Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Neurobiology

Pavlovian valuation systems in learning and decision making

Jeremy J Clark[1], Nick G Hollon[1,2,3] and Paul EM Phillips[1,2,3]

Environmental stimuli guide value-based decision making, but can do so through cognitive representation of outcomes or through general-incentive properties attributed to the cues themselves. We assert that these differences are conferred through the use of alternative associative structures differing in computational intensity. Using this framework, we review scientific evidence to discern the neural substrates of these assumed separable processes. We suggest that the contribution of the mesolimbic dopamine system to Pavlovian valuation is restricted to an affective system that is only updated through experiential feedback of stimulus–outcome pairing, whereas the orbitofrontal cortex contributes to an alternative system capable of inferential reasoning.
Finally we discuss the interactions and convergence of these systems and their implications for decision making and its pathology.

Addresses
[1] Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA 98195, United States
[2] Graduate Program in Neurobiology and Behavior, University of Washington, Seattle, WA 98195, United States
[3] Department of Pharmacology, University of Washington, Seattle, WA 98195, United States

Corresponding author: Phillips, Paul EM (pemp@uw.edu)

Current Opinion in Neurobiology 2012, **22**:1054–1061

This review comes from a themed issue on **Decision making**

Edited by **Kenji Doya** and **Michael N Shadlen**

For a complete overview see the Issue and the Editorial

Available online 28th June 2012

0959-4388/$ – see front matter, © 2012 Elsevier Ltd. All rights reserved.

http://dx.doi.org/10.1016/j.conb.2012.06.004

Introduction

Whether making a simple choice for dinner from the menu at your favorite restaurant or a complex decision such as what college to attend, choosing among competing options is a fundamental part of life. When faced with a decision, you often receive a barrage of advice including ''go with your gut'' while bearing in mind to ''look before you leap'' that might leave you wondering whether you should ''follow your heart'' or ''use your head.'' These clichés denote the intuitive separation of different types of valuation processes that are used to inform everyday decision making. Indeed, human behavior historically has been categorized as being the result of cognitive, deliberative processes or reflexive, affective, stimulus-driven processes. Such a dichotomy in the explanation of behavior has a strong philosophical foundation, and its influence also can be found in the interpretation of early work in experimental psychology [1]. Choosing advantageously among competing options is a consequence of individual subjective valuation that depends upon the learning processes required for accurately estimating rewarding outcomes based on predictive information [2]. To maintain full flexibility, individuals may use multiple systems differing in speed/accuracy tradeoffs that can be arbitrated to optimize behavior under different situations [3]. Indeed, the concept of multiple, parallel valuation systems has found a modern home in several subfields of neuroscience, psychology, and economics [2–6].

In the field of artificial intelligence and machine learning, reinforcement learning algorithms with different computational demands are often used in tandem [7]. Specifically, algorithms that are broadly classified as 'model-free' have low computational requirements and have been used in parallel with more computationally intensive 'model-based' algorithms. Derived partly from learning rules in experimental psychology [8], model-free algorithms such as temporal difference learning state that learning only occurs when the experienced value or current expectation of a particular motivational outcome deviates from that previously expected based upon environmental stimuli. Such stimuli are individually assigned values that are stored (or 'cached') and updated based upon reward-prediction errors, the degree to which this value deviates from the experienced reward. This comparison of a reward with its predictive stimulus must take place within a common currency, and the use of incentive value [9] for this currency permits predictive stimuli to elicit innate behaviors that are normally elicited by appetitive unconditioned stimuli [10]. Incentive value accounts for the motivational properties (either acquired or innate) of a stimulus but not its other discriminating sensory features such as color or shape that make up its 'identity'. Therefore, the core characteristics of a model-free valuation process are the computation of a prediction-error signal based upon experienced outcomes, the use of this signal to update the cached value assigned to the stimulus, and the ability of the stimulus to serve as an incentive based upon its cached value. As such, the computational requirements of this process are minimal. By contrast, model-based algorithms maintain a flexible model of the structure of the environment by storing specific stimulus attributes and transitions between 'states' that bind these elements in time and space. The model is accessed to make on-line inferences about associations between stimuli that have never been paired together, even when internal (goals) or external (environmental) factors change [3]. The generation of a model significantly increases the computational intensity of the process. This framework has been applied to the study of habitual and goal-directed instrumental behavior to learn optimal actions [2,3,11]. Here we highlight recent work expanding the application of a multiple-systems perspective to Pavlovian valuation to generate accurate predictive information.

