

Categories of rewards.

Erik J. Peterson
Dept. of Psychology
Colorado State University
Fort Collins, CO

Introduction

This introduction is of four parts. First I discuss “classical” rewards and the brain. Second I begin making a case for cognitive rewards, discussing as an example novelty, then moving onto other examples and finally arguing for the utility of flexible (synonymous here with categorical or generalizable or inferable) reward representations in the cognitive case as compared to simple exemplar representations (e.g. say a sip of raspberry juice taken after a long run on the afternoon of July the 2nd 1982 compared to the more general category of “drinking fruit juice”) . Third is a discussion of prior studies of instrumental generalization (in pigeons) and multi-valued decision making (in humans) - both of which can be viewed as evidence for categorical reward representations. I then outline the specific goals and methods of this proposed work. Finally is a diversion discussing other (non-rewarding) theories of phasic dopamine.

The classics.

Classically rewards and reinforcers have been linked to (or simply were) food (O’Doherty, Buchanan, Seymour, & Dolan, 2006), pain (Becerra & Borsook, n.d.; Schultz, 2007) and presumably sex though for, err, logistical reasons this is not often used in the laboratory; however physical contact with a loved one has been shown to activate the reward circuitry (Izuma, Saito, & Sadato, 2008; Fließbach et al., 2007)). These rewards are certainly potent, being used for over 50 successful years in studying learning in animal models (Iversen & Iversen, 2007) and people (H. Kim, Shimojo, & O’Doherty, 2010; Montague, King-Casas, & Cohen, 2006). However these kinds of rewards are not the only event that can lead to phasic firing the in

the VTA/SNc accompanied by activity changes in the striatum (i.e. activation of the reward circuitry). Novelty is an example.

It's novel

Anything in principle, and so far I am aware in practice, can be novel (i.e. contextually unexpected) but if anything can be novel and novelty is some kind of reward, this implies that every new experience is somehow (instantly) linked to another primary or secondary reinforcer (i.e. reward Björklund & Dunnett, 2007). This would require a flexible rapid abstract remapping of the novel experience to a previously learned reward. However it is difficult to see which reward episode would be used. It could be any really, say a sip of raspberry juice taken after a long run on the afternoon of July the 2nd 1982; however, memory is not nearly so precise and this seems unlikely. Perhaps instead the brain searches the current context for a relevant rewarding episode to bind the novelty to? If this is case, then does similarity of the current situation to the past affect the reward's value? It is unclear. In any case though it would be simpler if novelty were conceptualized as a cognitive reward, one that requires evaluation of current sense experience and past memories, but is ultimately derived endogenously. Reviewing again the literature with these types of cognitive rewards in mind led to other examples (see *Introduction*). In fact, that are more than enough examples to assert that cognitive rewards are fact. So then I began to wonder what other under-appreciated or undiscovered complexities reward representations might possess. Cognitive rewards could certainly be more flexible than evolutionarily primitive rewards; perhaps instead of being coded as individual exemplars (e.g., each tasty sip of juice is a wholly independent neural entity), rewards are actually represented as categories. That is while rewards can facilitate category learning perhaps, quasi-paradoxically, they are categories themselves....

Thinking about other rewards

Tricomi & Fiez, 2008, showed ventral striatum BOLD signal changes in a declarative memory task in which subjects were initially trained with feedback ("Right" or "Wrong") to distinguish 60 correct from incorrect word pairs. In the subsequent two rounds explicit feedback was withheld but activity in the caudate was observed when correct pairings were matched based on memory alone. Correct matches, that is goal achievement, led to strong activity. In two economic decision making tasks strong ventral striatum signals were observed when participants were required merely to imagine or consider alternative outcomes (Hayden, Pearson, & Platt, 2009;

Lohrenz, McCabe, Camerer, & Montague, 2007). Information about the future is rewarding as well; Bromberg-Martin & Hikosaka, 2009, showed that complex visual clues about an upcoming outcome were in themselves sufficient to cause bursts of firing in the VTA/SNc. Even the relatively simple cases of temporal discounting of rewards and the assessment of their uncertainty likely requires cognitive intervention, which is reflected in several reports of complex, multi-valued, reward-related signals in both dorsal and ventral-medial prefrontal cortices (Tobler, Christopoulos, O'Doherty, Dolan, & Schultz, 2009; Wallis & Kennerley, 2010; S. Kim, Hwang, Seo, & Lee, 2009; Seymour & McClure, 2008).

