Surprise Beyond Prediction Error

Justin R. Chumbley,[1]* Christopher J. Burke,[1] Klaas E. Stephan,[2] Karl J. Friston,[3] Philippe N. Tobler,[1] and Ernst Fehr[1]

[1]*Laboratory for Social and Neural Systems Research, University of Zurich, Switzerland*
[2]*Translational Neuromodeling Unit, ETH, Zurich, Switzerland*
[3]*Wellcome Trust Centre for Neuroimaging, UCL, London*

Abstract: Surprise drives learning. Various neural "prediction error" signals are believed to underpin surprise-based reinforcement learning. Here, we report a surprise signal that reflects reinforcement learning but is neither un/signed reward prediction error (RPE) nor un/signed state prediction error (SPE). To exclude these alternatives, we measured surprise responses in the absence of RPE and accounted for a host of potential SPE confounds. This new surprise signal was evident in ventral striatum, primary sensory cortex, frontal poles, and amygdala. We interpret these findings via a normative model of surprise. *Hum Brain Mapp 35:4805–4814, 2014.* © 2014 The Authors. Human Brain Mapping Published by Wiley Periodicals, Inc.

Key words: reward; prediction error; learning

INTRODUCTION

The concept of prediction error has taken center stage in many theories of learning, most notably in reinforcement learning. In "model-free" reinforcement learning, reward prediction errors (RPEs) learn the value of being in some context or state (Balleine et al., 2008; Dayan and Niv, 2008; McClure et al., 2003; O'Doherty et al., 2004). In "model-based" reinforcement learning, state prediction errors (SPEs) can learn an internal model of probable consequences of being in some state—i.e., they learn state transition probabilities. In both cases, PEs capture how "surprising" a reward/state is and how to adjust expectations accordingly (see Information Box). PE theories are appealing because of their conceptual simplicity: they simply learn from unexpected events. Yet it is unclear whether all surprise is reducible to some un/signed PE. This is because most experiments confound different forms of surprise: events far from the average, "expected" value are also improbable. We therefore looked for evidence of improbability-based surprise not reducible to RPE or SPE. We specifically asked whether two identical rewards, with identical RPE, could evoke different brain responses based on their relative probability, while taking care to exclude SPE mechanisms.

In our paradigm, some cues predicted bimodal rewards (one or three coins arose frequently, while two coins were rare; see Fig. 1). Thus subjects' seldom observed the average number of coins and rarely received the average monetary payment (the "expected reward"). Instead they usually received the extreme payments of one and three coins. Because RPEs reflect the difference between observed and expected reward, and the average reward corresponds to the expectation, RPEs are zero when participants observe two coins (i.e., $\delta=0$, see Information Box). However, surprise should be highest for these very same trials where RPEs are zero on average. Other cues predicted unimodal rewards: rewards for which this expected value—two coins—was frequent and unsurprising (see Fig. 1). According to model-free reinforcement learning, no

Additional Supporting Information may be found in the online version of this article.

Contract grant sponsor: CoMPLEX, Switzerland, by the Swiss National Science Foundation; Contract grant number: PP00P1_128574 and CRSII3_141965; Contract grant sponsor: Swiss National Centre of Competence in Research in Affective Sciences

*Correspondence to: Justin R. Chumbley, Blümlisalpstrasse 10, Zurich 8006, Switzerland. E-mail: justin.chumbley@econ.uzh.ch

Received for publication 8 November 2013; Revised 4 February 2014; Accepted 18 March 2014.

DOI 10.1002/hbm.22513
Published online 3 April 2014 in Wiley Online Library (wileyonlinelibrary.com).

learning takes place in the absence of a RPE; i.e., these theories provide no mechanism whereby the subject can learn that the two-coin outcome is surprising in one case but not the other.

In contrast, a “model-based system,” which encodes how likely each possible outcome is, may exploit SPEs to learn this discrimination (see Information Box). For this reason, differential brain responses to the two-coin outcome must reflect the model-based system, being either un/signed SPE or surprise *per se*, i.e., conditional improbability. We seek to identify surprise *per se*, by contrasting hemodynamic responses to the improbable versus probable “expected reward”, i.e., the two-coin outcome under bimodal versus unimodal distributions, while including SPE covariates in our statistical analysis.

Information Box: Model-Free RPEs and Model-Based SPEs

Surprise as captured by prediction error (PE or δ) has played an essential role in the interpretation of data from single cell recording and from neuroimaging studies (Friston, 2009; Glimcher, 2011; Rescorla and Wagner, 1972; Schultz and Dickinson, 2000; Schultz et al., 1997; Sutton and Barto, 1998): PE is defined as the difference between observed and expected quantities. A scalar RPE features in theories of “model-free” reinforcement learning and permits subjects to calibrate their reward expectations (Rescorla and Wagner, 1972; Schultz et al., 1997; Sutton and Barto, 1998). Following cue i, the RPE δ_{R_i} simply codes the difference between received and expected reward, $\delta_{R_i}=R_i-\eta_{R_i}$. This RPE is signed, meaning that more reward than expected, corresponding to positive RPE, has a different meaning from (i.e., is “better than”) less reward than expected, which corresponds to a negative prediction error. During learning, the expected reward η_{R_i} may be updated on each trial according to $\eta_{R_i} \leftarrow \eta_{R_i}+\alpha\delta_{R_i}$, where α is a learning rate parameter. One could argue though that the amount of surprise should not depend on the sign of the RPE. This notion can be captured with unsigned RPEs which are simply the absolute value of δ_{R_i}, denoted $|\delta_{R_i}|$. Unsigned RPE can be used to guide attention.

