

Categories of rewards.

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

Introduction

Birds will peck repeatedly, as mice will push levers, monkeys will hit buttons, and men will buy flowers, if each of these actions is followed by a primary reward – food, drink or sex. Buttons, levers and flowers have no value alone, so reinforcement theory goes, it is only by the *statistically regular pairing* with primary rewards that value is transferred (?). This is the classical view, and it has, in general, held up many years now (Iversen & Iversen, 2007). In fact, the neural basis reinforcement learning currently receives substantial attention, with notable progress is being made (?; Montague, King-Casas, & Cohen, 2006). However, reinforcement learning theories can't at current account for two recent key findings in the neural correlates of human learning. (1) the neural correlates of reward can appear by cognition alone (Hayden, Pearson, & Platt, 2009; Lohrenz, McCabe, Camerer, & Montague, 2007; Tricomi & Fiez, 2008; Jimura, Locke, & Braver, 2010). (2) Value can be transferred by inference, no pairing is needed (Bromberg-Martin, Matsumoto, Hong, & Hikosaka, 2010; Hampton, Bossaerts, & O'Doherty, 2006). Using a mixture of fMRI and computational modeling, I examine possible mechanisms of both of these aspects by modeling the reward representations as a kind of category; categories, i.e. abstract generalizable representations, are exclusively cognitive and require inference.

This introduction has six parts. First I discuss classical rewards and their neural correlates and what is known of their anatomical bases, including criticisms of the reward prediction account of dopamine function that underlies this work. Second I make a case for cognitive rewards, discussing as an example novelty, then moving onto other examples, and finally arguing for the necessity of generalizable reward representations. Third is a discussion of prior studies of value generalization in pigeons and other non-human animals as well as in humans, though the literature on

the latter is sparse. Fourth is a brief introduction to formal models of categorization, which leads into the fifth and final section, the specific goals and methods of this work.

Classics, Expectations and Tissues

A pleasurable start. 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 sex though for, err, logistical reasons this is less often used in the laboratory, especially in human subjects; They’re certainly potent, being used for over 50 successful years to study learning in animal models (Iversen & Iversen, 2007) and people (Kim, Shimojo, & O’Doherty, 2010; Montague et al., 2006). In the 1950’s the first clue how food and water cause reinforcement arose in the electrical self-stimulation studies of ?, ?, for a classic review see ?, ?. Olds and colleagues observed that when electrodes, which could be activated by a self-determined button press, were placed in the midbrain and in limbic areas animals would vigorously and repeatedly self-stimulate. By the 1970s data from pharmacological studies of rats, electrochemical recordings, knowledge of the signalling mechanisms of dopamine receptors, as well as neuroleptic drug actions in Schizophrenic patients, along with Old’s shocking work, led to the first major theoretical proposal for dopamine’s role - a signal for pleasure (i.e. the anhedonia hypothesis, ?, ?). However within 10 years it became clear that dopamine’s role extends beyond signaling primary rewards. Activity was seen following secondary rewards, novelty, salience, and others (?, ?, ?, ?). And, more importantly, dopamine depleted animals continued to enjoy rewards, i.e. they still developed taste preferences, enhanced response vigor (?, ?) and continued to respond to opiates (?, ?)¹. However in 1994 there was a surprise that would eventually explain many of anhedonia’s deficits. Schultz *et al* reported that dopaminergic firing depended on how expected a reward was (Mirenowicz & Schultz, 1994), which blossomed into the reward prediction error theory under review here.

Expectations Matter. As I said Mirenowicz & Schultz, 1994, was the first paper to demonstrate that phasic activity in the midbrain was dependent not just on the hedonic value of the reward, but was qualitatively related to both the value of the reward and how expected that reward was. Hollerman & Schultz, 1998 more fully