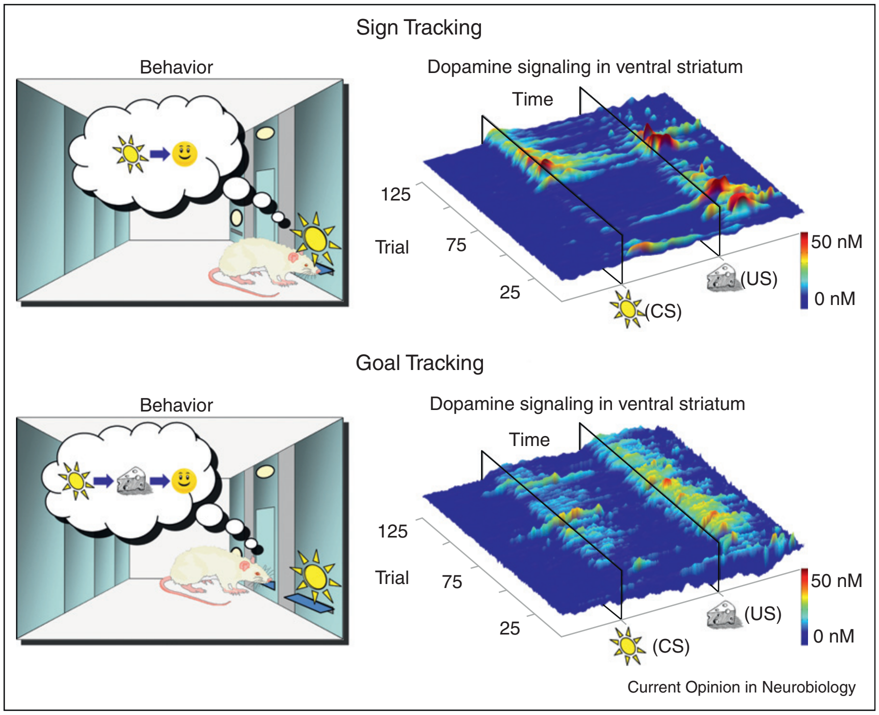
Pavlovian valuation systems

Animals, including humans, adapt their behavior to the environment by learning temporally contiguous and contingent relationships between stimuli. Associations may develop between arbitrary stimuli experienced through stimulus–stimulus co-occurrence. Alternatively, neutral cues that come into association with biologically relevant outcomes (e.g. food, water, potential mates) can co-opt innate consummatory and preparatory behavior through classical (Pavlovian) conditioning. These stimuli have potent influences over instrumental behavior and decision making. For instance, an appetitive conditioned stimulus (CS) can invigorate previously acquired instrumental behavior (Pavlovian-to-instrumental transfer), reinforce the acquisition of novel behaviors (conditioned reinforcement), or even drive behavior that is contrary to the acquisition of optimal outcomes (e.g. omission insensitivity [12]). Indeed, marketers have long taken advantage of how powerful conditioned stimuli can be in guiding valuation and decision making. Recall the song of the ice cream truck blaring through your neighborhood when you were a child. Did this evoke the image of an ice cream cone or a general feeling of happiness that drove you out the door? This distinction illustrates something that psychologists have long debated: the content of a Pavlovian association. Previous attempts to include motivational concepts in this type of learning have proposed two separable mechanisms [13,14]. One system, comprised of a direct link between a CS and conditioned response, approximates to model-free valuation algorithms; the desire to run out the door at the sound of the ice cream truck is thought to reflect preparatory behavior elicited by general incentive value. A second associative structure in which the CS elicits an explicit representation of the unconditioned stimulus (e.g. the image of an ice cream cone) aligns with model-based algorithms, as it is identity specific. The former associative structure can invigorate appetitive behavior generally, whereas the latter can bias actions selectively toward the specific outcome represented. Accordingly, Pavlovian-to-instrumental transfer (PIT) can be separated into a general invigoration of instrumental behavior and a specific bias of action selection toward the CS-associated reward [14,15,16••]. The Pavlovian influence on action selection requires a model-based representation because the effect is specific to the identity of the outcome paired with the Pavlovian stimulus. By contrast, simple model-free valuation could be sufficient to support the invigoration process. Another example of alternative Pavlovian valuation systems comes from studies of conditioned approach behavior when reward-predictive stimuli are spatially displaced from the reward-delivery location, as animals can adopt different conditioned responses [17]. During CS presentation, some animals approach the CS itself (‘sign tracking’) whereas others approach the location where the reward will be delivered (‘goal tracking’; Figure 1). These individual differences have been observed across species including humans [18]; and, importantly, both sign and goal tracking have been shown to be Pavlovian conditioned responses rather than ‘superstitious’ instrumental behaviors [17]. Sign-tracking behavior is consistent with a low-computational assessment of the reward-predictive stimulus where general incentive value is attributed to the stimulus, so it is treated similarly to unconditioned appetitive stimuli (rewards). However, the generation of goal-tracking responses requires more information than is provided by a simple incentive-attribution process, given that spatial information of reward delivery is represented upon presentation of the stimulus. A CS-evoked conditioned response that takes this form is indicative of a representational process, as the animal is approaching the location where food will be in the future. Thus, the development of sign-tracking versus goal-tracking conditioned responses signifies alternative associative processes in the use of predictive information [4,17].

Dopamine and model-free Pavlovian valuation

A role for dopamine in reward learning is suggested by data obtained from multiple paradigms [19,20,21•]. Electrophysiological studies during appetitive processing have shown transient increases in the firing rate of midbrain dopamine neurons that are time locked to unexpected, but not expected rewards, and to reward-predictive stimuli [22]. These responses can exhibit latencies so short that they do not permit significant cortical processing [23], consistent with the minimal computation characteristic of model-free valuation. The similarity between these dopaminergic responses and the reward-prediction errors used by model-free algorithms was articulated by Montague and colleagues [24]. Subsequently, the reward-prediction-error hypothesis of dopamine function has been supported by experimental evidence obtained in many scenarios examining appetitive processing [25–27,28••], particularly in more medial aspects of dopaminergic midbrain [29•] that project to limbic forebrain structures. Consistent with the hypothesis that the mesolimbic dopamine pathway transmits reward-prediction-error signals, qualitatively similar signals have been extracted from blood-oxygen-level-dependent (BOLD) hemodynamic response in the human ventral striatum, following the presentation of rewards or reward-predictive cues [30], signals that are dependent upon dopamine signaling [31]. Importantly, reward-prediction-error-like BOLD signals also have