While RPEs learn the expected value of each cue i, SPEs learn the probability of each specific outcome (see Ludvig et al., 2012; Sutton and Barto, 1990). Assuming that one of J discrete outcome states may follow cue i, a model-based system may express J signed SPEs, each denoted $\delta_{S_{ij}}$, and J unsigned SPEs, denoted $|\delta_{S_{ij}}|$, in response to the attained outcome. Each SPE has the form $\delta_{S_{ij}}=s_{ij}-\eta_{S_{ij}}$, where S_{ij} indicates a binary transition (1 for yes/ 0 for no) from cue i to outcome j and $\eta_{S_{ij}}$ is the expected probability of this transition. The expected state transition probabilities $\eta_{S_{ij}}$ may then each be updated according to $\eta_{S_{ij}} \leftarrow \eta_{S_{ij}}+\alpha\delta_{S_{ij}}$.

In applying these definitions to our task (see Fig. 1), we assume that the reward R on each trial—which drives model-free RPE learning—is simply equal to the magnitude of financial payoff, i.e., 1, 2, or 3 Swiss francs (CHF, see Fig. 1). Regarding model-based SPE learning, note that there are nine transition probabilities in total in our task: three outcomes j for each possible cue i (see Fig. 1). In our task, $j \in \{CHF\ 1, CHF\ 2, CHF\ 3\ \}$, $i \in \{cue\ 1, cue\ 2, cue\ 3\}$. To take a concrete example, imagine a trial in which three coins followed cue 1, then $S_{13}=1$ while $S_{11}=0$ and $S_{12}=0$. The model-based system then expresses three signed SPEs $\delta_{S_{1j}}$ and three unsigned SPEs $|\delta_{S_{1j}}|$ in response to the outcome. The expected state transition probabilities $\eta_{S_{1j}}$ may then each be updated according to $\eta_{S_{1j}} \leftarrow \eta_{S_{1j}}+\alpha\delta_{S_{1j}}$. Because this model-based system may learn that two coins are likely to follow cue 2 but not cue 1 or cue 3, it can learn discriminations that the model-based system cannot (see Introduction and Fig. 1).

Both the models considered above learn about the rewards/states and express some form of mismatch between prediction and observation. While un/signed PE expresses the (un/signed) arithmetic difference between some expectation and observation (Dayan et al., 2000; Friston et al., 2006; Pearce and Hall, 1980; Roesch et al., 2012), the present study looks for signals which code the conditional surprise or improbability of an event but are not reducible to PE (MacKay, 2003).

METHODS AND MATERIALS

Participants

All subjects had normal or corrected-to-normal vision and were screened to exclude those with a previous history of neurological or psychiatric disease. All gave informed consent and the study was approved by the Ethics Committee of the Canton of Zurich. After completing a consent form and MR safety questionnaire, participants were invited to read the task instructions.

Procedure and Rationale

Naive subjects viewed visual stimuli presented against a black background on a computer monitor while in an fMRI scanner. On each trial one of three visual cues (fractals) was presented at random on the left or the right of the screen. After 2 ± 1 s this cue was replaced by coin(s) in the center of the screen indicating a monetary reward of 1, 2, or 3 Swiss francs (CHF). Subjects stood to win the amount indicated if they correctly reported the side of the cue with a button-press. This task served only to maintain attention and was designed to be easy: on any one trial, the predictive cue was perceptibly either on the left or the right, 5 cm from the midline. In line with this, subjects performed this incidental laterality judgment task at ceiling, for all cues and reward levels. The reward following each cue was sampled

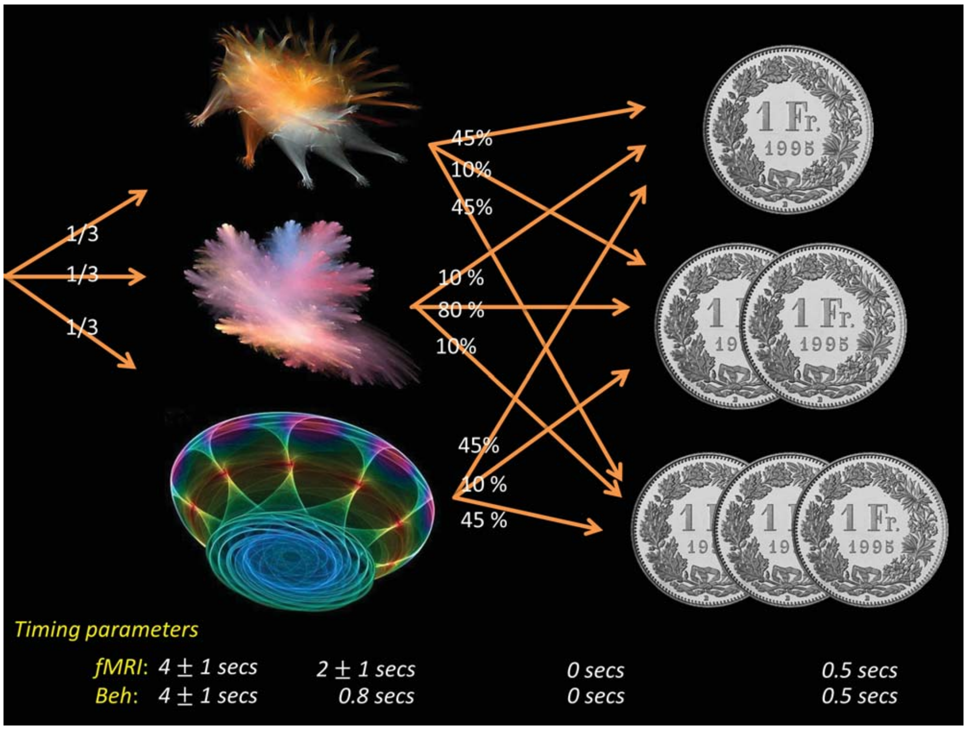

Figure 1.
The trial structure. With 1/3 probability one of three cues is randomly presented. Cues were presented for 0.8 s (behavior) or 1–3 s (fMRI), immediately followed by 1, 2, or 3 monetary units, presented with the indicated conditional probabilities. All cues/outcomes are presented the same number of times, the only predictable structure being in different probabilistic associations between each cue and the reward. Timings for the fMRI and pure behavioral studies are given above.