¹Though both these criticisms are constrained as dopamine deficient mice, sans caffeine, are extraordinarily lethargic. To pep them up, caffeine is administered. And caffeine, through a cascade driven by adenosine A2A and cannabinoid CB1 receptors in ventral striatum has biochemical effects similar to dopamine (Lazarus et al., 2011; Rossi et al., 2010)

explored this observation, showing that unexpected rewards lead to increases in the firing rate, strongly expected rewards elicited no response, while expected rewards that failed to arrive lead to a pause. Roesch, Calu, & Schoenbaum, 2007 found the same patterns in rats. O'Doherty, Dayan, Friston, Critchley, & Dolan, 2003, found them too in fMRI studies of humans. Meanwhile Fiorillo, Tobler, & Schultz, 2003, along with Bayer & Glimcher, 2005, quantified the relationship between the unexpectedness of the reward and phasic activity via a reward prediction error term derived from a reinforcement learning model fit to each animal's behavior. The reward prediction error from the model strongly correlated with the dopamine response - both its increases and decreases. A second key similarity between reinforcement learning models and the dopamenergic response was transfer. If a cue reliably predicts a reward the reinforcement learning equations require value to transfer from the primary reward to the cue, mimicking Pavlovian conditioning (which was the initial goal of these models). This same behavior was observed in the dopamine response (Roesch et al., 2007; McClure, Berns, & Montague, 2003). Furthermore RL models are statistically predictive of non-human animal's choice behaviors (Hampton & O'Doherty, 2007). Dopamenergic firing patterns are consistent with optimal formal learning theories (Waelti, Dickinson, & Schultz, 2001) and has been shown to mediate cortical-striatal coupling (Ouden, Daunizeau, Roiser, Friston, & Stephan, 2010). Single doses of dopamine antagonists and agonists have also demonstrated a causal relationship between dopamine levels and learning rate (Pizzagalli et al., 2008; Diaconescu, Menon, Jensen, Kapur, & McIntosh, 2010), which is broadly though not exclusively consistent with a reinforcement theory interpretation. These findings are the backbone of the reward prediction theory, which has since been extended substantially.

Based on novel findings about novelty (?, ?; Blatter & Schultz, 2006; Guitart-Masip, Bunzeck, Stephan, Dolan, & Duzel, 2010) the reward prediction hypothesis has been extended to incorporate activity observed following presentation of novel stimuli (Kakade & Dayan, 2002) as well as to explain reward anticipatory firing via an average reward prediction error (Knutson & Wimmer, 2007). Another variation allowed for the observation of simultaneous neural implementation of model-free and model-based reinforcement learning (A. Smith, Li, Becker, & Kapur, 2006; Daw, Gershman, Seymour, Dayan, & Dolan, 2011). Alternative but reconcilable accounts have also been offered that allow for dissociation of first and second order conditioning as well as pavlovian to instrumental transfer (O'Reilly, Frank, Hazy, & Watz, 2007). The reward prediction hypothesis has also been incorporated into theoretical accounts of addiction (Redish, 2004) and to predict the salience/information of upcoming stimuli (?, ?), among others.

Additionally there are several findings which are, as yet, are unaccounted for theoretically. The dopamine response adaptively scales with past reward magnitudes, which appeared similar to reward value divided by the cumulative variance (Tobler, Fiorillo, & Schultz, 2005).

Matsumoto & Hikosaka, 2009, reported a very broad set of dopamenergic firing patterns. In the classical (bivalent) view dopamine neurons should fire more for rewards, negatively to reward omission, and positively to the omission of punishment, and negatively to aversive events (Kim, Shimojo, & O'Doherty, 2006). Matsumoto:2009p7219, found that some neurons respond as expected but many others responded positively to aversive stimuli (an air puff on the eye) increasing as the punishment grew larger (than expected), and decreasing as it grew smaller. Unexpectedly many neurons responded positively to *both* appetitive and aversive conditions when expectations were exceeded and negatively when they predictions were optimistic. Matsumoto & Hikosaka, 2009, further reported that neurons that responded positively to aversive events were clustered in the substantia nigra pars compacta while appetitively tuned neurons appeared more often in the ventral tagmental area. Likewise K. S. Smith, Berridge, & Aldridge, 2011 demonstrated *simultaneous* tunings to both reward value, reward expectancy, salience as well as to novelty. If the dopamenergic response is universally this complex, the bivalent view needs substantial refinement, as do perhaps our analysis techniques; many different yet related models may *correctly* fit the same data, an issue which has received some prior attention outside of neuroimaging² ?, ?.