Figure 1



Alternative Pavlovian valuation systems indicated by behavior and corresponding neurotransmission. When reward-predictive stimuli are spatially displaced from the reward-delivery location, animals can adopt different conditioned responses (left panels). The initial delivery of a primary reward elicits the appropriate unconditioned response such as approach and consumption. After repeated pairing, some animals approach the conditioned stimulus during its presentation ('sign tracking' – top panel; left) whereas others approach the location where the reward will be delivered in the future ('goal tracking' – bottom panel; left). The surface plots depict trial-by-trial fluctuations in dopamine concentration during the twenty-second window around conditioned-stimulus (CS) and unconditioned-stimulus (US) presentation over five days of training (one session per day) in both sign-tracking (top panel; right) and goal-tracking animals (bottom panel; right). Dopamine signaling in sign trackers is consistent with the reporting of a reward prediction error, as US signaling decreases across trials when it becomes predicted, in parallel with the acquisition of approach behavior. However, dopamine signaling in goal-tracking animals is not consistent with the reporting of a reward prediction error, as US signaling is maintained even when approach behavior reaches asymptote. These data are a replication of those reported in reference [34••]. Importantly, the use of selectively bred animals allowed for the demonstration that learning is dopamine dependent in animals where reward prediction errors are encoded but not in animals that lack such encoding [34••]. These findings favor the reward-prediction-error hypothesis of dopamine function in model-free learning.

been observed in dopaminergic nuclei of the midbrain [32]. These studies collectively demonstrate that dopamine signaling resembles the critical teaching signal from a model-free valuation algorithm.

If dopamine does act as a model-free teaching signal, during the implementation of the valuation process dopamine should encode a reward-prediction-error signal, learning should be dopamine dependent, and predictive stimuli should be attributed with incentive value. Moreover, these characteristics should not generalize to situations where learning is not compatible with a model-free valuation process. As previously discussed, during a Pavlovian conditioned approach task, sign tracking is indicative of a situation where the CS is attributed with incentive value. During this behavior, dopamine release in the nucleus accumbens does indeed resemble a reward-prediction-error signal [33,34••] (Figure 1), and learning is dependent upon dopamine [34••,35]. Remarkably, however, animals that adopt a goal-tracking response do not exhibit incentive attribution to the CS [36], do not encode reward-prediction-error signals by mesolimbic dopamine [34••] (Figure 1), and do not require dopamine for learning to occur [34••]. Thus, these data support the notion that dopamine is used as a teaching signal selectively when model-free valuation is utilized. Additionally, dopamine is not necessary for outcome-specific action selection bias during Pavlovian-to-instrumental transfer, even though dopamine is required for the invigoration component of this Pavlovian influence on instrumental behavior [16••]. These studies demonstrate that the role of dopamine in appetitive Pavlovian conditioning is selective to a process that resembles model-free valuation.

Orbitofrontal cortex and model-based Pavlovian valuation

In contrast to the mesolimbic dopamine system, the orbitofrontal cortex (OFC) plays a critical role when animals must flexibly update and use expectations of specific outcomes to guide their behavior [37]. Subpopulations of OFC neurons in rodents and primates encode a variety of task parameters related to the value and identity of outcomes, both actual and expected [38–45]. Many researchers emphasize the correlates of subjective expected value as a 'common neural currency' observed in a subset of OFC neurons [42,44], and consistent findings have been observed in human orbitofrontal and ventromedial prefrontal BOLD signals [46,47]. Electrophysiology studies typically find additional populations of OFC neurons that encode other parameters such as probability [44], delay [43], response requirement [44], or specific sensory features of the outcome independent of its value, that is, identity [38,42]. OFC perturbations preferentially disrupt model-based inferences about specific stimulus–outcome associations. OFC-lesioned monkeys and rats are unimpaired in the acquisition of simple Pavlovian associations for which model-free cached values may be sufficient [48,49]. They also remain capable of avoiding food that has been paired with illness or fed to satiety. However, unlike intact controls, OFC-lesioned animals continue responding to stimuli that predict these devalued outcomes. These reinforcer devaluation studies demonstrate that the OFC contributes to the animal's ability to derive the updated expected value of a cue by linking the previously learned CS-reward association with the current incentive value of that outcome without having yet directly experienced the pairing of this cue with the devalued outcome.

