Rewards are categories.

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 transfered (Rescorla, 1988). This is the classical view, and it has, in general, held up many years now (Iversen & Iversen, 2007). And indeed, the neural mechanisms of reinforcement learning are becoming increasingly clear following years of exciting and intense inquiry (?, ?; Montague, King-Casas, & Cohen, 2006). However, due to their reliance on primary and secondary reward concepts, reinforcement learning theories can't account for two important facts. One, rewarding effects are observed in the absence of primary and secondary reinforcers (Hayden, Pearson, & Platt, 2009; Lohrenz, McCabe, Camerer, & Montague, 2007; Tricomi & Fiez, 2008; Jimura, Locke, & Braver, 2010). Two, value can be transfered by inference, no pairing is needed (Bromberg-Martin, Matsumoto, Hong, & Hikosaka, 2010; Hampton,

Bossaerts, & O'Doherty, 2006). Based on fMRI and behavioral data, I examine possible mechanisms of both of these aspects by modeling rewards as categories. Categories are intrinsically cognitive and inferential.

This introduction has five parts. First I review classical rewards and the reward prediction error account of dopamine function, along with its anatomical basis. Second, I switch hats and (non-fatally) criticise these accounts. Third, I make a case for cognitive rewards – rewarding activity observed outside primary and secondary reinforcement. Fourth I argue for the necessity of generalizable reward representations, covering stimulus generalization and categorization along the way. In the fifth and final section, the exact goals and methods of this work are laid out.

Classics, Expectations and Tissues

A pleasurable start. Classically rewards and reinforcers have been linked to (or simply were) food (J. P. O'Doherty, Buchanan, Seymour, & Dolan, 2006), pain (?, ?; Schultz, 2007) and sex though for, err, logistical reasons this is less often used in the laboratory; They're certainly potent, being used for over 50 successful years to study learning in animal models (Iversen & Iversen, 2007) and people (H. Kim, Shimojo, & O'Doherty, 2010; Montague et al., 2006). In the 1950's the first clue how food and water neuronally cause reinforcement arose in the electrical self-stimulation studies of OLDS & MILNER, 1954, and, Crow, 1972. Olds and colleges observed that when electrodes were placed in the dopamenergic midbrain animals would vigorously and repeated self-stimulate, using the available button. By the 1970s data from pharmacological studies of rats, electrochemical recordings, knowledge of the signalling

mechanisms of dopamine receptors, as well as neoleptic drug actions in Schizophrenic patients, along with Old's shocking work, lead to the first major theoretical proposal for dopamine's role - a signal for pleasure, i.e. the anhedonia hypothesis, (Wise, 1978). However within 10 years it became clear that dopamine's role extend beyond signaling primary rewards [1] and pleasure. Activity was seen following secondary rewards, novelty, salience, and others (Spanagel & Weiss, 1999; Salamone, Correa, Mingote, & Weber, 2005; Bromberg-Martin, Matsumoto, & Hikosaka, 2010b). And, more importantly, dopamine depleted animals continued enjoy rewards, i.e. they still developed taste preferences, enhanced response vigor (Cannon & Palmiter, 2003) and continued to respond to opiates (Hnasko, Sotak, & Palmiter, 2005). However in 1994 there was a surprise that would eventually explain many of anhedonia's deficits. Mirenowicz & Schultz, 1994, reported that dopamengeric firing depended on how expected a reward was, which blossomed into the reward prediction error theory under review here.

Expectations Matter. Hollerman & Schultz, 1998 more fully explored, Mirenowicz & Schultz, 1994, observation showing that unexpected rewards lead to increases in the firing rate, fully expected rewards elicit no response, while expected rewards that failed to arrive lead to dip in the baseline firing rate. Roesch, Calu, & Schoenbaum, 2007 found the same patterns in rats while J. P. O'Doherty, Dayan, Friston, Critchley, & Dolan, 2003, found them too in fMRI studies of the human striatum; Striatal BOLD changes reflect phasic dopamine activity (Schonberg et

---

[1]To which the hypothesis was bafflingly limited

al., 2009; Surmeier, Ding, Day, Wang, & Shen, 2007) [2]. In ground breaking work, Waelti, Dickinson, & Schultz, 2001, showed that contrary to behaviorist expectations stimulus-response learning as well as the dopamenergic signal was maximized not be reward reliability (i.e. more rewards are not always better) but instead that learning was maximized when rewards were intermittent. Waelti et al., 2001, along with, Fiorillo, Tobler, & Schultz, 2003 and Bayer & Glimcher, 2005, successfully modelled changes in reward expectancy with a reward prediction error term derived from a reinforcement learning model [3] fit to each animal's behavior. The reward prediction error from the model strongly correlated with the dopamine response - both its increases and decreases. However expectancy related changes are not the only important prediction reinforcement learning models make for dopamenergic activity. Reward value must transfer. If an initially neutral cue reliable predicts a reward the reinforcement learning equations require prediction error (and thus the dopamenergic response) to transfer to the cue, thus mimicking Pavlovian conditioning. This behavior was observed in the dopamine response as well (Roesch et al., 2007; McClure, Berns, & Montague, 2003). In sum, every quantitative prediction made by reinforcement learning models has been observed in the dopamenergic signal.

