Behavioral/Systems/Cognitive

Serotonin Selectively Modulates Reward Value in Human Decision-Making

Ben Seymour,[1] Nathaniel D. Daw,[2,3] Jonathan P. Roiser,[4] Peter Dayan,[3] and Ray Dolan[1]

[1]Wellcome Trust Centre for Neuroimaging, University College London (UCL), London WC1N 3BG, United Kingdom, [2]Center for Neural Science, Department of Psychology, New York University, New York, New York 10012, and [3]Gatsby Computational Neuroscience Unit and [4]Institute of Cognitive Neuroscience, UCL, London WC1N 3BG, United Kingdom

Establishing a function for the neuromodulator serotonin in human decision-making has proved remarkably difficult because if its complex role in reward and punishment processing. In a novel choice task where actions led concurrently and independently to the stochastic delivery of both money and pain, we studied the impact of decreased brain serotonin induced by acute dietary tryptophan depletion. Depletion selectively impaired both behavioral and neural representations of reward outcome value, and hence the effective exchange rate by which rewards and punishments were compared. This effect was computationally and anatomically distinct from a separate effect on increasing outcome-independent choice perseveration. Our results provide evidence for a surprising role for serotonin in reward processing, while illustrating its complex and multifarious effects.

Introduction

The function of serotonin in motivation and choice remains a puzzle. Existing theories propose a diversity of roles that include representing aversive values or prediction errors, behavioral flexibility, delay discounting, and behavioral inhibition (Soubri, 1986; Deakin and Graeff, 1991; Daw et al., 2002; Doya, 2002; Robbins and Crockett, 2009; Boureau and Dayan, 2011; Cools et al., 2011; Rogers, 2011). Various of these theories appeal to interactions between reward and punishment, with serotonin acting as an opponent to the neuromodulator dopamine, whose involvement in appetitive motivation and choice is rather better established (Schultz et al., 1997; Bayer and Glimcher, 2005). However, although 5-HT2c receptors exert a direct inhibitory influence on dopamine neurons (Di Matteo et al., 2002), facilitative effects of serotonin have been observed via other receptor subtypes (including 5HT2a receptors) (Di Matteo et al., 2008) and several studies manipulating central serotonin levels have reported effects on reward processing (Evers et al., 2005; Roiser et al., 2006; Cools et al., 2008a,b; Crockett et al., 2009; Tanaka et al., 2009), along with perseveration in choosing options that offer dwindling returns (or even intermittent punishment) (Cools et al., 2008a). A formidable problem in this domain is that existing tasks have not succeeded in probing the distinct computational aspects of reward and punishment learning in a selective manner. The goal of the present study was to dissociate these different aspects of decision-making and to test the role of serotonin.

Here, we studied simultaneous reward and avoidance learning in a four-armed bandit paradigm (Fig. 1) and probed the contribution of central serotonin loss induced by acute dietary tryptophan depletion (Carpernter et at., 1998). On each trial, subjects ($n = 30$) selected one of four possible options, each associated with a chance of reward (20 pence) and a chance of punishment (a painful electric shock). Importantly, on each trial, each rewarding or punishing outcome was delivered probabilistically according to an independent stochastic function, allowing us to determine their effects on choice behavior and neural responses unambiguously. That is, the probabilities of reward and punishment were independent from one another, as well as independent between options. These probabilities evolved slowly over time between bounds of 0 and 0.5 according to separate random diffusions.

Materials and Methods

Subjects. The study was approved by the Joint Ethics Committee of the Institute of Neurology and National Hospital for Neurology and Neurosurgery, and all subjects gave informed consent before participating. We studied 30 healthy subjects, recruited by local advertisement. We excluded subjects according to the following criteria (numbers in parentheses refer to the number of exclusions for subjects answering our initial advertisement): standard exclusion criteria for MRI scanning (2 subjects), any history of neurological (including any ongoing pain) or psychiatric illness (6 subjects), history of depression in a first-degree relative (6 subjects). In female subjects, the time around menses was avoided. Once selected, no subjects were subsequently excluded for any reason.

Experimental procedure. Subjects fasted on the night before the study and were asked to avoid high-tryptophan-containing foods on the day before the study. They attended in the morning when consent was gained and blood taken for baseline (T_0) quantification of serum amino acid

Received Jan. 5, 2012; revised Feb. 7, 2012; accepted Feb. 28, 2012.

Author contributions: B.S., N.D.D., J.P.R., P.D., and R.D. designed research; B.S., N.D.D., J.P.R., and P.D. performed research; B.S., N.D.D., J.P.R., and P.D. analyzed data; B.S., N.D.D., J.P.R., P.D., and R.D. wrote the paper.

This work was supported by a Wellcome Trust Programme Grant to R.D. We thank Trevor Robbins, Wako Yoshida, and Quentin Huys for useful discussions.

Correspondence should be addressed to Ben Seymour, Wellcome Trust Centre for Neuroimaging, Leopold Muller Functional Imaging Lab, 12 Queen Square, London WC1N 3BG, UK. E-mail: bjs49@cam.ac.uk.

DOI:10.1523/JNEUROSCI.0053-12.2012

 0270-6474/12/325833-10$15.00/0

Table 1. Amino acids constituents by mass

Isoleucine	4.2 g
Leucine	6.6 g
Lysine	4.8 g
Methionine	1.5 g
Phenylalanine	6.6 g
Threonine	3.0 g
Valine	4.8 g
Tryptophan or placebo	3 or 0 g

Table 2. Bond and Larder scale (1974)

Alert/Drowsy
Calm/Excited
Strong/Feeble
Clear-Headed/Muzzy
Well-coordinated/Clumsy
Energetic/Lethargic
Contented/Discontented
Tranquil/Troubled
Quick-witted/Mentally slow
Relaxed/Tense
Attentive/Dreamy
Proficient/Incompetent
Happy/Sad
Amicable/Antagonistic
Interested/Bored
Gregarious/Withdrawn

concentrations. Subjects received a computerized tutorial explaining in detail the nature of the task, including explicit instruction on the independence of reward and punishment, the independence of each bandit from each other, and the nonstationarity of the task. Each of these points were supported by demonstrations and componential practice tasks, after which subjects moved on to perform a genuine practice task, with only the absence of shock delivery (still, however, displayed on the screen) differing it from the subsequent experimental task. At this time, subjects also underwent a pain thresholding procedure (see Delivery of painful shocks, below). Subjects then ingested the amino acid tablets and relaxed in a quiet area until 5 h had elapsed, at which time blood was taken again (T_5). The subjects then entered the scanner to perform the task.

After the amino acid ingestion, during the waiting period, subjects completed the Cloninger tridimensional personality questionnaire. Subscales for novelty-seeking, harm-avoidance, and reward dependence did not correlate with behavioral parameters for average reward or reward–aversion trade-off, and as such the data are not reported.

Tryptophan-depletion procedure and serum assay. We used a randomized, placebo-controlled, double-blind, tryptophan-depletion design (Fernstrom and Wurtman, 1972). This involved ingestion of a tryptophan-depleted (TRP−) or sham amino acid (TRP+) mixture (for amounts, see Table 1).

The amino acid mixture was commercially mixed (DHP Pharma) and capsulated in 500 mg capsules ($n = 76$ capsules) and labeled according to the blinding protocol. For subjects administered the sham depletion mixture, six capsules contained lactose (total 3 g). This procedure allows subjects to fully ingest all the amino acids without significant gastrointestinal side effects, notably nausea, common with standard-dose tryptophan depletion in which the mixture is prepared as a suspension in water. No subject suffered from such side effects in the present study. The unblinding code was supplied in sealed envelopes, opened after all 30 subjects had completed the study.

Immediately after venupuncture, blood was centrifuged at 3000 rpm for 5 min, and serum was separated by centrifugation and stored at −20°C. Plasma total amino acid concentrations (tyrosine, valine, phenylalanine, isoleucine, leucine, and tryptophan) were measured by means of high-performance liquid chromatography with fluorescence end-point detection and precolumn sample derivatization adapted from the methods of Fürst et al. (1990). Norvaline was used as an internal standard. The limit of detection was 5 nmol/ml using a 10 μl sample volume, and interassay and intraassay coefficients of variation were <15% and <10%, respectively.

