NeuroImage 115 (2015) 214–223

Contents lists available at ScienceDirect

NeuroImage

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

Expectation modulates neural representations of valence throughout the human brain

Ashwin G. Ramayya, Isaac Pedisich, Michael J. Kahana *

Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104, USA

ARTICLE INFO

Article history:
Received 24 October 2014
Accepted 19 April 2015
Available online 30 April 2015

Keywords:
Intracranial electroencephalography
ECoG
iEEG
High frequency activity
HFA
Reward
Value
Valence
Reinforcement learning

ABSTRACT

The brain's sensitivity to unexpected gains or losses plays an important role in our ability to learn new behaviors (Rescorla and Wagner, 1972; Sutton and Barto, 1990). Recent work suggests that gains and losses are ubiquitously encoded throughout the human brain (Vickery et al., 2011), however, the extent to which reward expectation modulates these valence representations is not known. To address this question, we analyzed recordings from 4306 intracranially implanted electrodes in 39 neurosurgical patients as they performed a two-alternative probability learning task. Using high-frequency activity (HFA, 70–200 Hz) as an indicator of local firing rates, we found that expectation modulated reward-related neural activity in widespread brain regions, including regions that receive sparse inputs from midbrain dopaminergic neurons. The strength of unexpected gain signals predicted subjects' abilities to encode stimulus–reward associations. Thus, neural signals that are functionally related to learning are widely distributed throughout the human brain.

Introduction

Theories of reinforcement learning postulate that greater learning occurs following unexpected outcomes than following expected outcomes (Rescorla and Wagner, 1972; Pearce and Hall, 1980; Sutton and Barto, 1990). How the brain represents these unexpected gains and losses has been the focus of considerable research. For example, functional neuroimaging studies have identified a specialized group of brain regions that encode reward prediction errors (Berns et al., 2001; McClure et al., 2003; Pessiglione et al., 2006; Montague et al., 2006; Rutledge et al., 2010; Bartra et al., 2013). Several of these regions (e.g., ventral striatum, medial prefrontal cortex) receive prominent inputs from midbrain dopaminergic (DA) neurons, a neural population known to be functionally important for reinforcement learning in animals (Schultz et al., 1997; Reynolds et al., 2001) and humans (Zaghloul et al., 2009; Ramayya et al., 2014a).

Recent evidence raises the possibility that the neural processes that support reinforcement learning may extend beyond regions that are heavily innervated by dopamine neurons. Vickery et al. (2011) used multi-voxel pattern analysis to decode outcome valence from activity in almost every cortical and subcortical region in the human brain. However, because this study did not assess reward expectation, the extent to which these widespread valence signals reflect reward prediction errors that are functionally important for learning is not known. If reinforcement learning is a widespread brain process, one would predict that valence representations throughout the brain would be modulated by reward expectation.

To test this hypothesis, we obtained intracranial electroencephalography (iEEG) recordings from the cortex and medial temporal lobe (MTL) of 39 patients with drug-refractory epilepsy as they performed a two-alternative probability learning task. We studied changes in high-frequency activity (HFA; 70–200 Hz) at individual electrodes, an established indicator of local spiking activity (Manning et al., 2009; Ray and Maunsell, 2011) that can be used to study heterogeneous patterns of activity within a region (Bouchard et al., 2013a). We identified putative valence signals that demonstrated differential HFA following positive and negative outcomes and we then assessed their relation to trial-by-trial estimates of reward expectation. In this way, we sought to characterize the anatomical distribution of expectation-modulated valence signals and assess their functional relevance for learning.

Materials and methods

Subjects

Patients with drug-refractory epilepsy underwent a surgical procedure in which grid, strip, and depth electrodes were implanted in order to localize epileptogenic regions. Clinical circumstances alone

* Corresponding author at: Department of Psychology, University of Pennsylvania, 3401 Walnut St., Room 303C, Philadelphia, PA 19104, USA.
E-mail address: kahana@psych.upenn.edu (M.J. Kahana).

http://dx.doi.org/10.1016/j.neuroimage.2015.04.037
1053-8119/

determined number of implanted electrodes and their placement. Data were collected from Thomas Jefferson University Hospital (TJUH) and the Hospital of University of Pennsylvania (HUP) in collaboration with the neurology and neurosurgery departments at each institution. Our research protocol was approved by the Institutional Review Board at each hospital and informed consent was obtained from the participants. In total, we recorded neural activity from 39 subjects (12 females, seven left-handed, mean age 37 years).

Reinforcement learning task

Subjects performed a two-alternative probability learning task, which has been previously used to study reinforcement learning and value-based decision making (Fig. 1; Frank et al., 2004, 2007; Zaghloul et al., 2012). During the task, subjects selected between pairs of Japanese characters ("items") and received positive or negative feedback following each choice. Subjects were informed that one item in each pair carried a higher probability of positive feedback than the other item, and were asked to select items that maximized their probability of obtaining positive feedback. On a given trial, the items were simultaneously displayed on the screen; one on the left side and one on the right side. They were presented on a dark gray background in white font. The items remained on the screen until subjects responded by pressing the left or right "SHIFT" button on a keyboard (to select the item on the left or right side of the screen, respectively). Once a response was registered by the computer, the selected item was highlighted in blue, and feedback was provided immediately. In the event of positive feedback, we presented a green screen and the sound of a cash register. In the event of negative feedback, we presented a red screen and the sound of an error tone. The colored screen was presented for 2 s. There was a 0–400 ms jitter between successive trials. Items were randomly arranged on the left or right side of the screen from trial to trial.

During a session, subjects were presented with up to three novel item pairs, each carrying a distinct relative reward rate (80/20, 70/30, or 60/40). This feature of the task allows for the study of value-based decision making in a subsequent stage of the experiment that is not considered in this study (Frank et al., 2007; Zaghloul et al., 2012). Distinct item pairs were presented in a randomly interleaved manner. Reward rates associated with each item were determined randomly prior to