Furthermore reinforcement learning models are statistically predictive of non-human animal's choice behaviors (Hampton & O'Doherty, 2007). Single doses of dopamine antagonists and agonists have also demonstrated a casual relationship between dopamine levels and learning rate (Pizzagalli et al., 2008; Diaconescu, Menon,

---

[2]Though this has recently come under question in rat models comparing single unit and field recordings to high resolution blood flow changes (Mishra et al., 2011)

[3]Specifically the Rescorla-Wagner model

Jensen, Kapur, & McIntosh, 2010), which is broadly though not exclusively consistent with a reinforcement theory interpretation. Reward prediction terms have also been shown to mediate cortical-striatal coupling (Ouden, Daunizeau, Roiser, Friston, & Stephan, 2010). Not limited to the above predictions, the reinforcement learning account has extended substantially, both theoretically and empirically.

Based on novel findings about novelty (Bunzeck & Düzel, 2006; Blatter & Schultz, 2006; Guitart-Masip, Bunzeck, Stephan, Dolan, & Duzel, 2010b) 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 (Behrens, Woolrich, Walton, & Rushworth, 2007), among others.

Additionally there are several findings which are, as yet, are unaccounted for theoretically. 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 pun-

ishment, and negatively to aversive events. Instead H. Kim, Shimojo, & O'Doherty, 2006; Matsumoto & Hikosaka, 2009, found that some neurons respond as expected but many others responded positively to aversive stimuli increasing as the punishment grew larger than expected, and decreasing as it grew smaller than expected. Rather then demonstrating an inverse coding scheme (as above), many neurons also responded positively to *both* appetitive and aversive conditions when expectations were exceeded and negatively when they predictions were optimistic. In a separate experiment K. S. Smith, Berridge, & Aldridge, 2011 demonstrated *simultaneous* tunings to both reward value, reward expectancy, salience as well as to novelty. The dopamine response also appears to adaptively scale with past reward magnitudes, similar to the reward value divided by the cumulative variance (Tobler, Fiorillo, & Schultz, 2005) If these new reports hold up and the dopamenergic response is this complex, the bivalent view needs substantial refinement, as do perhaps our analysis techniques; many different models or neural coding schemes may *correctly* fit the same data, an issue which has received some prior attention outside of neuroimaging [4] Chamberlin, 1965.

Networked Plausibility. The dopamenergic firing patterns outlined above originate in the VTA/SNc, a small brainstem nucleus whose neurons project strongly to both the striatum, prefrontal cortex and the hippocampus. Electrophysiological recordings of VTA/SNc neurons show two firing modes – tonic and phasic (?, ?). The phasic mode is of interest here, as it is this that putatively reflects the reward predictions error. A reward prediction *error* signal of course requires predictions. In

---

[4]For my take on solving this problem in fMRI, see the methods section.

this case predictions of total future reward, which itself requires both estimates of reward value, as well as the probability of receiving a reward. How exactly such predictions are made is only partly understood. Candidate regions for this calculation include the striatum, the limbic system routed through the habenula, as well at the orbital, ventral medial and the dorsal lateral frontal cortices. I present each in turn, leaving the totality largely unintegrated, thus accurately reflecting the literature's state.

Selecting Striatum. The striatum is the input area of the basal ganglia, a brain region involved in categorization, logical inference, habit formation, working memory and feedback mediated stimulus-response learning (Frank, Loughry, & O'Reilly, 2001; Jin & Costa, 2010; Schmitzer-Torbert & Redish, 2004; Seger, 2008; Seger & Miller, 2010; Yin & Knowlton, 2006). In stimulus-response 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 reward prediction error signal (Haruno & Kawato, 2006; Seger & Miller, 2010). The remaining three regions (the body and tail of the caudate and the putamen) are involved with stimulus-response 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 reward prediction error (Seger & Miller, 2010); dopamine's exact role in these

areas is less clear. In addition to projecting to basal ganglia, 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). Thus the striatum may form a value evaluation *and* action selecting loop, similar to the classic actor-critic reinforcement learning architecture (Bornstein & Daw, 2011; Ito & Doya, 2011) though this view has recently received some strong criticisms (Joel, Niv, & Ruppin, 2002).

Intact dopamine projections and complete striatal function is necessary for rapid stimulus-response learning. Administering dopamine antagonists to human and non-human animals adversely affects stimulus-response learning (Pizzagalli et al., n.d.), as does lesioning the VTA/SNc. Complete lesions of the striatum also prevent stimulus-response 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 un-related 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. Un-medicated Parkinson's patients, who have low striatal dopamine levels, show marked decreases in stimulus-response 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).

