t+1},a_{t+1}) - Q(s_t,a_t)\right]$$

where α is a free learning rate parameter that determines to what extent the newly acquired information will override the old information. The discount parameter γ weighs the extent to which anticipated reward from next state s_{t+1} is taken into account when computing the TD error term, and was fixed to $\gamma = 0.9$. This is a standard TD learning rule and the TD error term $(r_t + \gamma \max_{a_{t+1}} Q(s_{t+1},a_{t+1}) - Q(s_t,a_t))$ was applied to update only the chosen action. This was the only update used in the standard model to which the FPE model was compared.

In the second stage of valuation processing, participants observe the counterfactual outcome associated with a trial-specific optimal bet (3 € for winning trials; 0 € for losing trials). The counterfactual outcome is associated with an improved consequence relative to the experienced outcome. Hence, for example in state 1, with a bet of 1 € and a winning outcome, the participant would earn +20 €. On this trial there are two counterfactual outcomes that produce an improved consequence, so we use the best counterfactual alternative. In this case, the counterfactual outcome is associated with a bet of 3 € producing earnings of +60 €. We calculate the fictive error signal on a winning trial (referred to as $f+$) as the difference between experienced and counterfactual outcome, which is equal to −40 € for this particular example, and represents a counterfactual loss. The counterfactual gain is computed in a similar manner. Again using state 1 as an example, a bet of 2 € on a losing trial would produce a loss of 20 €, but a bet of 0 is most optimal, and so the difference is again computed to reveal that the fictive error

from a losing trial ($f-$) is +20 €. These fictive error signals are defined the same as the two fictive error signals, $f+$ and $f-$, in Lohrenz et al. (2007). The $f+$ and $f-$ were applied to update action-specific values differently according to:

(1) when the market went up and, if less than the maximum of 3 Euros was invested, participants could have won more if they had invested more, i.e. they experienced an $f+$ (a counterfactual loss/missed opportunity). In this case, $f+ = (3 - a_t) \cdot$ market development, where 3 Euros is the maximum amount that subjects can bet at each trial.
(2) when the market went down and, if more than the minimum of 0 Euros was invested, participants could have lost less if they had invested less i.e. they experienced an $f- = (a_t - 3) \cdot$ market development (a counterfactual gain).

Then, the Q value is updated again prior to the next trial, this time with the FPE to promote the actions which would have invested more when the market goes up (for all $a \geq a_t$) and the actions which would have invested less when the market goes down (for all $a \leq a_t$), as shown in the formulae below:

$$Q(s_t, a) = Q(s_t, a) + \alpha_{\mathrm{FPE}}\left[f + \gamma \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) - Q(s_t, a)\right]$$

where $\alpha_{\mathrm{FPE}} = \{\alpha_+, \alpha_-\}$ and $f = \{f+, f-\}$. α_+ is the learning rate over counterfactual loss and α_- is the learning rate over counterfactual gain $f-$. The introduction of these two additional parameters to the standard Q-learning model enables the model to update expected values with counterfactual gains and counterfactual losses differently. Positive and negative counterfactual information were updated differently since both behavioral and neuroimaging data suggest that they might impact decisions and neural activity differently (Chandrasekhar et al., 2008; Fujiwara et al., 2009; Lohrenz et al., 2007; Zeelenberg and Pieters, 2007).

After this two-staged belief update, actions are selected stochastically according to the probabilities determined by the state-action values through a softmax distribution:

$$P(s_t, a_t) = \frac{\exp(\beta \times Q(s_t, a_t))}{\sum_{n=1}^{4} \exp(\beta \times Q(s_t, n))}.$$

In total, this model contains 4 free parameters: standard Q learning rate α, $f+$ learning rate α_+, $f-$ learning rate α_-, and the inverse temperature parameter β. It nests the standard Q model ($\alpha_+ = 0$ and $\alpha_- = 0$). The goodness of fit for the FPE model was compared with that from the standard Q learning model using the pseudo-r^2 statistic from each model's fit to the data.

MR protocol

All MR images were acquired with a 3 T whole-body MR system (Magnetom TIM Trio, Siemens Healthcare) using a 32-channel receive-only head coil. Structural MRI was recorded from each participant using a T1 weighted magnetization-prepared rapid gradient-echo (MPRAGE) sequence with a voxel resolution of 1 × 1 × 1 mm^3, coronal orientation, phase-encoding in the left–right direction, FoV = 192 × 256 mm, 240 slices, 1100 ms inversion time, TE = 2.98 ms, TR = 2300 ms, and 9° flip angle. Functional MR time series were recorded using a T2* GRAPPA EPI sequence with TR = 2380 ms, TE = 25 ms, anterior–posterior phase encode, 40 slices acquired in descending (non-interleaved) axial plane with 2 × 2 × 2 mm^3 voxels (204 × 204 mm FoV; skip factor = .5), with an acquisition time of approximately 8 min per scanning run.

MR data processing

Structural and functional MR image analyses were conducted in SPM8 (Wellcome Department of Cognitive Neurology, London, UK). Anatomical images were segmented and transformed to Montreal Neurological Institute (MNI) standard space, and a group average T1 custom anatomical template image was generated using DARTEL. Prior to generating the group template, we conducted a voxel-based morphometry (VBM) analysis to ensure the absence of gross anatomical differences associated with the dietary depletion protocol. Results of the VBM analysis showed no statistically significant morphological differences, even at a liberal threshold. Functional images were corrected for slice-timing acquisition offsets, realigned and corrected for the interaction of motion and distortion using unwarp toolbox, co-registered to anatomical images and transformed to MNI space using DARTEL, and finally smoothed with an 8 mm FWHM isotropic Gaussian kernel.

Functional images were analyzed using the general linear model (GLM) implemented in SPM8. First level analyses included onset regressors for each stimulus event excluding the anticipation phase (see section above), and a set of parametric modulators corresponding to trial-specific task outcome variables and computational model parameters. Trial-specific task outcome variables (and their corresponding stimulus event) include the choice value (0–3 Euros) of the investment (choice phase) and the total value of gains/losses over each round (corresponding to multi-trial feedback event). Model derived parametric modulators included the time series of Q-values for the selected action (choice phase), TD (outcome phase), and $f+$ or $f-$, the fictive error signals (counterfactual outcome phase). The fictive error signal was used as the regressor because it was common to the computation of each FPE update to selected and unselected actions. As with the computational model, it was divided into two sets of trials for FPE+ and FPE− (all trials are accounted for in the model) although the FPEs themselves were not used as regressors (because there could be more than one on each trial corresponding to the number of unselected actions requiring a valuation update). Reward/punishment value was not modeled as a parametric modulator because the TD error time series and trial-by-trial reward values were strongly correlated (all $rs > .7$; $ps < .001$). The configuration of the first-level GLM regressors for the standard Q-learning model was identical to that employed in the FPE model except that winning and losing counterfactual outcome onsets were modeled as a single event category (counterfactual outcome phase), and parametric modulators for the trial-by-trial counterfactual gain and loss values were not included.

All regressors were convolved with a canonical hemodynamic response function. Prior to model estimation, coincident parametric modulators were serially orthogonalized as implemented by default in SPM (i.e., the Q-value regressor was orthogonalized with respect to the choice value regressor). This was done to prevent the first level GLM from allowing variance that was common to both regressors to go undetected. In addition, we included a set of regressors for each participant to censor EPI images with large, head movement related spikes in the global mean.

Second level analyses consisted of a one-way analysis of variance (ANOVA). To control for false positives at the group level, AlphaSim (Forman et al., 1995) implemented with AFNI (Cox, 1996) was used to determine two different thresholds to apply to cortical and subcortical clusters. The simulation for cortical clusters included all brain voxels (whole-brain correction). The simulation for subcortical clusters (subcortical volume correction) was performed inside a mask (2870 voxels) of the caudate (head, body, tail), nucleus accumbens, and putamen. Both simulations used a single-voxel threshold of $p < .005$ and a smoothness of 8 mm^3. Results of the simulation showed that a minimum cluster size of 156 and 32 contiguous voxels yielded a corrected $p < .05$ for cortical and subcortical clusters, respectively. These empirically derived thresholds are more conservative with respect to false positive results compared to those recommended by

Lieberman and Cunningham (2009), which were shown to provide an appropriate balance between Type I and Type II error rates for whole-brain corrections. As such, the subcortical correction threshold was applied to all subcortical clusters even if they were not included in the simulation mask (e.g., amygdala). The fMRI results from the standard Q-learning model are not shown. Standardized (MNI) coordinates [x y z] are reported with the z-scored peak voxel value and cluster sizes (n).