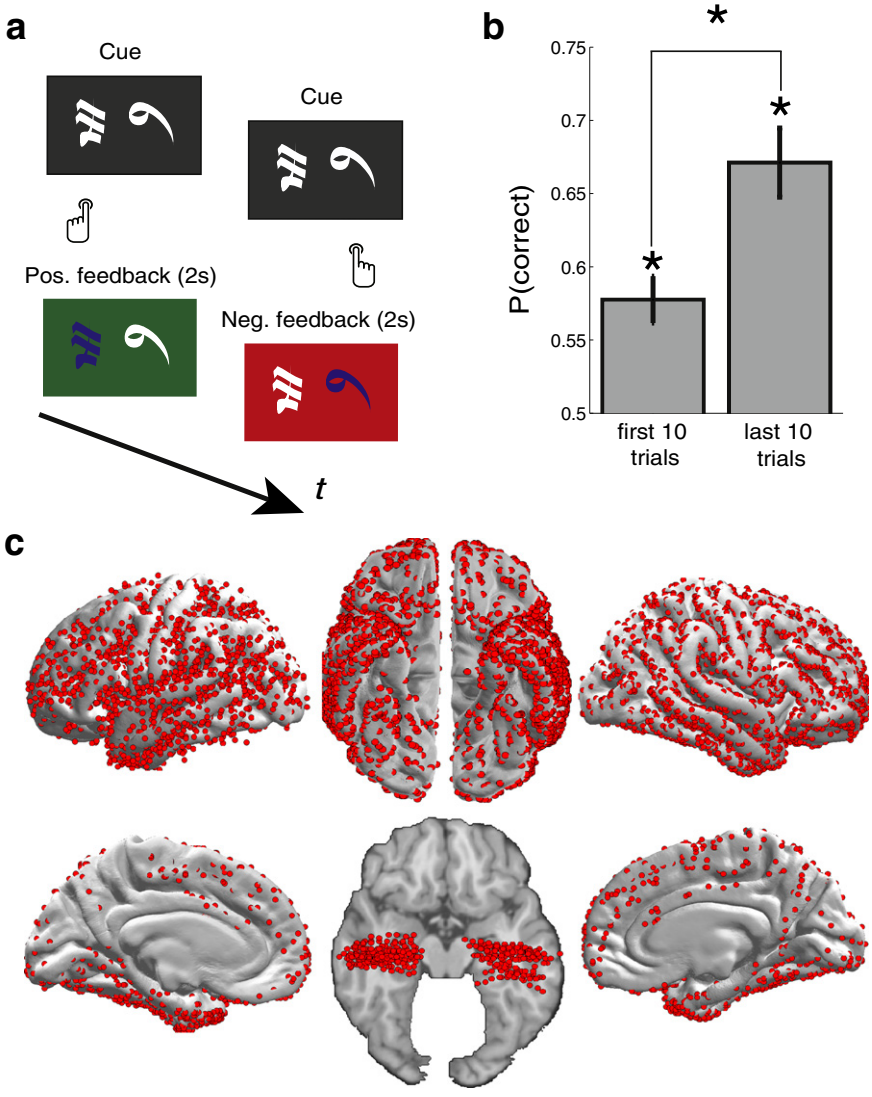

Fig. 1. Reinforcement learning task, and subjects' behavior, and electrode locations. a. Subjects selected between pairs of Japanese characters on a computer screen and probabilistically received positive or negative audio-visual feedback following each choice. b. Average tendency towards selecting the high-probability item during the first and last 10 trials of each item pair. Error bars represent s.e.m. across subjects. c. iEEG electrodes from each subject were localized to a common anatomical space (see Materials and methods section). We show strip and grid electrodes on the cortical surface, and depth electrodes targeting the medial temporal lobe on the axial slice. On rare occasions, depth electrodes were placed in the frontal and parietal lobes to supplement surface recordings (not shown).

each session and fixed throughout the experiment. Each session began with the exclusive presentation of a single item pair (random selection of a relative reward rate). If participants met a minimum performance criteria on the given item pair over a block of 10 trials (i.e., accuracy ≥60% for 80/20 or 70/30 pairs, or ≥50% for the 60/40 pair), a second item pair was introduced and randomly interleaved along with the first item pair. A third item pair was only introduced in subjects that met the performance criteria on the two item pairs already introduced. Participants performed a total of 107 sessions (each subject performed an average of 2.82 sessions), with an average of 130 trials per session.

iEEG recordings

Subdural (grids and strips) and depth electrodes were spaced 10 mm and 8 mm apart, respectively. iEEG was recorded using a Nihon-Kohden (TJUH) or Nicolet (HUP) EEG system. Based on the amplifier and the discretion of the clinical team, signals were sampled at either 512, 1024, or 2000 Hz. Signals were converted to a bipolar montage by taking the difference of signals between each pair of immediately adjacent electrodes on grid, strip, or depth electrodes. The resulting bipolar signals were treated as new virtual electrodes (henceforth referred to as "electrodes" throughout the text), originating from the midpoint between each electrode pair (Burke et al., 2013). Analog pulses synchronized the electrophysiological recordings with behavioral events.

Extracting high-frequency activity from iEEG recordings

We convolved segments of iEEG recordings (1000 ms before feedback onset to 2000 ms after onset, plus a 1000 ms flanking buffer) with 30 complex valued Morlet wavelets (wave number 7) with center frequencies logarithmically spaced from 70 to 200 Hz (Addison, 2002). We first squared and then log-transformed the wavelet convolutions, resulting in a continuous representation of log-power surrounding each feedback presentation. We averaged these log-power traces in 200 ms epochs with 190 ms overlap surrounding feedback presentation (−1000–2000 ms), yielding 281 total time intervals surrounding feedback presentation. To identify HFA, we averaged power across all frequencies (ranging from 70 to 200 Hz). We *z*-transformed HFA power values within each session by the mean and standard deviation of task-related HFA recorded from that session (0–500 ms post-stimulus, −750–0 ms pre-choice, and 0–2000 ms post-feedback). Henceforth, we refer to *z*-transformed HFA values as HFA.

Assessing HFA differences between positive and negative outcomes

For each electrode, we identified temporally-contiguous HFA differences between positive and negative feedback by performing a cluster-based permutation procedure that accounts for multiple comparisons (Maris and Oostenveld, 2007). As suggested by Maris and Oostenveld (2007), we began by performing an unpaired *t*-test at each time interval comparing HFA distributions associated with all positive and negative feedback trials performed by the subject. Using an uncorrected $p = 0.05$ as a threshold, we identified the largest cluster of temporally adjacent windows that showed positive *t*-statistics (greater HFA following positive compared to negative outcomes), and the largest cluster of temporally adjacent windows that showed negative *t*-statistics (greater HFA following negative compared to positive outcomes). By taking the sum within each of these clusters, we computed positive and negative "cluster statistics", respectively. To assess the statistical significance of each cluster statistic, we generated a null distribution of cluster statistics based on 1000 iterations of shuffled data (on each iteration, positive and negative feedback labels were randomly assigned to HFA traces recorded during the session). Based on where each cluster statistic fell on the null distribution, we generated a one-tailed *p*-value for each effect. We considered an effect to be significant if it was associated with a one-tailed p-value < 0.025, thus, the false-positive rate of identifying either a positive or negative cluster at a given electrode was set at 5%.

Assessing the frequency of a particular effect across subjects

To assess whether a particular effect was more frequently observed by chance across subjects, we performed the following procedure ("counts *t*-test"). In each subject, we counted the number of significant electrodes that we observed ("true counts"), and generated a binomial distribution of counts values expected by chance ("null counts distribution"), based on the number of available electrodes in that subject and the false-positive rate associated with the test. We obtained a *z*-scored counts value in each subject by comparing the true counts value to the null counts distribution. We then assessed whether distribution of *z*-scores across subjects deviated from zero via a one-sample paired *t*-test; positive *t*-statistics suggest that the effect was more frequently observed than by chance, and negative *t*-statistics suggest that the effect was less frequently observed than by chance. When comparing the frequencies of two-effects across subjects (e.g., reward and penalty effects), we performed a paired counts *t*-test in the following manner. Within each subject, we obtained *z*-scored counts values for reward and penalty effects based on the null counts distribution as described earlier, and compared the distributions of reward- and penalty-related *z*-values across subjects (via paired *t*-test). Positive *z*-values indicate that reward effects occurred more frequently than penalty effects, whereas negative values indicate that penalty effects occurred more frequently than reward effects. We corrected for multiple comparisons using a false discovery rate (FDR) procedure (Benjamini and Hochberg, 1995).

Electrode localization

Surface electrodes (strips and depths) were manually identified on each post-operative CT scans and transformed to a common cortical surface representation to allow for comparisons across subjects. We employed FreeSurfer (Dale et al., 1999) to generate a cortical surface representation that was representative of our patient population, which includes individuals undergoing intracranial EEG monitoring for drug-refractory epilepsy. We did this by generating cortical surface reconstructions for a large group of patients who volunteered to participate in our research studies. We included patients for whom a pre-operative MRI was available from which a cortical surface could be modeled ($n = 62$). Along with subjects who participated in the current study, this group included subjects who participated in previous studies conducted by our group (e.g., Burke et al., 2013). We aggregated these surfaces to generate an average cortical surface representation, which was co-registered to the MNI152 brain (Fischl et al., 1999). Each point on this surface representation was automatically assigned an anatomical label based on a manually-labeled anatomical atlas (Desikan et al., 2006). To map electrode coordinates from the CT scan onto the cortical surface, we registered each post-operative CT scan to the average cortical surface using a rigid-body 6 degrees-of-freedom affine transformation algorithm, and manually adjusted each transform such that electrodes were positioned as close to the cortical surface as possible. Finally, electrodes were "snapped" to the cortical surface by moving each electrode to the nearest point on the gyral surface (mean deviation of all electrodes was 2.16 mm; 95% of electrodes were moved less than 5.53 mm). We assigned an anatomical label to each bipolar pair of electrodes based on the location on the cortical surface that was closest to the midpoint between the two electrodes. Depth electrodes were manually localized by a neuroradiologist using a post-operative MRI scan. To visualize these depth electrodes in a common anatomical space, we transformed them to MNI coordinates using the same CT-to-average surface transformation described above. However, we did not snap depth electrodes to the cortical surface. Depth electrodes were visualized on a MNI brain slice generated using the WFU pick

atlas toolbox (Maldjian et al., 2003). We categorized bipolar electrodes into several regions of interest (ROIs) based on their associated anatomical labels (Table 1). We defined ROIs in order to segregate regions that might be expected to demonstrate distinct functional patterns based on prior fMRI studies of reinforcement learning (Vickery et al., 2011; Kahnt et al., 2011; Bartra et al., 2013), while ensuring an adequate number of electrodes within each region for across-subject group analyses.