Given that (some) reward representations are mediated by complex cognition, notably including goal-achievement and imagined outcomes, then it is quite possible that these same reward representations may have a categorical (i.e. generalizable) aspect. Pigeons have long been used to study perceptual categorization in rewarding contexts, including many studies where the effect of altering a previously conditioned stimuli was assessed. For example, Guttman... , 1956, varied a preconditioned 570 nm light from 480-610, showing that while the bird's pecking rate (i.e. response vigor) decreased as one moved farther from 570, the birds still responded, that they is generalized. When novel variations of conditioned stimuli from two sensory modalities were mixed similar graded changes in vigor were observed. However, not all combinations were effective; some combinations produced no responses at all (Blough, 2001; Simmons... , 2008; Urcuioli, 2001). Unfortunately though much of this behavioral work has never been examined in other model systems nor with modern neural recording and imaging technologies. One of the few (perhaps the only) fMRI studies to examine conditioned stimuli in new contexts was that of Kahnt, Heinzle, Park, & Haynes, 2010. They first trained participants on the independent values of several colors, shapes and patterns (e.g. a white diamond, a green background, and a set of leftward moving dots). Then they presented a pair of distinct combinations of the initial stimuli. The participants then had to select the most valuable based on each stimuli's combined value. By employing machine learning methods they showed that the combined value was encoded in the ventro-medial PFC while the variation in value among the combined stimuli of was encoded in the dorsolateral prefrontal cortex (dlPFC). Unfortunately they did not examine activity in either the striatum or VTA/SNc. It has however been shown that when a novel combination of previously studied options is presented to rats in a stimulus-response learning task the resulting dopaminergic firing correlates with the better option, indicating generalization of reward knowledge to the new context (Roesch, Calu, & Schoenbaum, 2007).

Goals

Finally, the goals of this proposal are two fold. Goal one (which is arguably complete, see *Behavioral Results*) is to establish that categorical reward representations can be used to infer rewarding properties for never before seen exemplars and mediate successful learning. That is, to establish whether reward categories are generalizable, in the proposed task at least. Goal two will be approached from two directions. Overall the goal is to examine the computational underpinnings of reward inference: is reward value assigned as a function of category alone or instead does degree of similarity of an exemplar to the category prototype affect its value? Said concretely, if a participant is trained on a perceptual category consisting of sinusoidal gratings (as is the case here) are gratings from the category “+\$1” equivalent irrespective of their distance from the category center or does the value of a grating change depending on classification certainty (i.e. distance from center). In the first approach, the fit of two reinforcement learning models to BOLD data from the dopaminergic midbrain and ventral striatum will be compared. Model 1 will use binomial (constant) rewards. In model 2 reward value will be diminished as similarity to the category prototype decreases. The second approach will be to examine how the RPEs from the two Rescorla-Wagner models mediate the coupling between the (dorsal-lateral and ventral-medial) PFC regions of interest and the ventral and dorsal striatum. Recent electrophysiological and fMRI studies have suggested that reinforcement value is calculated (and perhaps) stored in the aforementioned areas of the PFC (Daw, Gershman, Seymour, Dayan, & Dolan, 2011; Bornstein & Daw, 2011; Frank, 2011; O’Reilly, 2010, 2006; O’Reilly & Frank, 2006). The exact underlying functional connectivity of striatal, VTA/SNc, and cortical regions remains uncertain (Daw et al., 2011; Bornstein & Daw, 2011; Frank, 2011). It is well established however that PFC directly influences striatal activity and these regions are coupled or mediated by VTA/SNc. As such several simple (3-node) mediation models, using both constant and similarity adjusted RPEs, will be compared.

A salient diversion

Methods and Analyses

Task

As discussed at the end of the introduction, the experimental procedure consists of two parts or tasks. Depicted in Fig 1. (top), the first is a passive classical con-

ditioning task where participants will learn reward categories by viewing randomly selected (without replacement) black and white sinusoidal gratings followed by “Gain \$1” or “Lose \$1” in, respectively, green or red letters. Reward categories will be derived from a information integration parameter distribution (Fig. 2, borrowed from (Spiering & Ashby, 2008)). The grating is onscreen for 0.5 seconds, followed by 0.5 seconds gap, with the outcome displayed for another 0.5 seconds with a 1 second fixation cross between each trial. Task 1, which will be completed outside the fMRI scanner, lasts just over 6 minutes, and includes 175 stimuli approximately evenly divided among the two categories. Each participant will be instructed to “Attend to the screen in order to learn which types of gratings lead to winning money and which types lead to losses”. The category distribution to value (gain or loss) mapping will be randomized for each participant.