Networked Plausibility. The dopamenergic firing patterns outline above originate in the VTA/SNc is a small brainstem nucleus whose dopamine-releasing neurons project strongly to both the striatum and the hippocampus. Electrophysiological recordings of VTA/SNc neurons show two firing modes – tonic and phasic (?, ?). The phasic mode is of interest here, at it reflect theoretically a reward predictions error signal. A reward predictions *error* signal of course requires predictions. In this case predictions of future value. Though how exactly such predictions are made is only partly understood. Candidate regions for this calculation include the striatum, the limbic system (via the habenula) as well as the orbital frontal and ventral medial cortices. I present each in turn, leaving the totality unintegrated, thus accurately reflecting the literature's state – tempting and incomplete.

²for my take on solving this problem in fMRI, see the methods section.

And In The Striatum. The striatum is an input area of the basal ganglia, a brain region highly involved in categorization, logical inference, habit formation, working memory and feedback mediated S-R learning (Frank, Loughry, & O'Reilly, 2001; Jin & Costa, 2010; Schmitzer-Torbert & Redish, 2004; Seger, 2008; Seger & Miller, 2010; Yin & Knowlton, 2006). In S-R learning, two of the five striatal subregions (the head of the caudate and the ventral striatum) process reward information (Yin, Ostlund, Knowlton, & Balleine, 2005; Yin, Ostlund, & Balleine, 2008; Schonberg et al., 2009). These two are highly innervated by projections from the VTA/SNc, but only the ventral striatum correlates with the RPE signal (Haruno & Kawato, 2006; Seger & Miller, 2010). The remaining three regions (the body and tail of the caudate and the putamen) are involved with S-R pair formation, visual categorization and response selection, respectively (Seger, 2008; Seger & Miller, 2010). Though these three also receive VTA/SNc projections and are sometimes sensitive to reward level (Bischoff-Grethe, Hazeltine, Bergren, Ivry, & Grafton, 2009), the BOLD signal does not correlate with the RPE (Seger & Miller, 2010); dopamine's exact role in these areas is less clear. Overall though, intact dopamine projections and complete striatal function is necessary for rapid S-R learning.

Administering dopamine antagonists to human and non-human animals adversely affects S-R learning (Pizzagalli et al., n.d.), as does lesioning the VTA/SNc. Complete lesions of the striatum also prevent S-R learning (Packard & Knowlton, 2002). Administering dopamine agonists or the readily converted precursor L-DOPA leads to increases in response vigor and the ability of a Pavlovian-conditioned stimulus to bias unrelated instrumental responses (i.e. pavlovian instrumental transfer) (Winterbauer & Balleine, 2007). Both pavlovian instrumental transfer and response vigor are, in part, facilitated by phasic dopamine increasing activity in the ventral striatum. Unmedicated Parkinson's patients, who have low striatal dopamine levels, show marked decreases in S-R learning with rewarding outcomes when compared to patients on medication and healthy age and intellect matched controls (Pizzagalli et al., n.d.). These same patients show an enhanced capability to learn from negative feedback which suggests that decreases in dopamine convey negative outcome information (Frank, Seeberger, & O'Reilly, 2004). Finally, there is a solid body of evidence suggesting that phasic dopamine alters the plasticity of neurons in the striatum which presumably facilitates stimulus-response learning (Calabresi, Picconi, Tozzi, & Filippo, 2007).

Limbic Integration. VTA/SNc also receives input from the internal globus pallidus or GPi (a major output structure of the basal ganglia with widespread cortical