Estimating reward expectation

To obtain trial-by-trial estimates of reward expectation, we fit a standard reinforcement learning model to each subjects' behavioral data (Sutton and Barto, 1990). Because our goal was to model choice behavior during learning, we only considered behavioral data from item pairs where subjects demonstrated evidence of learning (we required >70% accuracy on the last 10 trials and >50% accuracy overall). The *Q*-model maintains independent estimates of reward expectation (Q) values for each option i at each time t (Sutton and Barto, 1990). The model generates a choice on each trial by comparing the Q values of available options on that trial according to:

$$P_i(t) = \frac{\exp(Q_i(t)/\beta)}{\sum_j \exp(Q_j(t)/\beta)}, \quad (1)$$

where β is a parameter that controls the level of noise in the decision process (Daw et al., 2006). When $\beta = 0$ the model deterministically chooses the highest value option; when $\beta = \infty$ the model will randomly choose among the set of possible options. Once an item is selected by the model, feedback is received, and Q values are updated using the following learning rule: $Q_i(t+1) = Q_i(t) + \alpha[R(t) - Q_i(t)]$, where $R(t) = 1$ for correct feedback, $R(t) = 0$ for incorrect feedback, and α is the learning rate parameter that adjusts the manner in which previous reinforcements influence current Q values ($0 \geq \alpha \leq 1$). Large α values heavily weight recent outcomes when estimating Q, whereas small α values incorporate reinforcements from many previous trials. We identified the best-fitting parameters for each subject by performing a grid-search through the two dimensional parameter space (α, learning rate, and β, noise in the choice policy, 0.01 to 1, with a step size of 0.1) and selected the set of parameters that minimized the mean squared error between the model's predictions of subjects' choices (i^*), and subjects' actual choices. To quantify the model's goodness-of-fit, we compared each subject's mean squared error value to a null distribution of mean squared errors generated for that subject's data based on a random guessing model ($p = 0.5$ for all choices, 10,000 iterations). Based on this comparison, we obtained a p-value describing the false-positive rate associated with the observed mean squared error for that subject. In all subjects, the best-fitting parameters provided a better prediction of choice behavior than the random guessing model (FDR-corrected p's < 0.001). We describe mean best-fitting parameters and goodness-of-fit data in Table 3.

Table 1
Regions of interest. Anatomical labels used to define regions of interest.

Region of interest	Desikan–Killiany Atlas labels
Orbitofrontal cortex (OFC)	Medialorbitofrontal, lateralorbitofrontal
Dorsolateral prefrontal cortex (dlPFC)	Rostralmiddlefrontal, caudalmiddlefrontal
Ventrolateral prefrontal cortex (vlPFC)	Parstriangularis, parsopercularis, parsorbitalis
Anterior medial frontal	Superiorfrontal, rostralanteriorcingulate, caudalanteriorcingulate
Posterior medial frontal	Paracentral, posteriorcingulate, isthmuscingulate
Sensorimotor	Precentral, postcentral
Parietal	Superiorparietal, supramarginal, inferiorparietal
Temporal	Banksts, transversetemporal, banksts, middletemporal, inferiortemporal, superiortemporal
Fusiform	Fusiform
Occipital	Cuneus, lateraloccipital, lingual, pericalcarine
Medial temporal lobe (MTL)	Entorhinal, parahippocampal; depth electrodes labeled as hippocampal, enterorhinal, perirhinal, or parahippocampal by neuroradiologist

Data sharing

The behavioral and neural data used in this study are freely available online at (http://memory.psych.upenn.edu/ElectrophysiologicalData).

Results

39 subjects selected between pairs of Japanese characters ("items") and received positive or negative feedback following each choice (Fig. 1a). Subjects were informed that one item in each pair carried a higher reward probability than the other, and that their goal was to maximize their probability of obtaining positive feedback. During each session, subjects were presented with multiple item pairs in an interleaved manner, with each item pair carrying distinct relative reward rates (see Materials and methods section). To assess whether subjects demonstrated learning during the task, we tested the null hypothesis that subjects did not demonstrate a tendency towards selecting the high probability item. We found that subjects demonstrated a tendency towards choosing the high-probability item both during the first 10 trials ($t(38) = 4.84$, $p < 0.001$) and during the last 10 trials of an item pair ($t(38) = 7.24$, $p < 0.001$). Furthermore, we found that subjects were more likely to select the high probability item during the last 10 item pair presentations as compared to the first 10 item pair presentations ($t(38) = 5.11$, $p < 0.001$; Fig. 1b), suggesting that subjects demonstrated learning during the task.

Theories of reinforcement learning posit that individuals alter decisions based on learning signals which integrate information about outcome valence and reward expectation (Rescorla and Wagner, 1972; Sutton and Barto, 1990). To characterize the neural representations of these learning signals, we first identified neural populations that demonstrated distinct activity following positive and negative outcomes. We refer to these signals as "putative valence" signals because in addition to valence, positive and negative feedback conditions also differ in low-level sensory features. We obtained intracranial electroencephalography (iEEG) recordings from 4306 surface and depth electrodes located throughout the cortex and MTL (Fig. 1c). We focused our analyses on HFA (70–200 Hz), an iEEG feature that has been correlated with local neural firing rates (Manning et al., 2009; Ray and Maunsell, 2011), and thereby provides a spatio-temporally precise measure of local neuronal activity (Buzsaki et al., 2012; Burke et al., 2014). Rather than averaging activity within regions of interest, we studied HFA changes at individual electrodes in order to extract information from regions that may demonstrate heterogeneous representations of outcome valence and reward expectation (Bouchard et al., 2013b).

We identified electrodes that showed significant HFA differences between positive and negative feedback (cluster-based permutation procedure; Materials and methods section). We found that 2121 electrodes (49.3%) demonstrated HFA differences between positive and negative outcomes; 860 electrodes (19.9%) showed positive effects (relatively greater HFA following positive feedback, "reward electrodes") and 1012 electrodes (23.5%) showed negative effects (relatively greater HFA following negative feedback, "penalty electrodes," Fig. 2a). We also observed a small subset of electrodes (n = 249, 5.78%) that demonstrated both positive and negative effects during distinct time intervals. To assess whether a particular effect was more frequently observed across subjects than expected by chance, we performed an across-subject t-test on z-transformed counts values ("counts t-test," Materials and methods section). Across subjects, we observed reward and penalty electrodes at above-chance frequencies ($t(38) > 8.94$, $p < 0.001$, each effect was associated with a false-positive rate of 5%). We focus the remainder of our

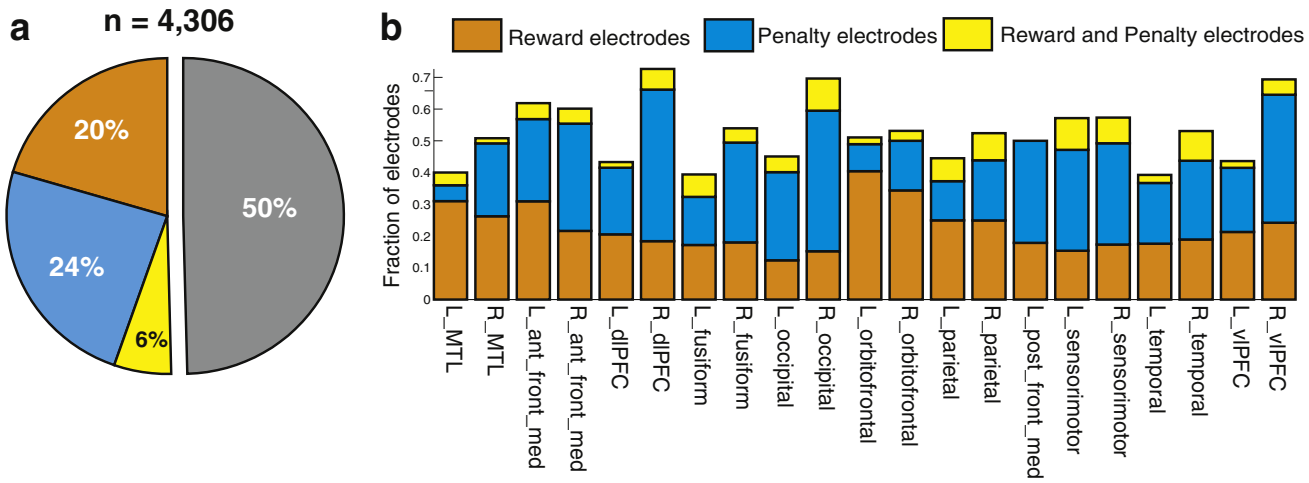

Fig. 2. Anatomical distribution of positive and negative outcome signals. a. Fraction of reward (orange) and penalty (blue) electrodes among all recorded electrodes. b. Fraction of positive and negative electrodes in each ROI. See Table 2 for statistics.

analyses on electrodes that exclusively showed a positive or a negative effect (henceforth, "putative valence-encoding electrodes").