The second task (Fig 1, bottom) is an abstract deterministic category learning task that replaces direct verbal feedback or reward with an appropriate grating from task 1; gratings associated with monetary wins will be used for positive reinforcement, and gratings associated with losses for negative reinforcement. Each trial begins with an abstract black and white “tree” stimuli, which belongs to one of two arbitrarily named categories (“q” or “w”). Subjects will indicate their categorization response by button press using either the right index or middle finger (respectively) on a magnet compatible response box placed on the participant’s thigh. The response window lasts up to 2.5 seconds. During the response period the “tree” remains onscreen. Once either time has elapsed or the participant responds, the “tree” disappears and 1-8 seconds later (as defined by the jitter optimization routine, below) is replaced by a sinusoidal grating. If the response was correct a new, that is never before experienced, exemplar grating from the “gain” distribution is used; if it was incorrect, a new “loss” grating appears. The feedback grating stays onscreen for 1 second and is then immediately replaced by a fixation cross, whose duration is again governed by the jitter optimization routine. Participants will learn to classify 6 “trees” (randomly selected at the start of the experiment out of a pool of 16). Each of the 6 are experienced a total of 40 times for total of 240 trials/scanning session. In this task participants are in part instructed to, “Use what you learned about the rewarding properties of the gratings to try and earn as much money as possible in this portion of the experiment”. Participants will be instructed about both tasks orally by the experimenter using Fig. 1 as a visual aid.

In fMRI (as in time-series signal analysis in general) there is an intrinsic tradeoff between simply detecting that a signal has occurred in the presence of noise and then estimating the timecourse (i.e. shape) of that signal (Dale, 1999; Birn, Cox, &

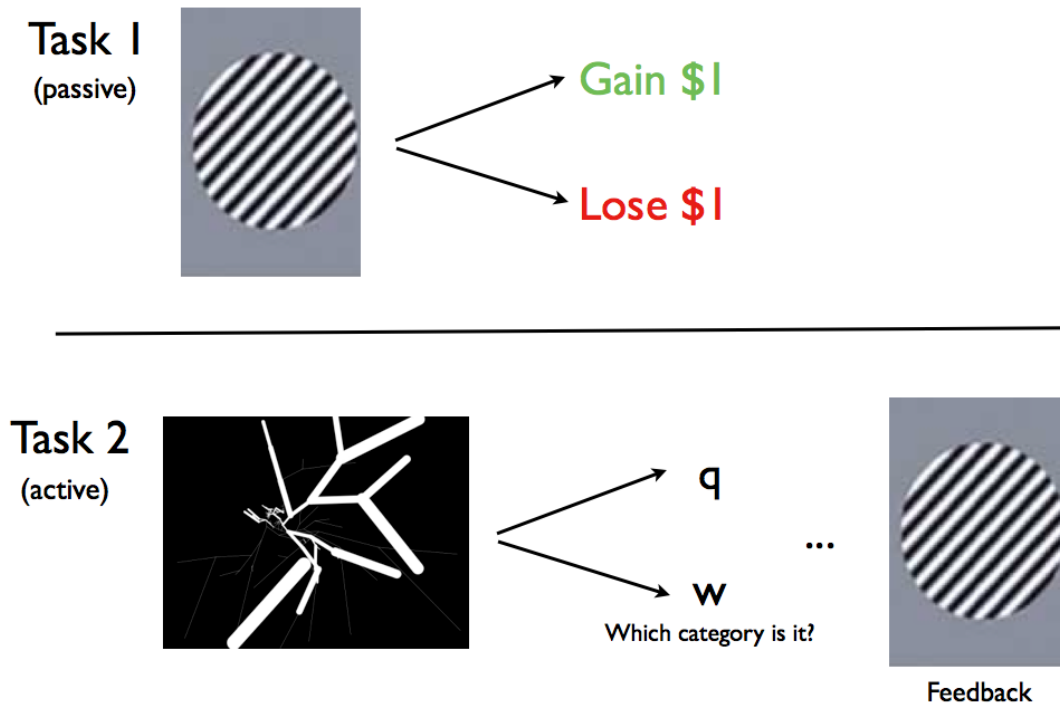


Figure 1. Depiction of the behavioral task. The top is the (passive) classical conditioning participants learn the reward categories. The bottom is the active abstract two-choice category learning.

