Published in final edited form as:

Neuropharmacology. 2014 January ; 76(0 0): . doi:10.1016/j.neuropharm.2013.05.033.

Transition from 'model-based' to 'model-free' behavioral control in addiction: involvement of the orbitofrontal cortex and dorsolateral striatum

Federica Lucantonio^{1,2}, Daniele Caprioli^{3,4}, and Geoffrey Schoenbaum²

¹Department of Anatomy and Neurobiology, University of Maryland School of Medicine, Baltimore, MD, 21201

²NIDA, Intramural Research Program, Baltimore, MD, 21224

³Behavioural and Clinical Neuroscience Institute, University of Cambridge, Cambridge, CB2 3EB, UK

⁴Department of Experimental Psychology, University of Cambridge, Cambridge CB2 3EB, UK2

Abstract

Cocaine addiction is a complex and multidimensional process involving a number of behavioral and neural forms of plasticity. The behavioral transition from voluntary drug use to compulsive drug taking may be explained at the neural level by drug-induced changes in function or interaction between a flexible planning system, associated with prefrontal cortical regions, and a rigid habit system, associated with the striatum. The dichotomy between these two systems is operationalized in computational theory by positing model-based and model-free learning mechanisms, the former relying on an "internal model" of the environment and the latter on pre-computed or cached values to control behavior. In this review, we will suggest that model-free and model-based learning mechanisms appear to be differentially affected, at least in the case of psychostimulants such as cocaine, with the former being enhanced while the latter are disrupted. As a result, the behavior of long-term drug users becomes less flexible and responsive to the desirability of expected outcomes and more habitual, based on the long history of reinforcement. To support our specific proposal, we will review recent neural and behavioral evidence on the effect of psychostimulant exposure on orbitofrontal and dorsolateral striatum structure and function.

Addiction is characterized by a failure of the decision-making system (Redish et al., 2008, Lucantonio et al., 2012). As outlined in the DSM IV, drug addicts lose control over drug seeking and drug taking behaviors, pursuing drug use despite negative and adverse consequences or even an explicit desire to stop doing so. These decision-making deficits may result from drug-induced neuroadaptations in brain areas important for controlling judgment and decision-making. The brain contains multiple, interacting systems for controlling decision-making (Cardinal et al., 2002, Balleine and O'Doherty, 2010). More flexible behavior, controlled by representations of future outcomes, is preferentially governed by regions of the prefrontal cortex (O'Doherty, 2011, Rushworth et al., 2011), while automatic or habitual behavior, depending on pre-computed or cached values, is

Correspondence should be addressed to G.S. (schoenbaumgm@nida.nih.gov).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

principally dependent on the dorsolateral striatum (Jog et al., 1999, Balleine et al., 2007). Both systems are of obvious benefit and the combination of their integrative and differential roles contribute to the selection of the appropriate behavior from potential options, which is the goal of the decision-making (Dickinson, 2002, van der Meer et al., 2012).

The dichotomy between these two systems is captured in computational neuroscience by the concepts of model-based and model-free reinforcement learning (Daw et al., 2005, Dayan and Daw, 2008). Importantly, this computational distinction encompass and describe both instrumental and pavlovian behaviors, aiming to understand how our choices and behavior are driven through basic learning mechanisms. In particular, model-based learning relies on a learned 'internal model' of the world to make decisions. The internal model includes predictions about the immediate consequences to follow each state in a sequence. Thus, this model can be used to mentally simulate events and outcomes and compute values for each available option 'on the fly'. This approach is computationally expensive, but it endows the actor with the ability to predict specific features of upcoming events and their current value, thereby allowing greater behavioral precision and flexibility. The prefrontal cortex has been shown to be crucial for implementing this model (Doll et al., 2012).