Results

Dietary depletion efficacy

Pairwise statistical tests of the difference in targeted amino acid: LNAA ratio confirmed the expected between group effects of the depletion procedure (Fig. 2, left). The D− group showed significantly decreased levels of phenylalanine and tyrosine compared to the P and S− groups. The S− group showed significantly decreased tryptophan compared to the P and D− groups. These results indicate that the depletion protocol selectively manipulated plasma free concentrations for either DA or 5HT precursors, and did not impair the P group. Random assignment to depletion groups did not result in groups composed of participants with significantly different IQ scores ($F(2,79) = .3, p = .7453$; $F(2,79) = .21$, $p = .8076$), personality traits measured by the TCI (all Fs < 1.38, all ps > .2589), or pre-existing biases for risk or loss aversion, ps = .9308, .8777, respectively. Importantly, dietary depletion did not interfere with participants' ability to associate winning and losing with each specific state. A 3 (group) × 7 (state) factorial ANOVA showed only a significant main effect of state on ratings of win expectancies, $F(6,539) = 88.88$, $p < .05$, and neither a significant effect of group nor significant group × state interaction. Follow-up comparisons showed that all three groups rated states 4 and 6 (both winning states) as winning more frequently than each of the other states, and state 4 was rated higher than state 6, all ps < .05. A separate repeated measures ANOVA showed that the ratings for each state were stable over the 8 individual measurements (i.e., not significantly different over blocks).

Task performance

The S− and D− groups did not differ significantly in any aspect of task performance from the P group. Successful learning of the task was defined as reaching the most lucrative state (state 4) in at least 7 of the last 10 rounds, and investing there at least 2 Euros. A total of 45 out of 80 participants exploited the state space successfully in the last 10 rounds: 17 in the P group, 15 in the S− group, and 13 in the D− group. The proportion of learners did not differ significantly between groups ($\chi^2(2) = .163$, $p = .9217$). A depletion group × learning success ANOVA revealed that learners accumulated significantly more reward than non-learners (41.9 +/− 7.8 and 14.2 +/− 25.5 Euros, $F(1) = 70.3$, $p < 0.0001$) but there was no difference between depletion groups as the interaction was not significant ($F(2) = 1.1$, $p = 0.35$). In addition, all visits to state 4 by all participants during the final 10 rounds were characterized by maximal investments (i.e., 3 Euros), suggesting that all participants recognized the value of the state, and that the difference between learners and non-learners was their ability to make strategic decisions (i.e., preference to accept the interim losses) that would grant them access to state 4.

Q-learning models

The model's free parameters were individually fitted to each subject by maximum likelihood estimation. Given the actual choice by the subject at each trial, the fitness of the model is measured by the action selection probability predicted by the model, which is called the likelihood and it's a function of the model parameters (Daw, 2011). Because the FPE model nests the standard Q model ($\alpha_+ = 0$ and $\alpha_- = 0$), we can compare their goodness of fit with the likelihood ratio test (LRT). Pseudo-r^2 shows how much better the model captures the behavioral data than a null model of random choices. Pseudo-r^2 is computed as $1-\frac{L}{R}$, for each subject, where L is either log data likelihood of the standard Q-learning model or of the FPE model and R is the log data likelihood under chance. Both models fit the behavioral data significantly better than chance, $p < 0.05$ for all 80 subjects (likelihood ratio test). But, the FPE model fits the behavioral data significantly better than standard Q model (likelihood ratio test statistic and p value averaged across subjects: $\chi^2 = 63$, $p = 2e-14$). A similar reduction in −LL was observed for the FPE model within all three groups independently as well (Table 1). However, the model fit did not differ among groups for either the FPE or standard Q-model (FPE model: pDP = 0.79, pDS = 0.26, pPS = 0.47; Q-learning model: pDP = 0.44, pDS = 0.36, pPS = 0.13; Wilcoxon test).

The FPE model explained the behavioral data in the initial trials much better than the standard Q model. Fig. 2 (right) compares the model performance over consecutive blocks of the task, each composed of an incremented subset of 30 trials. It shows that the FPE model was significantly better, ps < .001 (paired t-tests), when considering a

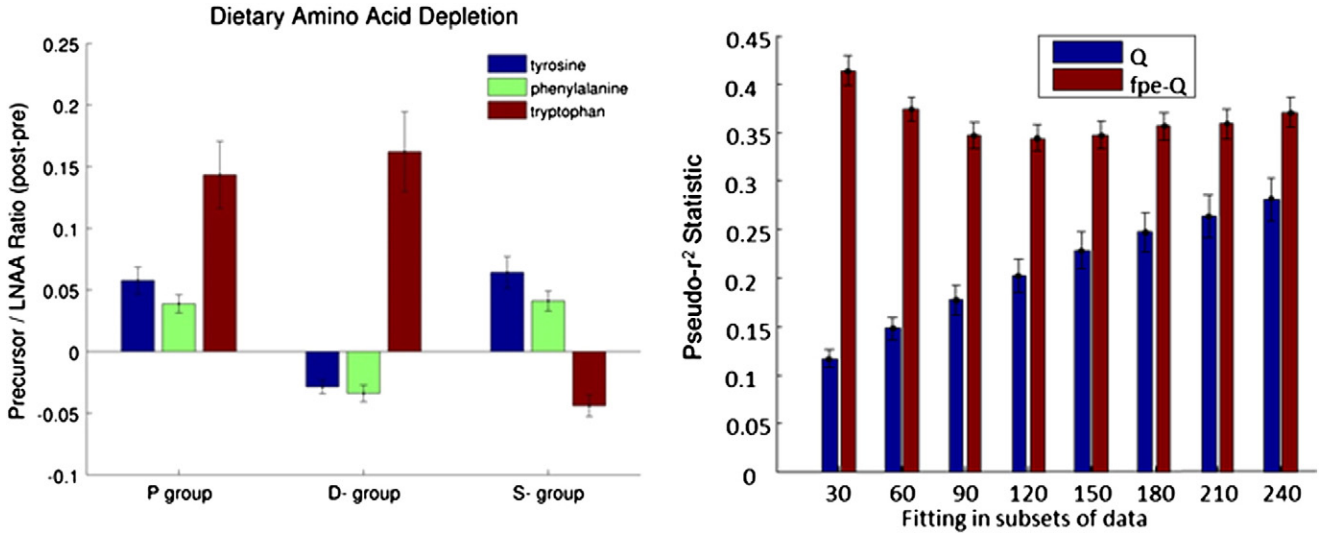

Fig. 2. Depletion efficacy and model comparison. Left: the bar graph shows the selectivity and efficacy of the depletion procedure. Right: the FPE model explains choice behavior better than the standard model after only a small subset of trials, and continues to outperform the standard model when considering larger portions of the data.

Table 1
Computational model best fitting parameters.

		β	−LL	α	α_+	α_-
		Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)	Mean (SE)
FPE-Q	S−	7.9071 (.516)	199.3099 (10.717)	0.1947 (.022)	0.0229* (0.007)	0.4726** (0.093)
	D−	14.8004 (2.339)	218.3173 (8.706)	0.1072 (0.020)	0.0271* (0.005)	0.2502** (0.079)
	PL	10.4962 (3.389)	188.7058 (11.421)	0.1927 (0.028)	0.0178 (0.012)	0.0801 (0.075)
	PNL	8.9845 (2.003)	224.1 (10.838)	0.1194 (0.015)	0.0052 (0.003)	0.6575 (0.078)
Q	S−	5.5689 (0.059)	226.0131 (14.704)	0.2206 (0.024)	–	–
	D−	9.1697 (2.135)	240.9172 (11.066)	0.1779 (0.019)	–	–
	P	6.4749 (1.59)	248.6354 (12.346)	0.2867 (0.050)	–	–

Note: * indicates that the 5HT depletion and DA depletion groups were significantly different from each other at $p < .05$, but neither was significantly different from the placebo group; and ** indicates that the 5HT depletion and DA depletion groups were significantly different from each other at $p < .05$, but neither was significantly different from the placebo group. S − = 5HT depletion; D− = DA depletion; PL = placebo learners; PNL = placebo non-learners.

small sample of behavior (first 30 trials) and remained better in each incremented subset of the data.