The tryptophan:large neutral amino acid (TRP:ΣLNAA) ratio is an indicator of central serotonergic function (Carpenter et al., 1998) and is shown below for each group at T_0 and T_5. Although the efficacy of acute tryptophan depletion has been questioned (van Donkelaar et al., 2011), there is a consensus that it is a reliable index of serotonergic signaling in the brain (for recent discussion, see Crockett et al., 2012). Depletion permits inferences about whether serotonin may be necessary for a specified task or task component, but it does not allow inferences of sufficiency. Other methods, such as selective serotonin reuptake inhibitor administration or loading tryptophan to achieve elevated levels of serotonin can emulate this, but are not used in this task. Here, tryptophan depletion robustly decreased the TRP:ΣLNAA ratio relative to sham depletion (group*time interaction: $p < 0.001$). Note that contrasting precontrast and postcontrast TRP:ΣLNAA controls for individual variation in baseline levels of tryptophan. The mean (SD) preprocedure (T_0) and postprocedure (T_5) of this measure are as follows: control group preprocedure = 0.1452 (0.0578), depleted group preprocedure = 0.1419 (0.0582), control group postprocedure = 0.1742 (0.0948), depleted group postprocedure = 0.0210 (0.0179).

To assess the side effects of the tryptophan-depletion procedure, we administered standard 10-point visual analog scales (VAS) assessing subjective states, as shown in Table 2 (Bond and Lader, 1974).

Subjects scored higher on the aggregate VAS at the end of the experiment (mean increase in VAS score, 0.34 per item; SE, 0.21), but there was no correlation with TRP:ΣLNAA ratio ($r = -0.056$, $p = 0.77$).

We also administered the Hamilton Depression Rating Scale [12 question version: mood, guilt, suicide, work, retardation, agitation, anxiety (psychological and somatic), depersonalization, paranoia, obsessiveness] before ingestion of the amino acids and before the task itself. This showed no evidence of existing depression in any subject at baseline, and no effect on mood of the tryptophan-depletion procedure.

Experimental task. Subjects performed a probabilistic instrumental learning task involving aversive (painful electric shocks) and appetitive (financial rewards) outcomes. This equated to a four-armed bandit decision-making task, with nonstationary, independent outcomes. Each trial commenced with the presentation of the four bandits (Fig. 1), following which subjects had 3.5 s to make a choice. If no choice was made (which occurred either never or very rarely across subjects), the trial would skip to the next trial automatically. After a choice was made, the chosen option was highlighted, all options remained on the screen, and an interval of 3 s elapsed before presentation of the outcome. If subject won the reward, "20p" appeared overlain on the chosen option. If the subject received a painful shock, the word "shock" appeared overlain on the chosen option, and a shock was delivered to the hand (see Delivery of painful shocks, below) simultaneously. If both shock and reward were received, both "20p" and "shock" appeared overlain on the chosen option, one above the other, and the shock was delivered. The outcome was displayed for 1 s, after which the options extinguished and the screen was blank for a random interval of 1.5–3.5 s.

Delivery of painful shocks. Two silver chloride electrodes were placed on the back of the left hand, through which a brief current was delivered to cause a transitory aversive sensation, which feels increasingly painful as the current is increased. Current was administered as a 1 s train of 100 Hz pulses of direct current, with each pulse being a 2 ms square waveform, administered using a Digitimer DS3 current stimulator, which is fully certified for human and clinical use. The stimulator was housed in an aluminum-shielded and MRI-compatible box within the scanner room, from which the electrode wires emerged and traveled to the subject. The equipment configuration was optimized by extensive testing to minimize radio frequency noise artifact during stimulation.

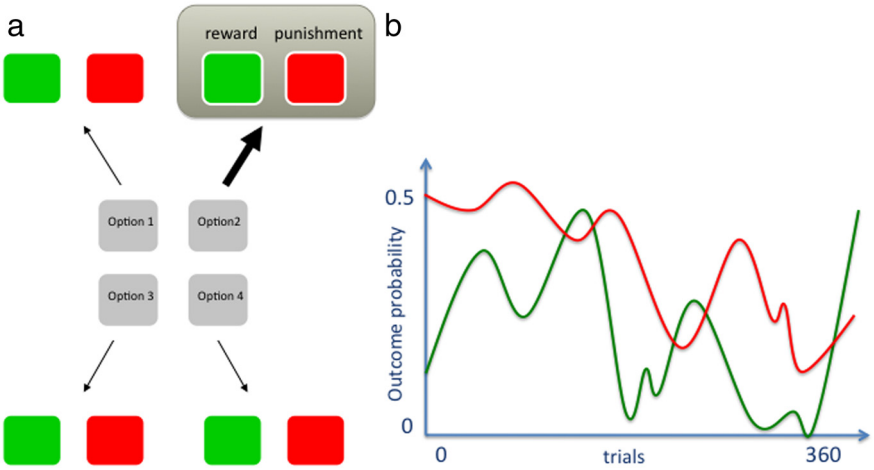

Figure 1. Task design. ***a***, Subjects were required to pick one of four options on each trial. Three seconds after selection, the option yield and outcome that included possible reward (20p) or punishment (painful shock) was given. ***b***, Exemplar graph for one option indicating that the probability of reward and punishment was independent and varied slowly over trials. All options were independent of each other. Outcomes were delivered simultaneously and followed by a variable intertrial interval of 2–6 s.

Painful shock levels were calibrated to be appropriate for each participant. Participants received various levels of electric shocks to determine the range of current amplitudes to use in the actual experiment. They rated each shock on a visual analog scale of 0–10, from "no pain at all" to "the worst possible pain". This allowed us to use subjectively comparable pain levels for each participant in the experiment.

We administered shocks, starting at extremely low intensities, and ascending in small step sizes, until subjects reached their maximum tolerance. No stimuli were administered above the participant's stated tolerance level. Once maximum tolerance was reached, they received a random selection of 14 subtolerance shocks, which removed expectancy effects implicit in the incremental procedure. We statistically fitted a Weibull (sigmoid) function to participants' rating for the 14 shocks and estimated the current intensities that related to a level 8/10 VAS score of pain (Fig. 2*b*). The participants rated another set of 14 random subtolerance shocks at the end of experiment. Subjects' VAS ratings of the pain habituated slightly by the end of the experiment, by a mean of 0.859 on the 0–10 scale (SE, 0.742). There was no correlation between the degree of habituation and ΔTRP: ΣLNAA ratio ($r = 0.056$, $p = 0.78$).

fMRI scanning details. Functional brain images were acquired on a 1.5T Sonata Siemens AG scanner. Subjects lay in the scanner with foam head-restraint pads to minimize any movement. Images were realigned with the first volume, normalized to a standard echo-planar imaging template, and smoothed using a 6 mm full-width at half-maximum Gaussian kernel. Realignment parameters (see fMRI analysis, below) were inspected visually to identify any potential subjects with excessive head movement. This was satisfactory in all subjects, and so none were excluded.

The task was displayed on a computer monitor projected into the head coil and visible on a screen at the end of the magnet bore, visible by the subjects by way of an angled head-coil mirror. Subjects made their choices using a four-button response pad held by their right side.

Behavioral analysis and reinforcement learning model. The logit regression analysis determines the individual influence that rewards, punishments, and previous choices at successively distant choices have on future choices. For reward and punishment, it estimates the decision weight, which represents the bias to choose a particular option according to the outcome experienced from selection of that option in the recent past. This formalizes the effect of reward and punishment on choice, and hence models their independent influence on choice in a way that simply looking at the repetition frequencies does not (Fig. 3*a*).

Thus, the net reward weight for option i is given by the history of rewards in the recent past:

$$\sum_t \omega^{\text{reward}} = x_{t-1}\omega_{t-1}^{\text{reward}} + x_{t-2}\omega_{t-2}^{\text{reward}} + x_{t-3}\omega_{t-3}^{\text{reward}} + x_{t-4}\omega_{t-4}^{\text{reward}} + x_{t-5}\omega_{t-5}^{\text{reward}} \quad (1)$$

where ω specifies the weights across trial lags t, and $x = 1$ if a reward was delivered, and 0 otherwise. The net punishment is determined similarly, given the recent history of punishments. We also incorporated the inherent tendency to repeat choices (ω^{choice}) regardless of outcomes (i.e., perseveration) (Lau and Glimcher, 2005).