randomly from a cue-specific probability distribution—which was unknown to the subjects (see Fig. 1). Each cue yielded two CHF on average but with a different probability distribution over monetary rewards. In this context, conventional model-free RPE learning algorithms (Rescorla and Wagner, 1972) can only learn the average or "expected" reward that is constant over cues, while more recent theories permit subjects to discriminate based on reward variance, risk or precision (Preuschoff and Bossaerts, 2007; Schultz et al., 2008). In theory, the model-based system may exploit SPEs to discriminate cue-specific outcome probabilities even when RPEs cannot help them e.g. when RPE is zero.

We wanted to ensure that all surprise responses in this task were cue-specific; i.e., that they reflected discrimination learning and not some other improbable feature of the outcome. We therefore arranged that, over all trials, cues, and rewards were presented with the same (marginal) frequency (Fig. 1). This ensured that novelty/familiarity of cues and outcomes were controlled because subjects saw each cue and reward the same number of times throughout the experiment. This is important because novelty may also elicit responses in the midbrain dopaminergic system implicated in RPE-processing (Ljungberg et al., 1992). Recency effects were constant because the presentation *rate* of each cue or reward was the same. Thus if subjects failed to discriminate cue-specific reward distributions, they would be equally surprised by all rewards.

To further control for PE, our regression included trial-specific unsigned and signed PEs as covariates, see (Pearce and Hall, 1980; Roesch et al., 2012). To assess behavioral evidence for learning, we asked subjects to report the probabilistic contingencies explicitly after the fMRI session: this probed their declarative "model" following learning. We also conducted a separate behavioral study with the identical design—except that different subjects were required to report the magnitude of rewards at the end of each trial. The purpose of this study was to provide additional behavioral evidence for the relevance of surprise. In particular we asked if response times increased on conditionally surprising trials.

Behavioral Study 1

In behavioral study 1, we studied sixteen healthy male volunteers (age range: 20–25 years). The purpose of this study was to establish the behavioral relevance of surprise. Subjects observed cue (fractal)—reward (coins) associations on the computer screen and reported the number of coins via key press, as quickly and accurately as possible (timing parameters given in Fig. 1). If subjects correctly reported the number of coins within a 500 ms time-window, they stood the possibility to win the equivalent money (a subset of **10** attempts were randomly selected and paid at the end). By experimental design, cues preceded the financial reward, CHF 1, 2, or 3 (see Fig. 1). There were three sessions separated by a 3 min break. All rewards were independent samples from the conditional distributions shown in Figure 1. In each trial of sessions two and three, the cue was drawn randomly with probability 1/3. Session 1 cues were presented in sequence i.e., 40 presentations of cue 1, then 40 of cue 2 then 40 of cue 3. We used all three sessions for the behavioral analysis. The actual frequencies presented to subjects were forced to be the same as those illustrated in Figure 1: we achieved this by drawing the outcome on each trial without replacement from an "Urn" containing 40 outcomes arranged in the proportions given in Figure 1, i.e., [18/40, 4/40, 18/40] and [4/40, 32/40, 4/40]. While this technically introduces a little dependence in trial-by-trial realizations—draws are not identically and independently distributed – it ensures consistent surprise responses between subjects with a relatively small number of trials.

fMRI

Using fMRI we studied 19 different male participants (age range: 20–25 years), presenting exactly the same cue-reward contingencies as above—but asked subjects to report the laterality (left/right) of the cue on each trial. This incidental behavioral task was the same for all cues and therefore independent of the cue-specific reward associations of interest. This meant that reaction time and response inhibition are not confounded with subjective surprise (as it was in the preceding, strictly behavioral, task). Task instructions introduced subjects to the visual cues (fractals) and outcomes (1, 2, or 3 coins) and informed subjects that each cue would be followed by 1, 2, or 3 coins that were "available to win" (1 coin = CHF 1). On each presentation of a cue, subjects were asked to report the position of the (fractal) cue on the screen by left/right button-press. They were told that success in this task determined their final monetary payoff. Specifically, a random subset of 10 trials per block would be selected after the experiment for payment: If subjects had successfully reported the cue-location within time, the corresponding money would be paid out. Subjects were told that they could not predict which cue would appear on any trial but that there "may be a relationship between the cue and the number of coins available." Participants' earnings were calculated for each session of the experiment.

Task and contingencies

Each trial started with a variable ITI with only a fixation cross visible in the center of the screen. The ITI length was sampled uniformly from the interval 4–6 seconds. The ITI was followed by the presentation of one out of three visual (fractal) cues, randomly on the left or right of the screen, for 1–3 seconds. At the offset of this cue, 1, 2, or 3 coins were presented, indicating money available to win. Following the presentation of coins, participants were shown the fixation cross again. There were three sessions separated by a break. All rewards were independent samples from the conditional distributions shown in Figure 1. In each trial of sessions two and three, the cue was drawn randomly with probability 1/3. In session 1, cues were presented in sequence i.e., forty presentations of cue1-reward, then forty of cue 2-reward then forty of cue 3-reward. The cue-outcome assignments, as well as the order of blocks in session 1, were counterbalanced across subjects. To preclude brain responses based on novelty, familiarity or recency effects, we excluded session one from the fMRI analysis. The fMRI results below therefore report on sessions two and three. Each session was 10-min long with two 3-min breaks in between.