By contrast, the model-free system relies on cached or stored values to influence behavior. These values are determined by the reinforcement history of each available option, and thus are simply scalar numbers, blind to the identity or other defining features of the expected future outcome. This approach is simple and fast but relatively inflexible, because stored values do not immediately update if the current value or desirability of the outcome changes. Resultant decisions are based on past history more than future consequences. As a result of this inflexibility, it has been proposed that model-free learning underlies habitual behavior, and the dorsolateral striatum, already demonstrated to be involved in automatic behavior (Balleine et al., 2007), seems to be the most promising candidate as a neural correlate of exclusively model-free signals (Montague et al., 1996, Wunderlich et al., 2012).

A rich body of work using these computational tools has already broadened our understanding of the microstructure of behavior and the neural circuits underlying value-based decision-making (O'Doherty et al., 2003, Seymour et al., 2004). However this computational framework may also shed light on behavior and learning in conditions where the entire computational machine is going awry, such as in drug addiction (Redish, 2004). In particular, a common theme running through addiction theories is that the initial drug use is mediated by the positive reinforcing effects of the drug, but the transition to addiction may be linked to a reduced behavioral control with a concomitant development of drug-seeking habits (Jentsch and Taylor, 1999, Volkow and Fowler, 2000, Everitt and Robbins, 2005, Berridge et al., 2009).

Here we attempt to understand this shift by using the distinction between model-free and model-based behavioral control. We suggest that these two learning mechanisms appear to be differentially affected by cocaine. In particular, model-based representations dependent on orbitofrontal cortex appear to be impaired and model-free representations in dorsolateral striatum enhanced by exposure to cocaine and other psychostimulants. As a result, behavior - both decisions and learning - is less flexible or reflective of new goals and more driven by past experience or habitual. This results in behavior that appears poorly considered or maladaptive relative to expressed desires in the present and a failure to learn normally from new or unexpected outcomes to improve decision-making in the future.

To support our specific proposal, we will review recent neural and behavioral evidence in humans and animals on the effect of cocaine exposure on orbitofrontal and dorsolateral

striatum structure and function, although we will include also other psychostimulants when appropriate.

Differential roles of the dorsolateral striatum and orbitofrontal cortex in model-free versus model-based behavior

The model-free system endows cues with pre-computed value based on their history of reinforcement. As a result, while model-free learning is fast and efficient on one hand, it is also limited in its flexibility (Niv et al., 2006). This limitation is easily captured by experiments in which decisions or learning are especially dependent on new or specific information regarding expected outcome rather than that pre-computed value. Typically such experiments involve reinforcer devaluation (Holland and Straub, 1979). In this procedure, after an animal learns to obtain food, the food is devalued by pairing it with lithium chloride, which induces nausea and aversion to the food, or by giving free access to the food to induce satiation. Subsequently, the animal's conditioned responding is tested in extinction conditions. Normally, a spontaneous decrease in responding for the devalued outcome is observed during the probe test. This initial change in behavior cannot be explained by the model-free system, since model-free representations require experience to modify their cached values; their predictions contain no information about the specific features of upcoming rewards or their current value. Since the devalued food is never experienced again, changes in behavior in this situation cannot be based on the experience. Rather it requires access to an internal model of the previously acquired associations and the current value of the food, from which a new value can be computed on the fly to drive behavior. Notably, in instrumental contexts, overtraining makes test performance insensitive to reinforcer devaluation: these animals no longer change learned responding as a result of reinforcer devaluation (Holman, 1975, Colwill and Rescorla, 1988) (Balleine and Dickinson, 1998, Killcross and Coutureau, 2003). This finding has also been replicated in humans (Tricomi et al., 2009b). Overtrained subjects do not show any devaluation effect and, interestingly, a portion of the putamen (homologous human region of the murine DLS) manifested an increased activation compared to undertrained subjects. Thus, when subjects have extensive practice in a stable domain, decision-making become automated and need no longer be based on a complex representation of the world and model-free representations seem to be the most accurate. Importantly, inactivation of DLS in overtrained rats reverses this effect, rendering animals again sensitive to alteration of the outcome value (Yin et al., 2004). This result indicates that the DLS is important for implementing model-free or habitual behavior.

