Rewards are categories.

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

Discussion

To review, I wanted to know whether or not cognitive rewards are represented as categories in the human brain. And whether such a representation might impact the reinforcement learning process. To start to answer these two interrelated questions, I collected fMRI data while participants completed a stimulus-response task using with pre-trained perceptual categories as rewards, one category for gains and one for losses. The behavioral and neural findings of this work, which I'll now discuss in detail, show that cognitive rewards can be categories, categories which do substantively impact reinforcement learning signals. I'll further argue that category representations would be a reasonable mechanistic explanation for the generalization of (classical) secondary reinforcers. Synthesizing all of the above, I'll ultimately conclude that rewards are categories. To build this case, I'll now step through both confirmatory and inconsistent results, beginning with the behavioral and ending with select regions of interest.

Taking us to can

In the behavioral task reward were categories, information integration (II) categories to be specific. II is classic category structure, much studied in humans and other animals (J. D. Smith et al., 2011; Ashby & Maddox, 2011; J. D. Smith, Beran, Crossley, Boomer, & Ashby, 2010). II categories are distinct from their contemporaries by requiring integration of multi-dimensional stimulus information, and so are difficult to verbally describe. II learning recruits procedural memory, which relies heavily on the dorsal striatum (Ashby, Alfonso-Reese, Turken, & Waldron, 1998). This lack of verbalizeability and multi-dimensional structure, make the reward categories irreconcilable with the classical rewards almost universally used human studies of reward (e.g "Win \$1", "Correct!", "Yes!"). Despite this large difference, participants easily and rapidly learned using the II categories. Performance, as measured by both accuracy and reaction times, were nearly identical to similar tasks using verbal rewards (p**??**).

Further arguing for homology between the reward kinds, the overall pattern of BOLD activity, i.e. all trials compared to the rest trials (p**??**), was also markedly similar to that observed in nearly identical tasks using classical rewards (for several examples see, Lopez-Paniagua and Seger (2011); Seger, Peterson, Cincotta, Lopez-Paniagua, and Anderson (2010); Cincotta and Seger (2007); Seger and Cincotta (2006, 2005)).

The behavioral and neurological consistency observed in stimulus-response learning using classical and II reward categories means that perceptual categories

can act as rewards and so, reversing that logic, rewards *can be* categories. Which leads naturally to the next analysis, whether the same neural algorithm(s) that mediate classical reward learning facilitate reward category learning as well.

Are, Reflected in Error(s)

A Known Pair's Logic. However before drawing any conclusions from the modeling data, I need to get some logical preliminaries out of the way. Many of the models of interest are both covariate and dependent. Under generic statistical circumstance it would be difficult, or even impossible, to compare such models. However in limited cases strong, even causal, conclusions are possible. Inside the same family and coding scheme, there is a single change between many of the models. For example, "rpe_acc" and "rpe_acc_guass" differ only by the similarity adjustment of the reward (i.e. Eq **??** and **??**). Because both models are fit to the same data[1] and so have identical signal-to-noise ratios, the 1.5^2 fold increase in information that comes from using "rpe_acc_guass" in the dorsal caudate *must* be caused by that single change (Pearl, 2010). So while 1.5 would be small increase when comparing two noisy random variables (Anderson, Burnham, & Thompson, 2000; Forster, 2000), I argue that, (1) because uncertainty is constant between the fits, and (2) because we also know the exact relation between two models, and (3) as the model's predictions only sometimes diverge (compare columns in Figure **??**), 1.5 should instead be considered strong evidence.

---

[1] Using the same deterministic loss function
[2] Bilateral average

Categories, in All the Right Spots. In most of the regions of interest, the reward prediction family ("rpe") was the most informative, ranging from 2.3-5.1 times more likely than the non-parametric "boxcar" model, p**??**. This alone strongly suggests that like classical rewards, the learning driven by reward categories are also mediated a dopaminergic reward prediction signal. Even more important is the fact that many of the most reward sensitive areas are best described by the Gaussian-similarity adjusted reward ("rpe_acc_gauss" in Figures **??**, **??**, **??**, **??**, and **??**), demonstrating that category parameters (i.e. the similarity metrics) directly effect reward valuation. This is a direct confirmation of my hypothesis that cognitive rewards have an underlying category representation.