Learning rates varied depending on the computational model. A one-way ANOVA with TD learning rate as dependent measure indicated a significant main effect, $F(2,79) = 4.9276$, $p < .01$. Post-hoc comparisons revealed that the group average learning rate was greater in S− (.1947; SD = .118) than D− (.1072; SD = .097), although neither was significantly different from P (.1338; SD = .089). A paired *t*-test over the entire sample of participants showed that the TD learning rate from the FPE model (.1445; SD = .1062) was significantly reduced compared to the TD learning rate from the standard (.2321; SD = .1969) model, $t(79) = 3.7005$, $p < .05$. The learning rates associated with both FPE+ (.0233; SD = .0286) and FPE− (.3733; SD = .4159) were significantly greater than zero over the entire sample of participants, $ts(79) = 7.29$ and 8.03, $ps < .01$, respectively. Expected value (Q) was also significantly different between the models, $t(79) = 4.89$, $p < .001$ (paired *t*-test), with the FPE model (.4756, +/−.174) yielding a significantly greater expected value for the chosen action than the standard model (.3115, +/−.236).

We also conducted an unplanned exploratory analysis based on a subdivision of the participants in the P group for parameters from the FPE model. Learners from the P group had a greater learning rate from the TD error term than non-learners (.1539 +/− .11 vs. .1075 +/− .06), although this did not reach statistical significance. The behavioral responsiveness to counterfactual losses was associated with more optimal performance. Learners (.0267 +/− .03 vs .0120 +/− .003) showed a significantly greater learning rate from counterfactual losses (FPE+) than non-learners, $t(16) = 1.83$, $p = .0425$ (one-tailed, unequal variances). In contrast, non-learners (.7402 +/− .15 vs. .1277 +/− .31) showed a greater learning rate from counterfactual gains (FPE−) than learners, $t(23) = -7.2$, $p < .001$ (two-tailed, unequal variances). This indicates that learners updated their expected values with counterfactual losses more so, and with counterfactual gains less so, than non-learners.

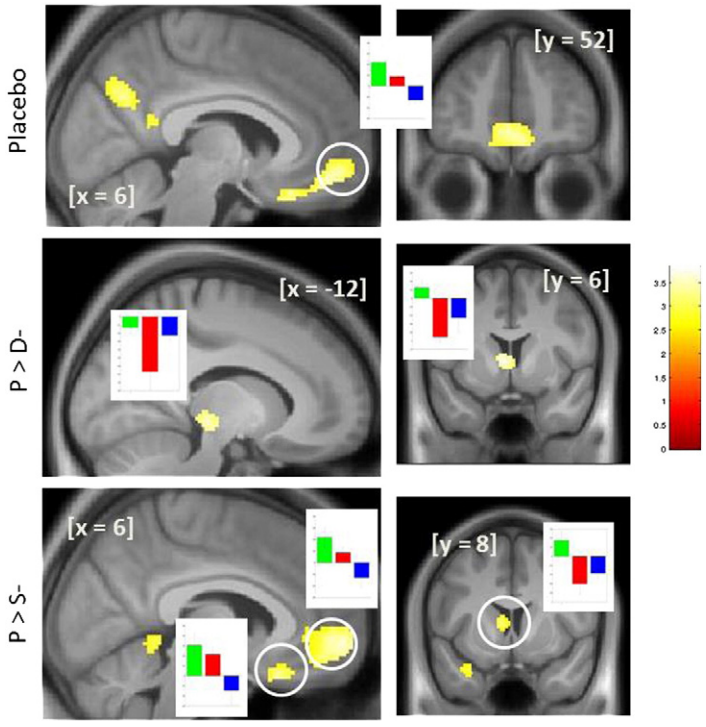

Fig. 3. BOLD signal change correlated with expected value. Top row: activity in the vmPFC and OFC was positively modulated by Q-values in the placebo group. Middle row: modulation of the BOLD signal by Q-values in the midbrain (left) and caudate (right) was significantly different in the placebo group than the dopamine depletion group. Bottom row: modulatory effects of Q-values were significantly different in the placebo group compared to the serotonin depletion group in the medial PFC and posterior OFC (left), and the caudate (right). Whereas the modulatory effect was positive for the placebo group, it was negative for the serotonin depletion group. Insets: bar graphs showing the beta values for P, D− and S− groups in green, red and blue, respectively.

Neuroimaging results: neural substrates for expected value

The FPE model-based fMRI analyses revealed several significant correlations among model parameters and BOLD signal changes that differed according to depletion group. Fig. 3 displays the results for these effects. Expected value was significantly correlated with BOLD signal change in the vmPFC as well as subcortical and midbrain structures. Q-values modulated the vmPFC in the P group ([6 52–10], z = 3.3, n = 960) and there were significant effects of depletion on the neural representation of expected value. The correlation between Q-values and vmPFC activity reversed its sign in the S− group compared to the P group in an overlapping region of the vmPFC ([−6 56 −4], z = 3.8, n = 1958), the posterior orbitofrontal cortex ([2 22 −24], z = 3.51, n = 180), and also the caudate ([−8 8 8], z = 3.19, n = 246). In addition, expected value was affected by serotonin depletion in the bilateral orbital gyrus ([−28 38 −14], z = 3.33, n = 287; [24 26 −12], z = 3.49, n = 302), ventral occipital cortex ([14 −54 −10], z = 3.06, n = 581) and posterior cingulate ([−8 −46 −6], z = 3.72, n = 581). Q-values were affected by DA depletion in the left caudate ([−6 6 4], z = 3.11, n = 222) and bilateral posterior thalamus and substantia nigra ([−10 −22 −4], z = 2.83, n = 126; [12 −30 −6], z = 2.91, n = 88). The Q-values derived from the standard Q-learning model failed to predict significant BOLD signal changes throughout the entire brain. Together, these results indicate that counterfactual learning signals are incorporated into a distributed representation of expected value across cortical and subcortical, as well as neuromodulatory systems.