To study the anatomical distribution of putative valence signals, we registered electrodes from each subject to a common anatomical space (Materials and methods section). In several ROIs (Table 1), we assessed whether putative valence-encoding electrodes were more frequently observed than chance (Fig. 2). We only considered ROIs where we recorded neural data from at least five subjects. In 13 of the 21 ROIs that met this criteria (including lateral temporo-parieto-prefrontal regions, anterior medial prefrontal cortex, and the fusiform gyrus), we found that subjects showed both reward and penalty electrodes more frequently than expected by chance (counts t-test, FDR-corrected p's < 0.05; see Table 2 for statistics). In four regions (left and right orbitofrontal cortices, left MTL, and left parietal lobe), we observed reward electrodes more frequently than expected by chance. In two regions (right occipital and left ventrolateral prefrontal cortices), we only observed penalty electrodes more frequently than expected by chance. Overall, we frequently observed putative valence-encoding electrodes in 19 of the 21 ROIs that we studied, suggesting that valence representations are widely distributed throughout the cortex and MTL.

If these putative valence-encoding signals represented learning signals, then one would expect their activity to modulate by subjects' reward expectation during the task. To assess whether this was the case, we studied the relation between reward expectation and mean HFA during time intervals that we observed significant valence-related differences in activity (identified using our cluster-based permutation procedure, Materials and methods section). Because our goal was to study neural processes related to learning, we only considered neural and behavioral data from item pairs in which subjects demonstrated evidence of learning (>70% accuracy on last 10 trials, and >50% accuracy overall). 1315 valence-encoding electrodes (from 26 subjects) were recorded during trials which met this criteria. We did not exclude any periods of time (e.g., early vs. late trials) by applying this learning criterion; rather, we excluded individual stimulus pairs that particular patients were unable to learn. We obtained qualitatively similar results when replicating the analyses described below without applying any learning criteria.

Table 2
Frequency of valence-encoding electrodes. For each region, we list the number of electrodes (column 1), number of subjects (column 2), frequency of reward electrodes (column 3), and frequency of penalty electrodes (column 4). Positive t-statistics indicate frequencies that are greater than expected, whereas negative t-statistics indicate frequencies that are lower than expected. Bold texts in columns 3 and 4 indicate regions that showed valence-encoding electrodes more frequently than expected by chance (FDR-corrected $p < 0.05$).

Region of Interest	Number of electrodes	Number of subjects	Frequency of reward electrodes; counts t-test results	Frequency of penalty electrodes; counts t-test results
L. OFC	48	15	**0.33**; $t(14) = 3.72$; $p = 0.002$	0.08; $t(14) = 0.460$; $p > 0.5$
R. OFC	67	16	**0.33**; $t(15) = 3.10$; $p = 0.007$	0.13; $t(15) = 1.44$, $p = 0.17$
L. dlPFC	223	21	**0.23**; $t(20) = 5.97$, $p < 0.001$	**0.22**; $t(20) = 4.46$, $p < 0.001$
R. dlPFC	246	19	**0.19**; $t(18) = 3.86$, $p = 0.001$	**0.47**; $t(19) = 5.32$, $p < 0.001$
L. vlPFC	92	18	0.20; $t(17) = 2.15$, $p = 0.046$	**0.21**; $t(17) = 3.39$, $p = 0.003$
R. vlPFC	65	16	**0.22**; $t(15) = 2.33$, $p = 0.034$	**0.37**; $t(15) = 3.12$, $p = 0.007$
L. anterior medial frontal	138	16	**0.30**; $t(15)2.91$, $p = 0.010$	**0.25**; $t(15) = 3.89$, $p = 0.001$
R. anterior medial frontal	149	18	**0.22**; $t(17) = 4.05$, $p < 0.001$	**0.33**; $t(17) = 3.30$, $p = 0.004$
L. posterior medial frontal	28	7	0.18; $t(6) = 1.60$, $p = 0.16$	0.32; $t(6) = 2.19$, $p = 0.07$
L. sensorimotor	277	23	**0.15**; $t(22) = 3.97$, $p < 0.001$	**0.32**; $t(22) = 4.06$, $p < 0.001$
R. sensorimotor	262	20	**0.17**; $t(19) = 2.68$, $p = 0.015$	**0.32**; $t(19) = 3.67$, $p = 0.002$
L. parietal	373	26	**0.25**; $t(25) = 5.53$, $p < 0.001$	0.11; $t(25) = 2.14$, $p = 0.042$
R. parietal	267	19	**0.24**; $t(18) = 3.70$, $p = 0.002$	**0.19**; $t(18) = 2.66$, $p = 0.016$
L. temporal	677	28	**0.18**; $t(27) = 5.05$, $p < 0.001$	**0.20**; $t(18) = 2.17$, $p = 0.052$
R. temporal	457	27	**0.16**; $t(26) = 3.72$, $p = 0.001$	**0.23**; $t(26) = 4.54$, $p < 0.001$
L. fusiform	98	23	**0.17**; $t(22) = 3.42$, $p = 0.002$	**0.13**; $t(22) = 4.87$, $p < 0.001$
R. fusiform	97	17	**0.18**; $t(16) = 3.13$, $p = 0.007$	**0.28**; $t(16) = 3.06$, $p = 0.008$
L. occipital	162	20	**0.13**; $t(19) = 2.55$, $p = 0.020$	**0.28**; $t(19) = 3.60$, $p = 0.002$
R. occipital	84	19	0.13; $t(18) = 1.19$, $p = 0.25$	**0.37**; $t(18) = 3.26$, $p = 0.004$
L. MTL	100	19	**0.32**; $t(18) = 4.36$, $p < 0.001$	0.05; $t(18) = 0.25$, $p > 0.5$
R. MTL	52	12	0.27; $t(11) = 2.17$, $p = 0.052$	0.17; $t(11) = 1.51$, $p = 0.157$

Given the heterogeneity in the neural data observed in our previous analysis, we sought to identify the subset of valence-encoding electrodes that were modulated by reward expectation. To obtain trial-by-trial estimates of reward expectation, we fit a standard-reinforcement learning model to each subject's behavioral data (Sutton and Barto, 1990; Materials and methods section; Table 3). Because distinct item pairs were presented in an interleaved manner, reward expectation estimates were dissociated from time during the task (Fig. 3a). For each valence encoding electrode, we studied the relation between HFA and reward expectation, separately following positive and negative feedback, using the following regression model :$Y = \beta_0 + \beta_Q Q + \beta_t T$. Here, Y is a vector containing HFA values, Q is a vector containing expectation values, and T tracked number of times a given item pair had been previously presented in order to account for any novelty-related changes in HFA. We considered an electrode to show an expectation-related effect if there was a significant β_Q coefficient (t-statistic, $p < 0.05$) associated with HFA following positive or negative feedback. This linear model, that was applied to each electrode's neural data, assumes the following: 1) HFA (Y) demonstrates a linear with each independent variable (Q and T), 2) each trial provides an independent observation of behavioral and neural data, 3) homoscedastic error distributions associated with each independent variable, and 4) normality of the error distribution.