Linking with the Limbic. Recent work has highlighted the habenula (a small nucleus posterior to the thalamus) as being especially important in generating reward prediction error-like phasic activity in VTA/SNc. Besides its limbic connections (to the amygdala, hippocampus and the serotonergic dorsal raphe nucleus (Hikosaka, Sesack, Lecourtier, & Shepard, 2008)) the lateral habenula has reciprocal connections to the GPi and projections into the VTA/SNc. Based on this anatomy, it has been suggested that the habenula could serve a point of intersection between the striatum and the limbic system. Inline with this proposed role, the habenula can act to tonically inhibit or disinhibit dopamine release in VTA/SNc neurons, making VTA/SNc activity inversely correlate compared to the habenula. As habenula activity decreases, burst firing in VTA/SNc increases. Likewise as habenula firing increases, VTA/SNc activity temporarily pauses. Reversible chemical inhibition of habenula also increases VTA/SNc phasic activity (Hikosaka et al., 2008). As lesions to the habenula result in marked increases in dopamine levels in the dorsal and ventral striatum (Bromberg-Martin, Matsumoto, & Hikosaka, 2010a). Detailed dual recording studies of both areas *hint* that combined these areas are calculating the value of stimulus-response pairs (Bromberg-Martin, Matsumoto, & Hikosaka, 2010a). In summary, the GPi, withs its access to cortical inputs via the striatum, and habenula with is , with its capability for altering VTA/SNc activity, may form the physiological loop necessary to calculate (portions) of the reward prediction error.

Front and Center, Lateral Too. Orbital frontal cortex has been repeatedly shown in neuroimaging (J. O'Doherty, Kringelbach, Rolls, Hornak, & Andrews, 2001) and lesion (Hornak et al., 2004) studies to encode the absolute value of rewarding or punishing outcomes, thus playing a pivtal role in the new field of neuroeconomics (Glimcher, Dorris, & Bayer, 2005). However orbital frontal areas are more than a simple value store. They also play a role in response and outcome recall as well as responses selection (Rudebeck et al., 2008; Furuyashiki, Holland, & Gallagher, 2008), motivation, pain and pleasure (Atlas, Bolger, Lindquist, & Wager, 2010), outcome anticipation and prediction (Tanaka et al., 2006; Roesch & Olson, 2007) and causal attribution (Tanaka, Balleine, & O'Doherty, 2008). Additionally, orbital function is highly dependent on the basolateral amygdala (J. O'Doherty, 2003) offering a path for orbital activities to inform reward prediction error calculation.

TODO: Add ventral medial section; read more on it first.

Besides value, estimating the likelihood a reward will occur is the other key reward prediction calculation. Correlations with both the chance of receiving a reward (Tobler, Christopoulos, O'doherty, Dolan, & Schultz, 2009), the variance of expected value (Kahnt, Heinzle, Park, & Haynes, 2010) as well as with risk-seeking behaviors (Tobler, O'Doherty, Dolan, & Schultz, 2007) have been reported in the dorsolateral prefrontal areas. These same areas have also bee implicated in inter-temporal choice, i.e. deciding between immediate and delayed rewards (S. Kim, Hwang, Seo, & Lee, 2009; S. Kim, Hwang, & Lee, 2008). However despite a lack of complete or formal unification between striatal, prefrontal and limbic areas, several efforts have been made combining frontal and striatal function, i.e. the aforementioned cortical-

striatal loops (Frank, 2011; Seger & Miller, 2010; Frank et al., 2001; ?, ?) as well as other normative (Bar-Gad, Morris, & Bergman, 2003; Botvinick et al., 2008) and data-driven approaches (Bogacz & Gurney, 2007).

Bad Prediction, No Cookie

My hypothesis that rewards represented categorically in the brain assumes that dopamine, specifically phasic projections from the VTA/SNc to striatal and cortical areas, acts to stamp in stimulus-response relationships. I'll now take a critical look at that assumption.

Not cortex, colliculus. Recent work by Dommett et al., 2005, is a *potentially* deadly issue for the reward prediction hypothesis. The reward prediction theory requires very specific timing in order to map rewards to states and actions (i.e to solve the temporal credit assignment problem). This requirement is satisfied, with dopamine activity peaking in a 100 ms long burst about 100 ms after the initial stimulus. However 100 ms is not much time for visual processing, let alone prefrontal examination. Concerned about the plausibility of such rapid processing, Dommet *et al* disinhibited neurons in the brains of anesthetized rats in both the superior colliculus, a brainstem visual processing area, and early visual cortex with sa GABA antagonist (whichs temporarily restores neural activity in the normally unresponsive neurons of an anesthetized animal) the exposed the animals opened eyes to set of 2 Hz light pulses while recording dopamine cells in the SNc/VTA as well as neurons in both visual cortex and the superior colliculus. Disinhibition of the visual cortex lead to no

changes in dopamine firing. While superior colliculus disinhibited resulted in about half the recorded cells in the VTA/SNc to display phasic firing similar in character to that typically observed following an unexpected rewarding event. Another third of the dopamine cells displayed a pause in activity, similar to that observed when an expected reward fails to arrive (Mirenowicz & Schultz, 1994). The remaining cells responded first positively then negatively. From this the authors concluded that the superior colliculus and not the visual cortex is an effective activator of VTA/SNC neurons [5]. This is a problem as the superior colliculus responds only to very limited range of visual stimuli – appearance, disappearance, or movement of objects as well as luminance changes. It does not respond to contrast, velocity, wavelength or the geometrical configuration of stimuli (Dommett et al., 2005). That is the superior colliculus couldn't realistically extract reward information from the visual stimuli used in nearly all the studies of reward to date. Instead, Dommett *et al* argue that all of the many reward studies "can be solved based on luminance changes and/or of the position of specific reward-related visual stimuli". In other words, the expectancy-of-reward related changes in phasic dopamine, is an artifact of the task design. However this interpretation is too strong. At best they have shown that there is not substantive direct connection between visual areas and the VTA/SNc. If there was only a single downstream synapse between visual cortex and SNc/VTA (which would add around 2-10 ms in lag) their protocol would not have disinhibited it and so would have failed to elicit a dopamine response during visual stimulation.