Behavior 2

After scanning, we elicited subjects' belief about the relative frequency of 1, 2, or 3 coins associated with each of the three cues, which probed their declarative knowledge of the probabilistic contingencies (bimodal versus unimodal). To elicit self-reported beliefs about the relative frequency of each outcome, subjects were given three sheets of paper, one for each cue. At the top of each page was a picture of the cue: along the bottom of the page were pictures of 1, 2, and 3 coins (the same pictures that reported outcomes during the task itself). Above each coin(s) was an empty space. For each coin outcome, subjects used a pencil to report, "the percentage of times this number of coins followed this cue." A histogram was deemed "bimodal" if and only if the probability assigned to outcome 2 was lower than the probability assigned to both outcome 1 and outcome 3. Otherwise the histogram was deemed "unimodal."

fMRI Data Acquisition

Images were acquired using a Philips Achieva 3T whole-body scanner with an eight channel SENSE head coil (Philips Medical Systems, Best, The Netherlands) at the Laboratory for Social and Neural Systems Research (SNS Lab), Zurich. Subjects viewed the stimuli through a mirror fitted on top of the head coil. We acquired gradient echo T2*-weighted echo-planar images (EPIs) with blood-oxygen-level–dependent (BOLD) contrast (slices/volume, 37; repetition time, 2.47 s). Approximately 350 volumes were collected in each session of the experiment. Scan onset times varied randomly relative to stimulus onset times. Volumes were acquired at a +15° tilt to the anterior commissure-posterior commissure line, rostral > caudal. Imaging parameters were the following: echo time, 30 ms;

field of view, 220 mm. The spatial resolution of the functional data was 3 × 3 × 3 mm. A T1-weighted 3D-TFE high-resolution structural image was also acquired for each participant. For this, the following parameters were used: Repetition Time (TR) = 7.4 s, Echo Time (TE) = 3.4 s, inversion time (TI) = 876.2 ms (minimum TI delay), Flip angle (deg) = 8, Field of view (FOV) = 250 × 250 (×180), matrix size = 240 (Reconstruction matrix), voxel size = 1 × 1 × 1 (1.041 reconstructed); Acquisition time 5.57 min.

fMRI Image Analysis

Statistical parametric mapping (SPM8; Functional Imaging Laboratory, University College London) was used to spatially realign functional data, and coregister them to the individual anatomical image before normalizing to standard MNI space and smoothing with an isometric Gaussian kernel with a full-width at half-maximum of 9 mm.

First-level design (within-subject)

For each subject, we used linear regression to model fMRI BOLD responses to each of the nine cue-conditional outcomes, i.e., one coin following cue 1, two coins following cue 1, three coins following cue 1, one coin following cue 2... etc. We used a standard rapid-event–related fMRI approach in which evoked hemodynamic responses to stimulus events are estimated separately by convolving a canonical hemodynamic response function with a stimulus function encoding the onsets for each event. These nine events were entered into a design matrix together with six movement parameters. Our main objective here was to contrast probable versus improbable rewards, in a condition which has zero RPE on average, i.e., at the expected reward of two coins. To exclude SPE explanations, we therefore added further control variables as follows.

Basic SPE model. We included un/signed SPEs as "parametric modulators," conditional on five different learning rates. Parametric modulators were derived from the learning models described in the Information Box. Specifically, they were

1. Signed SPEs associated with state-transitions on each trial, $\delta_{S_{ij}}^{\alpha_k}$ (see Information Box), conditional on five learning rates α_k=0.1, 0.3, 0.5, 0.7, 0.9. By extending the notation used in the Information Box, these can be written as $\delta_{S_{ij}}^{\alpha_k}$.
2. Unsigned SPEs for each learning rate, i.e., $|\delta_{S_{ij}}^{\alpha_k}|$.

We used five learning rates because of evidence that there may be many different learning rates in the brain, operating simultaneously in different areas (O'Doherty et al., 2003; Tobler et al., 2007). We did not take an independent behavioral or autonomic measure of "the learning rate" as a proxy for the neuronal learning rate. While this may be appropriate, it rests on the stronger assumptions that (1) There is a single neuronal learning rate, (2) the behavioral learning rate and the neuronal learning rate are identical. By including five different learning rates, we gave the PE model the best chance to explain BOLD activation.

Augmented model. As a secondary confirmation, to further exclude RPE based explanations, we confirmed that any effects remained significant in an augmented model which also contained un/signed RPEs. To specify this augmented model, we added two further sets of parametric modulators, also time locked to the outcome of each trial, to the above design

1. Signed RPE associated with the monetary outcome on each trial, conditional on five different learning rates α_k=0.1, 0.3, 0.5, 0.7, 0.9. These can be written as $\delta_{R_i}^{\alpha_k}$.
2. Unsigned RPEs for each of the five learning rates, i.e., $|\delta_{R_i}^{\alpha_k}|$.

In this way, even though RPEs equal 0 for 2 coins on average, we ensure that we are maximally conservative when we make the claim that our surprise responses are not RPE responses: i.e., they are not confounded with any residual component of an RPE signal.