Because of the previous neural descriptions of learning-related feedback signals (Bayer and Glimcher, 2007; Bromberg-Martin et al., 2010), we did not have any a priori hypotheses regarding the specific relation between HFA and reward expectation. We refer to electrodes that demonstrated any expectation-related modulation of post-reward or post-penalty HFA as "putative learning electrodes". We identified 433 putative learning electrodes (32.9% of valence-encoding electrodes), a more frequent occurrence than expected by chance (counts t-test, $t(25) = 6.10$, $p < 0.001$; false-positive rate = 10%). Two example putative learning electrodes are shown in Fig. 3b.

To characterize the anatomical distribution of putative learning electrodes, we studied the proportion of valence-encoding electrodes that were modulated by reward expectation in several ROIs (Fig. 4a). We only included regions in which we identified valence-encoding electrodes from at least five subjects (after filtering data based on our learning criteria). We found that putative learning electrodes were more frequently observed than expected by chance in several ROIs (Table 4, counts t-test, FDR-corrected $p < 0.05$). In addition to prefrontal regions, where they have previously been described, we also frequently observed putative learning electrodes in occipital, temporal and parietal regions, where they have rarely been described. We observed a trend towards these signals occuring more frequently in the right hemisphere than in the left hemisphere ($t(17) = -1.89$, $p = 0.076$). Thus, putative learning electrodes were widely distributed throughout the human brain and showed a trend towards greater prominence in the right hemisphere.

Having characterized the anatomical properties of putative learning electrodes, we sought to study their functional properties. Particularly, we wanted to study the manner in which HFA at valence-encoding electrodes was modulated by reward expectation, in order to shed light on the manner in which these neural signals integrate information about valence and reward expectation. Because previous monkey single-unit studies have shown that cortical neurons frequently encode unexpected outcomes with increases in firing rate (Asaad and Eskandar, 2011), one might expect to frequently observe post-reward HFA and post-penalty HFA to demonstrate opposing relations with reward expectation. Post-reward HFA should demonstrate a negative relation with reward expectation, indicating that HFA is greater when reward expectation is low (unexpected rewards), compared to when reward expectation is high (expected rewards). In contrast, post-penalty HFA should show a positive relation with reward expectation, indicating that HFA is greater when reward expectation is high (unexpected penalties), compared to when it is low (expected penalties). Consistent with this view, we found that post-reward HFA more frequently showed a negative relation with reward expectation (n = 222, 16.8%) than a positive relation with reward expectation (n = 59, 4.49%, counts t-test, $t(25) = 3.12$, $p = 0.004$), whereas post-penalty HFA more frequently showed a positive relation with reward expectation (n = 130, 9.89%t, $t(25) = 2.35$, $p = 0.027$, Fig. 4b). Thus, the most common patterns of expectation-related modulations in HFA were consistent with representations of unexpected rewards and penalties ("UR" and "UP," respectively). We observed minimal overlap between these groups of electrodes as only 1.7% of valence-encoding electrodes demonstrated both patterns of activity. We observed UR electrodes more frequently than expected by chance in several right hemisphere ROIs (occipital, fusiform, temporal, and ventrolateral prefrontal; FDR-corrected $p < 0.05$), and a trend towards this effect in the left temporal and right sensorimotor ROIs (uncorrected $p < 0.05$). We observed trends towards observing UP electrodes more frequently than expected by chance in the right temporal, parietal, and sensorimotor ROIs (uncorrected $p < 0.05$).

If UR and UP electrodes reflect neural signals that guide learning, one might expect to observe a correlation between the strength of expectation-related changes in these electrodes and subjects' learning during the task. To measure the strength of these signals in each subject, we averaged the t-statistics associated with post-reward β_Q among all UR electrodes and the post-penalty β_Q among all UP electrodes in that subject, respectively. To index learning during the task, we computed the mean tendency that each subject showed towards choosing the high-probability item during the last 10 trials of each item pair ("accuracy"). Across subjects, we observed a significant correlation between accuracy and the strength of UR representations ($r = 0.65$, $p < 0.001$, Fig. 4c), but did not observe such a correlation with UP representations ($p > 0.5$). These results demonstrate that the strength of UR neural signals was correlated with subjects' learning during the task, suggesting that these electrodes reflect neural processes that are functionally relevant for learning.

Table 3
Summary of Q model fits. Mean (± s.e.m. across subjects) shown for best-fitting parameter values and goodness-of-fit measures (see Materials and methods section).

α	β	Mean sq. error	Mean sq. error (null)
.20 (±0.04)	0.23 (±0.04)	0.14 (±0.01)	0.26 (±0.01)

Discussion

By measuring intracranially-recorded high-frequency activity (HFA) as neurosurgical patients performed a two-alternative probability learning task, we found a significant number of electrode sites for which HFA distinguished between rewards and penalties. The broad anatomical distribution of valence-encoding electrodes is consistent with the findings of a recent fMRI study that used multi-voxel pattern analysis to decode outcome valence from almost all human brain regions (Vickery et al., 2011). In most brain regions sampled, we observed strongly heterogeneous responses, with a mixture of recording sites exhibiting relative HFA increases following rewards and other recording sites exhibiting relative HFA increases following penalties. Because HFA is thought to reflect the summed activity of a large population of local neurons (Nir et al., 2007; Ray et al., 2008; Miller, 2010; Burke et al., 2015), our results suggest that neuronal populations in most human brain regions encode outcome valence in a heterogeneous manner (i.e., some neurons show relative increases following rewards, whereas others show relative increases following penalties). Such heterogeneous encoding patterns have been demonstrated in cortical regions by several monkey single-unit studies during reinforcement learning (Matsumoto et al., 2007; Asaad and Eskandar, 2011), and have been suggested as a reason why univariate functional neuroimaging studies may not be able to detect many cognitive signals when averaging activity within brain regions (Wallis and Kennerley, 2011). Such