---

[5]Another reason for more cautious interpretation: Dommet *et al* disinhibition experiments in visual cortex were very limited (N=4 cells compared to 30 for the superior colliculus experiments). Nor did they assess the extent of disinhibition in visual cortex.

That is, given the brain's high degree of inter-connectivity and probable small-world architecture (Bassett & Bullmore, 2006) a failure to find a direct anatomical relation is not in and of itself conclusive. That said, how the dopamine can be so quick and consistent remains an open and very important question.

Oh, wanting….. Standing in opposition to both the ahedonia and reward prediction hypotheses is the incentive salience account, which is derived from the brute fact that addicts often greatly want drugs of abuse, but once drugs are received addicts, or their animal model counterparts, do no report an excess of pleasure (Robinson & Berridge, 1993). And indeed pharmacological investigation of striatal areas supports distinction between wanting and liking (K. C. Berridge & Robinson, 2003). However recent experiments with dopamine deficient mice lead K. Berridge, 2007, to argue the putative dopamenergic reward signal instead signals degree of desire, i.e. wanting or in their parlance incentive salience. Tyrosine hydroxylase knockout (DD) mice, mice without detectable levels of dopamine in the brains, can learn a reward contingent T-maze tasks (where the animal must move left or right at the end of a corridor) – though they do not act on that learning till dopamine is restored (K. Berridge, 2007). DD mice do however display a reward preference when given the choice of a sweetened water versus untreated water. However unless dopamine is restored their overall desire for either is greatly decreased. Based on these findings K. Berridge, 2007, argue that dopamine is not necessary nor sufficient for reward driven learning. That is dopamine is not a casual agent in stimulus-response learning. Combined with studies of addicts (and there non-human animal equivalents) who display increased "wanting" of drugs but not "liking" (people rate

the experience as no more pleasurable than controls, mice consume no more of the substance that controls) K. Berridge, 2007, argue that phasic dopamine signals a stimulus' incentive salience, a synonym for wanting or desire. Interpretations based on DD mice though should be, I think, more constrained, as these mice 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, can have biochemical effects similar to dopamine (Lazarus et al., 2011; Rossi et al., 2010).

Taken in isolation these experiments in DD mice are quite damning to the notion of dopamine playing a casual role in stimulus-response learning. There is however a substantial body of work suggesting the opposite. Administering human and non-human animals dopamine antagonists adversely effects stimulus-response learning (Pizzagalli et al., n.d.), as does lesioning either the VTA/SNc or portions of the striatum. Complete lesions of the striatum prevent stimulus-response learning (Packard & Knowlton, 2002). This is relevant as it is the interaction between phasic dopamine and the striatum that is proposed to guide (drive) stimulus-response learning. Administering dopamine agonists, or readily converted dopamine precursor L-DOPA, leads to increased pavlovian instrumental transfer, as well as response vigor (Winterbauer & Balleine, 2007), both of which are thought to be facilitated by the interaction between phasic dopamine and activity in the ventral striatum. Parkinson's patients when taken off medication, and therefore are lacking striatal dopamine, show marked decreases in stimulus-response learning (Pizzagalli et al., n.d.). These same off-medication patients show an enhanced (compared to normal

age and intellect matched controls) capability to learn from negative feedback, suggesting their ability and desire to act is intact (Frank et al., 2004).

While it not clear how theoretically or functionally the evidence for incentive salience and reward prediction might be reconciled, it might not be necessary. K. S. Smith et al., 2011, showed distinct semi-overlapping tuning in VTA/SNc for both reward prediction, incentive salience as well as with measures of the animals enjoyment of the in-task reward (i.e. liking). I believe there may in fact be no one correct theoretical accounting; the neurons of the VTA/SNc may signal instead a family of functions - for other supporting examples see, Ito & Doya, 2011; K. S. Smith et al., 2011; Bornstein & Daw, 2011; Bromberg-Martin, Matsumoto, & Nakahara, 2010; Matsumoto & Hikosaka, 2009.

Thinking About Thinking Rewarding Thoughts

As I stated at the outset, cognition alone can generate activity similar in appearance and effect to that seen following primary and secondary rewards. For example, Tricomi & Fiez, 2008, showed ventral striatum BOLD signal changes in a declarative memory task in which subjects were initially trained with feedback 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; 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 et al., 2009; Lohrenz et al., 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 the in VTA/SNc. Inversly, neutral stimuli can prevent decreases in responding that normally accompany repeated delays in reward presentation (Reed, 1992), a phenomenon that was reproduced using a robotic rat, wherein the neutral stimuli were treated as intrinsically rewarding (Fiore et al., 2008).