The involvement of DLS in signaling cached values of cues, independent of specific information about outcomes, has also been shown using Pavlovian-to-instrumental transfer (PIT), a paradigm that examines the modulatory role of reward-paired stimuli on performance of an independently acquired reward-seeking behavior (Corbit and Janak, 2007). Inactivation of the DLS has been shown to affect transfer when it is based on general information about value but not when it is based on specific information about the outcome. Hence, the DLS appears not to represent the associative structure of the environment, but it is a good candidate for storing the pre-computed or cached values that characterize model-free behavioral control.

By contrast, although lesion, imaging and recording studies have implicated several regions in encoding model-based information, such as regions of the prefrontal cortex (Owen, 1997, de Wit et al., 2009) or the dorsomedial striatum (Yin et al., 2005, Shiflett et al., 2010), an extensive literature indicates that the OFC is critical for guiding behavior and learning in situations when specific information about expected outcomes is necessary, information that must be derived from a model-based representation of the structure of the task or

environment (McDannald et al., 2012). For example, normal performance in both Pavlovian and instrumental versions of the reinforcer devaluation task depends on the OFC (Gallagher et al., 1999, Pickens et al., 2003, Izquierdo et al., 2004, Pickens et al., 2005) (Gremel C.M. et al., Soc. Neurosc. Abst. 699.06/CCC40, 2012). Furthermore anticipatory neuronal activity in OFC changes when the value of an outcome is reduced by satiety (Critchley and Rolls, 1996, Tremblay and Schultz, 1999) and functional studies demonstrate that the blood oxygen level-dependent activity in OFC correlates with changes in the value of expected outcome (O'Doherty et al., 2000, Gottfried et al., 2003, Valentin et al., 2007). Taken together, these findings confirm the involvement of the OFC in computing the distinctive value of specific expected outcomes that are necessary for flexible, motivation-dependent evaluations of model-based representations.

The OFC has also been shown to be crucial for other tasks that require subjects to access model-based information for calculating the current value of a particular outcome, such as reversal learning (Schoenbaum et al., 2003, Ghods-Sharifi et al., 2008, Rudebeck and Murray, 2008, Rygula et al., 2010), differential outcome expectancy (Overmier et al., 1971), outcome-specific Pavlovian-to-instrumental transfer (Ostlund and Balleine, 2007), outcome-specific conditioned reinforcement (Burke et al., 2008), and most recently in sensory preconditioning (Jones et al., 2012). Moreover, information about outcomes signaled by OFC is critical both for guiding behavior and also for learning. This is demonstrated in the performance of OFC-lesioned or -inactivated animals in Pavlovian overexpectation and also when blocking or unblocking is driven by inferred or specific information about outcomes rather than pre-computed value (Burke et al., 2008, Takahashi et al., 2009, McDannald et al., 2011, Jones et al., 2012).

Pavlovian over-expectation is particularly noteworthy as it involves a form of extinction learning (Kremer, 1978). In this paradigm, several cues are separately trained as signals for reward. Subsequently, two of these cues are presented in compound followed by the same reward. This is the "overexpectation" stage in which animals expect to receive two rewards from the compound cue, one from each cue. Rats typically show increased responding initially but then learn to adjust their predictions in light of the delivery of only one reward. Subsequent test trials with the cues by themselves (in extinction) typically reveal a reduction in conditioned responding to the compounded cues. This effect is thought to reflect extinction induced by violation of summed expectations for reward during compound training. Inactivation of the OFC prevent both the increased responding during compound training and extinction in response to over-expectation. These results suggest that the OFC is crucial for integrating existing simple associative representations in order to create novel predictions or estimates about future outcomes. By contrast, rats tested on simple extinction exhibited normal performance in response to reward omission after OFC lesions or inactivation (Burke et al., 2009, Takahashi et al., 2009). Unlike over-expectation, extinction by reward omission can be accomplished by relying only on stored values of the individual cues, provided by the model-free system.