The overall tendency to choose a particular option is determined by a logistic choice (softmax) rule, the sum of these independent weights:

$$p(\text{choice} == i) = \frac{\exp\left(\sum_t \omega^{\text{reward}} + \sum_t \omega^{\text{punishment}} + \sum_t \omega^{\text{choice}}\right)}{\sum_i \exp\left(\sum_t \omega^{\text{reward}} + \sum_t \omega^{\text{punishment}} + \sum_t \omega^{\text{choice}}\right)} \quad (2)$$

We used a maximum likelihood method to determine the parameters given the data:

$$\text{LogLikelihood} = \sum_{\text{trial}} \log(p(\text{choice (trial)})) \quad (3)$$

We used an exponential kernel to parameterize the decay of the weights at successively distant trial lags. The parameters include outcome sensitivity and their corresponding decay rates. For example, the reward kernel becomes

$$\sum_t \omega^{\text{reward}} = x_{t-1}R + x_{t-2}\alpha R + x_{t-3}\alpha^2 R + x_{t-4}\alpha^3 R + x_{t-5}\alpha^4 R \quad (4)$$

where $x = 1$ if a reward was given n trials ago, and 0 otherwise, as previously. α is the decay rate, and R is the reward sensitivity.

The exponentially decaying conditional logit model emulates a reinforcement learning model, and we can use this model to estimate the prediction errors by which values are learned (in the fMRI analysis). Accordingly, we can specify an action-learning process with separate appetitive and aversive components, with the integrated action value for option i being the sum of independent action values:

$$Q_i = Q_i^{\text{reward}} + Q_i^{\text{punishment}} \quad (5)$$

Action values are learned using a prediction error, such that for the chosen option on each trial,

$$Q_i^{\text{reward}}(t+1) \leftarrow Q_i^{\text{reward}}(t) + \alpha^{\text{reward}}(R - Q_i^{\text{reward}}) \quad (6)$$

and the nonchosen options decay as follows:

$$Q_i^{\text{reward}}(t+1) \leftarrow Q_i^{\text{reward}}(t) + \alpha^{\text{reward}}(-Q_i^{\text{reward}}) \quad (7)$$

where r is the reward outcome, and α^{reward} is the reward-learning rate. Punishment learning proceeds in the same way. Choice is determined using the softmax learning rule:

$$p(\text{choice} == \text{i}) = \frac{\exp(Q_i + C_i)}{\sum_i \exp(Q_i + C_i)} \quad (8)$$

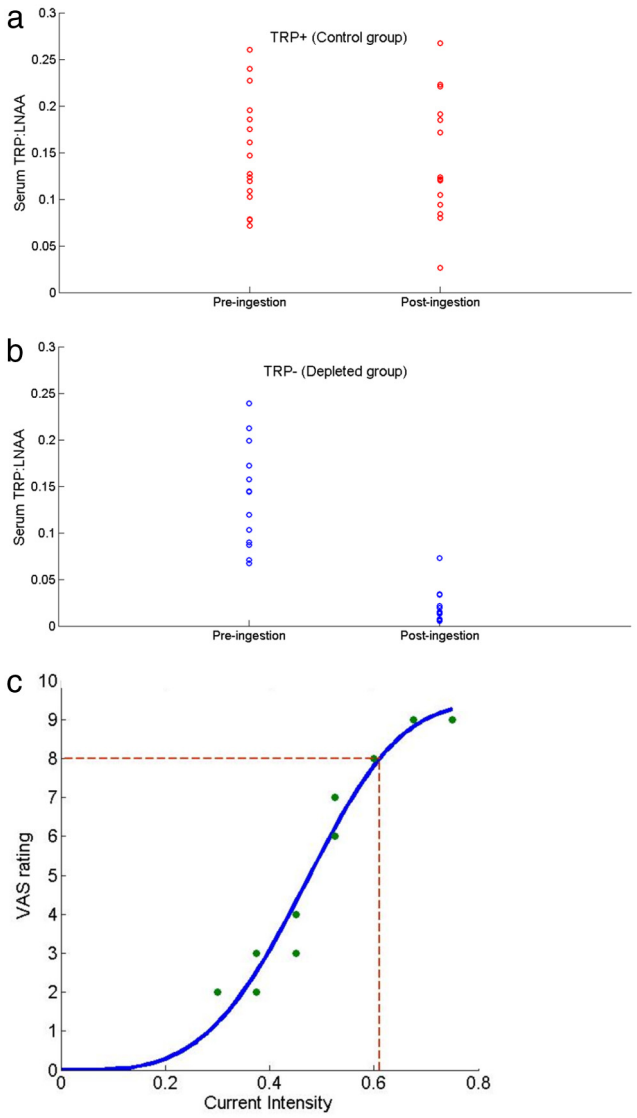

Figure 2. ***a***, ***b***, Pre-ingestion (***a***) and 5 h post-ingestion (***b***) of serum Tryp:LNAA in the TRP+ (Control) and TRP− (Depleted) groups. ***c***, Current—VAS responses showing a statistically fit sigmoid response function and estimated current intensity for VAS = 8 in an example subject.

This incorporates an exponentially decaying choice kernel *C*, in which positive values favor choice perseveration. Parameters were estimated using a maximum likelihood procedure, as previously, implemented using the optimization toolbox in Matlab (Mathworks).

We also considered whether a specific reward × punishment interaction might have an additive predictive role on choice by separately including reward, punishment, and reward + punishment as independent regressors. However, this did not identify a significant interaction effect.

fMRI analysis. Images were analyzed in an event-related manner using the general linear model in Statistical Parametric Mapping (SPM5), with the onsets of each outcome represented as a stick function to provide a stimulus function.

The regressors of interest were generated by convolving the stimulus function with a hemodynamic response function. For the first imaging analysis (of choice value), the regressors of interest were the reward, punishment, and choice values for the chosen option from the logit regression above. These were modeled at the onset of the cue (choice). For the second analysis (prediction error), we used the reward-specific and punishment-specific prediction errors modeled at two time points: the onset of the cue and the onset of the outcome. The choice kernel was modeled parametrically at the time of the choice (taken as cue onset time). We used the best fitting parameters at a group level across all subjects to test the null hypothesis that there was no difference between TRP+/TRP−.

Effects of no interest included the onsets of visual cues, the onsets of rewards, the onsets of the shocks, and the realignment parameters from the image preprocessing.

Single-subject parameter estimate maps were combined using the summary-statistics approach to random-effects analysis to determine (1) regions where parametric effects were expressed regardless of tryptophan status, using one-sample *t* tests; and (2) regions where parametric effects differed according to tryptophan status, by including ΔTRP:ΣLNAA ratio as a covariate in the analysis. This emulates the *t* test, but is more sensitive since it incorporates the subject-level differences in the efficacy of tryptophan depletion. Note however, it does not explain individual differences over and above that explained by the contrast between the two groups.

To emulate Daw et al. (2006), we also compared exploratory versus exploitative trials by defining those choices in which subjects chose an option with the estimated nonmaximum value (of the set of 4 options) as exploratory, and compared responses occurring at the time of the cue with trials in which subjects chose the estimated maximum value.