Informative, or to change terms to keep with other literatures, salient [6], stimuli have been observed to have rewarding-like effects in people as well, though supporting data is limited to fMRI experiments. Striatal BOLD increases have been observed in response to infrequently presented flashing images (Zink, Pagnoni, Martin, Dhamala, & Berns, 2003), and unexpected alarming tones (e.g. a siren replacing a constant 60Hz tone) (Zink, Pagnoni, Chappelow, Martin-Skurski, & Berns, 2006). Activity in the ventral striatum appeared in these tasks due to the stimuli alone, while dorsal activity was seen only when the stimuli had behavioral relevance. A control task suggested this increase was due not to the additional motor demands of the response but was, the authors argued, due to the increased saliency of the active versus passive condition. That is the context of behavioral response made the reward more salient. A similar dorsal to ventral division has been reported when comparing passive reward receipt to reward receipt requiring a response (J. P. O'Doherty et al., 2006) though these were attributed directly to the need for an instrumental response and not (necessarily) contextual salience. Additionally, like reward expectations salience-related activity scaled with intensity (Zink et al., 2006).

---

[6]Not to be confused with the "incentive salience", discussed above.

Novel (but not necessarily salient) stimuli also elicit reward prediction-like dopamenergic firing in monkeys (Blatter & Schultz, 2006) and in people (Bunzeck & Düzel, 2006). Indeed, reward and novelty appear interchangeable. Guitart-Masip, Bunzeck, Stephan, Dolan, & Duzel, 2010a, showed that when novel images proceed rewarding outcomes enhanced ventral striatal activity compared to reward alone. Rewards proceeding a visual stimulus or word-pair leads to enhanced memory for that pair (Lisman & Grace, 2005). Building on that Wittmann, Bunzeck, Dolan, & Düzel, 2007, showed enhanced recognition of natural scenes, compared to control, when images were proceeded by novel images. This effect was reproduced using high-resolution imaging of the VTA/SNc, with that area demonstrating a marked reward prediction signal during the task. This effect was extended to the anticipation of novel images, similar to reward anticipation studies of striatal function (Knutson, Adams, Fong, & Hommer, 2001). Novelty driven exploitation/exploration decision making relies on the striatum as well (Wittmann, Daw, Seymour, & Dolan, 2008).

Task completion, imagined rewards, neutrally valued informative cues, behaviorally salient but non-rewarding, as well as novel events have all been shown to act as reinforcers, and stimulate the dopamenergic midbrain into phasic firing. None of these, individually or as a group, can be explained parsimoniously as secondary rewards [7]. They were never *statistically regularly* paired with a primary reward.

---

[7]Nor as primary rewards, but this is a definitional problem.

Generally Generalizable.

In a recent review Wimmer, Daw, & Shohamy, 2012, made the argument that a stimulus' value must be generalizable to other similar stimuli: there are far too many possible states in the world to explore each individually, *and* you rarely if ever encounter the same stimulus twice. In essence, they make an argument that categories are necessary. I'll now extend that reasoning. There are at least as manys outcomes as there are stimuli: there are therefore too many possible outcomes to ever experience them all, *and* you'll rarely see the same outcome twice. As a result, you often can't predict exactly what will happen. Therefore, outcome expectations must generalize. As desirable outcomes are rewarding (see), rewards must generalize as well.

Needed similarities. And indeed many of the cognitive rewards outlined may employ or even require generalization, often in the form of similarity assessment. Tricomi & Fiez, 2008, finding that successful task completion is rewarding used a cued recall task, which in part relies on familiarity (Jacoby, 1991) and thus similarity (R. Nosofsky, 1988). Likewise Hayden et al., 2009; Lohrenz et al., 2007 studies of fictive rewards, rewards that were experienced only in the participants imaginations, must have been derived from past memories. Assessing both salience and incentive salience requires an understanding of the current context and an assessment general goals and desirable outcomes. In fact Zink, Pagnoni, Martin-Skurski, Chappelow, & Berns, 2004, showed that reward sans salience elicited no striatal BOLD activity. Even the relatively simple seeming cases of temporal discounting of rewards and the

assessment of reward uncertainty 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 et al., 2009; Seymour & McClure, 2008). Outside of similarity, Roesch et al., 2007, demonstrated rewarding inferences in rats. When novel stimuli were presented in the context of a familiar task with the same action options previously available, the dopamine spikes that resulted were significantly correlated with the most valued past action.