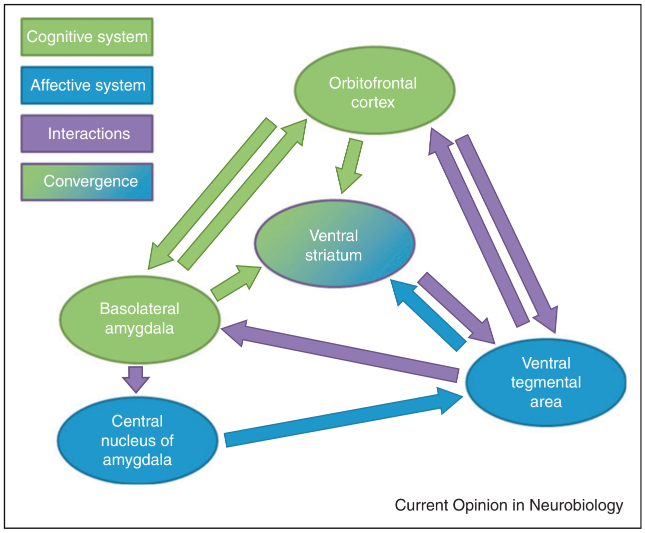
Additional behavioral procedures designed to isolate and probe the dissociable contributions of general affective value versus representations of specific outcome expectancies also have revealed that the OFC is critical for learning from changes in outcome identity. After initial acquisition, if a second cue is combined with a previously conditioned stimulus but the new compound predicts the exact same outcome, the second cue is said to be 'blocked,' and animals show minimal conditioned responding to the second cue when tested in isolation [50]. If, however, the compound stimulus yields more reward than the original CS, animals learn to respond to the new CS as well. Model-free learning algorithms are perfectly capable of accounting for such 'unblocking' owing to increases in value, and accordingly, OFC lesions do not disrupt this value-based learning [51••]. Model-free algorithms, however, cannot explain why new learning sometimes occurs when the second CS instead signals the presentation of a different outcome of equivalent value. OFC lesions disrupt such learning from changes in outcome identity [51••,52], interfere with the ability of these cue-evoked outcome expectancies to serve as conditioned reinforcers [52], and disrupt the cues' ability to selectively enhance responding for the specific outcomes the cues predict [53]. These findings collectively support the claim that the OFC is critical for representing specific features of expected outcomes signaled by predictive stimuli within a complex associative structure that permits inferences relating stimuli and states that have not been experienced [54,55].

Interaction and convergence

We have focused on the potential roles of the mesolimbic dopamine pathway and OFC in different types of Pavlovian valuation; however, these nodes are clearly part of broader circuits subserving these processes. Empirical evidence suggests that a circuit including the OFC, basolateral amygdala (BLA), and possibly the hippocampus contributes to the generation of Pavlovian values primarily through a model-based process, whereas a circuit including the central nucleus of the amygdala (CeA) and mesolimbic dopamine pathway [56] primarily generates model-free Pavlovian values. Lesions of the BLA disrupt sensitivity to devaluation and perturb specific but not general PIT, whereas amygdala CeA lesions disrupt general but not specific PIT and do not affect devaluation [57–59]. Although sign and goal tracking have not been compared in lesion studies, experiments examining sign tracking alone have demonstrated that CeA but not BLA lesions disrupt its acquisition [60] (but see [61]), and hippocampus lesions facilitate the acquisition of sign-tracking behavior [62], suggesting potential competition between these valuation systems. Such competition for control of behavior implies an arbitration process about which little is known at the neurobiological level [3] and is undoubtedly a salient topic of future investigation.

Even though these putative systems compete for the control of behavior, they are not entirely independent. In fact, the expected value correlates observed in the OFC may convey predictive information critical for the computation of reward-prediction errors (actual – *expected reward*) by dopamine neurons [37,63]. Consistent with this hypothesis, OFC stimulation predominately suppresses the activity of dopamine neurons in the ventral tegmental area (VTA) [64,65••], perhaps through activation of GABAergic VTA neurons whose collaterals form local inhibitory connections on dopamine neurons [66] and whose activity also scales with expected reward value [28••]. Following OFC lesions, dopamine neurons in the VTA remain responsive to reward outcomes and predictive cues, but do not show their typical reward-prediction-error-like modulation during learning or differential encoding of expected reward value [65••]. Modeling of these data revealed that, contrary to earlier predictions positing the OFC-derived input to provide expected value information required for calculation of prediction errors, the OFC may be critical for representing 'state' information that allows an animal to disambiguate overtly similar situations and derive more accurate outcome expectancies to guide action selection. This interpretation is consistent with the long-recognized but much-debated role of the OFC in reversal learning [37] and

Figure 2



Major neural loci of affective (cf. model-free) and cognitive (cf. model-based) Pavlovian systems and their interactions.

related decision-making tasks with dynamically changing reward contingencies [67,68•]. Recent work has also demonstrated that the firing patterns of dopamine neurons can be influenced by inferred states of the world as animals learn the latent structure of a serial reversal task in which an unexpected change in reward contingencies could itself signal a state transition [69•]. Indeed, theoretical work has provided computational accounts for such state representation and its influence on Pavlovian conditioning not generally captured by previous model-free reinforcement learning algorithms [70,71].

These circuits access common effector systems to influence motor behavior, with the mesolimbic pathway and OFC efferents converging in the ventral striatum (Figure 2), a structure previously identified as a 'limbic-motor interface' [72]. Indeed, there is evidence in humans for components of both model-free and model-based valuation being represented in this structure as measured by BOLD response [73•]. In addition, lesions of the ventral striatum disrupt learning from changes in either outcome identity or value [51••], disrupt sign-tracking behavior [74], and impair sensitivity to devaluation [75]. Lesions of the ventral striatum also disrupt PIT, with the core and shell regions of the nucleus accumbens mediating general and specific PIT, respectively [76].