Correction for multiple comparisons. We specifically hypothesized that choice value should modulate activity in ventromedial PFC (vmPFC), since this area has been consistently identified in both appetitive and aversive (avoidance) goal values in previous studies of instrumental decision making (Kim et al., 2006; Pessiglione et al., 2006; Plassman et al., 2010). To correct for multiple comparisons, we therefore specified an 8 mm radius region of interest encompassing this area [based on a priori coordinates: vmPFC: 3, 36, −18 (Plassmann et al., 2010); caudate nucleus: 9, 0, 18 (Daw et al., 2006); dorsal putamen: 21, 6, 12 (Schonberg et al., 2010)] and accept a family-wise error (FWE) correction of $p < 0.05$ within it. For the overlapping (conjunction) representation of reward and avoidance values, we required the representation of each to be individually significant at this threshold, i.e., both reward and avoidance value must satisfy this level of significance in isolation. For display purposes, this is shown at $p < 0.01$ (reward) × $p < 0.01$ (avoidance). For the representation of prediction error, we hypothesized that activity in the dorsal striatum should be modulated, as this has been consistently identified in the representation of prediction errors in instrumental (action-based) learning (O'Doherty et al., 2004). Again, the representation of overlapping error responses was required to satisfy the stringent requirement that both appetitive and aversive (avoidance) error should individually be significant at an FWE $p < 0.05$. The modulation of perservative tendency is more difficult to generate a serotonin-specific prior hypothesis on, and so we specified a whole-brain corrected FWE $p < 0.05$ as an acceptable level of significance.

Since it is important to report the potential role of other areas in task, we also identified areas at a lenient threshold of $p < 0.001$ (uncorrected) in regions previously well studied in decision-making tasks [dorsomedial prefrontal cortex, ventral putamen, nucleus accumbens, cerebellum, anterior insula, anterior cingulate cortex, brainstem (dorsal raphé nucleus)]. These are not areas that we specified in our principle a priori hypotheses, but are of potential interest in the context of the task.

Results

Our task entailed subjects having constantly to relearn the values of each option, and balance information acquisition (exploration) with reward acquisition and punishment avoidance (exploitation). Thus, in deciding what to choose, the task inherently required participants to balance the values of qualitatively distinct outcomes, namely a primary aversive outcome (pain) and a secondary appetitive outcome (money). For instance, in performing the task, subjects could concentrate solely on winning money and ignore the pain, or concentrate on avoiding pain and ignore the money, or somehow trade the two off against each other.

Subjects performed 360 trials, concatenated over three sessions. We manipulated brain serotonin using an acute dietary tryptophan depletion procedure in a between-subjects, random-

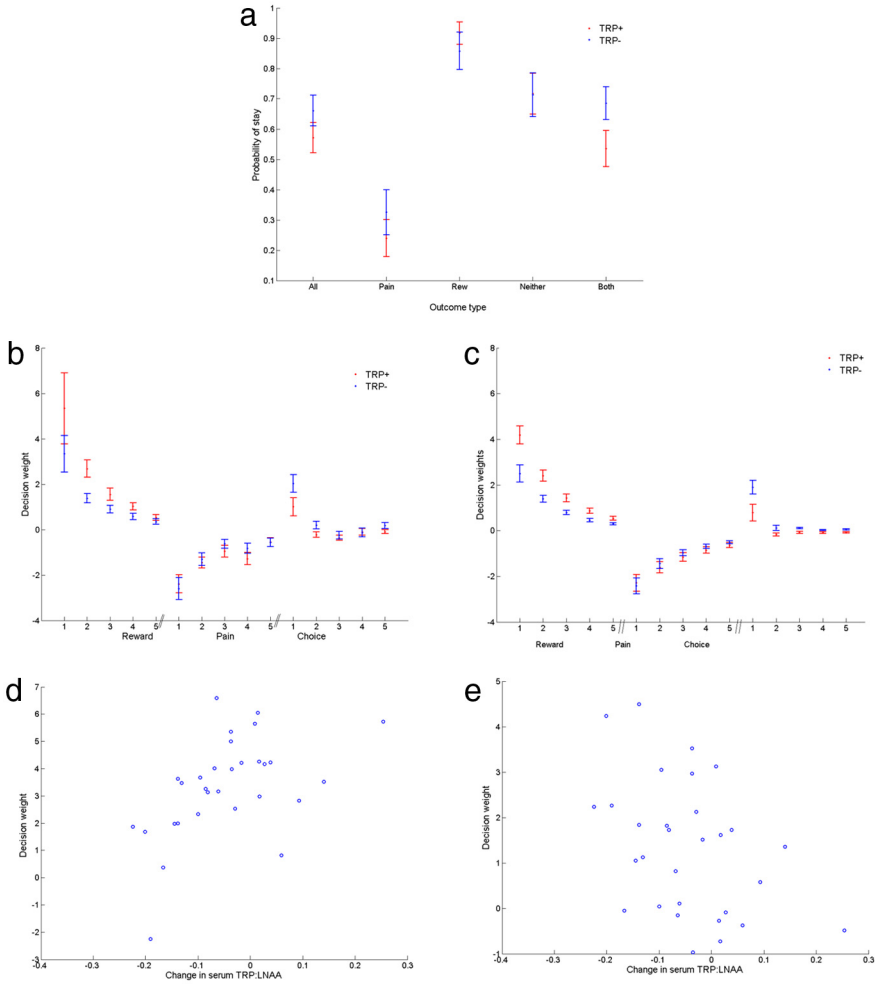

Figure 3. Behavioral results. ***a***, Probability (frequency/total number of trials) of repeating a choice on a particular option given any outcome (All) and given each of the four possible outcomes: reward alone (Rew), pain alone (Pain), reward and pain (Both), and neither reward nor pain (Neither). ***b***, Independent decision weights in TRP+ and TRP− groups. This shows the weights by which rewards, punishments, or previous choices governed the tendency to repeat the same option. Subjects were on average more sensitive to rewards (20p) than punishments (painful shock). ***c***, Decision weights parameterised as exponential kernels, which is equivalent to a truncated reinforcement learning model. Parameter estimates were TRP+ reward sensitivity = 4.18, punishment sensitivity = −2.29, choice repetition (perseveration) = 0.79, TRP− reward sensitivity = 2.50, punishment sensitivity = −2.43, choice repetition = 1.90. ***d***, ***e***, Correlation between change in serum tryptophan:large neutral amino acid ratio at 5 h post-amino acid ingestion and the parameters for reward sensitivity ($r = 0.50$, $p < 0.005$; ***d***) and choice perseveration ($r = -0.41$, $p < 0.01$; ***e***).

ized, placebo-controlled, and double-blind design. Of the 30 subjects who performed the task, 15 were randomized to the sham depletion (TRP+) group (and hence unimpaired brain serotonergic signaling) and 15 to the tryptophan-depleted (TRP−) group (with reduced brain serotonin signaling). Comparing these groups thus provides insight into the function of central serotonin (Carpenter et al., 1998).

Subjects learned to track rewards and avoid painful shocks during the task. Over all the trials, TRP+ subjects won money on 93.1 (2.9) and TRP− subjects on 98.8 (3.0) trials. TRP+ subjects received pain on 83.3 (1.9) and TRP− subjects on 81.3 (3.0) trials (SD in parentheses). TRP− subjects had a nonsignificant tendency to better overall performance, with a net win (total money minus pain) of 17.5 (3.7) compared with 9.8 (2.5) in TRP+ subjects ($p = 0.10$). Figure 3*a* describes the behavior in more detail, illustrating the probability of sticking with an option as function of different outcomes. It can be seen that TRP− subjects have a general tendency to be more perseverative (a tendency to repeat choices). There was no significant effect of preceding outcomes on response times on subsequent trials.

To examine the independent influence of rewards and punishments formally, we performed a conditional logit regression analysis (see Materials and Methods, above). High-positive decision weights indicate a strong tendency to repeat choosing an option on subsequent trials given a particular outcome. Negative weights indicate a tendency to switch options. We also included a choice weight, which simply looks at the inclination to repeat choices regardless of outcome, i.e., a basic perseverative tendency (choice stickiness) that is independent of rewards or punishments (Lau and Glimcher, 2005). Figure 3*b* shows the influence of these events over subsequent trials (see Materials and Methods, above): it can be seen that compared with TRP+ (control) subjects, TRP− (depleted) subjects show reduced decision weights for rewards, no clear difference in decision weights for punishments, and a greater tendency to perseverate.