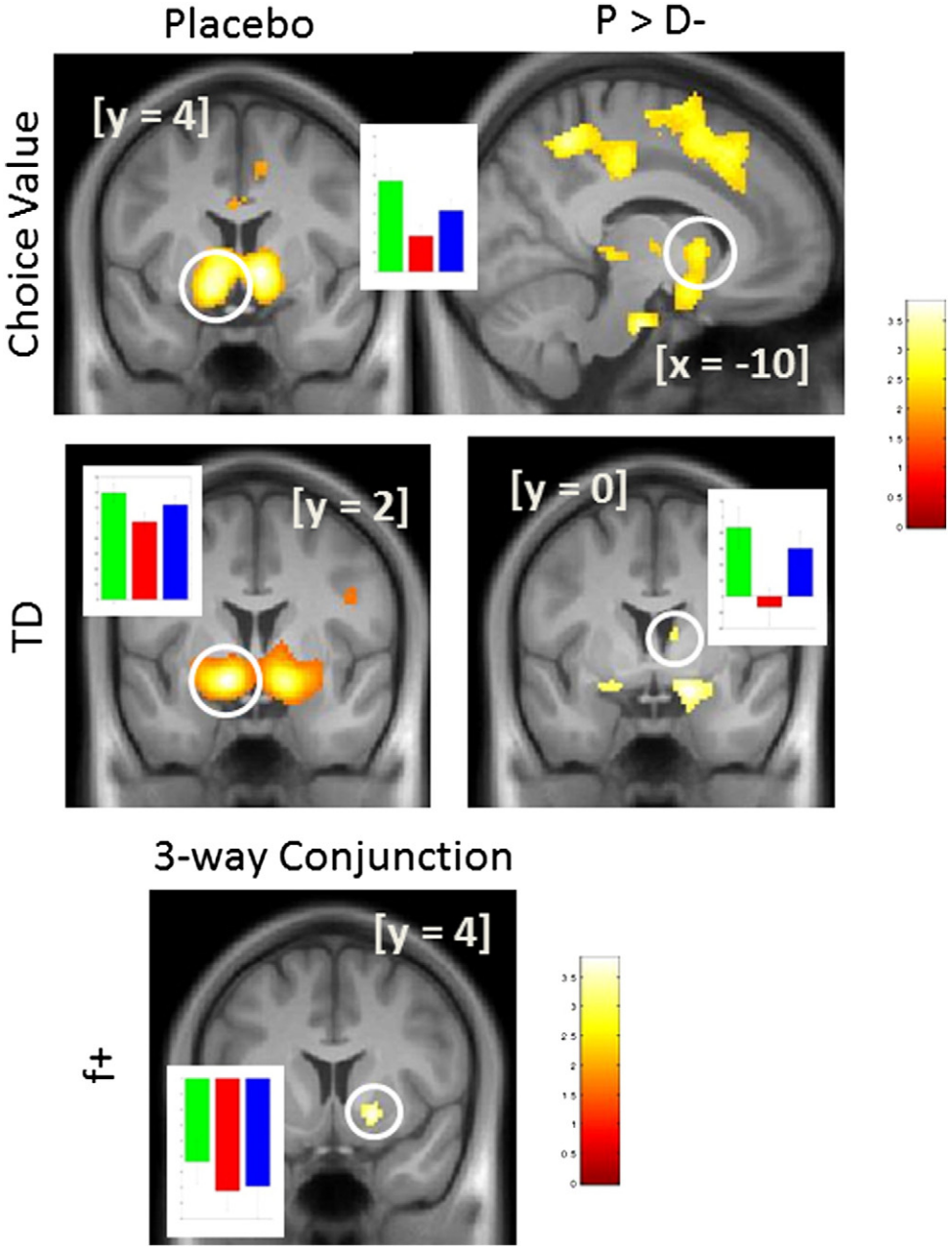

Fig. 4. Neural activity for choice, PE and FPE signals. Top row: choice value modulated activity in the ventral striatum (shown at $p < .05$ whole brain FWE corrected) in the placebo group (left). The modulation was significantly stronger in the placebo group than the dopamine depletion group (right). Middle row: reward prediction error (TD) modulated activity in the ventral striatum (shown at $p < .05$ whole brain FWE corrected) similarly in all three groups (left). The thalamus/caudate and bilateral amygdale were differentially modulated in the placebo and dopamine depletion groups (right). Bottom: the FPETD corresponding to counterfactual losses negatively modulated neural activity in the ventral striatum similarly for all three groups. Inset bar plots are as in Fig. 3.

Neuroimaging results: other modulations and effects of amino acid depletion

Fig. 4 displays the results from other model parameters and effects of depletion on BOLD signal. The nominal choice value (i.e., 0–3 Euros) robustly modulated activity of the ventral striatum ([10 4 4], z = 7.8, n = 157) in the P group, as did the TD error term ([−12 2 −12], z = 11.9, n = 669). Activity in the right ventral striatum ([18 4 −8], z = 3.35, n = 63) was negatively modulated by the *f*+ (i.e., missed opportunities; counterfactual losses). This effect was present across depletion groups as illustrated by the conjunction analysis in Fig. 4 (bottom). No regions of the brain were significantly modulated by *f*− on losing trials (fictive error from counterfactual gains).

The effect of DA depletion on brain activity in comparison to the P group is shown in Fig. 4 (right column). The D− group showed less activity modulation by the nominal choice value in the ventral and mid-dorsal striatum ([−10 8 2], z = 2.9, n = 46; [12 8 0], z = 3.3, n = 58), as well as supplementary and primary motor areas. The P group showed significantly greater modulation by TD error than the D− group in the left thalamus, slightly posterior to the caudate ([10 0 12], z = 2.8, n = 84), and bilateral amygdalae ([−22 −4 −14], z = 2.97, n = 90; [18 0 −16], z = 3.3, n = 54). The TD error derived from the standard Q-model modulated activity in the same set of regions; however, there were no significant group differences for any voxel. Also, no significant differences were observed between the P and S− groups for modulatory effects of choice value in the ventral striatum or ventral PFC, although the right DLPFC ([46 42 34]; z = 4.4, n = 1200) and anterior cingulate cortex ([2 26 32], z = 2.8, n = 828) showed stronger modulation for P group than S−. There were no other differences between the depletion and placebo groups.

The exploratory subdivision of the P group into learners and non-learners (see behavioral results above) also yielded interesting effects in the fMRI data that lend themselves to further interpretation of the neural mechanism of valuation processing with counterfactual learning signals. Since only the P group demonstrated a significant correlation with expected value in their fMRI data from our planned analyses, we only examined these participants for this analysis of sub-groups. Whereas learners from the P group showed a statistically significant correlation with Q-values in the vmPFC ([0 50 −8], z = 2.85, n = 210), non-learners showed a very weak representation of expected value at the time of choice indicated by a non-significant correlation with Q-values (Fig. 5). In addition, the neural response to the *f*+ was significantly different between the learners and non-learners in the P group. A region of the right ventral striatum ([8 −4 −6], z = 3.8, n = 171) and medial OFC ([2 26 −16], z = 3.3, n = 171) was negatively modulated by *f*+ for counterfactual losses in the learners, but positively modulated in the non-learners. Finally, non-learners showed a stronger correlation with the TD error signal (not shown in Fig. 5) in the OFC ([2 48 0], z = 3.4, n = 73).

Discussion

The results of this experiment demonstrate that counterfactual learning signals improve Q-learning model fit, and this improved model predicted BOLD signal changes correlated with expected value and reward PE that were sensitive to dietary manipulations of both dopaminergic and serotonergic neuromodulators. Expected value computed from the FPE model robustly modulated activity in the vmPFC and OFC in the P group. On the other hand, expected value from the standard model failed to predict BOLD signal modulations throughout the brain

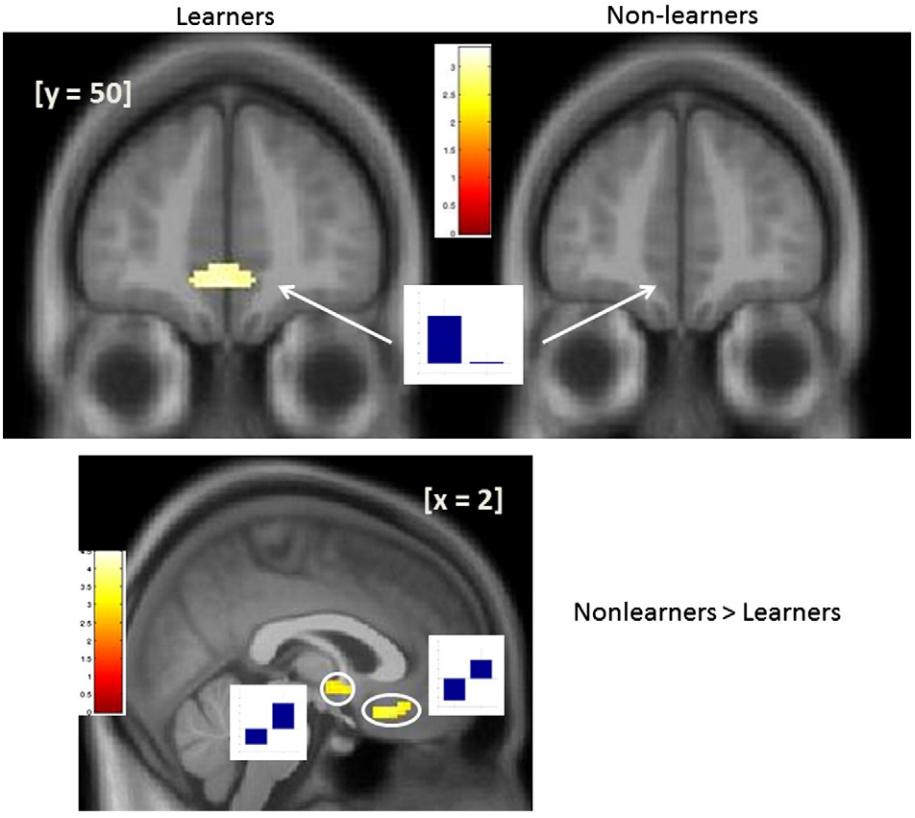

Fig. 5. Expected value and counterfactual losses in successful learning. Top: learners (n = 17) showed a significant modulation in the vmPFC (left) whereas the non-learners (n = 13) did not. The bar plot inset shows the group averaged beta values taken from the peak voxel (MNI [0, 50, −8]) of the cluster shown in the learners. This cluster overlaps that shown in the circled region at the top of Fig. 3 for expected value in the P group only. Bottom: learners and non-learners demonstrated differential modulation by counterfactual losses in the ventral striatum and posterior OFC. The learner groups demonstrated a negative modulation in both regions, and the non-learner groups showed a positive modulation.

even at a liberal threshold of $p < .01$ (uncorrected). The standard model also did not show any effect of either depletion group in comparison with the P group, but the FPE model in conjunction with dietary depletion revealed additional neural structures involved in representing expected value during choice. Whereas dietary depletion was not better for task performance, it proved to be a useful tool for functional brain imaging. As such, taking into account counterfactual outcomes and incorporating them into a representation of action-specific value improved the model's ability to identify a potential neural mechanism of choice behavior that was not revealed by the standard Q-learning model. In addition, the FPE model showed that learners and non-learners differentially utilized counterfactual gains and losses, and it produced differential correlations with expected value and counterfactual losses in the vmPFC and ventral striatum, respectively.