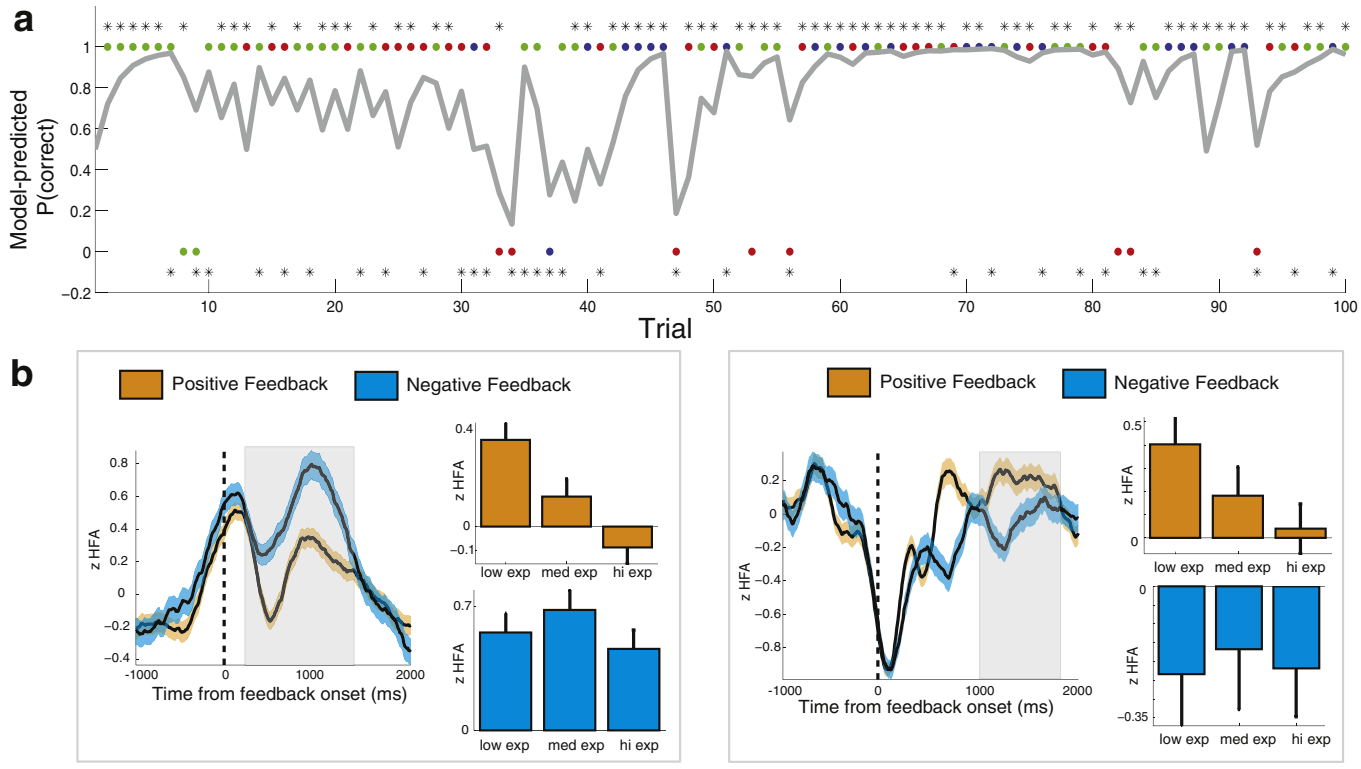

Fig. 3. Relating neural activity to reward expectation. a. Behavioral data from one example session. On the top of the figure, dots indicate when the subject chose the high-probability item. Color of the dots indicate the item pair that was presented (blue — 80/20, green — 70/30, red — 60/40). Asterisks indicate when positive feedback was provided following each choice. Bottom of the figure, dots indicate when the subject chose the low-probability item (same color scheme as a), whereas asterisks indicate when negative feedback was provided following each choice. Gray line indicates model-predictions of subjects' choices. b. Two example valence-encoding electrodes recorded from this subject that showed expectation-related changes in activity. Mean HFA response following positive (orange) and negative (blue) outcomes. Width indicates s.e.m. across trials. Shaded box indicates the time during which we observed significant valence-related HFA differences based on our cluster-based permutation procedure. During this time interval, we studied post-reward and post-penalty changes in HFA during terciles of reward expectation using a regression framework (see main text).

heterogeneous neural patterns may also explain recording sites that demonstrated relative reward- and penalty-related increases in HFA during distinct time intervals (Fig. 2).

Theories of reinforcement learning posit that individuals learn by encoding reward prediction errors that result in greater learning following unexpected outcomes than following expected outcomes (Bush and Mosteller, 1951; Rescorla and Wagner, 1972; Sutton and Barto, 1990). To assess whether these broadly distributed valence signals were related to reinforcement learning, we assessed the degree to which they were modulated by reward expectation. At each valence-encoding electrode, we correlated HFA during the time interval that HFA distinguished between rewards and penalties to trial-by-trial estimates of reward expectation (by applying a regression framework that controlled for variation in time on task and stimulus novelty). To obtain reliable trial-by-trial estimates of reward expectation, we only included behavioral and neural data from stimulus pairs for which subjects demonstrated evidence of learning. We wanted to test the hypothesis that electrodes that encoded both outcome valence and reward expectation (putative learning electrodes) were widely distributed throughout the brain.

We found that reward expectation reliably modulated valence signals in several regions of interest, including those in prefrontal, sensori-motor, parietal, temporal, and occipital cortices (Table 4). Whereas functional neuroimaging studies have primarily identified neural populations that encode putative learning signals in brain regions that receive prominent inputs from dopaminergic neurons (e.g., ventral striatum, medial prefrontal and orbitofrontal cortices; Berns et al., 2001; McClure et al., 2003; Pessiglione et al., 2006; Rutledge et al., 2010), we observed putative learning signals both in regions that receive prominent DA inputs (e.g., lateral prefrontal regions, and trends towards significance in medial and orbitofrontal cortices), and those that receive only sparse inputs from midbrain DA neurons (e.g., parietal, temporal, and occipital regions; Haber and Knutson, 2009). We did not observe a significant frequency of putative learning electrodes in the medial temporal lobe (recently linked to reinforcement learning; Foerde and Shohamy, 2011), however, this may be due to reduced power due to relatively low electrode counts. These results provide electrophysiological support for the emerging view that reinforcement learning is driven by widespread learning signals throughout the human brain.

Our results also shed light on the manner in which the brain encodes learning signals. At electrode sites that encoded learning signals, we found that post-reward HFA typically showed a negative relation with reward expectation (indicating greater HFA following unexpected compared to expected rewards), whereas post-penalty HFA typically showed a positive relation with reward expectation (indicating greater HFA following unexpected penalties to expected penalties). These results suggest that neural populations in the human brain typically encode unexpected outcomes with increases in firing rate, an encoding scheme that has been commonly demonstrated in cortical neural populations by monkey single-unit studies (Matsumoto et al., 2007; Asaad and Eskandar, 2011; Wallis and Kennerley, 2011). Because we typically observed representations of unexpected rewards and penalties on distinct electrodes, our results suggest that the brain may adopt a distributed and opponent-encoding scheme to represent unexpected outcomes — some neural populations encode unexpected rewards with increases in firing rate, whereas other populations encode unexpected penalties with increases in firing rate. Such an encoding scheme might emerge if unexpected rewards and penalty representations are generated by distinct neural systems (Daw et al., 2002). In contrast to neural

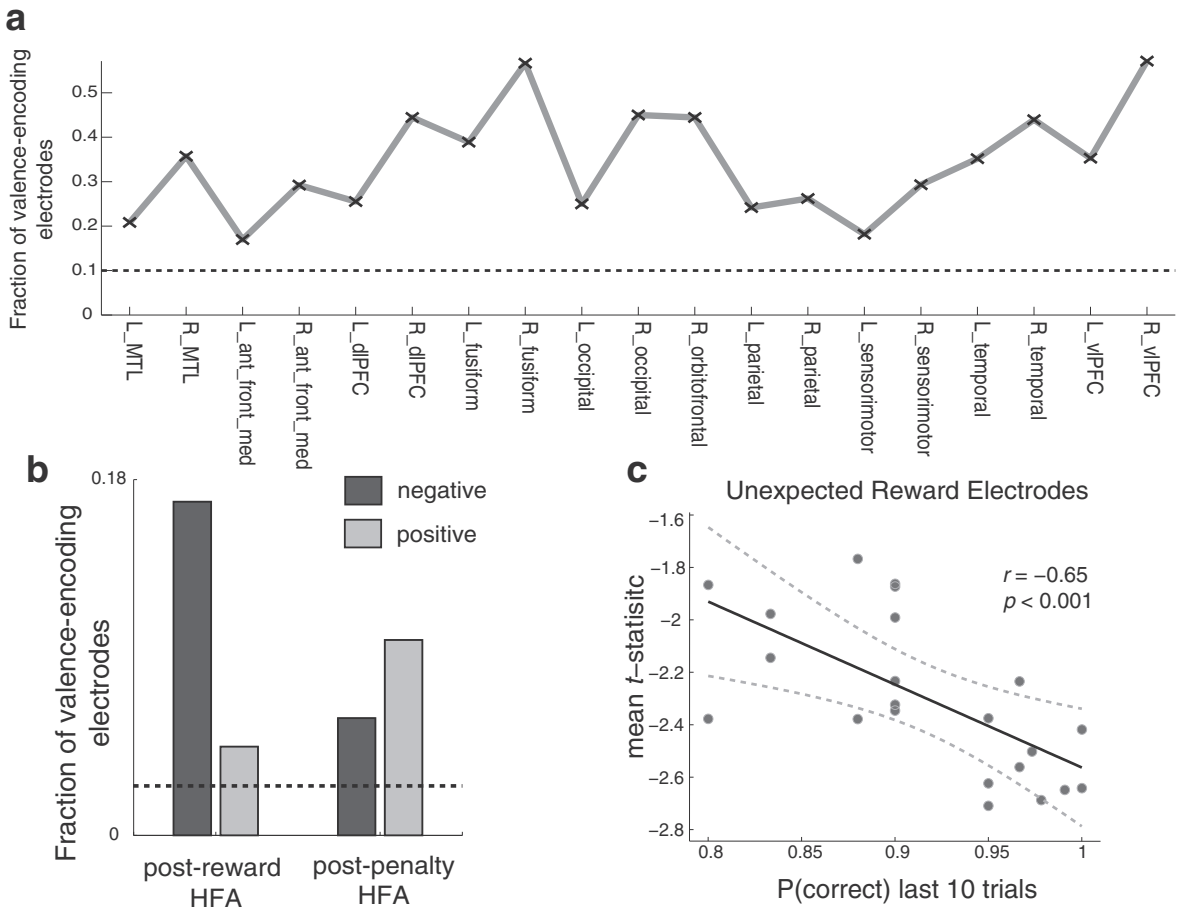

Fig. 4. Expectancy-related changes in activity among valence-encoding electrodes. a. Anatomical distribution of expectation-modulated valence electrodes. In several ROIs, we show the fraction of valence-encoding electrodes that were modulated by reward expectation. We only included regions from which we observed valence-encoding electrodes from at least five subjects. b. Patterns of HFA relations with reward expectation. Post-reward HFA most frequently showed a negative relation with reward expectation ("UR electrodes"), whereas post-penalty HFA most frequently showed a positive relation with reward expectation ("UP electrodes"). Dashed horizontal line indicates the false-positive rate. See main text for statistics. c. Correlating the signal strength of UR electrodes with subjects' behavioral performance (accuracy during last 10 trials of an item pair). Black line indicates regression line and dashed gray lines indicate 95% confidence intervals associated with the regression line. See main text for statistics.