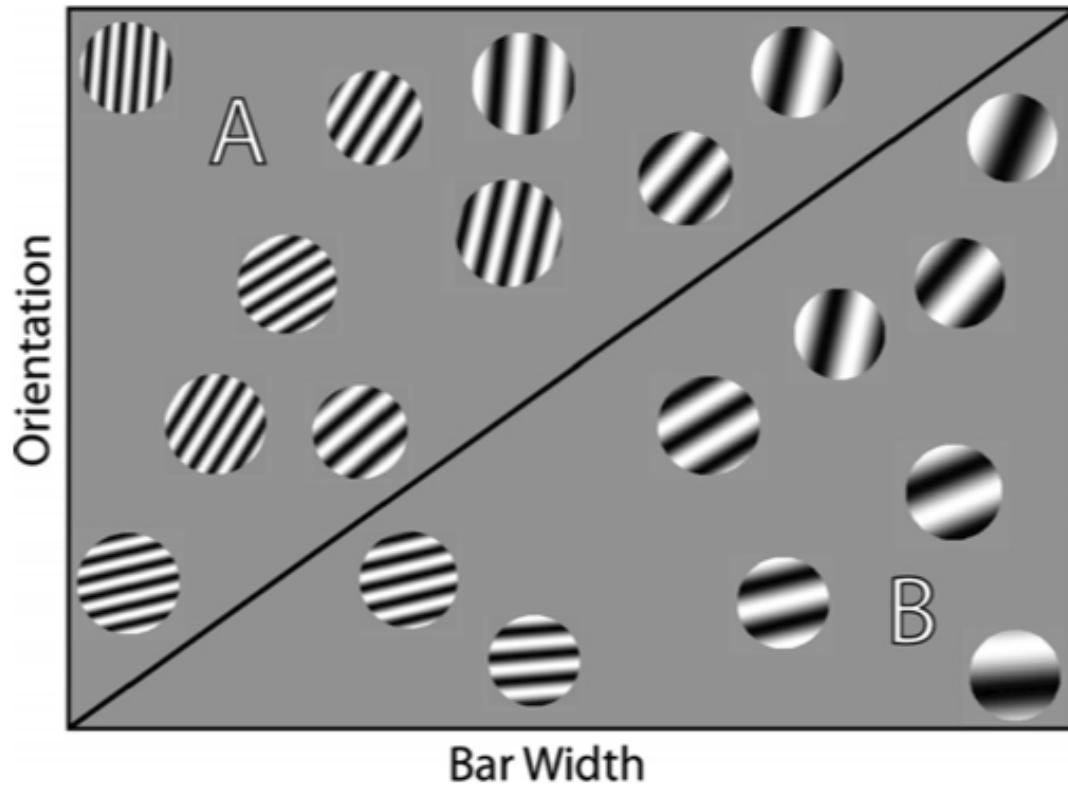


Figure 2. A diagram of sinusoidal grating distributions for an information integration (II) category. As II categories span the diagonal of the gratings parameter space (line width (W) and angle (θ)) successful learning requires consideration of both dimensions preventing participants from solving the categorization problem with simple rule based strategies (e.g. if the lines are wide is category “a”)

Bandettini, 2002; Liu, 2004). The state of the art method for setting trial ordering in an attempt to maximize both signal detection and estimation is a genetic algorithm design by Kao, Mandal, Lazar, & Stufken, 2009. Without extensive modification unfortunately their methodology cannot account for epoch-style designs¹ that this experiment requires; each trial needs to contain stimulus-response, jitter and feedback delivery periods. To account for this, Koa’s methods will be applied twice, once using a 6.5 second ISI (the estimated average length of a trial) and once with a 1 second ISI (the length of the feedback display in task 2). These two designs will then be manually interpolated to create a series of stimulus-response, jitter, feedback, and ITI events. To create the final trial ordering each stimulus-response event will then be subsequently randomly recoded to match one of the 6 “tree” stimuli. This randomization step will be performed independently for each subject so that stimulus assignment to events will be counterbalanced across subjects to control for specific item effects.

Finally, because not all participants can successfully learn simple two-choice category tasks like task 2 (see *Behavioral results* below) subjects will be prescreened with a variation of task 2, using colored fractals as stimuli, “a” and “b” as category labels, and written feedback (e.g. “Correct”). To be included in the study participants must reach 0.65 accuracy (measured with a rolling average) within 60 training examples during prescreening. Prescreening will occur immediately following Task 1 training and immediately preceding the scanning session.

fMRI acquisition and analyses

fMRI data for 12-15 participants will be acquired (approximately half female, age range 18-35, right handed). All participants will be prescreened for the typical exclusion factors (e.g. metal implants, mental disorders, etc) and will be compensated at a base rate of \$15.00 earning up to \$30 more depending on behavioral performance. 30 trials will be randomly selected from the 240 possible and the participant will be paid an additional dollar for every correct response and will lose a dollar for every incorrect response on these 30 trials. Scanning data will be acquired at the Intermountain Neuroimaging Consortium (INC) facility located at the University of Colorado at Boulder. The exact scanning parameters (TR, TE, in-plane resolution, etc) are to be determined as this is the Lab’s first use of the INC machine but will be typical of a whole-brain rapid event-related fMRI acquisition.

¹This flaw is common to all methods of stimulus timing optimization, so far as I am aware.