Counterfactual learning signals and expected value

Expectations are derived from experience. The variety of information that is incorporated when generating expected values can influence how other events are subsequently processed. There is accumulating evidence that humans do indeed incorporate counterfactual consequences into subsequent decisions, and that counterfactual consequences modulate neural activity (Bell, 1982; Boorman et al., 2009; Brassen et al., 2012; Buchel et al., 2011; Coricelli et al., 2005; Li and Daw, 2011; Loomes and Sugden, 1982; Nicolle et al., 2011). But none have incorporated FPE signals into valuation for strategic decisions that maximize long-term gains despite interim losses for action-specific valuation. For example, Li and Daw (2011) employed counterfactual outcomes in their study of value-based choices. They used a Rescorla–Wagner learning rule to update their model derived estimates of expected value, which by definition does not take into account future anticipated rewards. They found that neural activity associated with prediction errors was more consistent with a policy updating mechanism rather than a counterfactual valuation system. Also, neither Lohrenz et al. (2007) nor Chiu et al. (2008) included the fictive errors from counterfactual gains or losses in their Q-learning model of expected value. Instead, these two studies used a separate linear regression analysis to determine that only fictive errors from counterfactual losses ($f+$) predicted a change in subsequent choices.

Lohrenz et al. (2007) have previously demonstrated that fictive error signals associated with counterfactual losses explain the amount by which the immediately next bet is changed, but their study did not examine how this FPE might contribute to valuation processing. In their experiment, the $f+$ was computed as the difference between the obtained outcome (the factual reward) and the unobtained outcome (the counterfactual reward), yielding what they refer to as "$f+$", which is the counterfactual loss corresponding to the amount of reward not obtained due to a non-maximal bet on a winning trial. This $f+$ is similar to subjectively experienced regret (as noted in their discussion) for not having bet more after learning that they could have obtained more due to the winning outcome of the trial. Thus, the $f+$ can occur only on a winning trial. This $f+$ was used as a predictor variable in a multiple linear regression analysis to determine if it could predict the amount by which the next bet changed. Indeed, the results of their multiple linear regression analysis showed that including this variable as a predictor yielded a significant positive beta value, indicating that $f+$ predicted a significant increase in the next bet. Lohrenz et al. also included an "$f-$" in their study, which is a counterfactual gain occurring only on trials for which there was a losing outcome. They reported that it did not significantly predict the amount by which subjects changed their immediately next bet. Thus, their findings were congruent with regret-based theorizing in that a missed opportunity led to an increasingly risky bet, although they did not collect data concerning the emotional effect of this missed opportunity to fully state that they identified a "regret" phenomenon.

While Lohrenz et al. did include a Q-learning model in their experiment, it used a temporal difference (TD) prediction error for the factual reward only (as noted in the Method section of their manuscript) and is therefore the same as the standard model used for comparison in our experiment. In our computational model, we have taken their fictive error signals (computed as the difference between the obtained and unobtained outcomes) and further computed a TD error within a multi-stage Q-learning computational model of action-specific valuation for choice. We refer to this as an FPE, and it is the additional computation of this type of prediction error that takes future anticipated rewards into account, which was not included in the experiment by Lohrenz et al., that makes our model unique in its ability to identify whether FPE signals contribute to valuation, and not only to change in betting behavior. The fictive error signal is not counted twice, but rather it is computed and then used for further computation. Only in the second computation (the FPE formula) does it have an effect on valuation. Further, our model included an FPE+ (computed using the $f+$ signal as in Lohrenz et al.) and an FPE− (computed using the $f-$ signal as in Lohrenz et al.) as distinguishable types of error signals, just as reported by Lohrenz et al. However, whereas Lohrenz et al. found that only counterfactual losses ($f+$) significantly explained the amount of change in the next bet, our model shows that counterfactual gains (our FPE−) also significantly contributed to valuation.

In this study, the FPE model nested the standard Q model. If participants had not incorporated counterfactual information into their valuation processes then learning rates for the two FPE parameters would have been zero. The FPE model is then reduced to standard Q-learning and expected values should not differ between the two models. To the contrary, learning rates for both FPE parameters were significantly greater than zero, expected values (Q) were significantly different between the models, and the FPE model explained behavior better, suggesting that participants utilized counterfactual information for valuation computations.

A neural representation of expected value during choice behavior is strongly associated with vmPFC activation in humans. Gläscher et al. (2009) examined the neural representation of expected value during action- and stimulus-specific choices using a Q-learning model of learned expected value. Expected value for both types of choices was significantly correlated with BOLD signal changes in the vmPFC. In another study, Jocham et al. (2011) found that a DA antagonist (Amisulpride) enhanced performance on a value-based decision making task, and that learned expected value, which was also computed from a Q-learning model, modulated activity in the vmPFC. In our data, we identified a distributed neural system involved in the representation of expected value that is anchored in the vmPFC, which is consistent with findings noted above.

The depletion protocol allowed localization of dopaminergic and serotonergic processing in relation to expected value during choice, which expands the representation beyond the previously reported vmPFC. The difference between P and S− groups in the vmPFC was strongly significant due to the reversed modulatory effect of Q-values resulting from 5HT depletion (see Fig. 3). This reverse modulation in the PFC is the same effect of 5HT-depletion reported by Hindi Attar et al. (2012) in a study of Pavlovian prediction error processing with aversive consequences. Also, Seymour et al. (2012) studied the effects of 5HT depletion on choice valuation for both overt reward and pain avoidance choices. Although they did not report the direction of the effect that 5HT depletion had on neural activity, they showed that 5HT depletion altered the representation of expected value in the vmPFC, as well as the caudate. As such, our findings are consistent with these previous effects both in terms of the neural regions affected and the direction of the effect that 5HT depletion has in the PFC.

DA depletion revealed modulation by expected value in the dopaminergic midbrain, in the vicinity of the substantia nigra (SN). The SN has previously been implicated in novelty and memory processing (Bunzeck and Duzel, 2006), as well as reward anticipation (Kirsch

et al., 2003; Morris et al., 2006). While neural activity during the anticipation of a forthcoming reward does not unequivocally indicate involvement in the representation of expected value, it is consistent with the notion that the SN is involved in predictions of the sort that may require updating based on prediction errors. As such, it is likely that the SN plays multiple roles in prediction and reward anticipation, and our findings suggest that these roles are mediated by dopaminergic neuromodulation.

The overlapping effect of expected value in the caudate, for which both depletion groups were significantly different from the P group (see Fig. 3, middle right and bottom right), suggests a locus of integration between the DA and 5HT neuromodulatory systems. Other brain regions that were affected by 5HT depletion did not show an effect of DA depletion, and vice versa, suggesting that these two systems function independently for the most part in the representation of expected value. With this in mind, the fact that this part of the canonical reward system was affected by depletion of both types of neuromodulator implies that each influences the processing of these neurons. It cannot be concluded from this study, however, whether or not these effects are due to a local depletion within the caudate itself, or to a remote effect of depletion in neurons that provide input to the caudate. In fact, it might be that local DA depletion caused one effect, and remote 5HT depletion of caudate afferents caused the other effect. Further research is necessary to disentangle how these two systems interact in these neurons.