Table 4
Frequency of expectation-modulated valence-encoding electrodes. For regions in which we observed valence encoding electrodes in at least 5 subjects, we list the number of valence-encoding electrodes (column 1), number of subjects in which at least one valence-encoding electrode was observed (column 2), and the frequency that they were modulated by reward expectation. Positive *t*-statistics indicate frequencies that are greater than expected, whereas negative *t*-statistics indicate frequencies that are lower than expected. Bold texts in column 3 indicate regions that showed valence-encoding electrodes more frequently than expected by chance (FDR-corrected $p < 0.05$).

Region of interest	Number of valence-encoding electrodes	Number of subjects	Frequency of expectation-modulated electrodes; counts *t*-test results
R. OFC	18	9	0.44; $t(8) = 2.05$, $p = 0.075$
L. dlPFC	47	10	0.26; $t(9) = 2.02$, $p = 0.075$
R. dlPFC	126	12	**0.44**; $t(11) = 3.35$, $p = 0.006$
L. vlPFC	17	8	0.35; $t(7) = 1.87$, $p = 0.103$
R. vlPFC	28	9	**0.57**; $t(8) = 3.41$, $p = 0.009$
L. anterior medial frontal	53	9	0.17; $t(8) = 1.69$, $p = 0.130$
R. anterior medial frontal	65	10	0.29; $t(9) = 2.55$, $p = 0.031$
L. sensorimotor	77	11	**0.18**; $t(10) = 3.62$, $p = 0.005$
R. sensorimotor	116	12	**0.29**; $t(11) = 3.22$, $p = 0.008$
L. parietal	91	14	0.24; $t(13) = 2.37$, $p = 0.034$
R. parietal	103	12	**0.26**; $t(11) = 3.32$, $p = 0.007$
L. temporal	162	17	**0.35**; $t(16) = 4.83$, $p < 0.001$
R. temporal	132	17	**0.44**; $t(16) = 4.22$, $p < 0.001$
L. fusiform	18	10	0.39; $t(9) = 2.50$, $p = 0.034$
R. fusiform	30	10	**0.57**; $t(9) = 2.91$, $p = 0.017$
L. occipital	56	12	0.25; $t(11) = 1.40$, $p = 0.190$
R. occipital	40	10	**0.45**; $t(9) = 3.26$, $p = 0.009$
L. MTL	24	5	0.21; $t(4) = 1.56$, $p = 0.157$
R. MTL	14	9	0.36; $t(8) = 1.57$, $p = 0.192$

populations in the cortex, those in deep brain structures such as the midbrain dopaminergic nuclei have been shown to demonstrate more homogenous representations of unexpected outcomes (Schultz et al., 1997; Bayer and Glimcher, 2005; Bromberg-Martin et al., 2010; Glimcher, 2011; although, see Matsumoto and Hikosaka, 2009). Human electrophysiology studies in deeper brain structures have largely been consistent with these studies (Zaghloul et al., 2009; Patel et al., 2012; Lega et al., 2011; Ramayya et al., 2014b).

If electrodes encoding unexpected rewards and penalties reflect neural processes that are functionally related to learning, one might expect to observe a relation between the strength of these neural signals and subjects' behavioral performance during the task. We assessed whether there was a correlation between subjects' frequency of selecting the high reward probability item during the last 10 presentations of an item pair ("accuracy;" a measure of how well they encoded stimulus–reward associations), and the strength of unexpected reward and penalty signals, respectively. We found that the strength of unexpected reward representations was positively correlated with subjects' accuracy during the task. One interpretation of this result is that subjects who performed better during the task were better able to neurally represent unexpected rewards and penalties. Alternatively, it may be the case that we were better able to measure unexpected reward signals in subjects who performed well. In either case, this result provides evidence that widespread unexpected reward signals are functionally related to reinforcement learning. We did not observe a significant correlation between behavioral performance and the strength of unexpected penalty representations, however, this may reflect inadequate power as we observed fewer unexpected penalty than unexpected reward representations across the dataset.

Limitations

First, a subset of identified valence-encoding signals may reflect perceptual differences between reward and penalty feedback conditions (e.g., green vs. red screen, and cash-register vs. error tone). However, the widespread nature of these signals and their relation with reward expectation argue against this view (for a control analysis, see Supplementary material). Second, our analysis framework identifies putative learning signals by assessing the relation between valence-encoding neural signals and reward expectation. We are unable to assess whether these neural signals specifically represent reward prediction errors (Glimcher, 2011) because it is difficult to rule out the contribution of neural populations that encode other cognitive signals that may mimic reward prediction errors (e.g., salience; Pearce and Hall, 1980). Future studies may mitigate this issue by experimentally manipulating reward magnitude in addition to reward probability so as to apply more rigorous tests of specific reinforcement learning signals (e.g., reward prediction errors vs. salience; Rutledge et al., 2010).

Conclusions

Neural processes that encode both outcome valence and reward expectation were widely distributed throughout the human brain, and commonly observed in regions that receive sparse inputs from midbrain dopaminergic neurons (e.g., temporal, parietal, occipital). These neural processes typically showed increased activity following unexpected outcomes, as compared with expected outcomes, an encoding scheme which is consistent with previous findings from monkey single-unit studies (Asaad and Eskandar, 2011). The strength of neural processes that encoded unexpected rewards was correlated with behavioral performance during the task, suggesting a functional relevance for reinforcement learning. Our findings lend further support to the emerging view of reinforcement learning as a highly distributed brain function.

Acknowledgment

This work was supported in part by NIH Grants MH55687 and F30MH102030. We are grateful to the patients who selflessly volunteered in this study. We thank Drs. Michael Sperling, Ashwini Sharan, James Evans, Timothy Lucas and Kathryn Davis for patient recruitment. We thank John F. Burke and Ryan B. Williams, Dale H. Wyeth and Edmund Wyeth for technical assistance; Josh Gold, Kareem Zaghloul, Matt Nassar, Joe McGuire, and Katherine Hurley for insightful comments on this manuscript; and Nicole Long, Maxwell Merkow, Karl Healey for general discussion.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2015.04.037.