Optimal surprise model: In a third and final model we asked whether activations reflected *optimal* surprise, conditional on a Bayesian learner. Conditional surprise can be quantified mathematically by Shannon surprise, $-log(P(j|i)$, for which subjects must first learn the relative probability of rewards, denoted by $P(j|i)$, where again $j \in \{CHF\ 1, CHF\ 2, CHF\ 3\}$, $i \in \{cue\ 1, cue\ 2, cue\ 3\}$ (Dayan et al., 2000; Friston 2009; MacKay, 2003). We therefore looked for evidence of a hemodynamic signal that tracked the Shannon Surprise expressed by a model-based Bayesian learner. We trained a simple Bayesian model which learned the conditional probability of each reward state following each cue $P(j|i)$ and expressed Shannon surprise $-\log(P(j|i))$. We assumed that P was learnt by updating multinomial distribution over the random number of coins j, i.e., $P(j|i)=P(j|\theta_i)=\theta_i^j$, under i.i.d. assumptions. In this notation, each element of the 3-vector θ_i gives the probability of receiving 1, 2, or 3 coins following cue i: The superscript simply indexes these three elements. Assuming an uninformative (Dirichlet) prior $p(\theta_i)=Dir(a,b)$, with concentration parameter $a=1$ and uniform base distribution $b=(\frac{1}{3},\frac{1}{3},\frac{1}{3})$, the surprise at observing j coins then simply corresponds to $-\log\left(\frac{n_j+\frac{1}{3}}{\sum_{j=1}^{3} n_j+1}\right)$. Here n_j is the number of times that j coins have followed this cue to date, so $(n_j+\frac{1}{3})/(\sum_{j=1}^{3} n_j+1)$ just reports the (regularized) relative empirical frequency of j coins given the cue.

This procedure resulted in a trial-by-trial expression of Shannon surprise which we included as parametric

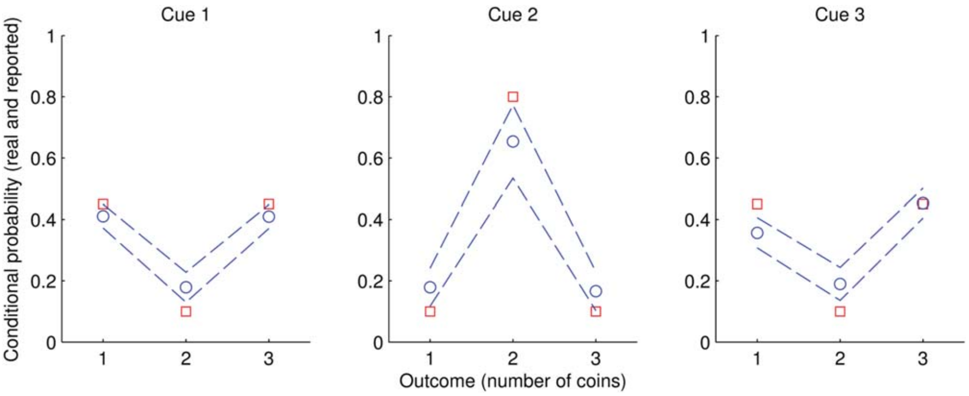

Figure 2.
Self-reported frequency of each outcome—1, 2, or 3 coins—conditional on each of the three cues. Blue circles indicate the mean frequency reported by subjects after the task. Dotted lines correspond to 95% confidence bounds. Red squares indicate the true frequency with which each outcome followed each cue, i.e., ground truth (see also Fig 1). The self-reported frequencies reflected the actual frequencies reasonably well.

modulator of the outcome for each trial. In addition to this, we included all of the un/signed RPEs and SPEs of the previous model as covariates of no interest. Our design included convolved stimulus events for each cue *i* and each outcome *j*, and movement parameters as covariates of no interest.

The inclusion of five learning rates increased the ability of RPE (and SPE) to explain variance otherwise attributable to a purely model-based surprise in all three models. In this way, we ensure that we are maximally conservative when we make the claim that our surprise responses are not RPE responses: i.e., they are not confounded with any residual component of an RPE signal.

Second-level design (between-subject)

We used the standard summary-statistic approach for inference. Namely, we treated subject-specific first-level contrast images as observations. To examine the consistency of our effects over subjects we used these contrast images to calculate a one-sample *t*-statistic. We first tested the contrast between hemodynamic responses to the improbable two CHF outcome versus the probable two CHF outcome. We then tested the group-level effect of trial-by-trial Shannon surprise, as elicited by our Bayesian learner.

RESULTS

Behavior 1

In the purely behavioral study subjects reported the number of coins presented on the screen. For each subject, we compared the average time it took to respond to the improbable two coin outcome (following bimodal cues) versus the probable two coin outcome (following unimodal cue). Using a one-sample summary-statistic approach, a t-test showed that subjects were on average 14 ms (95% CI = [2.5, 22.4]) slower in the improbable case ($P = 0.018$, df = 15). Supporting Information Figure 1 plots subjects' time to report the "expected reward" (i.e., the two CHF) following each cue.

Behavior 2

A different behavioral measure was taken from the 19 different subjects in the fMRI study (detailed below). In debriefing, we asked these subjects to draw a histogram over coins for each cue (the conditional probability distribution). Grading these as correct if they reported the true contingency (unimodal or bimodal), only six attempts out of $57{=}19\ subjects{\times}3\ cues$ were unsuccessful. Assuming (conservatively) that subjects chose unimodal and bimodal distributions with equal probability at chance (for each cue), a Binomial test gave $p < 0.00001$ $(N{=}57, \theta{=}0.5)$. Subjects therefore acquired an accurate declarative "model" of the contingencies.

As can be seen in Figure 2, the self-reported distributions qualitatively matched the real distributions. Apart from bi- versus unimodality, there was some suggestion of probability distortion (Tversky and Kahneman, 1992): small probabilities tended to be over-estimated and larger ones under-estimated.

fMRI Study

We first analyzed brain data with the basic model, in which SPE served as covariates of outcome-related responses. A between-subject (random effects) analysis contrasted hemodynamic responses to the improbable