This cortico-subcortical system for processing and representing expected value, including the vmPFC, striatum and dopaminergic midbrain, is further substantiated by direct white matter connectivity. There are DTI connections from regions of the SN to ventral, dorsal and lateral striatum (Chowdhury et al., 2013), and also the medial PFC (Menke et al., 2010). The ventral striatum (nucleus accumbens) has white matter fiber tracts connecting it to the vmPFC, and the integrity of these white matter fibers predicts delayed reward discounting (Peper et al., 2013). In addition, Motzkin et al. (2011) reported direct white matter connections from vmPFC to amygdala, which was involved in reward PE processing. Each of these regions was identified as playing an important, although different role in acquiring and representing expected value during choice in the SSIT. These monosynaptic connections suggest that the valuation and value representation network is a core cognitive behavioral network in the brain.

Participants who successfully exploited the task to maximize long-term gains demonstrated a different pattern of brain activity compared to those participants that failed to discover/exploit the task. According to Q-learning, participants that were able to exploit the task and select the optimal path (i.e., learners) did so by maximizing the long-term expected value of their actions. Their representation of expected value was more strongly influenced by counterfactual losses than in the group of non-learners. Previously, counterfactual losses (missed opportunities) have been associated with increased risk taking (Brassen et al., 2012; Buchel et al., 2011; Lohrenz et al., 2007) possibly due to the aversiveness of subjectively experiencing regret at the missed opportunity (Coricelli et al., 2005). In the SSIT, increased risk taking would lead through the optimal path and hence greater sensitivity to counterfactual losses is indeed advantageous.

A neural representation of expected value was present in the vmPFC for the group of participants who learned and exploited the task (Fig. 5). In contrast, the expected values of non-learners were more strongly influenced by counterfactual gains, which may suppress risk taking, and the activation in the vmPFC at the moment of choice was not present. Furthermore, the $f+$ signal was processed differently by learners and non-learners. As shown in Fig. 5, learners demonstrated a significant negative modulation by $f+$ in the ventral striatum, which is contradictory to the positively modulated response to TD/reward signals. This negative modulation is, however, consistent with effects reported by Buchel et al. (2011), as well as Brassen et al. (2012), for which the magnitude of negative modulation also predicted increased risk taking. The non-learners demonstrated a significant positive modulation, which resembles the neural response to the standard TD/reward signal. This suggests that the mismatch between responses to factual and counterfactual consequences in an overlapping region of the ventral striatum may be a potential neural mechanism for computing and incorporating counterfactual learning signals during valuation.

Despite the strong influence of counterfactual gains on expected value, we did not find neural activity that was significantly modulated by the $f-$ regressor for the P group. The first level GLM modeled this regressor as a parametric modulator occurring at the moment of counterfactual outcome presentation specifically on losing trials. Others have not temporally dissociated factual and counterfactual outcome events, and either measure fictive error signals implicitly (Lohrenz et al., 2007) or explicitly (Li and Daw, 2011) at the moment when the factual outcome is revealed. It may be that neural activity time-locked to a different stimulus event (the outcome presentation) may correlate with $f-$.

Previous literature concerning the effects of counterfactual consequences on choice behavior has focused on the interaction of cognitive and emotional effects of counterfactual losses in a variety of experimental paradigms (Sommer et al., 2009). For example, counterfactual losses lead to increased risk taking, and are strongly associated with subjectively experienced regret (Camille et al., 2004; Coricelli et al., 2005). Experiencing regret in the face of a missed opportunity is dependent on the structural and functional integrity of the ventral PFC (Camille et al., 2004), and it follows that adjusting behavior in order to strategically reduce anticipated regret (regret avoidance) involves activation of the posterior OFC (Coricelli et al., 2005). Moreover, healthy older adults that fail to adjust their behavior in response to missed opportunities report experiencing less (or no) regret, and also show differential sensitivity in the ventral striatum to missed opportunities compared to younger participants, or clinically depressed older adults (Brassen et al., 2012). In contrast, counterfactual gains ($f-$) are associated with subjective rejoice or relief, and they bias subsequent behavior toward more conservative choices. These differences in subjectively experienced emotions and behavioral biases suggest that counterfactual gains and losses contribute independently to valuation. Indeed, we found that learning rates associated with FPE+ and FPE− were significantly different, showing that FPE− had a greater effect on expected value than FPE+.

The differential effects of counterfactual gains and losses may be related to the volatility and/or risk inherent to the environment. Counterfactual losses may lead to increased riskiness when volatility is low, but may not exert an influence on choice when volatility or risk is high and ambiguous. Counterfactual gains may lead to more conservative choices when volatility and risk are high or unknown, with relatively small effects when volatility and risk are low and unambiguous (Fujiwara et al., 2009; Henderson and Norris, 2013). Risk and volatility were each ambiguous in the SSIT, and the nature of the environment involved frequent losses. This may explain why FPE− was a stronger influence on learning than FPE+ among our participant groups. Importantly however, the learning rates associated with the counterfactual learning signals dissociated learners from non-learners, with learners utilizing counterfactual losses more so than non-learners, who used counterfactual gains more so than learners. This is consistent with previously reported effects of missed opportunities and regret-related choices on subsequent decisions (Brassen et al., 2012; Buchel et al., 2011; Coricelli et al., 2005; Lohrenz et al., 2007) where more optimal decision making was associated with responsiveness to counterfactual losses specifically.

Differential effects of depletion on behavior and brain activity

Acute amino acid dietary depletion for DA and 5HT did not significantly affect any aspect of behavior in comparison to placebo depletion. Previous experiments report inconsistent effects of dietary depletion on

cognitive performance. For example, 5HT depletion selectively alters reward processing in some studies (Rogers et al., 2003; Schweighofer et al., 2008; Seymour et al., 2012; Tanaka et al., 2007), while others report that it selectively alters punishment processing (Cools et al., 2008; Evers et al., 2005; Robinson et al., 2012), and still others report that serotonergic depletion does not affect reversal learning or set shifting, and yet leads to enhanced decision making (Talbot et al., 2006). Similar inconsistencies exist in the literature regarding the effects of DA depletion for a variety of cognitive tasks (Harmer et al., 2001; McLean et al., 2004; Nagano-Saito et al., 2008, 2012; Robinson et al., 2010), although none have thoroughly investigated economic decision making.

Despite these inconsistent behavioral effects of dietary depletion, both DA and 5HT depletion reliably reduce dopaminergic (Leyton et al., 2004; McTavish et al., 1999; Montgomery et al., 2003) and serotonergic (Crockett et al., 2012; Yatham et al., 2001, 2012) neural activities, respectively. Moreover, depletion alters neural activity even in studies for which no effects on cognitive performance were observed (Evers et al., 2005, 2006). Thus, whereas cognitive-behavioral functioning may be robust to neurotransmitter depletion, neural activity itself shows greater sensitivity, and this allowed recovery of localized, neuromodulator-specific differences in BOLD signal changes in this experiment without confounds due to significantly different behavioral characteristics.

The fMRI analysis based on the FPE model parameters for the P group revealed a cortico-subcortical brain system that was not found in the results stemming from the standard model. The TD error term from the standard model accounted for BOLD signal changes in the ventral striatum and appeared nearly identical to the TD modulation effect from the FPE model shown in Fig. 4. Processing of the TD error from the FPE model was disrupted by DA depletion in the thalamus near the caudate, as well as bilateral amygdale. DA-mediated TD prediction error processing has previously been reported by Pessiglione et al. (2006) and Schonberg et al. (2010). In addition, DA depletion significantly disrupted processing of the nominal choice value on each trial. The effects of 5HT depletion were limited to expected value only. These selective effects of DA and 5HT depletion demonstrate a considerable degree of independence between the two neuromodulatory systems in that manipulating one system does not produce the same effects as manipulating the other system. But it also implies some possibly competitive interactions in that the two systems do not appear to compensate for each other (i.e., 5HT depletion does not produce hyper-responsive DA activity). This opponency-like interaction is further evident in the significantly different learning rates for FPE+ and FPE− between the two depletion groups (Table 1). Table 1 shows that the D− group showed greater sensitivity to counterfactual losses (significantly greater learning rate for the FPE+ than the S− group), whereas the S− group showed greater sensitivity to counterfactual gains, and resembles the differential effects of 5HT and DA on reward and punishment processing reported previously (Boureau and Dayan, 2010; Cools et al., 2011).