To capture this effect more formally, we note that the regression indicates an approximately exponential decay of the impact of reward and punishment outcomes over time. Such an influence is predicted by reinforcement learning models of action learning (see Materials and Methods, above), which formalize how action values are learned from experience and subsequently govern choice. This can be emulated within a (constrained) conditional logit regression by parameterizing the decision weights and decay rates as exponentially decaying kernels (Fig. 3*c*). This analysis confirms that first, TRP− subjects were significantly less sensitive to rewards, parameterized as the effect of reward magnitude (TRP+, 4.18; TRP−, 2.50; $p < 0.01$). Second, TRP− subjects were significantly more perseverative over choices (TRP+, 0.79; TRP−, 1.90; $p < 0.05$). There was no effect on the punishment sensitivity (TRP+, −2.29; TRP−, −2.43; $p = 0.58$), hence yielding a significant valence (reward/punishment) × TRP (+/−) interaction ($p < 0.05$). There was no effect of the decay rate governing the forgetting (learning) rate for either outcomes or for perseveration (reward: TRP+, 0.5915; TRP−, 0.4200; $p = 0.12$; punishment: TRP+, 0.7600; TRP−, 0.6446; $p = 0.25$; perseveration: TRP+, 0.3540; TRP−, 0.0116; $p = 0.14$). The selective effect on reward over punishment sensitivity (value) effectively controlled the exchange rate by which rewards and punishments are integrated to a common currency for decisions. Given that the reward magnitude was 20 pence, this allowed us to equate the equivalent pain cost as 11.0 pence for each shock in the TRP+ group, and 19.4 pence for each shock in TRP− group.

Next, we regressed these parameters against the serum change in TRP:ΣLNAA ratio determined from blood sampling of subjects before and after tryptophan or placebo depletion. This

Table 3. BOLD results

Reward value	**$p < 0.001$ (uncorrected)**
Medial prefrontal cortex	**−4, 54, −8 ($z = 3.74$); FWE $p < 0.05$ SVC**
Bilateral head of caudate	14, 2, 16 ($z = 4.12$); −8, 0, 14 ($z = 4.05$)
Anterior cingulate cortex	2, 44, 12 ($z = 3.72$)
Right lateral orbitofrontal cortex	44, 54, −8 ($z = 4.35$)
Avoidance value	**$p < 0.001$ (uncorrected)**
Medial prefrontal cortex	**4, 36, −8 ($z = 3.40$); FWE $p < 0.05$ SVC**
Nucleus accumbens	−4, 4, −10 ($z = 4.56$)
Secondary somatosensory cortex/posterior insula	−52, 20, 12 ($z = 5.45$); −38, −12, 16 ($z = 5.16$)
Precuneus	−14, −60, 18 ($z = 4.97$)
Posterior parietal cortex	0, −26, 52 ($z = 4.66$)
Overlapping avoidance and reward value	
Ventromedial prefrontal cortex	**2, 38, −12; −6, 52, −8; −6, 28, −12**
Dorsomedial prefrontal cortex	2, 60, 24; 16, 58, 34; −18, 60, 22; −16, 66, 14
Left posterior insula	−56, −2, 0
Bilateral head of caudate	12, 2, 16; −8, 2, 16
Medial parietal cortex	2, −20, 44
Precuneus	−8, −46, 34
Reward × serotonin (ΔTRP:ΣLNAA)	**$p < 0.001$ (uncorrected)**
Ventromedial prefrontal cortex	**12, 44, −8 ($z = 3.52$); FWE $p < 0.05$ SVC**
Left anterior insula	−54, 16, 6 ($z = 3.84$)
Precuneus	16, −54, 28 ($z = 3.23$)
Reward prediction error	**$p < 0.001$ (uncorrected)**
Bilateral head of caudate	**12, 0, 14 ($z = 4.11$); −4, −4, 12 ($z = 4.07$); FWE $p < 0.05$ SVC**
Cerebellum	20, −52, −24 ($z = 5.10$)
Dorsomedial prefrontal cortex	12, 64, 26 ($z = 4.24$)
Anterior cingulate cortex	−4, 36, 28 ($z = 3.45$)
Choice kernel	**$p < 0.001$ (uncorrected)**
Nucleus accumbens	−4, 10, −8 ($z = 4.29$)
Dosomedial prefrontal cortex	−6, 60, 34 ($z = 4.32$)
Ventromedial prefrontal cortex	−12, 44, −4 ($z = 3.88$)
Choice kernal x -serotonin (ΔTRP:ΣLNAA)	**$p < 0.001$ (uncorrected)**
Right head of caudate	18, 4, 14 (z = 4.91); FWE $p < 0.05$ (whole brain)
Anterior cingulate cortex	−2, 42, 16 ($z = 3.32$)
Anterior pole	16, 66, 10 ($z = 3.37$)*
Avoidance prediction error	**$p < 0.001$ (uncorrected)**
Right head of caudate	**12, 2, 18 ($z = 4.08$); FWE $p < 0.05$ SVC**
Bilateral dorsolateral putamen	**30, 0, 8 ($z = 3.76$); −22, 0, 8 ($z = 3.53$); FWE $p < 0.05$ SVC**
Ventrolateral putamen	30, −8, −6 ($z = 4.63$)
Bilateral amygdaloid complex	24, 2, −12 ($z = 3.84$); −28, −4, −14 ($z = 3.54$)
Overlapping reward and avoidance prediction error	**$p < 0.01 \times p < 0.01$ (uncorrected)**
Right head of caudate	10, 12, 6
Right dorsolateral putamen	28, 0, 6
Reward prediction error × serotonin (ΔTRP:ΣLNAA)	**$p < 0.001$ (uncorrected); FWE $p < 0.05$ SVC**
Right dorsolateral putamen	**34, −4, 4 ($z = 4.22$)**

Results in bold are a priori ROIs (ventromedial PFC, choice values; dorsal striatum, instrumental prediction errors; see Materials and Methods). SVC, Small volume correction.

*Not SVC as distinct from previous anterior pole activity related to exploration.

yielded a subject-by-subject ΔTRP:ΣLNAA, which is a peripheral indicator of central serotonergic availability. Consistent with the above analysis, this revealed a significant correlation of ΔTRP:ΣLNAA with both reward sensitivity ($r = 0.50$, $p < 0.005$; Fig. 3*d*) and choice perseveration ($r = -0.41$, $p < 0.01$; Fig. 3*e*).

Next, we assessed hemodynamic responses correlated with choice using a model-based functional magnetic resonance imaging (fMRI) approach (see Experimental procedures; (O'Doherty et al., 2007)). First, we regressed the estimated reward, punishment and choice values (derived from the constrained regression analysis (as illustrated Fig. 3*c*)) at the time of cue presentation with BOLD responses. Consistent with previous reports, we observed reward-specific BOLD responses in regions of ventromedial medial prefrontal cortex, orbitofrontal cortex, caudate nucleus, and anterior cingulate cortex (see results tables, Experimental Procedures). We observed no significant positive response to aversive value, but significant negative responses were seen in regions that included medial prefrontal cortex (Table 3). This suggests a reward-like representation of avoidance states, consistent with avoidance learning theory (Dinsmoor, 2001) and previous fMRI data (Kim et al., 2006; Pessiglione et al., 2006; Plassmann et al., 2010). Choice perseveration (the choice kernel parameter) was associated with responses in nucleus accumbens and dorsal and medial prefrontal cortex (Table 3).

Since choice depends on integrating independent reward and avoidance values, we sought neural responses in which the representation of both overlapped. This revealed a common response profile in ventromedial prefrontal cortex (Fig. 4*a*, circled). We also noted activity in caudate nucleus and dorsomedial prefrontal cortex (see Materials and Methods, Results, and Table 3). This signal is consistent with a common decision value currency across money and pain.