connections), thalamus, and the central nucleus of the amygdala (Botvinick, Niv, & Barto, 2008). Recent work though has highlighted the habenula (a small nucleus posterior to the thalamus) as being especially important in generating RPE-like phasic activity. The lateral habenula has reciprocal connections to the GPi and projections into the VTA/SNc. Based on this anatomy, it has been suggested that this nucleus serves a point of intersection between the striatum and the limbic system (e.g. the amygdala, hippocampus and the serotonergic dorsal raphe nucleus) (Hikosaka, Sesack, Lecourtier, & Shepard, 2008). The habenula acts to tonically inhibit or disinhibit dopamine release in VTA/SNc neurons. As habenula activity decreases, burst firing in VTA/SNc results; as habenula firing increases VTA/SNc firing is temporarily paused. Dual recordings of the habenula and GPi suggest they form a functional loop capable of calculating the value of S-R pairs (Bromberg-Martin, Matsumoto, & Hikosaka, 2010). Reversible chemical inhibition of the habenula also increases VTA/SNc phasic activity. Lesions to the habenula also result in marked increases in dopamine levels in the dorsal and ventral striatum (Bromberg-Martin, Matsumoto, & Hikosaka, 2010). In summary, the GPi (with its access to cortical inputs via the striatum) and habenula (with its capability for altering VTA/SNc activity) may form the physiological loop necessary to calculate the RPE; Bromberg-Martin, Matsumoto, & Nakahara, 2010, hint that this loop can signal both initial value estimates and rewarding outcomes.

Front and Center. Orbital frontal cortex has been repeatedly shown, both in neuroimaging (?, ?) and lesion (?, ?) studies to encode the absolute value of rewarding or punishing outcomes, thus playing a pivotal role in the new field of neuroeconomics (?, ?). However orbital frontal areas are more than a simple value store. It also plays a role in response and outcome recall (?, ?) and selection (?, ?) as well as in motivation, pain and pleasure (?, ?) and outcome anticipation (?, ?) and outcome prediction (?, ?) and causal attribution (?, ?). Additionally, orbital function is highly dependent on the basolateral amygdala (?, ?) offering a path for influencing reward prediction error calculation.

Estimating value is one part of a reward prediction but so is the expectation, that is the probability a reward will occur.

Let's Get Critical.. Roesch et al., 2007, also reported optimistic responses, that is when novel stimuli were presented in the context of a familiar task with the same action options previously available, the dopamine spike that resulted were significantly correlated with the best known past action. No simple modification to the

TD equation allows for the production of a prediction error to a single *specific* past outcome. It is possible though that this firing may have been driven by an alternative circuit, one with access to specific past events (for example the hippocampus which is known to both receive from and project to the VTA/SNc (Lisman & Grace, 2005)).

While

Bromberg-Martin & Hikosaka, 2009, demonstrated dopamine firing about informative (but not directly rewarding) cues.

References

- 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.
- Bayer, H., & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, *47*, 129–141.
- 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.
- Bischoff-Grethe, A., Hazeltine, E., Bergren, L., Ivry, R. B., & Grafton, S. T. (2009, Jan). The influence of feedback valence in associative learning. *Neuroimage*, *44*(1), 243–51.
- Blatter, K., & Schultz, W. (2006, Jan). Rewarding properties of visual stimuli. *Experimental Brain Research*, *168*(4), 541–6.
- Botvinick, M., Niv, Y., & Barto, A. (2008, Oct). Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition*.
- Bromberg-Martin, E. S., & Hikosaka, O. (2009). Midbrain dopamine neurons signal preference for advance information about upcoming rewards. *Neuron*, *63*(1), 119–126.
- Bromberg-Martin, E. S., Matsumoto, M., & Hikosaka, O. (2010, Jan). Distinct tonic and phasic anticipatory activity in lateral habenula and dopamine neurons. *Neuron*.
- Bromberg-Martin, E. S., Matsumoto, M., Hong, S., & Hikosaka, O. (2010, Aug). A pallidus-habenula-dopamine pathway signals inferred stimulus values. *J Physiol.*, *104*(2), 1068–76.
- Bromberg-Martin, E. S., Matsumoto, M., & Nakahara, H. (2010, Jan). Multiple timescales of memory in lateral habenula and dopamine neurons. *Neuron*.
- Calabresi, P., Picconi, B., Tozzi, A., & Filippo, M. D. (2007, May). Dopamine-mediated regulation of corticostriatal synaptic plasticity. *Trends Neurosci.*, *30*(5), 211–9.
- 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.
- Diaconescu, A. O., Menon, M., Jensen, J., Kapur, S., & McIntosh, A. R. (2010, Feb). Dopamine-induced changes in neural network patterns supporting aversive conditioning. *Brain Res*, *1313*, 143–61.
- Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003, Mar). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, *299*(5614), 1898–902.
- Frank, M. J., Loughry, B., & O’Reilly, R. C. (2001, Jun). Interactions between frontal cortex and basal ganglia in working memory: a computational model. *Cognitive, affective & behavioral neuroscience*, *1*(2), 137–60.
- Frank, M. J., Seeberger, L. C., & O’Reilly, R. C. (2004, Dec). By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science*, *306*(5703), 1940–3.
- Furuyashiki, T., Holland, P. C., & Gallagher, M. (2008, May). Rat orbitofrontal cortex