References

Addison, P.S., 2002. The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance. Institute of Physics Publishing, Bristol.
Asaad, W., Eskandar, E., 2011. Encoding of both positive and negative reward prediction errors by neurons of the primate lateral prefrontal cortex and caudate nucleus. J. Neurosci. 31 (49), 17772–17787.
Bartra, O., McGuire, J., Kable, J., 2013. The valuation system: a coordinate-based meta-analysis of bold fMRI experiments examining the neural correlates of subjective value. NeuroImage 76, 412–427.
Bayer, H., Glimcher, P., 2005. Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 47, 129–141.
Bayer, H., Glimcher, P., 2007. Statistics of midbrain dopaminergic neuron spike trains in the awake primate. J. Neurophysiol. 98 (3), 1428–1439.
Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57, 289–300.
Berns, G.S., McClure, S.M., Pagnoni, G., Montague, P., 2001. Predictability modulates human brain response to reward. J. Neurosci. 21 (8), 2793–2798.
Bouchard, K.E., Mesgarani, N., 2013a. Functional organization of human sensorimotor cortex for speech articulationNature 495, 327–332. http://www.nature.com/nature/journal/v495/n7441/abs/nature11911.html.
Bouchard, K.E., Mesgarani, N., Johnson, K., Chang, E.F., 2013b. Functional organization of human sensorimotor cortex for speech articulation. Nature 498 (7455).
Bromberg-Martin, E., Matsumoto, M., Hikosaka, O., 2010. Distinct tonic and phasic anticipatory activity in lateral habenula and dopamine neurons. Neuron 67 (1), 144–155.
Burke, J.F., Zaghloul, K.A., Jacobs, J., Williams, R.B., Sperling, M.R., Sharan, A.D., Kahana, M.J., 2013. Synchronous and asynchronous theta and gamma activity during episodic memory formation. J. Neurosci. 33 (1), 292–304.
Burke, J.F., Long, N.M., Zaghloul, K.A., Sharan, A.D., Sperling, M.R., Kahana, M.J., 2014. Human intracranial high-frequency activity maps episodic memory formation in space and time. NeuroImage 85 (Pt. 2), 834–843.
Burke, J.F., Merkow, M., Jacobs, J., Kahana, M.J., Zaghloul, K., 2015. Brain computer interface to enhance episodic memory in human participants. Front. Hum. Neurosci. 8.
Bush, R.R., Mosteller, F., 1951. A model for stimulus generalization and discrimination. Psychol. Rev. 58 (6), 413.
Buzsaki, G., Anastassiou, C., Koch, C., 2012. The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. Nat. Rev. Neurosci. 13, 407–419.
Dale, A.M., Fischl, B., Sereno, M., 1999. Cortical surface-based analysis I: segmentation and surface reconstruction. NeuroImage 9 (2), 179–194.
Daw, N., Kakade, S., Dayan, P., 2002. Opponent interactions between serotonin and dopamine. Neural Netw. 15 (4–6), 603–616.
Daw, N., O'Doherty, J., Dayan, P., Seymour, B., Dolan, R., 2006. Cortical substrates for exploratory decisions in humans. Nature 441, 876–879.
Desikan, R., Segonne, B., Fischl, B., Quinn, B., Dickerson, B., Blacker, D., Buckner, R.L., Dale, A., Maguire, A., Hyman, B., Albert, M., Killiany, N., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage 31 (3), 968–980.
Fischl, B., Sereno, M., Tootell, R., Dale, A.M., 1999. High-resolution inter-subject averaging and a coordinate system for the cortical surface. Hum. Brain Mapp. 8, 272–284.
Foerde, K., Shohamy, D., 2011. Feedback timing modulates brain systems for learning in humans. J. Neurosci. 31 (37), 13157–13167.
Frank, M.J., Seeberger, L.C., O'Reilly, R.C., 2004. By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 306, 1940–1943.
Frank, M., Samanta, J., Moustafa, A., Sherman, S., 2007. Hold your horses: impulsivity, deep brain stimulation, and medication in parkinsonism. Science 318, 1309–1312.
Glimcher, P., 2011. Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. Proc. Natl. Acad. Sci. U. S. A. 108 (3), 15647–15654.
Haber, S., Knutson, B., 2009. The reward circuit: linking primate anatomy and human imaging. Neuropsychopharmacology 35 (1), 4–26.
Kahnt, T., Heinzle, J., Park, S., Haynes, J., 2011. Decoding the formation of reward predictions across learning. J. Neurosci. 31 (41), 14624–14630.

Lega, B.C., Kahana, M.J., Jaggi, J.L., Baltuch, G.H., Zaghloul, K.A., 2011. Neuronal and oscillatory activity during reward processing in the human ventral striatum. NeuroReport 22 (16), 795–800.
Maldjian, J.A., Laurienti, P.J., Kraft, R.A., Burdette, J.H., 2003. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. NeuroImage 19 (3), 1233–1239.
Manning, J.R., Jacobs, J., Fried, I., Kahana, M.J., 2009. Broadband shifts in LFP power spectra are correlated with single-neuron spiking in humans. J. Neurosci. 29 (43), 13613–13620.
Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 164, 177–190.
Matsumoto, M., Hikosaka, O., 2009. Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature 459 (11), 837–841.
Matsumoto, M., Matsumoto, K., Abe, H., Tanaka, K., 2007. Medial prefrontal cell activity signaling prediction errors of action values. Nat. Neurosci. 10 (5), 647–656.
McClure, S.M., Berns, G.S., Montague, P.R., 2003. Temporal prediction errors in a passive learning task activate human striatum. Neuron 38 (2), 339–346.
Miller, K.J., 2010. Broadband spectral change: evidence for a macroscale correlate of population firing rate? J. Neurosci. 30 (19), 6477–6479.
Montague, P., King-Casas, B., Cohen, J.D., 2006. Imaging valuation models in human choice. Annu. Rev. Neurosci. 29, 417–448.
Nir, Y., Fisch, L., Mukamel, R., Gelbard-Sagiv, H., Arieli, A., Fried, I., Malach, R., 2007. Coupling between neuronal firing rate, gamma LFP, and BOLD fMRI is related to interneuronal correlations. Curr. Biol. 17 (15), 1275–1285.
Patel, S., Sheth, S., Gale, J.T., Greenberg, B., Dougherty, D., Eskandar, E.N., 2012. Single-neuron responses in the human nucleus accumbens during a financial decision-making task. J. Neurosci. 32 (21), 7311–7315.
Pearce, J., Hall, G., 1980. A model for Pavlovian conditioning: variations in the effectiveness of conditioned but not of unconditioned stimuli. Psychol. Rev. 87, 532–555.
Pessiglione, M., Seymour, B., Flandin, G., Dolan, R.J., Frith, C., 2006. Dopamine-dependent prediction errors underpin reward-seeking behavior in humans. Nature 442, 1042–1045.
Ramayya, A.G., Misra, A., Baltuch, G.H., Kahana, M.J., 2014a. Microstimulation of the human substantia nigra following feedback alters reinforcement learning. J. Neurosci. 34 (20), 6887–6895.
Ramayya, A.G., Zaghloul, K.A., Weidemann, C.T., Baltuch, G.H., Kahana, M.J., 2014b. Electrophysiological evidence for functionally distinct neuronal populations in the human substantia nigra. Front. Hum. Neurosci. 8, 1–9.
Ray, S., Maunsell, J., 2011. Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. PLoS Biol. 9 (4), e1000610.
Ray, S., Crone, N., Niebur, E., Franaszczuk, P., Hsiao, S., 2008. Neural correlates of high-gamma oscillations (60–200 Hz) in macaque local field potentials and their potential implications in electrocorticography. J. Neurosci. 28 (45), 11526.
Rescorla, R., Wagner, A., 1972. A theory of pavolvian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Black, A., Prokasy, W. (Eds.), Classical Conditioning II: Current Research and Theory. Appleton Century Crofts, New York, pp. 64–99.
Reynolds, J., Hyland, B., Wickens, J., 2001. A cellular mechanism of reward-related learning. Nature 413, 67–70.
Rutledge, R., Dean, M., Caplin, A., Glimcher, P., 2010. Testing the reward prediction error hypothesis with an axiomatic model. J. Neurosci. 30 (40), 13525–13536.
Schultz, W., Dayan, P., Montague, P.R., 1997. A neural substrate of prediction and reward. Science 275, 1593–1599.
Sutton, R., Barto, A., 1990. Time-derivative models of Pavlovian reinforcement. In: Gabriel, M., Moore, J. (Eds.), Learning and Computational Neuroscience: Foundations of Adaptive Networks. MIT Press, Cambridge, MA, pp. 497–537.
Vickery, T., Chun, M., Lee, D., 2011. Ubiquity and specificity of reinforcement signals throughout the human brain. Neuron 72 (1), 166–177.
Wallis, J.D., Kennerley, S., 2011. Contrasting roles of reward signals in the orbitofrontal and anterior cingulate cortex. Ann. N. Y. Acad. Sci. 1239, 33–42.
Zaghloul, K.A., Blanco, J.A., Weidemann, C.T., McGill, K., Jaggi, J.L., Baltuch, G.H., Kahana, M.J., 2009. Human substantia nigra neurons encode unexpected financial rewards. Science 323, 1496–1499.
Zaghloul, K.A., Lega, B.C., Weidemann, C.T., Jaggi, J.L., Baltuch, G.H., Kahana, M.J., 2012. Neuronal activity in the human subthalamic nucleus encodes decision conflict during action selection. J. Neurosci. 32 (7), 2453–2460.