Next, we compared these value responses between TRP+ and TRP− groups. We observed a positive correlation between re-

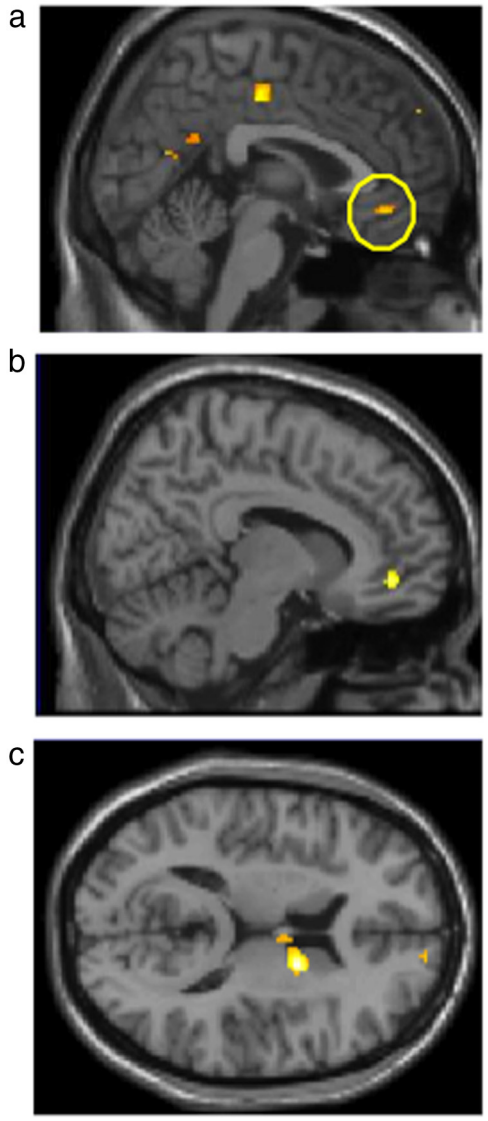

Figure 4. Neuroimaging results: choice values and choice perseveration. ***a***, BOLD responses corresponding to reward value and avoidance value (thresholded at $p < 0.01 \times p < 0.01$ for display) in ventromedial prefrontal cortex 2, 38, −12 (circled)— our region of a priori interest. Note that both appetitive and avoidance values individually satisfy correction for multiple comparisons in this region (FWE ROI $p < 0.05$, see Materials and Methods). ***b***, BOLD responses corresponding to covariation of reward value and change in serum tryptophan: long chain amino acid ratio following amino acid ingestion, threshold $p < 0.005$ for illustration (FWE ROI $p < 0.05$), showing response in ventromedial prefrontal cortex 12 44 −8 ($z = 3.52$). ***c***, BOLD responses corresponding to covariation of choice kernel and change in serum tryptophan: long chain amino acid ratio following amino acid ingestion, threshold $p < 0.005$ for illustration, showing responses in right head of caudate (18, 4, 14; FWE $p < 0.05$ whole brain) and right anterior pole (16, 66, 14).

ward value responses in ventromedial prefrontal cortex and TRP status. Figure 4*b* shows the correlation between reward-related brain responses and subject-specific serum ΔTRP:ΣLNAA. Thus, a neural response in vmPFC mirrored the behavioral results in showing a diminished representation of reward outcome in individuals whose serotonin was depleted.

No significant differences were observed for punishment avoidance value according to TRP status, as was also the case with the behavioral analysis. In the analysis of the regressor correlated with perseverative tendency, we observed a significant negative correlation in the head of caudate nucleus and a region in the anterior pole of the prefrontal cortex (Fig. 4*c*); that is, TRP+ subjects showed greater activity associated with perseveration, but an opposite behavioral tendency against perseveration. This region in anterior pole was near but distinct from a region identified by a classical analysis of exploration versus exploitation responses (Daw et al., 2006), which identified an uncorrected peak more posteriorly (MNI coordinates: −14, 56, 14; $z = 3.53$). Thus, perseverative activity does not appear to be coded as a component of value, but as a serotonin-sensitive ability to inhibit perseveration, manifest indirectly by greater caudate activity.

As mentioned above, the logit regression model (with exponentially decaying kernels) is equivalent to a reinforcement learning model (in which nonchosen options are treated as if they were chosen and yielded no outcome) with an additional choice kernel. By modeling this explicitly, we could derive the prediction errors relating to action updating. We found that reward prediction errors correlated with responses in striatal regions, as observed in numerous prior studies (Table 3) (O'Doherty et al., 2004). An aversive prediction error was negatively correlated with hemodynamic responses in regions of ventral and dorsal striatum. As was the case for outcome value, the negative correlation with the aversive prediction error implies a positive correlation with the same signal inverted: that is, oriented like a reward prediction error with omitted pain corresponding to increased responses and unexpected pain associated with decreased responses, i.e., an avoidance error signal. We then sought overlapping responses that correlated with both avoidance error and appetitive prediction error, since such a signal reflects a common neural error signal. This revealed specific striatal responses in the right head of caudate nucleus and right dorsolateral putamen (Fig. 5*a*).

Finally, we identified reward prediction error responses that correlated with TRP status, given a behavioral and neural modulation of reward value. This analysis identified a modulation of activity restricted to right dorsolateral putamen (but not right head of caudate). Figure 5*b* illustrates the peak correlation with subject-specific serum ΔTRP:ΣLNAA.

Discussion

In summary, our data provide converging behavioral and neural evidence that serotonin modulates (is necessary for) distinct behavioral and anatomical components of decision-making. Most surprising is our observation of a strongly positive dependence of reward outcome value on serotonin signaling, with corresponding cue-value-related activity in vmPFC and prediction-error-related activity in dorsolateral putamen (for errors). This value-dependent effect was behaviorally and anatomically distinct from an effect of serotonin on behavioral flexibility, as indicated by choice perseveration.

Previous behavioral results in humans have hinted at an effect of serotonin on reward learning (Rogers et al., 1999, 2003; Finger et al., 2007; Schweighofer et al., 2008), but the interpretation of these studies has been hampered by methodological issues concerning discriminating possibly distinct effects on behavioral flexibility and reward omission learning (which could involve aversive learning processes).

However, there is good evidence from nonhuman primates and rodents for a possibly facilitatory influence of serotonin on reward. First, a recent study in marmosets found that serotonergic lesions of ventromedial prefrontal cortex impair conditioned reinforcement and not extinction (Walker et al., 2009), arguing against an aversive (i.e., conveying omitted reward) mechanism

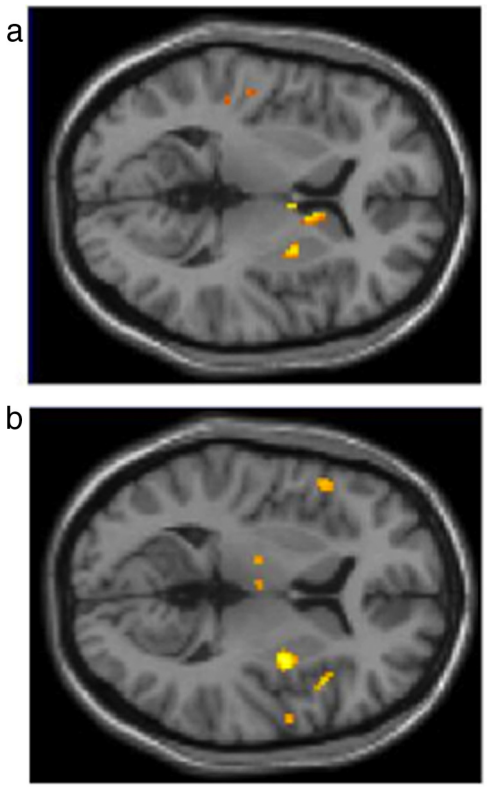

Figure 5. Neuroimaging results: prediction error. ***a***, Overlapping reward and punishment avoidance prediction errors (exclusively masked, $p < 0.01 \times p < 0.01$ for illustration) showing activity in right medial head of caudate (10, 12, 6) and right dorsolateral putamen (28, 0, 6). The appetitive and avoidance prediction error responses are individually significant at FWE ROI $p < 0.05$ ***b***, Reward prediction errors modulated by serum ΔTRP:ΣLNAA (covariate regression, $p < 0.001$ uncorrected for display; 34, −4, 4; FWE ROI $p < 0.05$).

underlying the reward deficits seen in reversal learning in previous experiments (Clarke et al., 2004).