Birds do it. While so far as I am aware my argument for cognitive-type rewards and their hypothesized category representation is novel, there are several non-human animal studies examining the generalizable properties of secondary reinforcers, though the neural mechanisms for such are unexamined. As an example, Guttman..., 1956, varied an instrumentally conditioned 570 nm light from 480-610, showing that while the bird's pecking rate (i.e. response vigor) decreased as one moved farther from 540, 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. Pigeons capability for generalization was not limited to simple cues, (Nakamura, Ito, Croft, & Westbrook, 2006) taught pigeons to discriminate categories of male and female birds. While J. D. Smith et al., 2011, taught pigeons to perceptually complex yet abstract categories. Interestingly though, not all experiments were effective at producing generalization behavior. Some combinations produced no responses at all (Blough, 2001; Simmons..., 2008; Urcuioli, 2001). The degree of generalization was not uniform, it depended on the

stimulus, e.g. generalization between shapes decreased more rapidly than colors Shepard, 1987. The source of this variability remained unknown until Shepard, 1987 seminal [8] insight.

Curves and categories. Shepard, 1987 demonstrated that the variability among instrumental response generalization curves in pigeons and other animals could be accounted for if one no longer measured generalized responding by an objective metric (e.g. wavelength) but instead estimated the psychological space of the animal's perception. In his own words: "I propose to start with the generalization data and to ask: Is there a unique monotone function whose inverse will transform those data into numbers interpretable as distances in some appropriate metric space". The metric space he proposed required "[...] a very strong additivity condition: For each subset of three points, the distance between the two most widely separated points equals the sum of the distances between those two points and the third point that lies between them". After making some additional assumptions [9] Shepard then goes onto show that the large variability among the many objective spaces he examined only minimally impacted the estimated (negative exponential) psychological space. In other words, in many cases the objective space can be mapped to psychological distances by simply taking the negative exponential of their euclidean distances [10].

———————

[8] ...not to mention painfully self-aggrandizing.

[9] 1. that a given stimulus is equiprobable over the psychological space and 2. that categories in the space are centrally symmetric and convex

[10] In fact, Shepard's work is correct free of this geometrical constraint, being we redone using information theory and kolmogorov complexity (Chatera & Vitányib, 2003); a lovely paper.

Based partly on Shepard's work the fields of perceptual and cognitive categorization have since made an every evolving study of how animals create and use categories. However there are still substantial controversies with, so far as I can tell, no one consensus view. The two extreme theoretical positions are exemplar models and prototype models (F. G. Ashby & Maddox, 2005). Exemplar models, in their simplest form, are feature counters, i.e. their categories are completely defined as the sum of previously observed information (R. M. Nosofsky, 1988). Prototype models instead hold only information about one ideal template (Rosch, 1973), often implemented as probability density (F. Ashby & Alfonso-Reese, 1995), i.e. a set of sufficient statistics over the relevant information. A common example is the mean and covariance, assuming prototype following matching a Gaussian distribution. In between the two extremes are hybrid approaches, of which there are many. These include using mixture models (Rossee, 2002), neural networks implementing self-organised maps (Love, 2004), and other clustering techniques (Kruschke, n.d.) as well as kernel methods (Jakel, Scholkopf, & Wichmann, 2008), among other increasingly esoteric approaches (Martin, Griffiths, & Sanborn, 2012); There are many more examples, as categorization has been an active area of mathematical psychologists investigation for over 40 years.

This deep mathematical literature is further complicated the Ashby Lab's empirical investigations of human category learning, which suggests, regardless of the underlying representation, humans treat different category structures in different ways (F. G. Ashby & Maddox, 2011). In their hands category learning has been subdivided into systems to handle rule-based (i.e. verbalizable), information-integration

(implicit, forces consideration of > 1 dimension of information), prototype-distortion (perceptual categorization), and unstructured category (categories without perceptual or semantic overlap, akin to sets in that their only necessary commonality is the fact they're grouped together). Like finding the neural correlates of these systems (F. G. Ashby & OBrien, 2005; F. Ashby & Ennis, 2006), integrating them with the many formal approaches is in its early days (F. G. Ashby & Maddox, 2011).

Why Are We Here Again?

I want to show whether cognitive rewards are represented as categories in the brain, and to examine what impact such representations have on the reinforcement learning process. To that end participants completed a stimulus-response task where classical rewards (e.g. "Correct!" or "Win $1.") were replaced with pre-trained visually and cognitively complex categories, one category for gains and one for losses. To learn the correct stimulus-response relations subjects then had to infer outcome and value. Behavioral and fMRI data collected under this task was then compared to a series of computational and functional connectivity models representing various possible algorithmic strategies for reward inference and assessing category membership. These analyses are spread across three chapters. The first covers the details of the task itself, the computational models and their parameter optimization. Chapter two covers the fMRI methods and basic brain-mapping analyses, as well as the fMRI-to-model compairson. Chapter three covers the functional analysis and its methods.

References

Ashby, F., & Alfonso-Reese, L. (1995, Jan). Categorization as probability density estimation. *Journal of Mathematical Psychology*, *39*(2), 216–233.

Ashby, F., & Ennis, J. (2006). The role of the basal ganglia in category learning. *Psychology of Learning and Motivation*, *46*, 1–36.

Ashby, F. G., & Maddox, W. T. (2005, Jan). Human category learning. *Annual review of psychology*, *56*, 149–78.

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

Ashby, F. G., & OBrien, J. B. (2005). Category learning and multiple memory systems. *Trends in Cognitive Science*, *9*(2), 83–89.

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.