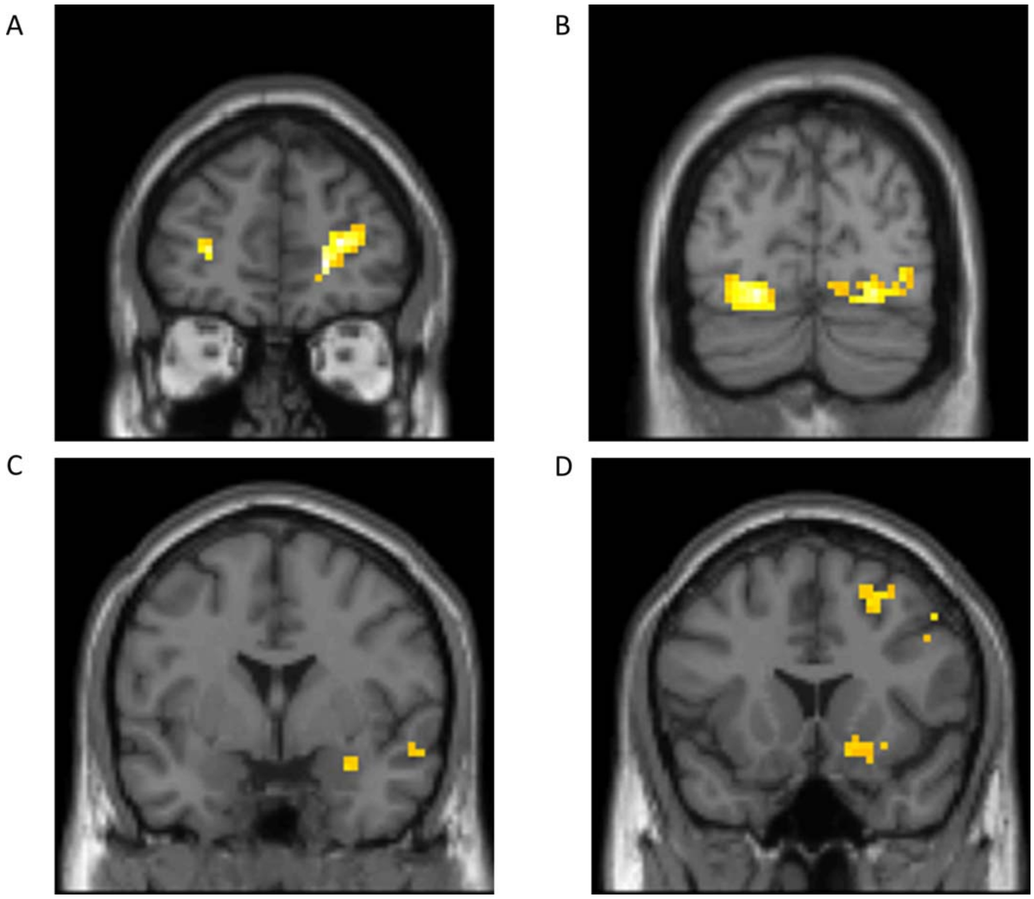

Figure 3.

Hemodynamic response to surprising, improbable rewards that carry no RPE. We used linear regression to assess hemodynamic responses to improbable versus probable rewards, under a condition with zero RPE on average. This statistical analysis controlled for un/signed SPEs. We found significant cluster activations in the right frontal pole (A), bilateral occipital lobe (B), the right amygdala (C), and the right mid frontal gyrus (D). The VS activation partly visible in (D) survived small volume correction using an anatomical definition of VS. These activations are consistent with surprise at the sensory properties of the outcome, i.e., the reward state, and/or surprise at the rewarding aspects of the outcome. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

versus probable two-coin outcome. This revealed four regions of activation. These effects were significant following multiple comparison correction across the whole brain [i.e., family-wise error (FWE) cluster-level whole-brain corrected with p=0.001 as cluster-inducing height threshold]. We found significant cluster activations in the right frontal pole, $P = 0.026$, x = (21, 47, 1), bilateral occipital lobe, $P < 0.001$, (−21, −79, −11) and (27, −79, −11) respectively, the right amygdala $P = 0.012$, (54, −4, −14) and the right mid frontal gyrus $P = 0.013$, (27, 8, 52), see Figure 3. Because the ventral striatum (VS) is strongly implicated in RPE, we wondered whether it would also be sensitive to conditional improbability. A small volume analysis using an anatomical definition revealed activation in the right VS, significant at cluster and peak level ($P = 0.037$ and $P = 0.04$). This is also visible in Figure 3D. Importantly, all of these activations were also significant in an augmented model which additionally controlled for un/signed RPE explicitly as a covariate (see "First-level design" section).

We next looked for evidence that the brain tracks trial-by-trial Shannon surprise (see the final model of "First-level design"). A between-subject (random effects) analysis examined the average effect of Shannon surprise, having controlled for un/signed RPE and SPE in the analysis. We again found strong bilateral occipital activation, $P < 0.001$, (−18, −91, −5) and (27, −79, −11) respectively and right frontal pole activation, $P = 0.017$, (21, 47, 1). Additionally, we found activation within the right superior parietal gyrus, $P = 0.012$, (30, −70, 49). We did not find ventral striatal activity following a small volume correction.

Following the request of a reviewer, we repeated all of the above analyses with 6 mm smoothing and observed a very similar pattern of significant activation in each case. Interestingly, this analysis now revealed a significant Shannon surprise activation in right VS following small volume correction.

DISCUSSION

We have studied reward learning in a passive learning situation. It is known that existing RPE schemes do not fully account for learning in this setting (Dayan and Niv, 2008; Schultz and Dickinson, 2000): For example, they have limited capacity for subjective uncertainty (Preuschoff and Bossaerts, 2007; Schultz et al., 2008) and simply associate each cue or "state" with a single value. Experimental evidence points to simple learning in the absence of RPEs, e.g. experiments in the conditioning literature on what is known as "identity unblocking," where a change in the identity of the rewarding stimulus leads to new learning, even when the amount of "reward" is properly controlled for (Burke et al., 2008; Bornstein et al., 2011; McDannald et al., 2012; Rescorla, 1999). In contrast, humans and animals can use environmental cues to predict the likelihood of specific outcomes (Balleine, 2005; Balleine et al., 2009; Dayan and Balleine, 2002; d'Acremont et al., 2013; Fletcher et al., 2007; Gläscher et al., 2010; Griffiths, 2007). There is evidence that such "internal models" are learned via other forms of SPE. To isolate the neuronal substrate of surprise—not attributable to RPE or SPE—we have used a simple Pavlovian task which the model-free RPEs cannot learn because RPEs in response to outcomes eliciting high versus low conditional surprise are zero and statistically controlled for SPE explanations.