Outside the of VTA/SNc, striatal BOLD activity has been, time after time, shown to reflect the dopaminergic reward prediction error signal making it a, if not the, key test of novel reward prediction hypotheses (see the *Introduction* for much supporting evidence on this point). The fact then that in the dorsal caudate the Gaussian-adjusted reward prediction error term offered a substantively more informative account than the unadjusted models is a is crucially important result (compare "rpe_acc_gauss" to "rpe_acc" in Figure **??**), combined that is with the fact that the dorsal caudate was strongly active (Figure **??**) and best described by the "rpe" family (Figure **??**).

The ventral striatum was also expected to play a strong role in this task, as it is both the ventral and dorsal striatum that have been most often correlated with reward prediction activity (O'Doherty, Dayan, Friston, Critchley, & Dolan, 2003; Knutson & Wimmer, 2007; Schönberg, Daw, Joel, & O'Doherty, 2007; Seger

et al., 2010). This is not to say dorsal and ventral areas are functionally homogeneous (Schonberg et al., 2009; O'Doherty et al., 2004; Atallah, Lopez-Paniagua, Rudy, & O'Reilly, 2007). The dorsal caudate has been repeatedly linked to more abstract kinds of rewarding activity (e.g. task outcomes, fictive rewards, money compared to juice (Tricomi and Fiez (2008); Lohrenz, McCabe, Camerer, and Montague (2007); Valentin and O'Doherty (2009), for a review see Grahn, Parkinson, and Owen (2008)). While ventral activity as been associated with hedonic valuations (O'Doherty et al., 2004). Given this functional divide, and the dorsal caudate's established role in II category learning (Ashby et al., 1998), in hindsight perhaps then it is no surprise that only dorsal striatum was found to be active.

The dorsal striatum and ACC have several telling similarities. Both, in part due to dopaminergic projections from the VTA/SNc that modulate LTP via D1 receptors (Schweimer & Hauber, 2006), are strongly involved in cognitive reward learning (Atlas, Bolger, Lindquist, & Wager, 2010; Hayden, Pearson, & Platt, 2009; Rudebeck et al., 2008; Rolls, McCabe, & Redoute, 2008; Quilodran, Rothé, & Procyk, 2008; Hampton & O'Doherty, 2007; Ernst et al., 2004), with the BOLD signal often reflecting prediction errors in higher-order conditioning experiments (Seymour et al., 2004) and fictive rewards (Hayden et al., 2009). The ACC though appears to specialize in mediating between competing *future* alternatives, especially in the context of effort required to achieve each option (Quilodran et al., 2008). The fact then the ACC also is most informatively described by the Guassian-adjusted reward prediction error is another strong piece of evidence supporting reward category representations.

While generally consistent with the reward category interpretation, the insula

was the one region that was equally well described by both reward codes (i.e. "acc": $1, 0$ or "gl": $-1, 1$). All others strongly preferred "acc" only. While it is not clear the functional role of both codes, which are quite different in their predictions (compare Figure **??** to **??**, see also Figure **??**), the "gl" coding is consistent insula's established role in the processing and prediction of aversive outcomes (CITE). Additionally, the finding of dual codes in the insula is the first confirmation of my secondary hypothesis (p**??**), that the reported complex reward codes shown single cell recordings of VTA/SNc (Kim, Shimojo, & O'Doherty, 2006; Matsumoto & Hikosaka, 2009; K. S. Smith, Berridge, & Aldridge, 2011) will be reflected in the BOLD signal.