At this point we hope to have made clear the operational difference between model-free and model-based learning and the role of the DLS as a model-free actor and the OFC as a crucial model-based critic (Hampton et al., 2006, Valentin et al., 2007, Hampton et al., 2008). Next, we will discuss effects of cocaine-exposure on behaviors that depend on these two brain areas.

Altered model-free and model-based representations by cocaine exposure

We have seen that overtraining makes performance insensitive to changes in reward value in reinforcer devaluation tasks (Holman, 1975, Colwill and Rescorla, 1988, Balleine and

Dickinson, 1998, Killcross and Coutureau, 2003, Tricomi et al., 2009a) and that the DLS is involved in this deficit (Yin et al., 2004). Nelson and Killcross (Nelson and Killcross, 2006) have demonstrated that animals sensitized with amphetamine show no devaluation effect, even after limited training (Fig. 1). Animals were sensitized with amphetamine for 7 consecutive days and after a week they were trained to press a lever for a reward and then tested after the instrumental outcome was devalued by specific satiety or LiCl-induced nausea. Amphetamine-sensitized animals failed to modify their lever-press performance in response to the changed value of the outcome, responding at equivalent levels to those seen in non-devalued controls. While this might reflect dysfunction in other model-based areas, in a recent study, Zapata and colleagues (Zapata et al., 2010) showed that such insensitivity to outcome devaluation can be reversed by inactivation of DLS in cocaine trained animals. They combined a drug seeking/taking chained schedule of intravenous cocaine administration and reward devaluation methods. In particular, responding on the designated drug seeking lever provides access to a drug taking lever. Responses on the drug taking lever result in cocaine infusions. With extended training on the chained schedule, devaluation of the drug taking lever by extinction, that is the outcome of the drug seeking lever, does not reduce drug seeking behavior, measured as lever presses on the drug seeking lever. However, when the DLS is transiently inactivated, outcome devaluation was effective in decreasing drug seeking.

As described earlier, the DLS is also required for general-affective but not outcome-specific PIT (Colwill, 1988). A key recent study (Shiflett, 2012) has demonstrated that drug exposure selectively enhances general PIT effects, while it abolishes specific transfer (Fig. 2a). Animals learned the associations between two different cues and rewards. Then, they received instrumental training in which two responses lead to the same outcomes. Following Pavlovian and instrumental training, rats were sensitized to amphetamine for 7 days. Finally, the two stimuli were tested in extinction conditions for specific or general transfer effects. Drug exposure abolished specific transfer in which each stimulus selectively enhanced responding for the same outcome, but enhanced general transfer in which each stimulus enhanced responding for both outcomes. A similar enhancement of the general transfer effect has also been shown by Wyvell and Berridge (Fig. 2b) (Wyvell and Berridge, 2001). They have shown that sensitization caused by previous amphetamine administration causes reward cues to trigger excessive pursuit of their associated reward, even when sensitized rats are tested in a drug-free state. These results suggest that drug-experienced rats' representations are blind to the identity of the expected future reward, as we have seen being a core feature of the model-free system, while perhaps over-representing the general or cached value that is common to both. These deficits are appropriately positioned for a critical role in drug relapse, in which drug cues trigger excessive and compulsive drug-seeking behavior even after long periods of abstinence (de Wit and Stewart, 1981, Robbins and Everitt, 1999).

Sensitization has also been shown to enhance extinction or acquisition of simple Pavlovian associations. For example, psychostimulant-exposed rats have been demonstrated to show enhanced conditioned responding (Harmer and Phillips, 1998, Taylor and Horger, 1999, Taylor and Jentsch, 2001), higher sensitivity to cues predicting large rewards (Roesch et al., 2007a), enhanced conditioned inhibition (Harmer and Phillips, 1999), and enhanced general-affective conditioned reinforcement (Di Ciano, 2008). This simple effect could easily be explained by stronger model-free representations.