Implications and conclusions

We have reviewed evidence that Pavlovian processes can be separated based upon computational intensity into at least two systems, subserved by discrete but related neural circuits. We have described an 'affective' learning system that is embedded in a serial circuit between the CeA, VTA, and the ventral striatum, and a 'cognitive' system embedded in a circuit linking OFC, BLA, and ventral striatum (Figure 2). For lucidity, we have framed these systems in the context of model-free and model-based reinforcement-learning algorithms respectively, although this designation may be an oversimplification. Additionally, a 'model-based' process that mediates Pavlovian-like associations may also be amenable to stimulus–stimulus associations where the unconditioned stimulus is not biologically significant (e.g. latent learning) and so may, in fact, represent a non-Pavlovian stimulus–stimulus associative learning process. Nonetheless, even if this system has more generalizable features, it can support associations between biologically relevant and arbitrary stimuli.

An important aspect of studying neural systems is the insight it may provide into pathological conditions. Pavlovian processes are integral to substance abuse and, in particular, Pavlovian cues can elicit drug craving, reinstatement of drug-seeking behavior, and relapse following drug abstinence [77]. Abused substances universally increase dopamine in the nucleus accumbens [78], which is considered to be antecedent to drug-related pathology [79]. Therefore, stimulus-driven behavior supported by model-free, affective Pavlovian processes is likely to be selectively upregulated by drugs [6] because of its dependence upon dopamine neurotransmission [16••,34••]. Furthermore, model-based Pavlovian processes are downregulated following drug-taking behavior [80], particularly in later stages of substance abuse when reduced metabolism in the OFC is observed [81]. Indeed, individuals who abuse drugs more strongly exhibit stimulus-driven affective behavior [82]. As mentioned above, there is little known about the biological process of arbitration between systems. Nonetheless, cognitive training in deliberative decision making, such as mindfulness, probably biases individuals toward OFC-derived behaviors and away from dopamine-dependent affective processes. Indeed, this practice is used with some success in treating drug addiction [83], presumably because it renders decision making under a biological domain, dominated by the OFC, where values assigned to predictive stimuli are not potentiated by drugs themselves and behaviors are less sensitive to stimulus-driven influences in general. Therefore, a more comprehensive understanding of the interaction, arbitration, and convergence of these valuation processes will probably provide answers to long-standing questions in the field of addiction research and more generally to the question of how individuals integrate affective and cognitive processes in value-guided decisions.

Acknowledgements

This work was supported by the National Science Foundation, the National Institute of Mental Health grant R01-MH079292 and the National Institute on Drug Abuse grant R01-DA027858. We thank Geoffrey Schoenbaum for useful feedback.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Tolman EC: **There is more than one kind of learning**. *Psychol Rev* 1949, **56**:144-155.

2. Rangel A, Camerer C, Montague PR: **A framework for studying the neurobiology of value-based decision making**. *Nat Rev Neurosci* 2008, **9**:545-556.

3. Daw ND, Niv Y, Dayan P: **Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control**. *Nat Neurosci* 2005, **8**:1704-1711.

4. Toates F: **The interaction of cognitive and stimulus-response processes in the control of behaviour**. *Neurosci Biobehav Rev* 1998, **22**:59-83.

5. Kahneman D: **Maps of bounded rationality: psychology for behavioral economics**. *Am Econ Rev* 2003, **95**:1449-1475.

6. Redish AD, Jensen S, Johnson A: **A unified framework for addiction: vulnerabilities in the decision process**. *Behav Brain Sci* 2008, **31**:415-437 discussion 437–487.

7. Sutton RS, Barto AG: *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press; 1998.

8. Rescorla RA, Wagner AR: **A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and non-reinforcement**. In *Classical Conditioning II: Current Research and Theory.* Edited by Black AH, Prokasy WF. New York: Appleton-Century-Crofts; 1972:64-99.

9. Berridge KC: **Motivation concepts in behavioral neuroscience**. *Physiol Behav* 2004, **81**:179-209.

10. McClure SM, Daw ND, Montague PR: **A computational substrate for incentive salience**. *Trends Neurosci* 2003, **26**:423-428.

11. Balleine BW, Daw ND, O'Doherty JP: **Multiple forms of value learning and the function of dopamine**. In *Neuroeconomics: Decision Making and the Brain.* Edited by Glimcher PW, Camerer CF, Fehr E, Poldrack RA. New York: Elsevier; 2009:367-387.

12. Dayan P, Niv Y, Seymour B, Daw ND: **The misbehavior of value and the discipline of the will**. *Neural Netw* 2006, **19**:1153-1160.

13. Konorski J: *Integrative Activity of the Brain*. Chicago: University of Chicago Press; 1967.

14. Rescorla RA: **Learning about qualitatively different outcomes during a blocking procedure**. *Learn Behav* 1999, **27**:140-151.

15. Kruse JM, Overmier JB, Konz WA, Rokke E: **Pavlovian conditioned stimulus effects upon instrumental choice behavior are reinforcer specific**. *Learn Motiv* 1983, **14**:165-181.

16. •• Ostlund SB, Maidment NT: **Dopamine receptor blockade attenuates the general incentive motivational effects of noncontingently delivered rewards and reward-paired cues without affecting their ability to bias action selection**. *Neuropsychopharmacology* 2012, **37**:508-519.

This work demonstrates that dopamine antagonists reduce the general invigoration of instrumental behavior by Pavlovian conditioned stimuli without biasing action selection based on the specific outcomes predicted by these reward-predictive cues.

17. Boakes RA: **Performance on learning to associate a stimulus with positive reinforcement**. In *Operant-Pavlovian Interactions.* Edited by Davis H, Hurwitz HMB. Hillsdale, NJ: Lawrence Erlbaum Associates; 1977:67-97.