We showed that there are surprise responses that cannot be accounted for PE. We also observed surprise signals in the right frontal pole (Fig. 3A), bilateral occipital lobe (Fig. 3B), the right amygdala (Fig. 3C) and the right mid frontal gyrus (Fig. 3D). A small volume analysis revealed significant surprise effects, beyond PE, in the VS. Primary visual and frontal polar activations were replicated across all of our analyses.

Primary visual responses are consistent with subjective surprise at sensory features of the outcome: i.e., reward *identity* as opposed to a scalar reward *value* or utility (Alink et al., 2010; Dayan and Niv, 2008; Kok et al., 2012). This response may reflect top–down attention effects that follow in the wake of surprise. In any case, a surprise effect in early visual cortex accords with theories holding that top–down predictions modulate the response of primary sensory regions to incoming sensory information. From this perspective our data emphasize that these predictions are probabilistically sophisticated: neither scalar nor unimodal (Gaussian). Our data may also cast light on earlier studies that showed PEs modulate visual cortex responses and its connectivity during associative learning (den Ouden et al., 2009, 2010; Summerfield and Koechlin, 2008; Summerfield et al., 2008;) but did not dissociate surprise. Our empirical dissociation of surprise serves as a reminder that prediction errors are not the only way to understand such learning.

Frontal polar responses occurred in a region implicated in sophisticated model-based capabilities, including goal-directed reasoning and general problem-solving (Genovesio et al., 2013). This region evolved after the split between New World and Old World primates, and may have specifically evolved during ape and human relation (Genovesio et al., 2013).

There are two distinct notions of surprise relevant to paradigms like ours. The *perceptual surprise* associated with perceptual state or "identity" of the outcome (based on probabilistic distributions over a perceptual space) which we have emphasized thus far is, at least conceptually, distinct from the *utility surprise* about how *rewarding* the outcome is. This latter would require a probability distribution over the scalar "utility" or "reward value." Crucially for us, in our task *neither* can be learned with simple RPE-based surprise mechanisms. To maximize the subjective and hemodynamic impact of surprising events *perceptual surprise* and the *utility surprise* were intentionally aliased or confounded in our design, i.e., a reward value or "utility" of one CHF is associated with a given visual percept (a circle/coin), a reward value of two CHF is associated with another percept (two overlapping circles/coins) and a reward value of three CHF is associated with a third percept (three overlapping circles/coins). In principle, by simultaneously evoking *perceptual surprise* and *utility surprise*, our design gains sensitivity to either effect at the cost of losing specificity about which effect is responsible. In practice, however previous literature suggests a significant disjunction between the brain regions involved in perceptual versus utility processing. That we observed surprise responses in primary (visual) perceptual regions has encouraged us to interpret this in terms of *perceptual surprise*, i.e., consequent from learned associations between the cue and the *perceptual properties of the reward*. Conversely, the observation of surprise effects in VS and amygdala points to *utility surprise*.

Our paradigm relates to the literature on implicit statistical learning in which state-state associations are learned without any feedback e.g. (Fiser and Aslin, 2001; Turk-Browne et al., 2005). Our task differs in that we can dissociate reward-independent learning that arises within a classical reward learning task and exclude common PE explanations. Previous studies have examined the neural bases of predictive or causal learning with neutral stimuli e.g. (Corlett et al., 2007; d'Acremont et al. 2013; Fletcher et al., 2001; Gläscher et al., 2010; Turner et al., 2004). Several brain structures appear to code prediction errors in relation to such learning (Boly et al., 2011; Corlett et al., 2010; Friston et al., 2006; Friston, 2009; Gläscher et al., 2010; Schultz and Dickinson, 2000). Our analysis revealed surprise responses beyond PE responses. In other words, these responses were based on the conditional improbability of events, which could not be explained by the most straightforward formulation of PEs.

ACKNOWLEDGMENTS

The authors thank Peter Dayan for helpful comments.

REFERENCES

Alink A, Schwiedrzik CM, Kohler A, Singer W, Muckli L (2010): Stimulus predictability reduces responses in primary visual cortex. J Neurosci 30:2960–2966.

Balleine BW (2005): Neural bases of food-seeking: Affect, arousal and reward in corticostriatolimbic circuits. Physiol Behav 86: 717–730.

Balleine BW, Daw ND, O'Doherty JP (2008): Multiple forms of value learning and the function of dopamine. Neuroeconomics: decision making and the brain 367–385.

Balleine B, Daw N, O'Doherty J, Balleine B (2009): Multiple forms of value learning and the function of dopamine. Adv Virus Res 367.

Boly M, Garrido MI, Gosseries O, Bruno MA, Boveroux P, Schnakers C, Massimini M, Litvak V, Laureys S, Friston K (2011): Preserved feedforward but impaired top-down processes in the vegetative state. Science 332:858.

Bornstein AM, Nylen EL, Steele SA (2011): Unblocking the neural substrates of model-based value. J Neurosci 31:10117–10118.

Burke KA, Franz TM, Miller DN, Schoenbaum G (2008): The role of the orbitofrontal cortex in the pursuit of happiness and more specific rewards. Nature 454:340–344.