While, as reviewed in the *Introduction*, the middle frontal (i.e. dorsallateral) cortex plays role a in estimating future reward probabilities, the singular relation between activity in this region and the "rpe_acc" model (p**??**, see also Figure **??**) is best explained another of this regions well established roles, encoding of abstract rules (Wallis, Anderson, & Miller, 2001). While prefrontal regions have been previously shown to reflect prediction errors (Ramnani, Elliott, Athwal, & Passingham, 2004), I speculate that the reward categories are transformed in dorsallateral PFC into reward rules, something akin to "this category of gratings is worth \$1.". And that these rule encoded rewards have their own reward prediction errors calculations.

A fit inconsistency. Both "rpe_acc" and "rpe_gl" fit the behavioral data better either of the corresponding similarity-adjusted models (Figure **??**). If rewards are in fact categories the opposite pattern would be expected. This inconsistency though has an strong alternative explanation. Even with perfect performance, the largest

possible value estimate is smaller for the adjusted models (as suggested by Figure **??**, compare the maximum value peaks for "rpe_acc" compared to "rpe_acc_exp" and "rpe_acc_exp"). These smaller value estimates result in lower probability estimates (via the softmax transform, Eq **??**) and so result in lower log-likelihood scores (i.e. worse fits). Despite this inherent limitation the adjusted models could be modified to give equivalent performance. As the task is deterministic, once the optimal choices were learned the models could switch strategies and rely on a "working memory" strategy: just do what you did last time. This kind of working memory has recently been shown to be quite entangled with human reinforcement learning (Collins & Frank, 2012). Alternatively the reward prediction errors could be renormalized based on the cumulative variance, following observations of just such behavior (Tobler, Fiorillo, & Schultz, 2005).

Back to the Secondary. When conditioned as secondary reinforcers simple stimuli generalize well in humans and animals (for a review see p**??**). This generalization is by inference, i.e no direct reinforcement is needed (Guttman, 1956; Nakamura, Ito, Croft, & Westbrook, 2006; J. D. Smith et al., 2011). Mechanistically how such generalization occurs has not been studied. Based on the success of the similarity-adjusted reward prediction errors above, I speculate that the even simple stimuli have fundamentally categorical representations. And that these representations, via similarity-adjusted prediction errors, facilitate stimulus generalization. In addition to perfectly matching Shepard's (1987) theoretical predictions for exponential of Gaussian decays in the degree of generalization (p**??**), categorical representations of the kind studied here (p**??**) for secondary rewards would implicitly allow animals to

generalize on the first new example, matching the observed behavior. A categorical basis for even simple stimuli is advantageous in non-generalization trials as well. The intrinsic noise in neuronal encoding means that second viewing a stimulus must have a (slightly) different representation (Ashby & Townsend, 1986). A categorical representation would easily overcome such noisy encodings.

The big conclusion

Based on the consistency between classical and reward categories, both behaviorally and in overall BOLD activity patterns, I first concluded that rewards *can* be categories. This, combined with the fact that reward categories generate reward predictions errors, and these errors strongly reflect category structure, and that categories offer a powerful and parsimonious explanation for the generalization of secondary reinforcers, I finally conclude that rewards *are* categories.

Future Work

TODO – flesh these out, in some cases lots

If rewards are categories, the next question is whether rewards are *only* categories. While generalization (and so categories) may be universal (Shepard, 1987), specifics matter too. In fact memory for specifics is at odds with generalizable (i.e. abstract) memories (Atallah, Frank, & O'Reilly, 2004). Given item and categories must always diverge, both item and category reward representations would be useful. However these is a marked degree of overlap between the reward processing and

category learning systems (Seger & Miller, 2010; Ashby & Maddox, 2011). While this overlap might be due to the fact that most categories (in the lab) are learned using rewards, there is an alternative if rewards are just a kind of category. The overlapped activity could reflect one category building another (dissimilar) category, though see the discussion of dorsallateral PFC above.

Behavioral predictions (however see scaling, WM).

Reward categories would be useful for robots.