Standard fMRI data preprocessing (motion correction, slice-time correction, drift correction, temporal, spatial smoothing and Tailarach normalization) will be carried out in BrainVoyager, version 2.1. The main focus of this work will be to test how reward categories are represented and processed in specific regions of interest (ROIs). *A priori* regions of interest are the VTA/SNc, ventral and dorsal striatum, and ventromedial and dorsolateral PFC (the latter two have been show to correlate with reinforcement learning value computation Kahnt et al., 2010). All regions will be defined by anatomical tracings on an average image formed from of all participants' normalized anatomical MRI scans. The regions of interest will be used to compare computational models of categorical reward representations as described in *Computational analyses* below.

Since this is, so far as I am aware, the first time rewards have been treated as categories in an fMRI experiment, whole brain activation mapping is also of interest. This will be done in BrainVoyager with the final maps reported as a set of two overlaid probability maps - heat maps where the value in each voxel is the fraction of participants who had significant activity (defined here as either $p < 0.05$ or $p < 0.01$; RFX correction) at that voxel. This is an improvement over the typical thresholded t-value maps as probability maps include information not just about the strength of the effect but also its consistency. *A priori* whole brain contrasts of interest are all trials compared to fixation (the most powerful contrast which will provide an overall picture of activity patterns), all stimulus-response periods compared to fixation as well as all feedback periods compared to fixation, and feedback period contrasted to stimulus-response period (and the reverse). In addition the effect of outcome valance (gain or loss) will be compared for each of the prior contrasts. Finally probability maps will be qualitatively compared to results from a similar experiment (that used verbal feedback - "correct" or "incorrect") carried out by Lopez-Paniagua, PhD, for his masters work (Lopez-Paniagua & Seger, 2011).

Computational analyses

Reinforcement learning measures will be derived from a set of Rescorla-Wagner models (Eq. 1 and 2.), with decision making approximated by the logistic function (i.e. softmax, Eq. 7) with the parameters (α and β) minimized by maximum log-likelihood. Two separate Rescorla-Wagner models will be considered to examine two different reward representations (Eq. 3 and 4). The first reward representation is all or none, whereas the other incorporates the categorical structure by incorporating a distance parameter, D . D is the Euclidean distance between the average angle (θ)

and width (\bar{W}) in one of the two reward categories and the angle and width of the individual example grating (i.e. θ and w). I will test whether the BOLD signal in the ventral striatum and VTA/SNc is best described by the prediction error term (Eq 2) combined with with Eq. 3 or with Eq. 4. Similarly BOLD changes in the dorsal striatum and in the two prefrontal regions of interest (see *fMRI*) will be modeled by $V(s, t)$ (Eq 1), again comparing the two methods for calculating $r(t)$.

To restate, Rescorla-Wagner value updates are defined by,

$$V(s, t) \leftarrow V(s, t) + \alpha * \delta \quad (1)$$

where

$$\delta = r(t) - V(s, t) \quad (2)$$

with $r(t)$ calculated as either

$$r_c(t) = \{1, 0\} \quad (3)$$

or

$$r_d(t) = \frac{r_c(t)}{D} \quad (4)$$

where

$$D = \sqrt{(\bar{\theta} - \theta)^2 + (\bar{W} - w)^2} \quad (5)$$

and in all cases

$$V_{initial}(s, t) = 0. \quad (6)$$

$$p(s_1) = \frac{e^{\beta V(s_1, t)}}{e^{\beta V(s_1, a)} + e^{\beta V(s_2, a)}}; \quad s_1 = (s_i, q), \quad s_2 = (s_i, w). \quad (7)$$

Prior to regression analysis, for each regions of interest outlined above, all reinforcement measures will be convolved with the canonical (“double-gamma”) hemodynamic response function and, to be consistent with the treatment of BOLD data, z-scored and 4Hz high-pass filtered. Model fits will be compared by AIC and BIC². Assuming AIC and BIC agree (as is likely) the best fit model will only be accepted as valid if it is also significant predictor in the regression.

Eight mediation models, representing the complete set of unique combinations of PFC (either ventral or dorsal) connected to striatum (ventral or dorsal) with the

²There is so far as I can tell some debate over the best criterion, in general, and it seems there is no harm in calculating both and hoping for agreement. In case of disagreement I propose using their F1 score to decide the best model fit ($F1 := \frac{AIC * BIC}{AIC + BIC}$)

two RPE models (Eq. 3 and 4) as mediators will be compared. It is of particular interest whether the best models for ventral and dorsal striatal activity will correspond well to regression fits outline above. While only half of these models are theoretically or empirically motivated, recent simulations evaluating the robustness and reliability of structural equation models (of which mediation models are a subset) demonstrated that if BOLD models of functional architecture are to be meaningfully compared all possible models need be considered, not just those believed best *a priori* (Lohmann, Erfurth, Müller, & Turner, 2011).