18. Wilcove WG, Miller JC: **CS-USC presentations and a lever: human autoshaping**. *J Exp Psychol* 1974, **103**:868-877.

19. Wise RA: **Dopamine, learning and motivation**. *Nat Rev Neurosci* 2004, **5**:483-494.

20. Zweifel L, Parker J, Lobb C, Rainwater A, Wall V, Fadok J, Darvas M, Kim M, Mizumori S, Paladini C *et al.*: **Disruption of NMDAR-dependent burst firing by dopamine neurons provides selective assessment of phasic dopamine-dependent behavior**. *Proc Natl Acad Sci USA* 2009, **106**:7281-7288.

21. • Tsai HC, Zhang F, Adamantidis A, Stuber GD, Bonci A, de Lecea L, Deisseroth K: **Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning**. *Science* 2009, **324**:1080-1084.

This research provides the first demonstration using optogenetics that phasic stimulation selectively of VTA dopamine-containing neurons is sufficient for appetitive conditioning, as measured using a conditioned place preference procedure.

22. Ljungberg T, Apicella P, Schultz W: **Responses of monkey dopamine neurons during learning of behavioral reactions**. *J Neurophysiol* 1992, **67**:145-163.

23. Dommett E, Coizet V, Blaha CD, Martindale J, Lefebvre V, Walton N, Mayhew JE, Overton PG, Redgrave P: **How visual stimuli activate dopaminergic neurons at short latency**. *Science* 2005, **307**:1476-1479.

24. Montague PR, Dayan P, Sejnowski TJ: **A framework for mesencephalic dopamine systems based on predictive Hebbian learning**. *J Neurosci* 1996, **16**:1936-1947.

25. Waelti P, Dickinson A, Schultz W: **Dopamine responses comply with basic assumptions of formal learning theory**. *Nature* 2001, **412**:43-48.

26. Bayer HM, Glimcher PW: **Midbrain dopamine neurons encode a quantitative reward prediction error signal**. *Neuron* 2005, **47**:129-141.

27. Pan WX, Schmidt R, Wickens JR, Hyland BI: **Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network**. *J Neurosci* 2005, **25**:6235-6242.

28. •• Cohen JY, Haesler S, Vong L, Lowell BB, Uchida N: **Neuron-type-specific signals for reward and punishment in the ventral tegmental area**. *Nature* 2012, **482**:85-88.

This work used optogenetics to more conclusively identify dopaminergic and GABAergic neurons in the VTA. In addition to confirming that burst firing of dopamine neurons resembles reward-prediction errors, they found that identified GABAergic cells show persistent firing that ramps throughout CS-US intervals and scales with expected reward value, providing a possible route by which expected value signals could feed into the computation of prediction errors in dopamine neurons.

29. • Matsumoto M, Hikosaka O: **Two types of dopamine neuron distinctly convey positive and negative motivational signals**. *Nature* 2009, **459**:837-841.

This study found that a subset of dopamine neurons in dorsolateral substantia nigra pars compacta show phasic excitation to both appetitive and aversive stimuli, whereas more medial nigral and VTA dopamine neurons primarily showed valence sensitivity consistent with reward-prediction error coding.

30. O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ: **Temporal difference models and reward-related learning in the human brain**. *Neuron* 2003, **38**:329-337.

31. Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD: **Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans**. *Nature* 2006, **442**:1042-1045.

32. D'Ardenne K, McClure SM, Nystrom LE, Cohen JD: **BOLD responses reflecting dopaminergic signals in the human ventral tegmental area**. *Science* 2008, **319**:1264-1267.

33. Day JJ, Roitman MF, Wightman RM, Carelli RM: **Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens**. *Nat Neurosci* 2007, **10**:1020-1028.

34. •• Flagel SB, Clark JJ, Robinson TE, Mayo L, Czuj A, Willuhn I, Akers CA, Clinton SM, Phillips PE, Akil H: **A selective role for dopamine in stimulus-reward learning**. *Nature* 2011, **469**:53-57.

This work demonstrated that rats attributing incentive value directly to a reward-predictive cue do so in a dopamine-dependent manner and have dopamine signaling that resembles reward-prediction errors used in model-free learning algorithms. For other rats that instead learned to approach the site of reward delivery during CS presentation, dopamine antagonists did not prevent such learning, and these rats showed a pattern of dopamine transmission that was inconsistent with reward prediction error signaling.

35. Parkinson JA, Dalley JW, Cardinal RN, Bamford A, Fehnert B, Lachenal G, Rudarakanchana N, Halkerston KM, Robbins TW, Everitt BJ: **Nucleus accumbens dopamine depletion impairs both acquisition and performance of appetitive Pavlovian approach behaviour: implications for mesoaccumbens dopamine function**. *Behav Brain Res* 2002, **137**:149-163.

36. Robinson TE, Flagel SB: **Dissociating the predictive and incentive motivational properties of reward-related cues through the study of individual differences**. *Biol Psychiatry* 2009, **65**:869-873.

37. Schoenbaum G, Roesch MR, Stalnaker TA, Takahashi YK: **A new perspective on the role of the orbitofrontal cortex in adaptive behaviour**. *Nat Rev Neurosci* 2009, **10**:885-892.

38. Rolls ET, Baylis LL: **Gustatory, olfactory, and visual convergence within the primate orbitofrontal cortex**. *J Neurosci* 1994, **14**:5437-5452.