Corlett P, Murray G, Honey G, Aitken M, Shanks D, Robbins T, Bullmore E, Dickinson A, Fletcher P (2007): Disrupted prediction-error signal in psychosis: Evidence for an associative account of delusions. Brain 130:2387–2400.

Corlett P, Taylor J, Wang X, Fletcher P, Krystal J (2010): Toward a neurobiology of delusions. Prog Neurobiol 345–369.

d'Acremont M, Schultz W, Bossaerts P (2013): The human brain encodes event frequencies while forming subjective beliefs. J Neurosci 33:10887–10897.

Dayan P, Balleine BW (2002): Reward, motivation, and reinforcement learning. Neuron 36:285–298.

Dayan P, Niv Y (2008): Reinforcement learning: The good, the bad and the ugly. Curr Opin Neurobiol 18:185–196.

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

den Ouden HEM, Friston KJ, Daw ND, McIntosh AR, Stephan KE (2009): A dual role for prediction error in associative learning. Cereb Cortex 19:1175–1185.

den Ouden HEM, Daunizeau J, Roiser J, Friston KJ, Stephan KE (2010): Striatal prediction error modulates cortical coupling. J Neurosci 30:3210.

Fiser J, Aslin RN (2001): Unsupervised statistical learning of higher-order spatial structures from visual scenes. Psychol Sci 12:499–504.

Fletcher P, Anderson J, Shanks D, Honey R, Carpenter T, Donovan T, Papadakis N, Bullmore E (2001): Responses of human frontal cortex to surprising events are predicted by formal associative learning theory. Nat Neurosci 4:1043–1048.

Friston K (2009): The free-energy principle: A rough guide to the brain? Trends Cogn Sci 13:293–301.

Friston K, Kilner J, Harrison L (2006): A free energy principle for the brain. J Physiol Paris 100:70–87.

Genovesio A, Wise SP, Passingham RE (2013): Prefrontal–parietal function: From foraging to foresight. Trend Cogn Sci 2014;18: 72–81.

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

Glimcher PW (2011): Understanding dopamine and reinforcement learning: The dopamine RPE hypothesis. Proc Natl Acad Sci 108(Suppl 3):15647–15654.

Griffiths TL, Canini KR, Sanborn AN, Navarro DJ (2007): Unifying rational models of categorization via the hierarchical Dirichlet process. Proceedings of the Twenty-Ninth Annual Conference of the Cognitive Science Society 323–328.

Kok P, Jehee JF, de Lange FP (2012): Less is more: Expectation sharpens representations in the primary visual cortex. Neuron 75:265–270.

Ljungberg T, Apicella P, Schultz W (1992): Responses of monkey dopamine neurons during learning of behavioral reactions. J Neurophysiol 67:145–163.

Ludvig EA, Sutton RS, Kehoe EJ (2012): Evaluating the TD model of classical conditioning. Learn Behav 40:305–319.

MacKay D (2003): Information Theory, Inference, and Learning Algorithms. Cambridge University Press.

McClure SM, Berns GS, Montague PR (2003): Temporal prediction errors in a passive learning task activate human striatum. Neuron 38:339–346.

McDannald MA, Takahashi YK, Lopatina N, Pietras BW, Jones JL, Schoenbaum G (2012): Model-based learning and the contribution of the orbitofrontal cortex to the model-free world. Eur J Neurosci 35:991–996.

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

O'Doherty 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.

Pearce JM, Hall G (1980): A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychol Rev 87:532.

Preuschoff K, Bossaerts P (2007): Adding prediction risk to the theory of reward learning. Ann N Y Acad Sci 1104:135–146.

Rescorla R, Wagner A (1972): Variations in the Effectiveness of Reinforcement and Nonreinforcement. New York: Classical Conditioning II: Current Research and Theory, Appleton-Century-Crofts:113–123.

Rescorla RA (1999): Learning about qualitatively different outcomes during a blocking procedure. Anim Learn Behav 27: 140–151.

Roesch MR, Esber GR, Li J, Daw ND, Schoenbaum G (2012): Surprise! Neural correlates of Pearce–Hall and Rescorla–Wagner coexist within the brain. Eur J Neurosci 35:1190–1200.

Schultz W, Dickinson A (2000): Neuronal coding of prediction errors. Ann Rev Neurosci 23:473–500.

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

Schultz W, Preuschoff K, Camerer C, Hsu M, Fiorillo C, Tobler P, Bossaerts P (2008): Explicit neural signals reflecting reward uncertainty. Phil Trans R Soc Lond, Ser B: Biol Sci 363:3801.

Summerfield C, Koechlin E (2008): A neural representation of prior information during perceptual inference. Neuron 59:336–347.

Summerfield C, Monti JMP, Trittschuh EH, Mesulam MM, Egner T (2008): Neural repetition suppression reflects fulfilled perceptual expectations. Nat Neurosci 11:1004.

Sutton R, Barto A (1990): Time-derivative models of Pavlovian reinforcement. Learning and computational neuroscience: Foundations of adaptive networks:497–537.

Sutton R, Barto A (1998): Reinforcement learning: An introduction. Cambridge: The MIT press.

Tobler PN, Fletcher PC, Bullmore ET, Schultz W (2007): Learning-related human brain activations reflecting individual finances. Neuron 54:167–175.

Turk-Browne NB, Jungé JA, Scholl BJ (2005): The automaticity of visual statistical learning. J Exp Psychol Gen 134:552–563.

Turner DC, Aitken MR, Shanks DR, Sahakian BJ, Robbins TW, Schwarzbauer C, Fletcher PC (2004): The role of the lateral frontal cortex in causal associative learning: Exploring preventative and super-learning. Cereb Cortex 14:872–880.

Tversky A, Kahneman D (1992): Advances in prospect theory: Cumulative representation of uncertainty. J Risk Uncertainty 5:297–323.