Overall, these results suggest a generalized drug-induced increase in the efficacy of neural processing in the DLS. Accordingly, exposure to psychostimulants has been shown to affect activity of single units recorded in DLS in awake, behaving animals. We have recorded single-unit activity in ventral striatum and dorsal striatum in rats exposed to cocaine

approximately 1–3 months earlier (Takahashi et al., 2007). Neurons were recorded as rats learned and reversed odor discrimination problems in which unique odor cues were paired with a sucrose reward and a quinine punishment (Schoenbaum et al., 1999). In both groups neurons fire to the cues based on their associations with subsequent responses and/or outcomes. However, cocaine shifts the balance of associative encoding between these two regions, abolishing the strong cue selectivity normally present in ventral striatum while enhancing the weak cue selectivity normally present in dorsal striatum. It is worth noting that the dorsal striatum is unique among the brain regions where we have recorded in cocaine-treated rats in that it maintains or even enhances its normal pattern of cue-evoked activity (Fig. 3) (Stalnaker et al., 2006, Stalnaker et al., 2007).

At the same time that DLS-dependent behaviors seem enhanced, a striking parallel exists between the behavioral effects of OFC lesions or inactivation and the long-lasting effects of chronic psychostimulant exposure. The first example comes from the Pavlovian reinforcer devaluation task, described earlier (Schoenbaum and Setlow, 2005). Cocaine-experienced animals fail to change their conditioned responding in response to devaluation of the food, showing a deficit identical to that seen in OFC-lesioned animals (Fig. 4). These animals show normal conditioning and taste aversion learning, and they also extinguish conditioned responding normally in the absence of the food reward; however, the devaluation effect is altered, suggesting that cocaine exposure impairs the ability to represent and use information about expected outcomes.

This is consistent with recent results obtained using the sensory preconditioning task in cocaine exposed animals (Wied H. et al., Winter Conference on Brain Research Abst. P37). In sensory preconditioning, a subject learns to pair two cues together and one of these cues predicts a reward (Brogden, 1939). A conditioned response to both the reward-paired cue and the preconditioned cue is manifested. The response to the preconditioned cue must be inferred based on the knowledge of the associative structure of the task while the response to the reward-paired cue can be computed using stored values acquired during past experience. Rats previously trained to self-administer cocaine, as we have seen earlier for OFC-inactivated animals, fail to show conditioned response to the preconditioned cue, while leaving unimpaired the response driven by the reward-paired cue (Jones et al., 2012). These results confirm that the performance of drug exposed animals is impaired in tasks in which expected values must be inferred from model-based representations of the task.

Other behavioral procedures sensitive to OFC dysfunction give more evidence that psychostimulants impair the ability to guide behavior on the basis of specific predictions about expected future outcomes. For example, several studies in rats, monkeys and humans demonstrate that reversal learning is impaired after psychostimulant exposure (Jentsch et al., 2002, Ersche et al., 2008, Schoenbaum and Shaham, 2008). Moreover, delay discounting tasks have shown that psychostimulant exposure enhances impulsive choices in humans and rats (Coffey et al., 2003, Kirby and Petry, 2004, Heil et al., 2006, Roesch et al., 2007b, Simon et al., 2007). Overall, these results suggest that cocaine exposure alters the ability of animals to compute the distinctive value of specific expected outcomes that are necessary for flexible, motivation-dependent evaluations of model-based representations.