- separately encodes response and outcome information during performance of goal-directed behavior. *J Neurosci*, 28(19), 5127–38.
- Glimcher, P. W., Dorris, M., & Bayer, H. (2005, Aug). Physiological utility theory and the neuroeconomics of choice. *Games and economic behavior*, 52(2), 213–256.
- Guitart-Masip, M., Bunzeck, N., Stephan, K., Dolan, R. J., & Duzel, E. (2010, Feb). Contextual novelty changes reward representations in the striatum. *Journal of Neuroscience*, 30(5), 1721.
- Hampton, A. N., Bossaerts, P., & O’Doherty, J. P. (2006, Aug). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *J Neurosci*, 26(32), 8360–7.
- 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.
- Haruno, M., & Kawato, M. (2006, Feb). Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus-action-reward association learning. *J Physiol.*, 95(2), 948–59.
- Hayden, B. Y., Pearson, J. M., & Platt, M. L. (2009, May). Fictive reward signals in the anterior cingulate cortex. *Science*, 324(5929), 948–50.
- Hikosaka, O., Sesack, S. R., Lecourtier, L., & Shepard, P. D. (2008, Nov). Habenula: crossroad between the basal ganglia and the limbic system. *J Neurosci*, 28(46), 11825–9.
- Hollerman, J., & Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neuroscience*, 1(4), 304–309.
- Hornak, J., O’Doherty, J., Bramham, J., Rolls, E. T., Morris, R. G., Bullock, P. R., et al. (2004, Apr). Reward-related reversal learning after surgical excisions in orbito-frontal or dorsolateral prefrontal cortex in humans. *Journal of cognitive neuroscience*, 16(3), 463–78.
- Iversen, S. D., & Iversen, L. L. (2007). Dopamine: 50 years in perspective. *Trends Neurosci.*, 30(5), 188–193.
- Jimura, K., Locke, H. S., & Braver, T. S. (2010, May). Prefrontal cortex mediation of cognitive enhancement in rewarding motivational contexts. *Proc Natl Acad Sci USA*, 107(19), 8871–6.
- Jin, X., & Costa, R. M. (2010, Jul). Start/stop signals emerge in nigrostriatal circuits during sequence learning. *Nature*, 466(7305), 457–462.
- Kakade, S., & Dayan, P. (2002, Jan). Dopamine: generalization and bonuses. *Neural Networks*, 15(4-6), 549–559.
- 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.
- 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*.