Bar-Gad, I., Morris, G., & Bergman, H. (2003, Dec). Information processing, dimensionality reduction and reinforcement learning in the basal ganglia. *Progress in Neurobiology*, *71*(6), 439–73.

Bassett, D. S., & Bullmore, E. (2006). Small-world brain networks. *The Neuroscientist*, *12*(6), 512–523.

Bayer, H., & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, *47*, 129–141.

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007, Sep). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214–21.

Berridge, K. (2007). The debate over dopamine's role in reward: the case for incentive

salience. *Psychopharmacology*, *191*(3), 391–431.

Berridge, K. C., & Robinson, T. E. (2003, Sep). Parsing reward. *Trends Neurosci*, *26*(9), 507–13.

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.

Blough, D. (2001, Jan). The perception of similarity. *Avian visual cognition*.

Bogacz, R., & Gurney, K. (2007, Feb). The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural Computation*, *19*(2), 442–77.

Bornstein, A. M., & Daw, N. D. (2011, Mar). Multiplicity of control in the basal ganglia: computational roles of striatal subregions. *Current opinion in neurobiology*.

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. (2010a, Jan). Distinct tonic and phasic anticipatory activity in lateral habenula and dopamine neurons. *Neuron*.

Bromberg-Martin, E. S., Matsumoto, M., & Hikosaka, O. (2010b, Dec). Dopamine in motivational control: rewarding, aversive, and alerting. *Neuron*, *68*(5), 815–34.

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*.

Bunzeck, N., & Düzel, E. (2006, Aug). Absolute coding of stimulus novelty in the human substantia nigra/vta. *Neuron*, *51*(3), 369–79.

Calabresi, P., Picconi, B., Tozzi, A., & Filippo, M. D. (2007, May). Dopamine-mediated regulation of corticostriatal synaptic plasticity. *Trends Neurosci.*, *30*(5), 211–9.

Cannon, C. M., & Palmiter, R. D. (2003, Nov). Reward without dopamine. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *23*(34), 10827–31.

Chamberlin, T. C. (1965, May). The method of multiple working hypotheses: With this method the dangers of parental affection for a favorite theory can be circumvented. *Science*, *148*(3671), 754–9.

Chatera, N., & Vitányib, P. M. (2003). The generalized universal law of generalization. *Journal of Mathematical Psychology*, *47*(3), 346–369.

Crow, T. J. (1972, Nov). Catecholamine-containing neurones and electrical self-stimulation. 1. a review of some data. *Psychol Med*, *2*(4), 414–21.

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.

Dommett, E., Coizet, V., Blaha, C., Martindale, J., Lefebvre, V., Walton, N., et al. (2005, Mar). How visual stimuli activate dopaminergic neurons at short latency. *Science*, *307*(5714), 1476.

Fiore, V., Mannella, F., Mirolli, M., Gurney, K., Baldassarre, G., & Ricerche, C. delle. (2008). Instrumental conditioning driven by apparently neutral stimuli: A model

tested with a simulated robotic rat. *International Conference on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*, 139.

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. (2011, Jun). Computational models of motivated action selection in corticostriatal circuits. *Current opinion in neurobiology*, *21*(3), 381–6.

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. (2010a). Contextual novelty changes reward representations in the striatum. *Journal of Neuroscience*, *30*(5), 1721.

Guitart-Masip, M., Bunzeck, N., Stephan, K., Dolan, R. J., & Duzel, E. (2010b, Feb). Contextual novelty changes reward representations in the striatum. *Journal of Neuroscience*, *30*(5), 1721.

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

Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2006, Aug). The role of the ventro-

medial 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.

Hnasko, T. S., Sotak, B. N., & Palmiter, R. D. (2005, Dec). Morphine reward in dopamine-deficient mice. *Nature*, *438*(7069), 854–7.

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.

Ito, M., & Doya, K. (2011, Jun). Multiple representations and algorithms for reinforcement learning in the cortico-basal ganglia circuit. *Current opinion in neurobiology*, *21*(3), 368–73.

Iversen, S. D., & Iversen, L. L. (2007). Dopamine: 50 years in perspective. *Trends Neurosci.*, *30*(5), 188–193.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language*, *30*, 513–541.

Jakel, F., Scholkopf, B., & Wichmann, F. A. (2008, Apr). Generalization and similarity in exemplar models of categorization: Insights from machine learning. *Psychonomic Bulletin & Review*, *15*(2), 256–271.

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.

Joel, D., Niv, Y., & Ruppin, E. (2002, Jan). Actor-critic models of the basal ganglia: new anatomical and computational perspectives. *Neural Networks*, *15*(4-6), 535–47.

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.

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*.

Kim, S., Hwang, J., & Lee, D. (2008, Jul). Prefrontal coding of temporally discounted values during intertemporal choice. *Neuron*, *59*(1), 161–72.

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.

Knutson, B., Adams, C. M., Fong, G. W., & Hommer, D. (2001, Aug). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *J Neurosci*, *21*(16), RC159.

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.

Kruschke, J. (n.d.). Alcove: An exemplar-based connectionist model of category learning. *PSYCHOLOGICAL REVIEW*, *99*(1), 22–44.