Two quick notes: First, all models in this work assume reward related activity in both VTA/SNc and ventral striatum is bivalent, i.e. positive outcomes lead to an increase in firing and negative outcomes lead to a suppression (D’Ardenne, McClure, Nystrom, & Cohen, 2008; Cooper & Knutson, 2008; Menon et al., 2007). This assumption has been made by nearly all RPE work to date so is not unjustified. However recent recordings in monkeys suggest that there are multiple sub-populations of cells, some of which are bivalent in the predicted direction, but others of which fire positively to both both positive and negative outcomes, and yet others with an inverted bivalent pattern – suppression to positive outcomes and increased firing to negative outcomes (Matsumoto & Hikosaka, 2009; Levita et al., 2009). Even among these subpopulations individual neurons showed trial-by-trial variance. So without further information it would be difficult, if not impossible, to model these complex and opposing patterns as a single timecourse model as is necessary for regression analysis. Second, data will be exported from BrainVoyager in the NIFTI format (<http://nifti.nih.gov/>) and all computational analyses will be carried out using custom Python (2.7.1) code developed for this proposal, as well as ongoing unrelated machine learning (MVPA) and fMRI simulation experiments. Code is complete and ready for use, excepting that to be used for statistical mediation testing, which will be ported from the Wager Lab’s Matlab toolbox (available at <http://wagerlab.colorado.edu/tools>).

Behavioral results

The experiment protocol outlined above was completed by 33 participants (2/3 female; see Fig 3 and 4). Using 63% as the cutoff which is near the $p < 0.05$ threshold of the binomial test, learning significantly exceeded chance at trial 19. This learning rate is consistent with past work in the lab using just verbal or monetary feedback. Also consistent with prior work was the mean reaction time remaining steady (near 750 ms) as learning proceeded. In summary, the categorical rewards

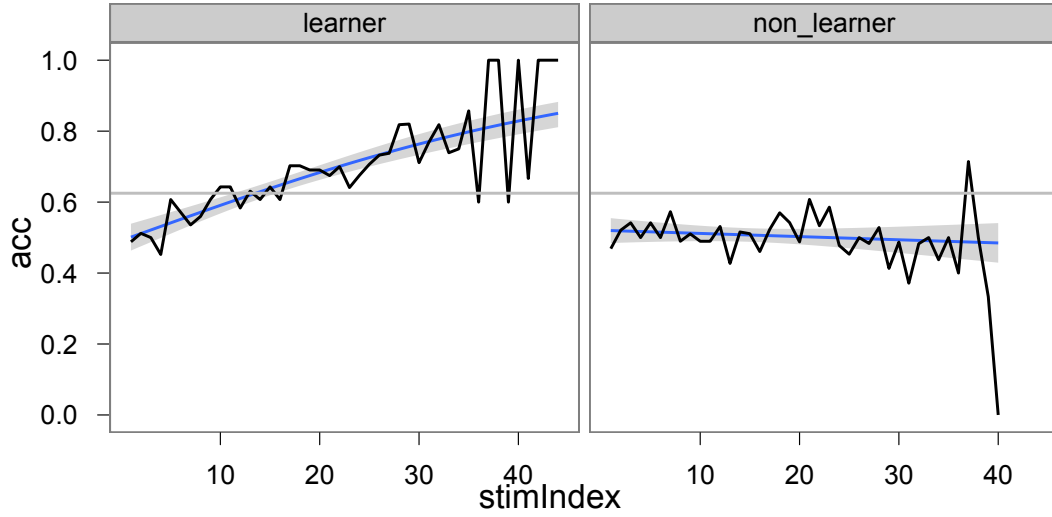


Figure 3. Mean accuracy (black), averaged for all 6 stimuli by trial, blue line and grey represent a binomial regression fit of the data and bootstrapped 95% confidence intervals, respectively. 53% were learners (left) defined by a mean accuracy in the last 10 trials (for all 6 stimuli) greater than 63%. Note: the trial order randomization procedure in this pilot did not enforce equal number of trials in for each of the 6 stimuli ($M=38$) leading to artifactual increases in variability above trial 35 or so. This oversight will be corrected prior to fMRI data acquisition.

appear behaviorally very similar to direct verbal or monetary rewards; the next step is fMRI data acquisition.