- 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.
- Lazarus, M., Shen, H.-Y., Cherasse, Y., Qu, W.-M., Huang, Z.-L., Bass, C. E., et al. (2011, Jul). Arousal effect of caffeine depends on adenosine a2a receptors in the shell of the nucleus accumbens. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 31(27), 10067–10075.
- Lisman, J. E., & Grace, A. A. (2005, Jun). The hippocampal-vta loop: controlling the entry of information into long-term memory. *Neuron*, 46(5), 703–13.
- 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.
- Matsumoto, M., & Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature*, 459(7248), 837–841.
- McClure, S. M., Berns, G. S., & Montague, P. R. (2003, Apr). Temporal prediction errors in a passive learning task activate human striatum. *Neuron*, 38(2), 339–46.
- Mirenowicz, J., & Schultz, W. (1994). Importance of unpredictability for reward responses in primate dopamine neurons. *J Physiol.*, 72(2), 1024.
- Montague, P., King-Casas, B., & Cohen, J. (2006). Imaging valuation models in human choice. *Annu Rev Neurosci*, 29, 417–448.
- O’Doherty, J. (2003, Aug). Can’t learn without you: predictive value coding in orbitofrontal cortex requires the basolateral amygdala. *Neuron*, 39(5), 731–3.
- O’Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J., & Andrews, C. (2001, Jan). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nat Neurosci*, 4(1), 95–102.
- 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’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’Reilly, R., Frank, M. J., Hazy, T., & Watz, B. (2007, Jan). Pvlv: the primary value and learned value pavlovian learning algorithm. *Behav. Neurosci.*
- Ouden, H. den, Daunizeau, J., Roiser, J., Friston, K., & Stephan, K. (2010). Striatal prediction error modulates cortical coupling. *Journal of Neuroscience*, 30(9), 3210.
- Packard, M. G., & Knowlton, B. J. (2002, Jan). Learning and memory functions of the basal ganglia. *Annu Rev Neurosci*, 25, 563–93.
- Pizzagalli, D. A., Evins, A. E., Schetter, E. C., Frank, M. J., Pajtas, P. E., Santesso, D. L., et al. (n.d.). Single dose of a dopamine agonist impairs reinforcement learning in humans: Behavioral evidence from a laboratory-based measure of reward responsiveness. *Psychopharmacology*, 196(2), 221.
- 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.
- Redish, A. D. (2004, Dec). Addiction as a computational process gone awry. *Science*, 306(5703), 1944–7.
- 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.
- Roesch, M. R., & Olson, C. R. (2007, Dec). Neuronal activity related to anticipated reward in frontal cortex: does it represent value or reflect motivation? *Annals of the New York Academy of Sciences*, 1121, 431–46.
- Rossi, S., Chiara, V. D., Musella, A., Mataluni, G., Sacchetti, L., Siracusano, A., et al. (2010). Effects of caffeine on striatal neurotransmission: Focus on cannabinoid cb1 receptors. *Molecular Nutrition & Food Research*, 54(4), 525–531.
- 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.
- Schmitzer-Torbert, N., & Redish, A. D. (2004, May). Neuronal activity in the rodent dorsal striatum in sequential navigation: separation of spatial and reward responses on the multiple t task. *J Physiol.*, 91(5), 2259–72.
- 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.
- Seger, C. A. (2008, Jan). How do the basal ganglia contribute to categorization? their roles in generalization, response selection, and learning via feedback. *Neuroscience and Biobehavioral Reviews*, 32(2), 265–78.
- Seger, C. A., & Miller, E. K. (2010, Jan). Category learning in the brain. *Annu Rev Neurosci*, 33, 203–19.
- Smith, A., Li, M., Becker, S., & Kapur, S. (2006, Mar). Dopamine, prediction error and associative learning: a model-based account. *Network*, 17(1), 61–84.
- 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.
- Tanaka, S. C., Balleine, B. W., & O’Doherty, J. P. (2008, Jun). Calculating consequences: brain systems that encode the causal effects of actions. *J Neurosci*, 28(26), 6750–5.
- Tanaka, S. C., Samejima, K., Okada, G., Ueda, K., Okamoto, Y., Yamawaki, S., et al. (2006). Brain mechanism of reward prediction under predictable and unpredictable environmental dynamics. *Neural Networks*, 19(8), 1233–1241.

- 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.
- Waelti, P., Dickinson, A., & Schultz, W. (2001, Jul). Dopamine responses comply with basic assumptions of formal learning theory. *Nature*, 412(6842), 43–8.
- Winterbauer, N. E., & Balleine, B. W. (2007, Jan). The influence of amphetamine on sensory and conditioned reinforcement: evidence for the re-selection hypothesis of dopamine function. *Frontiers in integrative neuroscience*, 1, 9.
- Yin, H. H., & Knowlton, B. J. (2006, Jun). The role of the basal ganglia in habit formation. *Nat Rev Neurosci*, 7(6), 464–76.
- Yin, H. H., Ostlund, S. B., & Balleine, B. W. (2008, Oct). Reward-guided learning beyond dopamine in the nucleus accumbens: the integrative functions of cortico-basal ganglia networks. *Eur J Neurosci*, 28(8), 1437–48.
- Yin, H. H., Ostlund, S. B., Knowlton, B. J., & Balleine, B. W. (2005, Jul). The role of the dorsomedial striatum in instrumental conditioning. *Eur J Neurosci*, 22(2), 513–23.