Concluding remarks

The OFC and vmPFC, modulated by Q-values in this study, are often cited as part of a valuation system, however, they are each recently acknowledged as important nodes in a long-term memory system for associative information (Euston et al., 2012; Rushworth et al., 2011). Also, we observed modulation in the retrosplenial cortex, which is a region implicated in contextual associative memory processing (Ranganath and Ritchey, 2012). It may be that valuation, decision-making and episodic memory systems interact (or share functional anatomy), which is consistent with the type of processing necessary for learning associations among context, action, events, and consequences in the SSIT paradigm. As such, seemingly incompatible models of memory and decision making may be mutually informative in the development of neurobiologically plausible models of large-scale neurocognitive brain function.

In summary, the neural systems for choice and valuation with counterfactual learning signals include cortical and subcortical structures that involve an interaction of DA and 5HT processing. Model comparison demonstrated that counterfactual processing occurs during reward-based action-specific value learning when such information is available. The depletion procedure proved to be a useful tool for fMRI research because it was able to identify circumscribed neural tissue where a particular type of neuromodulator was selectively involved in processing specific aspects of the task: 5HT is involved in expected value representation by the vmPFC, OFC, and caudate, and, DA is involved for expected value representation in the striatum and midbrain, and is also important for reward PE processing in the striatum. These findings show that the effects of counterfactual consequences on choice can be mediated by a direct effect on action-specific expected values, and contribute to a growing body of research aimed at dissecting the neural substrates of reward-based value learning for optimal choice behavior. Although fictive error signal processing was unimpaired by depletion, these results demonstrated that FPE signals are an important component of valuation and reward-based learning in a computational model, and revealed a possible neural mechanism for incorporating fictive error signals into a more optimal value representation for fiscal decision making.

Acknowledgments

We would like to thank two anonymous reviewers for their insightful comments. Authors M.J.T. and R.G. contributed equally to this work. This research was supported by a Bernstein Prize for Computational Neuroscience BMBF 01GQ1006 to J.G., as well as BMBF 01GQ0912, BMBF 01GQ0911 and DFG GRK 1589/1.

References

Bell, D.E., 1982. Regret in decision making under uncertainty. Oper. Res. 30, 961–981.
Boorman, E.D., Behrens, T.E., Woolrich, M.W., Rushworth, M.F., 2009. How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. Neuron 62, 733–743.
Boorman, E.D., Rushworth, M.F., Behrens, T.E., 2013. Ventromedial prefrontal and anterior cingulate cortex adopt choice and default reference frames during sequential multi-alternative choice. J. Neurosci. 33, 2242–2253.
Boureau, Y.L., Dayan, P., 2010. Opponency revisited: competition and cooperation between dopamine and serotonin. Neuropsychopharmacology 36, 74–97.
Brassen, S., Gamer, M., Peters, J., Gluth, S., Buchel, C., 2012. Don't look back in anger! Responsiveness to missed chances in successful and nonsuccessful aging. Science 336, 612–614.
Buchel, C., Brassen, S., Yacubian, J., Kalisch, R., Sommer, T., 2011. Ventral striatal signal changes represent missed opportunities and predict future choice. Neuroimage 57, 1124–1130.
Bunzeck, N., Duzel, E., 2006. Absolute coding of stimulus novelty in human substantia nigra/VTA. Neuron 51, 369–379.
Camille, N., Coricelli, G., Sallet, J., Pradat-Diehl, P., Duhamel, J.R., Sirigu, A., 2004. The involvement of the orbitofrontal cortex in the experience of regret. Science 304, 1167–1170.
Chandrasekhar, P.V., Capra, C.M., Moore, S., Noussair, C., Berns, G.S., 2008. Neurobiological regret and rejoice functions for aversive outcomes. Neuroimage 39, 1472–1484.
Chiu, P.H., Lohrenz, T.M., Montague, P.R., 2008. Smokers' brains compute, but ignore, a fictive error signal in a sequential investment task. Nat. Neurosci. 11, 514–520.
Chowdhury, R., Lambert, C., Dolan, R.J., Duzel, E., 2013. Parcellation of the human substantia nigra based on anatomical connectivity to the striatum. Neuroimage 81, 191–198.
Cools, R., Robinson, O.J., Sahakian, B., 2008. Acute tryptophan depletion in healthy volunteers enhances punishment prediction but does not affect reward prediction. Neuropsychopharmacology 33, 2291–2299.
Cools, R., Nakamura, K., Daw, N.D., 2011. Serotonin and dopamine: unifying affective, activational, and decision functions. Neuropsychopharmacology 36, 98–113.
Coricelli, G., Critchley, H.D., Joffily, M., O'Doherty, J.P., Sirigu, A., Dolan, R.J., 2005. Regret and its avoidance: a neuroimaging study of choice behavior. Nat. Neurosci. 8, 1255–1262.
Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173.
Crockett, M.J., Clark, L., Roiser, J.P., Robinson, O.J., Cools, R., den Ouden, H., et al., 2012. Converging evidence for central 5HT effects in acute tryptophan depletion. Mol. Psychiatry 17, 121–123.