39. Critchley HD, Rolls ET: **Hunger and satiety modify the responses of olfactory and visual neurons in the primate orbitofrontal cortex**. *J Neurophysiol* 1996, **75**:1673-1686.

40. Schoenbaum G, Chiba AA, Gallagher M: **Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning**. *Nat Neurosci* 1998, **1**:155-159.

41. Tremblay L, Schultz W: **Relative reward preference in primate orbitofrontal cortex**. *Nature* 1999, **398**:704-708.

42. Padoa-Schioppa C, Assad JA: **Neurons in the orbitofrontal cortex encode economic value**. *Nature* 2006, **441**:223-226.

43. Roesch MR, Taylor AR, Schoenbaum G: **Encoding of time-discounted rewards in orbitofrontal cortex is independent of value representation**. *Neuron* 2006, **51**:509-520.

44. Kennerley SW, Wallis JD: **Evaluating choices by single neurons in the frontal lobe: outcome value encoded across multiple decision variables**. *Eur J Neurosci* 2009, **29**:2061-2073.

45. Kepecs A, Uchida N, Zariwala HA, Mainen ZF: **Neural correlates, computation and behavioural impact of decision confidence**. *Nature* 2008, **455**:227-231.

46. Hampton AN, Bossaerts P, O'Doherty JP: **The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans**. *J Neurosci* 2006, **26**:8360-8367.

47. Hare TA, O'Doherty J, Camerer CF, Schultz W, Rangel A: **Dissociating the role of the orbitofrontal cortex and the striatum in the computation of goal values and prediction errors**. *J Neurosci* 2008, **28**:5623-5630.

48. Izquierdo A, Suda RK, Murray EA: **Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency**. *J Neurosci* 2004, **24**:7540-7548.

49. Gallagher M, McMahan RW, Schoenbaum G: **Orbitofrontal cortex and representation of incentive value in associative learning**. *J Neurosci* 1999, **19**:6610-6614.

50. Kamin LJ: **Predictability, surprise, attention and conditioning**. In *Punishment and Aversive Behavior.* Edited by Campbell BA, Church RM. New York: Appleton-Century-Crofts; 1969:279-296.

51. •• McDannald MA, Lucantonio F, Burke KA, Niv Y, Schoenbaum G: **Ventral striatum and orbitofrontal cortex are both required for model-based, but not model-free, reinforcement learning**. *J Neurosci* 2011, **31**:2700-2705.

This research demonstrated that ventral striatum, but not OFC, is necessary for learning from changes in reward value. Using a trans-reinforcer blocking procedure (see also [14,52]) to isolate the specific sensory features of reward outcomes from their general affective value, they find that a value-insensitive subset of control rats learn from changes in outcome identity, whereas lesions of either OFC or ventral striatum disrupt such learning.

52. Burke KA, Franz TM, Miller DN, Schoenbaum G: **The role of the orbitofrontal cortex in the pursuit of happiness and more specific rewards**. *Nature* 2008, **454**:340-344.

53. Ostlund SB, Balleine BW: **Orbitofrontal cortex mediates outcome encoding in Pavlovian but not instrumental conditioning**. *J Neurosci* 2007, **27**:4819-4825.

54. Ostlund SB, Balleine BW: **The contribution of orbitofrontal cortex to action selection**. *Ann N Y Acad Sci* 2007, **1121**:174-192.

55. Schoenbaum G, Takahashi Y, Liu TL, McDannald MA: **Does the orbitofrontal cortex signal value?** *Ann N Y Acad Sci* 2011, **1239**:87-99.

56. Cardinal RN, Parkinson JA, Hall J, Everitt BJ: **Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex**. *Neurosci Biobehav Rev* 2002, **26**:321-352.

57. Hatfield T, Han JS, Conley M, Gallagher M, Holland P: **Neurotoxic lesions of basolateral, but not central, amygdala interfere with Pavlovian second-order conditioning and reinforcer devaluation effects**. *J Neurosci* 1996, **16**:5256-5265.

58. Corbit LH, Balleine BW: **Double dissociation of basolateral and central amygdala lesions on the general and outcome-specific forms of Pavlovian-instrumental transfer**. *J Neurosci* 2005, **25**:962-970.

59. Johnson AW, Gallagher M, Holland PC: **The basolateral amygdala is critical to the expression of Pavlovian and instrumental outcome-specific reinforcer devaluation effects**. *J Neurosci* 2009, **29**:696-704.

60. Parkinson JA, Robbins TW, Everitt BJ: **Dissociable roles of the central and basolateral amygdala in appetitive emotional learning**. *Eur J Neurosci* 2000, **12**:405-413.

61. Chang SE, Wheeler DS, Holland PC: **Effects of lesions of the amygdala central nucleus on autoshaped lever pressing**. *Brain Res* 2012, **1450**:49-56.

62. Ito R, Everitt BJ, Robbins TW: **The hippocampus and appetitive Pavlovian conditioning: effects of excitotoxic hippocampal lesions on conditioned locomotor activity and autoshaping**. *Hippocampus* 2005, **15**:713-721.

63. Takahashi YK, Roesch MR, Stalnaker TA, Haney RZ, Calu DJ, Taylor AR, Burke KA, Schoenbaum G: **The orbitofrontal cortex and ventral tegmental area are necessary for learning from unexpected outcomes**. *Neuron* 2009, **62**:269-280.