References

Anderson, D. R., Burnham, K. P., & Thompson, W. L. (2000). Null hypothesis testing: Problems, prevalence, and an alternative. *The Journal of Wildlife Management*, *64*(4), 912–923.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998, Jul). A neuropsychological theory of multiple systems in category learning. *Psychol Rev*, *105*(3), 442–81.

Ashby, F. G., & Maddox, W. T. (2011, Apr). Human category learning 2.0. *Ann N Y Acad Sci*, *1224*, 147–61.

Ashby, F. G., & Townsend, J. T. (1986, Apr). Varieties of perceptual independence. *Psychol Rev*, *93*(2), 154–79.

Atallah, H. E., Frank, M. J., & O'Reilly, R. C. (2004, Nov). Hippocampus, cortex, and basal ganglia: insights from computational models of complementary learning systems. *Neurobiol Learn Mem*, *82*(3), 253–67.

Atallah, H. E., Lopez-Paniagua, D., Rudy, J. W., & O'Reilly, R. C. (2007, Jan). Separate neural substrates for skill learning and performance in the ventral and dorsal striatum. *Nat Neurosci*, *10*(1), 126–31.

Atlas, L. Y., Bolger, N., Lindquist, M. A., & Wager, T. D. (2010, Sep). Brain mediators of predictive cue effects on perceived pain. *J Neurosci*, *30*(39), 12964–77.

Cincotta, C. M., & Seger, C. A. (2007, Feb). Dissociation between striatal regions while learning to categorize via feedback and via observation. *Journal of cognitive neuroscience*, *19*(2), 249–65.

Collins, A. G. E., & Frank, M. J. (2012, Apr). How much of reinforcement learning is working memory, not reinforcement learning? a behavioral, computational, and

neurogenetic analysis. *Eur J Neurosci*, *35*(7), 1024–35.

Ernst, M., Nelson, E. E., McClure, E. B., Monk, C. S., Munson, S., Eshel, N., et al. (2004, Jan). Choice selection and reward anticipation: an fmri study. *Neuropsychologia*, *42*(12), 1585–97.

Forster, M. (2000, Mar). Key concepts in model selection: Performance and generalizability. *Journal of Mathematical Psychology*, *44*(1), 205–231.

Grahn, J. A., Parkinson, J. A., & Owen, A. M. (2008, Nov). The cognitive functions of the caudate nucleus. *Progress in Neurobiology*, *86*(3), 141–55.

Guttman, N. (1956, Jan). Discriminability and stimulus generalization. *Journal of Experimental Psychology*.

Hampton, A. N., & O'Doherty, J. P. (2007, Jan). Decoding the neural substrates of reward-related decision making with functional mri. *PNAS*, *104*(4), 1377–82.

Hayden, B. Y., Pearson, J. M., & Platt, M. L. (2009, May). Fictive reward signals in the anterior cingulate cortex. *Science*, *324*(5929), 948–50.

Kim, H., Shimojo, S., & O'Doherty, J. P. (2006, Jul). Is avoiding an aversive outcome rewarding? neural substrates of avoidance learning in the human brain. *PLoS Biology*, *4*(8), e233.

Knutson, B., & Wimmer, G. E. (2007, May). Splitting the difference: how does the brain code reward episodes? *Annals of the New York Academy of Sciences*, *1104*, 54–69.

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 between corticostriatal loops during component processes of category learning. *Journal of cognitive neuroscience*, *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.

Nakamura, T., Ito, M., Croft, D. B., & Westbrook, R. F. (2006, Nov). Domestic pigeons (columba livia) discriminate between photographs of male and female pigeons. *Learning & Behavior*, *34*(4), 327–339.

O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003, Apr). Temporal difference models and reward-related learning in the human brain. *Neuron*, *38*(2), 329–37.

O'Doherty, J. P., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004, Apr). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, *304*(5669), 452–4.

Pearl, J. (2010, Jan). An introduction to causal inference. *Int J Biostat*, *6*(2), Article 7.