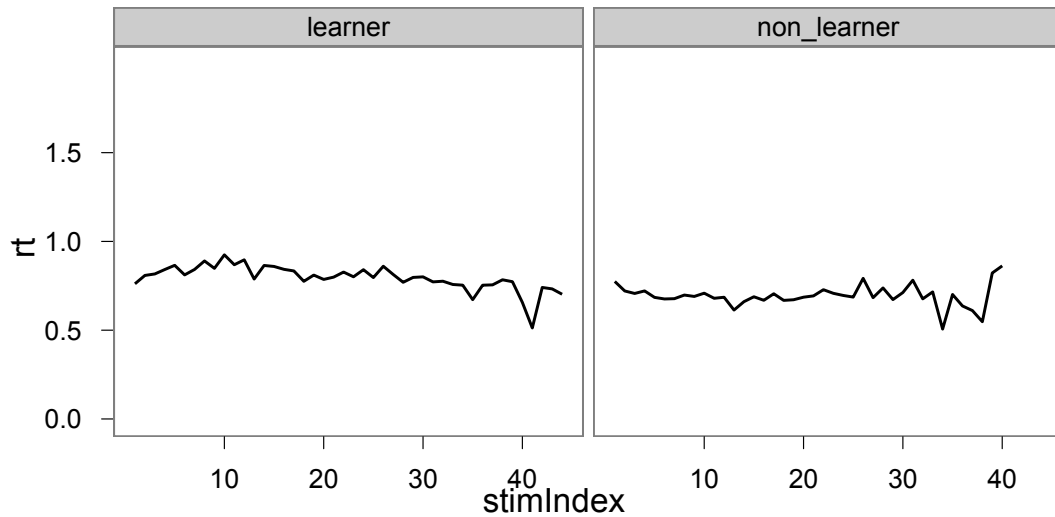


Figure 4. Mean reaction time (black), averaged for all 6 stimuli by trial, blue line and grey represent a linear regression fit of the data and bootstrapped 95% confidence intervals, respectively. See Fig 3 for learning criterion and other relevant details.

References

- Becerra, L., & Borsook, D. (n.d.). Signal valence in the nucleus accumbens to pain onset and offset. *European journal of pain (London, England)*, 12(7), 866.
- Birn, R. M., Cox, R. W., & Bandettini, P. A. (2002, Jan). Detection versus estimation in event-related fmri: choosing the optimal stimulus timing. *Neuroimage*, 15(1), 252–64.
- Björklund, A., & Dunnett, S. B. (2007, May). Fifty years of dopamine research. *Trends Neurosci.*, 30(5), 185–7.
- Blough, D. (2001, Jan). The perception of similarity. *Avian visual cognition*.
- Bornstein, A. M., & Daw, N. D. (2011, Mar). Multiplicity of control in the basal ganglia: computational roles of striatal subregions. *Current opinion in neurobiology*.
- Bromberg-Martin, E. S., & Hikosaka, O. (2009). Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron*, 63(1), 119–126.
- Cooper, J. C., & Knutson, B. (2008, Jan). Valence and salience contribute to nucleus accumbens activation. *Neuroimage*, 39(1), 538–47.
- Dale, A. M. (1999, Jan). Optimal experimental design for event-related fmri. *Hum Brain Mapp*, 8(2-3), 109–14.

- D'Ardenne, K., McClure, S. M., Nystrom, L. E., & Cohen, J. D. (2008, Feb). Bold responses reflecting dopaminergic signals in the human ventral tegmental area. *Science*, *319*(5867), 1264–7.
- Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011, Mar). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–15.
- Fliessbach, K., Weber, B., Trautner, P., Dohmen, T., Sunde, U., Elgar, C., et al. (2007). Social comparison effects reward-related activity in the human ventral striatum. *Science*, *318*, 1305–1308.
- Frank, M. J. (2011, Jun). Computational models of motivated action selection in corticostriatal circuits. *Current opinion in neurobiology*, *21*(3), 381–6.
- Guttman, H. A. (1956, Jan). Discriminability and stimulus generalization. *Journal of Experimental Psychology*.
- Hayden, B. Y., Pearson, J. M., & Platt, M. L. (2009, May). Fictive reward signals in the anterior cingulate cortex. *Science*, *324*(5929), 948–50.
- Iversen, S. D., & Iversen, L. L. (2007). Dopamine: 50 years in perspective. *Trends Neurosci.*, *30*(5), 188–193.
- Izuma, K., Saito, D. N., & Sadato, N. (2008, Apr). Processing of social and monetary rewards in the human striatum. *Neuron*, *58*(2), 284–94.
- Kahnt, T., Heinzle, J., Park, S. Q., & Haynes, J.-D. (2010, May). Decoding different roles for vmPFC and dlPFC in multi-attribute decision making. *Neuroimage*.
- Kakade, S., & Dayan, P. (2002, Jan). Dopamine: generalization and bonuses. *Neural Networks*, *15*(4-6), 549–559.
- Kao, M.-H., Mandal, A., Lazar, N., & Stufken, J. (2009, Feb). Multi-objective optimal experimental designs for event-related fMRI studies. *Neuroimage*, *44*(3), 849–56.
- Kim, H., Shimojo, S., & O'Doherty, J. P. (2010, Aug). Overlapping responses for the expectation of juice and money rewards in human ventromedial prefrontal cortex. *Cereb Cortex*.
- Kim, S., Hwang, J., Seo, H., & Lee, D. (2009, Apr). Valuation of uncertain and delayed rewards in primate prefrontal cortex. *Neural Networks*, *22*(3), 294–304.
- Levita, L., Hare, T. A., Voss, H. U., Glover, G., Ballon, D. J., & Casey, B. J. (2009, Feb). The bivalent side of the nucleus accumbens. *Neuroimage*, *44*(3), 1178–87.
- Liu, T. T. (2004, Jan). Efficiency, power, and entropy in event-related fMRI with multiple trial types. part ii: design of experiments. *Neuroimage*, *21*(1), 401–13.
- Lohmann, G., Erfurth, K., Müller, K., & Turner, R. (2011, Sep). Critical comments on dynamic causal modelling. *Neuroimage*.
- Lohrenz, T., McCabe, K., Camerer, C., & Montague, P. (2007). Neural signature of fictive learning signals in a sequential investment task. *Proceedings of the National Academy of Sciences*, *104*(22), 9493.
- Lopez-Paniagua, D., & Seger, C. A. (2011). Interactions within and be-