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.

Love, B. C. (2004, Jan). Sustain: A network model of category learning. *PSYCHOLOGICAL REVIEW*, 309—332.

Martin, J. B., Griffiths, T. L., & Sanborn, A. N. (2012, Jan). Testing the efficiency of markov chain monte carlo with people using facial affect categories. *Cogn Sci*, *36*(1), 150–62.

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.

Mishra, A. M., Ellens, D. J., Schridde, U., Motelow, J. E., Purcaro, M. J., DeSalvo, M. N., et al. (2011, Oct). Where fmri and electrophysiology agree to disagree: corticothalamic and striatal activity patterns in the wag/rij rat. *J Neurosci*, *31*(42), 15053–64.

Montague, P., King-Casas, B., & Cohen, J. (2006). Imaging valuation models in human choice. *Annu Rev Neurosci*, *29*, 417–448.

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.

Nosofsky, R. (1988, Oct). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 700–708.

Nosofsky, R. M. (1988, Jan). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 54—65.

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.

OLDS, J., & MILNER, P. (1954, Dec). Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *J Comp Physiol Psychol*, *47*(6), 419–27.

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.

Reed, P. (1992). Signalled delay of reward: Overshadowing versus sign- tracking explanations. *Learning and Motivation*, *23*(1), 27–42.

Rescorla, R. A. (1988, Mar). Pavlovian conditioning. it's not what you think it is. *Am*

Psychol, *43*(3), 151–60.

Robinson, T. E., & Berridge, K. C. (1993, Jan). The neural basis of drug craving: an incentive-sensitization theory of addiction. *Brain Res Brain Res Rev*, *18*(3), 247–91.

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.

Rosch, E. H. (1973). Natural categories. *Cognitive Psychology*, *4*(3), 328–350.

Rossee, Y. (2002). Mixture models of categorization. *Journal of Mathematical Psychology*, *46*(2), 178–210.

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.

Salamone, J. D., Correa, M., Mingote, S. M., & Weber, S. M. (2005, Feb). Beyond the reward hypothesis: alternative functions of nucleus accumbens dopamine. *Curr Opin Pharmacol*, *5*(1), 34–41.

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.

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.

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

Simmons..., S. (2008, Jan). Individual differences in the perception of similarity and difference. *Cognition*.

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, 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, 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.

Spanagel, R., & Weiss, F. (1999, Nov). The dopamine hypothesis of reward: past and

current status. *Trends Neurosci.*, *22*(11), 521–7.

Surmeier, D. J., Ding, J., Day, M., Wang, Z., & Shen, W. (2007, May). D1 and d2 dopamine-receptor modulation of striatal glutamatergic signaling in striatal medium spiny neurons. *Trends Neurosci.*, *30*(5), 228–35.

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., Christopoulos, G. I., O'doherty, J. P., Dolan, R. J., & Schultz, W. (2009, Apr). Risk-dependent reward value signal in human prefrontal cortex. *Proceedings of the National Academy of Sciences*, *106*(17), 7185–7190.

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.

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

Tobler, P. N., O'Doherty, J. P., Dolan, R. J., & Schultz, W. (2007, Feb). Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems. *J Physiol.*, *97*(2), 1621–32.

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. . . . .*

Waelti, P., Dickinson, A., & Schultz, W. (2001, Jul). Dopamine responses comply with

basic assumptions of formal learning theory. *Nature*, *412*(6842), 43–8.

Wallis, J. D., & Kennerley, S. W. (2010, Apr). Heterogeneous reward signals in prefrontal cortex. *Current Opinion in Neurobiology*, *20*(2), 191–198.

Wimmer, G. E., Daw, N. D., & Shohamy, D. (2012, Apr). Generalization of value in reinforcement learning by humans. *Eur J Neurosci*, *35*(7), 1092–104.

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.

Wise, R. A. (1978, Aug). Catecholamine theories of reward: a critical review. *Brain Res*, *152*(2), 215–47.

Wittmann, B. C., Bunzeck, N., Dolan, R. J., & Düzel, E. (2007, Oct). Anticipation of novelty recruits reward system and hippocampus while promoting recollection. *Neuroimage*, *38*(1), 194–202.

Wittmann, B. C., Daw, N. D., Seymour, B., & Dolan, R. J. (2008, Jun). Striatal activity underlies novelty-based choice in humans. *Neuron*, *58*(6), 967–73.

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.

Zink, C. F., Pagnoni, G., Chappelow, J., Martin-Skurski, M., & Berns, G. S. (2006, Feb). Human striatal activation reflects degree of stimulus saliency. *Neuroimage*, *29*(3), 977–83.

Zink, C. F., Pagnoni, G., Martin, M. E., Dhamala, M., & Berns, G. S. (2003, Sep). Human striatal response to salient nonrewarding stimuli. *J Neurosci*, *23*(22), 8092–7.

Zink, C. F., Pagnoni, G., Martin-Skurski, M. E., Chappelow, J. C., & Berns, G. S. (2004, May). Human striatal responses to monetary reward depend on saliency. *Neuron*, *42*(3), 509–17.