Quilodran, R., Rothé, M., & Procyk, E. (2008, Jan). Behavioral shifts and action valuation in the anterior cingulate cortex. *Neuron*, *57*(2), 314–25.

Ramnani, N., Elliott, R., Athwal, B. S., & Passingham, R. E. (2004, Nov). Prediction error for free monetary reward in the human prefrontal cortex. *Neuroimage*, *23*(3), 777–86.

Rolls, E. T., McCabe, C., & Redoute, J. (2008, Mar). Expected value, reward outcome, and temporal difference error representations in a probabilistic decision task. *Cereb Cortex*, *18*(3), 652–63.

Rudebeck, P. H., Behrens, T. E., Kennerley, S. W., Baxter, M. G., Buckley, M. J., Walton, M. E., et al. (2008, Dec). Frontal cortex subregions play distinct roles in choices between actions and stimuli. *J Neurosci*, *28*(51), 13775–85.

Schönberg, T., Daw, N. D., Joel, D., & O'Doherty, J. P. (2007, Nov). Reinforcement learning signals in the human striatum distinguish learners from nonlearners during

reward-based decision making. *J Neurosci*, *27*(47), 12860–7.

Schonberg, T., O'Doherty, J. P., 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*.

Schweimer, J., & Hauber, W. (2006, Jan). Dopamine d1 receptors in the anterior cingulate cortex regulate effort-based decision making. *Learn Mem*, *13*(6), 777–82.

Seger, C. A., & Cincotta, C. (2005). The roles of the caudate nucleus in human classification learning. *J Neurosci*, *25*(11), 2941–2951.

Seger, C. A., & Cincotta, C. M. (2006, Nov). Dynamics of frontal, striatal, and hippocampal systems during rule learning. *Cereb Cortex*, *16*(11), 1546–55.

Seger, C. A., & Miller, E. K. (2010, Jan). Category learning in the brain. *Annu Rev Neurosci*, *33*, 203–19.

Seger, C. A., Peterson, E. J., Cincotta, C. M., Lopez-Paniagua, D., & Anderson, C. W. (2010, Apr). Dissociating the contributions of independent corticostriatal systems to visual categorization learning through the use of reinforcement learning modeling and granger causality modeling. *Neuroimage*, *50*(2), 644–56.

Seymour, B., O'Doherty, J. P., Dyan, P., Koltzenburg, M., Jones, J. K., Dolan, R. J., et al. (2004). Temporal difference models describe higher-order learning in humans. *Nature*, *429*, 664 – 667.

Shepard, R. (1987, Sep). Toward a universal law of generalization for psychological science. *Science*, *237*(4820), 1317–1323.

Smith, J. D., Ashby, F. G., Berg, M. E., Murphy, M. S., Spiering, B., Cook, R. G., et al. (2011). Pigeons' categorization may be exclusively nonanalytic. *Psychonomic Belletin & Review*, *18*(2), 414–421.

Smith, J. D., Beran, M. J., Crossley, M. J., Boomer, J., & Ashby, F. G. (2010, Jan). Implicit and explicit category learning by macaques (macaca mulatta) and humans (homo sapiens). *J Exp Psychol Anim Behav Process*, *36*(1), 54–65.

Smith, K. S., Berridge, K. C., & Aldridge, J. W. (2011, Jul). Disentangling pleasure from incentive salience and learning signals in brain reward circuitry. *Proc Natl Acad Sci USA*, *108*(27), E255–64.

Tobler, P. N., Fiorillo, C. D., & Schultz, W. (2005, Mar). Adaptive coding of reward value by dopamine neurons. *Science*, *307*(5715), 1642–5.

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.

Valentin, V. V., & O'Doherty, J. P. (2009, Dec). Overlapping prediction errors in dorsal striatum during instrumental learning with juice and money reward in the human brain. *J Physiol.*, *102*(6), 3384–91.

Wallis, J. D., Anderson, K. C., & Miller, E. K. (2001, Jun). Single neurons in prefrontal cortex encode abstract rules. *Nature*, *411*(6840), 953–6.