- tween corticostriatal loops during component processes of category learning. [http://dx.doi.org/10.1162/jocn.2008.23\(10\).3068](http://dx.doi.org/10.1162/jocn.2008.23(10).3068) – –3083.
- Matsumoto, M., & Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature*, *459*(7248), 837–841.
- Menon, M., Jensen, J., Vitcu, I., Graff-Guerrero, A., Crawley, A., Smith, M. A., et al. (2007, Oct). Temporal difference modeling of the blood-oxygen level dependent response during aversive conditioning in humans: effects of dopaminergic modulation. *Biol Psychiatry*, *62*(7), 765–72.
- Montague, P., King-Casas, B., & Cohen, J. (2006). Imaging valuation models in human choice. *Annu Rev Neurosci*, *29*, 417–448.
- O’Doherty, J. P., Buchanan, T. W., Seymour, B., & Dolan, R. J. (2006, Jan). Predictive neural coding of reward preference involves dissociable responses in human ventral midbrain and ventral striatum. *Neuron*, *49*(1), 157–66.
- O’Reilly, R. C. (2006, Oct). Biologically based computational models of high-level cognition. *Science*, *314*(5796), 91–4.
- O’Reilly, R. C. (2010, Aug). The what and how of prefrontal cortical organization. *Trends Neurosci*, *33*(8), 355–61.
- O’Reilly, R. C., & Frank, M. J. (2006, Feb). Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation*, *18*(2), 283–328.
- Pizzagalli, D. A., Evins, A. E., Schetter, E. C., Frank, M. J., Pajtas, P. E., Santesso, D. L., et al. (2008, Feb). Single dose of a dopamine agonist impairs reinforcement learning in humans: Behavioral evidence from a laboratory-based measure of reward responsiveness. *Psychopharmacology (Berl)*, *196*(2), 221–232.
- Roesch, M. R., Calu, D. J., & Schoenbaum, G. (2007, Dec). Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. *Nat Neurosci*, *10*(12), 1615–24.
- Schonberg, T., O’Doherty, J., Joel, D., Inzelberg, R., Segev, Y., & Daw, N. D. (2009, Aug). Selective impairment of prediction error signaling in human dorsolateral but not ventral striatum in parkinson’s disease patients: evidence from a model-based fmri study. *Neuroimage*.
- Schultz, W. (2007). Behavioral dopamine signals. *Trends Neurosci*, *30*(5), 203–210.
- Seymour, B., & McClure, S. M. (2008, Apr). Anchors, scales and the relative coding of value in the brain. *Current Opinion in Neurobiology*, *18*(2), 173–8.
- Simmons, S. (2008, Jan). Individual differences in the perception of similarity and difference. *Cognition*.
- Spiering, B. J., & Ashby, F. G. (2008, Sep). Response processes in information-integration category learning. *Neurobiol Learn Mem*, *90*(2), 330–8.
- Tobler, P. N., Christopoulos, G. I., O’Doherty, J. P., Dolan, R. J., & Schultz, W. (2009, Apr). Risk-dependent reward value signal in human prefrontal cortex. *Proc Natl Acad Sci USA*, *106*(17), 7185–90.

- Tricomi, E., & Fiez, J. A. (2008, Jul). Feedback signals in the caudate reflect goal achievement on a declarative memory task. *Neuroimage*, *41*(3), 1154–67.
- Urcuioli, P. (2001, Jan). Categorization and acquired equivalence. *Avian visual cognition* [On-line]. Available: *www. . . .*
- Wallis, J. D., & Kennerley, S. W. (2010, Apr). Heterogeneous reward signals in prefrontal cortex. *Current Opinion in Neurobiology*, *20*(2), 191–198.