64. Lodge DJ: **The medial prefrontal and orbitofrontal cortices differentially regulate dopamine system function**. *Neuropsychopharmacology* 2011, **36**:1227-1236.

65. •• Takahashi YK, Roesch MR, Wilson RC, Toreson K, O'Donnell P, Niv Y, Schoenbaum G: **Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex**. *Nat Neurosci* 2011, **14**:1590-1597.

This work showed that OFC lesions disrupt the normal coding of reward prediction errors by VTA dopamine neurons. Modeling the effects of OFC lesions on dopamine neuron activity revealed that the OFC's contribution to optimal action selection may involve state representation rather than being required for signaling expected value information *per se*.

66. van Zessen R, Phillips JL, Budygin EA, Stuber GD: **Activation of VTA GABA neurons disrupts reward consumption**. *Neuron* 2012, **73**:1184-1194.

67. Tsuchida A, Doll BB, Fellows LK: **Beyond reversal: a critical role for human orbitofrontal cortex in flexible learning from probabilistic feedback**. *J Neurosci* 2010, **30**:16868-16875.

68. • Walton ME, Behrens TE, Buckley MJ, Rudebeck PH, Rushworth MF: **Separable learning systems in the macaque brain and the role of orbitofrontal cortex in contingent learning**. *Neuron* 2010, **65**:927-939.

Using a dynamic three-option decision-making task, this work demonstrated that OFC lesions impair monkeys' ability to accurately assign value to the specific stimuli responsible for a given reward outcome.

69. • Bromberg-Martin ES, Matsumoto M, Hong S, Hikosaka O: **A pallidus-habenula-dopamine pathway signals inferred stimulus values**. *J Neurophysiol* 2010, **104**:1068-1076.

This study found that in monkeys with extensive training on a serial reversal task, the activity of some dopamine neurons reflects the use of knowledge about the task structure to derive inferred values in addition to values updated through direct experience.

70. Redish AD, Jensen S, Johnson A, Kurth-Nelson Z: **Reconciling reinforcement learning models with behavioral extinction and**

renewal: implications for addiction, relapse, and problem gambling. *Psychol Rev* 2007, **114**:784-805.

71. Gershman SJ, Blei DM, Niv Y: **Context, learning, and extinction**. *Psychol Rev* 2010, **117**:197-209.

72. Mogenson GJ, Jones DL, Yim CY: **From motivation to action: functional interface between the limbic system and the motor system**. *Prog Neurobiol* 1980, **14**:69-97.

73. Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ: **Model-**
• **based influences on humans' choices and striatal prediction errors**. *Neuron* 2011, **69**:1204-1215.

This work in humans used a two-stage decision-making task where values could be updated according to either model-free or model-based learning strategies. Behavioral results revealed that participants used a combination of both strategies when making subsequent choices, and valuation signals incorporating both strategies were observed in the ventral striatum BOLD signal, indicating that this region processes both model-based and model-free value representations.

74. Parkinson JA, Willoughby PJ, Robbins TW, Everitt BJ: **Disconnection of the anterior cingulate cortex and nucleus accumbens core impairs Pavlovian approach behavior: further evidence for limbic cortical-ventral striatopallidal systems**. *Behav Neurosci* 2000, **114**:42-63.

75. Singh T, McDannald MA, Haney RZ, Cerri DH, Schoenbaum G: **Nucleus accumbens core and shell are necessary for reinforcer devaluation effects on Pavlovian conditioned responding**. *Front Integr Neurosci* 2010, **4**:126.

76. Corbit LH, Balleine BW: **The general and outcome-specific forms of Pavlovian-instrumental transfer are differentially mediated by the nucleus accumbens core and shell**. *J Neurosci* 2011, **31**:11786-11794.

77. Shaham Y, Shalev U, Lu L, De Wit H, Stewart J: **The reinstatement model of drug relapse: history, methodology and major findings**. *Psychopharmacology (Berl)* 2003, **168**:3-20.

78. Di Chiara G, Imperato A: **Drugs abused by humans preferentially increase synaptic dopamine concentrations in the mesolimbic system of freely moving rats**. *Proc Natl Acad Sci USA* 1988, **85**:5274-5278.

79. Kalivas PW, Volkow ND: **The neural basis of addiction: a pathology of motivation and choice**. *Am J Psychiatry* 2005, **162**:1403-1413.

80. Lucantonio F, Stalnaker TA, Shaham Y, Niv Y, Schoenbaum G: **The impact of orbitofrontal dysfunction on cocaine addiction**. *Nat Neurosci* 2012, **15**:358-366.

81. Volkow ND, Fowler JS, Wang GJ, Hitzemann R, Logan J, Schlyer DJ, Dewey SL, Wolf AP: **Decreased dopamine D2 receptor availability is associated with reduced frontal metabolism in cocaine abusers**. *Synapse* 1993, **14**:169-177.

82. Bickel WK, Marsch LA: **Toward a behavioral economic understanding of drug dependence: delay discounting processes**. *Addiction* 2001, **96**:73-86.

83. Bowen S, Chawla N, Collins SE, Witkiewitz K, Hsu S, Grow J, Clifasefi S, Garner M, Douglass A, Larimer ME *et al.*: **Mindfulness-based relapse prevention for substance use disorders: a pilot efficacy trial**. *Subst Abus* 2009, **30**:295-305.