Second, microdialysis from rat medial prefrontal cortex in instrumental reward tasks shows significant serotonin efflux compared with yoked conditions. Medial prefrontal cortex has dense 5-HT2a receptors (Martín-Ruiz et al., 2001), and pyramidal neurons in medial PFC and orbitofrontal cortex (in the rat) reciprocally and simultaneously project to dorsal raphé and ventral tegmental area, sources of the brain's serotonin and dopamine neurons, respectively (Vázquez-Borsetti et al., 2009, 2011). Stimulation of cortical 5-HT-2a receptors increases the release of dopamine in the mesocortical dopaminergic system (Pehek et al., 2006). Other serotonin receptor subtypes (including 5-HT1A, 5-HT1B, 5-HT3, and 5-HT4) have also been shown to have facilitatory effects on dopamine transmission, in contrast to the well documented inhibitory influence of 5HT2c and other receptor subtypes (Di Matteo et al., 2008). Evidence also exists that the rewarding effects of cocaine, as indicated in a two-choice discrimination learning procedure, is potentiated through coinvolvement of serotonergic mechanisms (Cunningham and Callahan, 1991; Kleven and Koek, 1998).

Perhaps the most notable recent data suggesting serotonin might play a role in coding reward value comes from electrophysiological recordings from macaque dorsal raphé neurons during a reward-based instrumental saccade task (Bromberg-Martin et al., 2010). Here, a majority of raphé neurons displayed a pattern of activity consistent with coding reward value toward reward cues and outcomes. In rats, recordings from dorsal raphé neurons have also been shown to correlate with reward in a delayed-reward task (Miyazaki et al., 2011). However, in both tasks, it is difficult to definitively conclude that the neurons identified were serotonergic, since their electrophysiological signatures are known to be diverse (Kocsis et al., 2006), and it remains possible that the activity represents nonserotonergic neurons.

We found that reduced central serotonin levels were associated both with more persistent responding and with greater persistence-related hemodynamic responses in the medial head of the caudate nucleus, suggesting that this tonic signal may modulate effective value outside the caudate. The modulation of striatal activity may be a downstream effect of serotoninergic manipulation, since selective serotoninergic caudate lesions do not disrupt reversal learning (in marmosets), in contrast to dopaminergic lesions (Clarke et al., 2011) and opposite to the effect in orbitofrontal cortex (Clarke et al., 2004).

Choice persistence could arise from multiple causes. One possibility is a modulation in the representation of average reward (an aspiration level), a signal that provides an estimate as to how good or bad an agent expects the environment to be (Daw et al., 2002). Accordingly, if average reward prediction is high, then the outcome of current options are likely to be judged less attractive than if the average reward prediction is low. In such a scenario, the tendency to switch actions and explore elsewhere in search of higher rewards will be stronger. Alternatively, if the average reward signal is low, then current options will seem marginally better, leading to a tendency to persist. In this way, perseveration may arise as a direct consequence of comparing immediate versus long-term predictions. The idea that serotonin might reflect long-term reward prediction can be seen as parallel to psychological observations of serotonin's well characterized involvement in antidepressant drug action (Harmer et al., 2009) and provides a possible mechanistic link to the distinct effect of reward value coding.

However, there are other factors that may also contribute to choice persistence, though these have not previously been linked to serotonergic function. For instance, it could result from a simple (Mackintosh, 1983) or sophisticated (Peters and Schaal, 2008) form of stimulus-response learning, in which previously taken choices are "stamped-in". Alternatively, it might be viewed as a process that encourages oversampling of information, a mechanism advantageous in highly variable environments in which reinforcement learning has a tendency to be oversensitive to the immediate past, leading to risk aversion (Denrell and March, 2001). Indeed, the control of one particular aspect of flexible behavior, namely the balance of exploration and exploitation (Daw et al., 2006), has previously been linked to frontal pole activity and it is interesting to note its involvement here too, although the precise locus of activity is somewhat distinct from that identified previously.

Our data help refine our understanding of the role played by the striatum in motivation. Previous Pavlovian punishment studies (in which punishments are delivered regardless of any action) have shown an aversive prediction error signal, oriented positively (opposite to that seen in the present study) in ventral and dorsal striatum (Jensen et al., 2003; Seymour et al., 2004, 2007), suggesting a site of convergence with a (putatively dopaminergic) reward prediction error. However, in the present study, the aversive signal becomes reward-signed. We suggest that the key difference between studies is the availability of choices in the present design. If so, this would be consistent with two-factor theories of instrumental avoidance, in which avoidance is mediated by the reward of attaining a safety state that

signals successful avoidance (Mowrer, 1960; Dinsmoor, 2001). It is possible that in passive studies on aversion, punishments are framed as punishments by an aversive system, but when control is possible through active choice, punishments are framed appetitively as missed opportunities to avoid aversive outcomes (Delgado et al., 2009). In fact, this is consistent with previous demonstrations of reference sensitivity in striatal activity, where the contextual valence is apparently set by predictive cues (Seymour et al., 2005).

Critically, by forcing independent representation of reward and avoidance, our data suggest that avoidance prediction, carried as an opponent reward-predictive signal, coactivates the same region of striatum (dorsolateral putamen and medial head of caudate) and ventrodmedial prefrontal cortex that signals predictions and values of standard rewards. This demonstrates a central role for these regions in integrating distinct motivational pathways. Whereas this appetitive-aversive integration (algorithmically, the addition of appropriately scaled excitatory and inhibitory values) (Dickinson and Dearing, 1979) is commonplace in everyday decision tasks, to our knowledge, this is the most direct experimental demonstration of its neuroanatomical substrate.

References

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

Bond A, Lader M (1974) The use of analogue scales in rating subjective feelings. Br J Med Psychol 47:211–218.

Boureau YL, Dayan P (2011) Opponency revisited: competition and cooperation between dopamine and serotonin. Neuropsychopharmacology 36:74–97.

Bromberg-Martin ES, Hikosaka O, Nakamura K (2010) Coding of task reward value in the dorsal raphe nucleus. J Neurosci 30:6262–6272.

Carpenter LL, Anderson GM, Pelton GH, Gudin JA, Kirwin PD, Price LH, Heninger GR, McDougle CJ (1998) Tryptophan depletion during continuous CSF sampling in healthy human subjects. Neuropsychopharmacology 19:26–35.

Clarke HF, Dalley JW, Crofts HS, Robbins TW, Roberts AC (2004) Cognitive inflexibility after prefrontal serotonin depletion. Science 304:878–880.

Clarke HF, Hill GJ, Robbins TW, Roberts AC (2011) Dopamine, but not serotonin, regulates reversal learning in the marmoset caudate nucleus. J Neurosci 31:4290–4297.

Cools R, Roberts AC, Robbins TW (2008a) Serotoninergic regulation of emotional and behavioural control processes. Trends Cogn Sci 12:31–40.

Cools R, Robinson OJ, Sahakian B (2008b) Acute tryptophan depletion in healthy volunteers enhances punishment prediction but does not affect reward prediction. Neuropsychopharmacology 33:2291–2299.

Cools R, Nakamura K, Daw ND (2011) Serotonin and dopamine: unifying affective, activational, and decision functions. Neuropsychopharmacology 36:98–113.

Crockett MJ, Clark L, Robbins TW (2009) Reconciling the role of serotonin in behavioral inhibition and aversion: acute tryptophan depletion abolishes punishment-induced inhibition in humans. J Neurosci 29:11993–11999.

Crockett MJ, Clark L, Roiser JP, Robinson OJ, Cools R, Chase HW, Ouden H, Apergis-Schoute A, Campbell-Meikeljohn D, Seymour B, Sahakian BJ, Rogers RD, Robbins TW (2012) Converging evidence for central 5-HT effects in acute tryptophan depletion. Mol Psychiatry 17:121–123.

Cunningham KA, Callahan PM (1991) Monoamine reuptake inhibitors enhance the discriminative state induced by cocaine in the rat. Psychopharmacology 104:177–180.

Daw ND, Kakade S, Dayan P (2002) Opponent interactions between serotonin and dopamine. Neural Netw 15:603–616.

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

Deakin JF, Graeff FG (1991) 5-HT and mechanisms of defence. J Psychopharmacol 5:305–315.

Delgado MR, Jou RL, Ledoux JE, Phelps EA (2009) Avoiding negative outcomes: tracking the mechanisms of avoidance learning in humans during fear conditioning. Front Behav Neurosci 3:33.