Daw, N., 2011. Trial-by-trial data analysis using computational models. In: Delgado, M.R., Phelps, E.A., Robbins, T.W. (Eds.), Decision Making, Affect, and Learning. Oxford University Press, Oxford, pp. 3–38.
Doya, K., 2008. Modulators of decision making. Nature Neuroscience 11, 410–416.
Euston, D.R., Gruber, A.J., McNaughton, B.L., 2012. The role of medial prefrontal cortex in memory and decision making. Neuron 76, 1057–1070.
Evers, E.A., Cools, R., Clark, L., van der Veen, F.M., Jolles, J., Sahakian, B.J., Robbins, T.W., 2005. Serotonergic modulation of prefrontal cortex during negative feedback in probabilistic reversal learning. Neuropsychopharmacology 30, 1138–1147.
Evers, E.A., van der Veen, F.M., van Duersen, J., Schmitt, J.A., Deutz, N.E., Jolles, J., 2006. The effect of acute tryptophan depletion on the BOLD response during performance monitoring and response inhibition in healthy male volunteers. Psychopharmacology (Berl) 187, 200–208.
Forman, S.D., Cohen, J.D., Fitzgerald, M., Eddy, W.F., Mintun, M.A., Noll, D.C., 1995. Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a cluster-size threshold. Magn. Reson. Med. 33, 636–647.
Fujiwara, J., Tobler, P.N., Taira, M., Iijima, T., Tsutsui, K., 2009. A parametric relief signal in human ventrolateral prefrontal cortex. Neuroimage 44, 1163–1170.
Gläscher, J., Hampton, A.N., O'Doherty, J.P., 2009. Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. Cereb. Cortex 19, 483–495.
Harmer, C.J., McTavish, S.F.B., Clark, L., Goodwin, G.M., Cowen, P.J., 2001. Tyrosine depletion attenuates dopamine function in healthy volunteers. Psychopharmacology (Berl) 154, 105–111.
Haruno, M., Kawato, M., 2006. Heterarchical reinforcement-learning model for integration of multiple cortico-striatal loops; fMRI examination in stimulus-action-reward association learning. Neural Networks 19, 1242–1254.
Henderson, S.E., Norris, C.J., 2013. Counterfactual thinking and reward processing: an fMRI study of responses to gamble outcomes. Neuroimage 64, 582–589.
Hindi Attar, C., Finkch, B., Buechel, C., 2012. The influence of serotonin on fear learning. PLoS ONE 7, e42397.
Jocham, G., Klein, T.A., Ullsperger, M., 2011. Dopamine-mediated reinforcement learning signals in the striatum and ventromedial prefrontal cortex underlie value-based choices. J. Neurosci. 31, 1606–1613.
Kirsch, P., Schienle, A., Stark, R., Sammer, G., Blecker, C., Walter, B., Ott, U., Burkart, J., Vaitl, D., 2003. Anticipation of reward in a nonaversive differential conditioning paradigm and the brain reward system: an event-related fMRI study. NeuroImage 20, 1086–1095.
Knutson, B., Taylor, J., Kaufman, M., Peterson, R., Glover, G., 2005. Distributed neural representation of expected value. J. Neurosci. 25, 4806–4812.
Leyton, M., Dagher, A., Boileau, I., Casey, K., Baker, G., Diksic, M., et al., 2004. Decreasing amphetamine-induced dopamine release by acute phenylalanine/tyrosine depletion: a PET/[11C]raclopride study in healthy men. Neuropsychopharmacology 29, 427–432.
Li, J., Daw, N.D., 2011. Signals in human striatum are appropriate for policy update rather than value prediction. J. Neurosci. 31, 5504–5511.
Lieberman, M.D., Cunningham, W.A., 2009. Type I and type II error concerns in fMRI research: re-balancing the scales. Soc. Cogn. Affect. Neurosci. 4, 423–428.
Lohrenz, T., McCabe, K., Camerer, C.F., Montague, P.R., 2007. Neural signature of fictive learning signals in a sequential investment task. Proc. Natl. Acad. Sci. U. S. A. 104, 9493–9498.
Loomes, G., Sugden, R., 1982. Regret theory: an alternative theory of rational choice under uncertainty. Econ. J. 92, 805–824.
McLean, A., Rubinzstein, J.S., Robbins, T.W., Sahakian, B.J., 2004. The effects of tyrosine depletion in healthy volunteers: implications for unipolar depression. Psychopharmacology (Berl) 171, 286–297.
McTavish, S.F.B., Cowen, P.J., Sharp, T., 1999. Effects of tyrosine-free amino acid mixture on regional brain catecholamine synthesis and release. Psychopharmacology (Berl) 141, 182–188.
Menke, R.A., Jbabdi, S., Miller, K.L., Matthews, P.M., Zarei, M., 2010. Connectivity-based segmentation of the substantia nigra in human and its implications for Parkinson's disease. Neuroimage 52, 1175–1180.
Montague, P.R., King-Cassas, B., Cohen, J.D., 2006. Imaging valuation models of choice. Annual Review of Neuroscience 29, 417–448.
Montgomery, A.J., McTavish, S.F., Cowen, P.J., Grasby, P.M., 2003. Reduction of brain dopamine concentration with dietary tyrosine plus phenylalanine depletion: an [11C] raclopride PET study. Am. J. Psychiatry 160, 1887–1889.
Morris, G., Nevet, A., Arkadir, D., Vaadia, E., Bergman, H., 2006. Midbrain dopamine neurons encode decision for future action. Nat. Neurosci. 9, 1057–1063.
Motzkin, J.C., Newman, J.P., Kiehl, K., Koenigs, M., 2011. Reduced prefontal connectivity in psychopathy. J. Neurosci. 31, 17348–17357.
Nagano-Saito, A., Leyton, M., Monchi, O., Goldberg, Y.K., He, Y., Dagher, A., 2008. Dopamine depletion impairs frontostriatal functional connectivity during a set-shifting task. J. Neurosci. 28, 3697–3706.
Nagano-Saito, A., Cisek, P., Perna, A.S., Shirtel, F.Z., Benkelfat, C., Leyton, M., Dagher, A., 2012. From anticipation to action, the role of dopamine in perceptual decision making: an fMRI-tyrosine depletion study. J. Neurophysiol. 108, 501–512.
Nicolle, A., Bach, D.R., Driver, J., Dolan, R.J., 2010. A role for the striatum in regret-related choice repetition. J. Cogn. Neurosci. 23, 845–856.
Nicolle, A., Fleming, S.M., Bach, D.R., Driver, J., Dolan, R.J., 2011. A regret-induced status quo bias. J. Neurosci. 31, 3320–3327.
O'Doherty, J.P., 2004. Reward representations and reward-related learning in the human brain: insights from neuroimaging. Curr. Opin. Neurobiol. 14, 769–776.
Peper, J.S., Mandl, R.C.W., Braams, B.R., de Water, E., Heijboer, A.C., et al., 2013. Delay discounting and frontostriatal fiber tracts: a combined DTI and MTR study on impulsive choices in healthy young adults. Cereb. Cortex 23, 1695–1702.
Pessiglione, M., Seymour, B., Flandin, G., Dolan, R.J., Frith, C.D., 2006. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature 442, 1042–1045.
Ranganath, C., Ritchey, M., 2012. Two cortical systems for memory-guided behaviour. Nature Reviews Neuroscience 13, 713–726.
Robinson, O.J., Standing, H.R., DeVito, E.E., Cools, R., Sahakian, B.J., 2010. Dopamine precursor depletion improves punishment prediction during reversal learning in healthy females but not males. Psychopharmacology (Berl) 211, 187–195.
Robinson, O.J., Cools, R., Sahakian, B.J., 2012. Tryptophan depletion disinhibits punishment but not reward prediction: implications for resilience. Psychopharmacology (Berl) 219, 599–605.
Rogers, R.D., 2011. The roles of dopamine and serotonin in decision making: evidence from pharmacological experiments in humans. Neuropsychopharmacology 36, 114–132.
Rogers, R.D., Tunbridge, E.M., Bhagwagar, Z., Drevets, W.C., Sahakian, B.J., Carter, C.S., 2003. Tryptophan depletion alters the decision-making of healthy volunteers through altered processing of reward cues. Neuropsychopharmacology 28, 153–162.
Rushworth, M.F., Noonan, M.P., Boorman, E.D., Walton, M.E., Behrens, T.E., 2011. Frontal cortex and reward-guided learning and decision-making. Neuron 70, 1054–1069.
Schonberg, T., O'Doherty, J.P., Joel, D., Inzelberg, R., Segev, Y., Daw, N.D., 2010. Selective impairment of prediction error signaling in human dorsolateral but not ventral striatum in Parkinson's disease patients: evidence from a model-based fMRI study. Neuroimage 49, 772–781.
Schweighofer, N., Bertin, M., Shishida, K., Okamoto, Y., Tanaka, S.C., Yamawaki, S., Doya, K., 2008. Low-serotonin levels increase delayed reward discounting in humans. J. Neurosci. 28, 4528–4532.
Seymour, B., Daw, N.D., Roiser, J.P., Dayan, P., Dolan, R., 2012. Serotonin selectively modulates reward value in human decision-making. J. Neurosci. 32, 5833–5842.
Sokol-Hessner, P., Hsu, M., Curley, N.G., Delgado, M.R., Camerer, C.F., Phelps, E.A., 2009. Thinking like a trader selectively reduces individuals' loss aversion. Proc. Natl. Acad. Sci. 106, 5035–5040.
Sommer, T., Peters, J., Glascher, J., Buchel, C., 2009. Structure–function relationships in the processing of regret in the orbitofrontal cortex. Brain Struct. Funct. 213, 535–551.
Sutton, R.S., Barto, A.G., 1981. Reinforcement Learning: An Introduction (Adaptive Computation and Machine Learning). MIT Press, Cambridge, MA.
Talbot, P.S., Watson, D.R., Barrett, S.L., Cooper, S.J., 2006. Rapid tryptophan depletion improves decision-making cognition n healthy humans without affecting reversal learning or set shifting. Neuropsychopharmacology 31, 1519–1525.
Tanaka, S.C., Schweighofer, N., Asahi, S., Shishida, K., Okamoto, Y., Yamawaki, S., Doya, K., 2007. Serotonin differentially regulates short- and long-term prediction of rewards in the ventral and dorsal striatum. PLoS One 2, e1333.
Watkins, C., Dayan, P., 1992. Q-learning. Mach. Learn. 8, 279–292.
Yatham, L.N., Liddle, P.F., Shiah, I.S., Lam, R.W., Adam, M.J., Zis, A.P., Ruth, T.J., 2001. Effects of rapid tryptophan depletion on brain 5-HT(2) receptors: a PET study. Br. J. Psychiatry 178, 448–453.
Yatham, L.N., Liddle, P.F., Sossi, V., Erez, J., Vafai, N., Lam, R.W., Blinder, S., 2012. Positron emission tomography study of the effects of tryptophan depletion on brain serotonin(2) receptors in subjects recently remitted from major depression. Arch. Gen. Psychiatry 69, 601–609.
Young, S.N., Smith, S.E., Pihl, R.O., Ervin, F.R., 1985. Tryptophan depletion causes a rapid lowering of mood in normal males. Psychopharmacology (Berl) 87, 173–177.
Zeelenberg, M., Pieters, R., 2007. A theory of regret regulation 1.0. J. Consum. Psychol. 17, 3–18.