Denrell J, March JG (2001) Adaptation as information restriction: the hot stove effect. Organiz Sci 12:523–538.

Dickinson A, Dearing MF (1979) Appetitive-aversive interactions and inhibitory processes. In: Mechanisms of learning and motivation: a memorial volume to Jerzy Konorski (Dickinson A, Boakes RA, eds.), pp. 203–231. London: Psychology Press.

Di Matteo V, Cacchio M, Di Giulio C, Esposito E (2002) Role of serotonin2C receptors in the control of brain dopaminergic function. Pharmacol Biochem Behav 71:727–734.

Di Matteo V, Di Giovanni G, Pierucci M, Esposito E (2008) Serotonin control of central dopaminergic function: focus on in vivo microdialysis studies. In: Serotonin-dopamine interaction: experimental evidence and therapeutic relevance (Di Giovanni G, Di Matteo V, Esposito E, eds.), pp 7–44. Amsterdam: Elsevier.

Dinsmoor JA (2001) Stimuli inevitably generated by behavior that avoids electric shock are inherently reinforcing. J Exp Anal Behav 75:311–333.

Doya K (2002) Metalearning and neuromodulation. Neural Netw 15:495–506.

Evers EA, Cools R, Clark L, van der Veen FM, Jolles J, Sahakian BJ, Robbins TW (2005) Serotonergic modulation of prefrontal cortex during negative feedback in probabilistic reversal learning. Neuropsychopharmacology 30:1138–1147.

Fernstrom JD, Wurtman RJ (1972) Brain serotonin content: physiological regulation by plasma neutral amino acids. Science 178:414–416.

Finger EC, Marsh AA, Buzas B, Kamel N, Rhodes R, Vythilingham M, Pine DS, Goldman D, Blair JR (2007) The impact of tryptophan depletion and 5-HTTLPR genotype on passive avoidance and response reversal instrumental learning tasks. Neuropsychopharmacology 32:206–215.

Fürst P, Pollack L, Graser TA, Godel H, Stehle P (1990) Appraisal of four pre-column derivatization methods for the high-performance liquid chromatographic determination of free amino acids in biological materials. J Chromatogr 499:557–569.

Harmer CJ, Goodwin GM, Cowen PJ (2009) Why do antidepressants take so long to work? A cognitive neuropsychological model of antidepressant drug action. Br J Psychiatry 195:102–108.

Jensen J, McIntosh AR, Crawley AP, Mikulis DJ, Remington G, Kapur S (2003) Direct activation of the ventral striatum in anticipation of aversive stimuli. Neuron 40:1251–1257.

Kim H, Shimojo S, O'Doherty JP (2006) Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. PLoS Biol 4:e233.

Kleven MS, Koek W (1998) Discriminative stimulus properties of cocaine: enhancement by monoamine reuptake blockers. J Pharmacol Exp Ther 284:1015–1025.

Kocsis B, Varga V, Dahan L, Sik A (2006) Serotonergic neuron diversity: identification of raphe neurons with discharges time-locked to the hippocampal theta rhythm. Proc Natl Acad Sci U S A 103:1059–1064.

Lau B, Glimcher PW (2005) dynamic response-by-response models of matching behavior in rhesus monkeys. J Exp Anal Behav 84:555–579.

Mackintosh NJ (1983) Conditioning and associative learning. London: Clarendon.

Martín-Ruiz R, Puig MV, Celada P, Shapiro DA, Roth BL, Mengod G, Artigas F (2001) Control of serotonergic function in medial prefrontal cortex by serotonin-2A receptors through a glutamate-dependent mechanism. J Neurosci 21:9856–9866.

Miyazaki K, Miyazaki KW, Doya K (2011) Activation of dorsal raphe serotonin neurons underlies waiting for delayed rewards. J Neurosci 31:469–479.

Mowrer OH (1960) Learning theory and behavior. New York: Wiley.

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, 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.

Pehek EA, Nocjar C, Roth BL, Byrd TA, Mabrouk OS (2006) Evidence for the preferential involvement of 5-HT2A serotonin receptors in stress- and drug-induced dopamine release in the rat medial prefrontal cortex. Neuropsychopharmacology 31:265–277.

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, Schaal S (2008) Natural actor-critic. Neurocomputing 71: 1180–1190.
Plassmann H, O'Doherty JP, Rangel A (2010) Appetitive and aversive goal values are encoded in the medial orbitofrontal cortex at the time of decision making. J Neurosci 30:10799–10808.
Robbins TW, Crockett MJ (2009) Role of serotonin in impulsivity and compulsivity: comparative studies in experimental animals and humans. In: Handbook of the behavioral neurobiology of serotonin, vol. 21 (Muller CP, Jacobs B, eds.), pp 415–428. London: Elsevier.
Rogers RD (2011) The roles of dopamine and serotonin in decision making: evidence from pharmacological experiments in humans. Neuropsychopharmacology 36:114–132.
Rogers RD, Blackshaw AJ, Middleton HC, Matthews K, Hawtin K, Crowley C, Hopwood A, Wallace C, Deakin JF, Sahakian BJ, Robbins TW (1999) Tryptophan depletion impairs stimulus–reward learning while methylphenidate disrupts attentional control in healthy young adults: implications for the monoaminergic basis of impulsive behaviour. Psychopharmacology 146:482–491.
Rogers RD, Tunbridge EM, Bhagwagar Z, Drevets WC, Sahakian BJ, Carter CS (2003) Tryptophan depletion alters the decision-making of healthy volunteers through altered processing of reward cues. Neuropsychopharmacology 28:153–162.
Roiser JP, Blackwell AD, Cools R, Clark L, Rubinsztein DC, Robbins TW, Sahakian BJ (2006) Serotonin transporter polymorphism mediates vulnerability to loss of incentive motivation following acute tryptophan depletion. Neuropsychopharmacology 31:2264–2272.
Schonberg T, O'Doherty JP, Joel D, Inzelberg R, Segev Y, Daw ND (2010) Selective impairment of prediction error signaling in human dorsolateral but not ventral straitum in Parkinson's disease: evidence from a model-based fMRI study. Neuroimage 49:772–781.
Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275:1593–1599.
Schweighofer N, Bertin M, Shishida K, Okamoto Y, Tanaka SC, Yamawaki S, Doya K (2008) Low-serotonin levels increase delayed reward discounting in humans. J Neurosci 28:4528–4532.
Seymour B, O'Doherty JP, Dayan P, Koltzenburg M, Jones AK, Dolan RJ, Friston KJ, Frackowiak RS (2004) Temporal difference models describe higher-order learning in humans. Nature 429:664–667.
Seymour B, O'Doherty JP, Koltzenburg M, Wiech K, Frackowiak R, Friston K, Dolan R (2005) Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. Nat Neurosci 8:1234–1240.
Seymour B, Daw N, Dayan P, Singer T, Dolan R (2007) Differential encoding of losses and gains in the human striatum. J Neurosci 27:4826–4831.
Soubrié P (1986) Reconciling the role of central serotonin neurons in human and animal behavior. Behav Brain Sci 9:319–335.
Tanaka SC, Shishida K, Schweighofer N, Okamoto Y, Yamawaki S, Doya K (2009) Serotonin affects association of aversive outcomes to past actions. J Neurosci 29:15669–15674.
van Donkelaar EL, Blokland A, Ferrington L, Kelly PA, Steinbusch HW, Prickaerts J (2011) Mechanism of acute tryptophan depletion: is it only serotonin? Mol Psychiatry 16:695–713.
Vázquez-Borsetti P, Cortés R, Artigas F (2009) Pyramidal neurons in rat prefrontal cortex projecting to ventral tegmental area and dorsal raphe nucleus express 5-HT2A receptors. Cereb Cortex 19:1678–1686.
Vázquez-Borsetti P, Celada P, Cortés R, Artigas F (2011) Simultaneous projections from prefrontal cortex to dopaminergic and serotonergic nuclei. Int J Neuropsychopharmacol 14:289–302.
Walker SC, Robbins TW, Roberts AC (2009) Differential contributions of dopamine and serotonin to orbitofrontal cortex function in the marmoset. Cereb Cortex 19:889–898.