While the above deficits might reflect processing changes in other areas, single unit recording shows that psychostimulant exposure affects signaling of outcome expectancies by OFC (Thorpe et al., 1983, Schoenbaum et al., 1998). For example, rats learning the contingencies between cues and outcomes exhibit distinct populations of neurons in OFC that respond selectively to one of the outcomes. After the rats learn the predictive associations, a subset of these neurons fire selectively for the odor that predicts their preferred outcome (Schoenbaum et al., 1998). Thus, the OFC activity seems to reflect a

representation of the expected outcome at the time of cue sampling, which can guide decision-making and successive learning if expectations or contingencies are violated. Recording in cocaine-exposed animals show that OFC neurons do not develop these neural correlates at the time of odor sampling (Fig. 5) (Stalnaker et al., 2006).

Of course, as reviewed above, outcome expectancies signaled by OFC are not only crucial for guiding behavior but they also contribute directly to learning, as when specific features of the expected reward are changed, but general cached value is stable. This function has been demonstrated in settings like overexpectation, described earlier. Consistent with the proposal that drug exposure disrupts this function of OFC, cocaine-experienced rats show deficits identical to those seen in OFC-inactivated animals (Lucantonio F. et al., Soc. Neurosc. Abst. 707.707/MMM709, 2010). Specifically cocaine-exposed animals tested in the overexpectation task fail to show summation during compound training and also fail to show the spontaneous reduction in responding to the compounded cue in the probe test. Thus, it seems that cocaine-experienced animals are unable to integrate reward expectations during compound training to drive learning. In contrast, the same animals show no deficit in extinction driven by simple omission of reward, which does not require any integration of expectancies but can rely on cached values. We have also shown that OFC signaling in this task reflects a real-time integration of the cue-evoked expectations for reward, necessary to generate novel estimates about future outcomes and drive learning (Takahashi Y. et al., Soc. Neurosc. Abst. 404.18/MMM19, 2010). However, preliminary work in cocaine-exposed animals has shown that this signal is altered, suggesting that cocaine disrupts the normal ability of networks in OFC to signal estimates about future outcomes, normally used for both adjusting current behavior and driving learning (Lucantonio F. et al., Soc. Neurosc. Abst. 198.16/WW39, 2011).

Conclusions

Cocaine addiction is characterized by poor judgment and maladaptive decision-making. Here we have reviewed evidence to support the proposal that such maladaptive behaviors associated in addicted subjects may be due, at least in part, to a drug-induced change in the balance between the model-based and model-free systems (Figure 6). The transition between the two systems is consistent with impairments in the ability to generate and retrieve specific information about expected outcomes and an enhanced utilization of inflexible, cached network of values as a consequence of drug exposure. These changes can affect behavior and learning in complex ways. In simple situations in which model-free information is sufficient, learning and behavior could appear even to be enhanced or more efficient in drug-exposed subjects. However, in more complex situations in which expected values must be inferred from model-based representations of the task, learning and behavior can be impaired. Hence, these computational tools offer potential insights into the puzzling inability of addicted people to learn to modify drug-seeking behaviors in complex, real-world situations. In particular, they shed light on our understanding of the role of pavlovian and instrumental interactions in the development of compulsive drug taking. We know that the drug acts as an instrumental reinforcer, while environmental cues associated with the effects of the drug gain incentive salience through pavlovian processes. Interaction between these pavlovian and instrumental processes lead to the transition from voluntary drug use to compulsive drug taking (Everitt and Robbins, 2005). Computational tools may be used to disentangle and analyze these phenomena from a more basic prospective, in order to study the shared learning mechanisms, and how they become aberrant with drug exposure.

Interestingly, similar to what we have shown with cocaine exposure, stress seems to modulate prefrontal cortex mechanisms and dorsolateral striatum processes in a way that biases towards model-free strategies (Shansky and Lipps, 2013). Indeed, evidence has

shown that stress disrupts synaptic plasticity and cognitive control processing in prefrontal cortex (Alexander et al., 2007, Diamond et al., 2007, Steinhäuser et al., 2007), although it facilitates striatum-dependent processes (Schwabe et al., 2007, Quirarte et al., 2009). The similarity between cocaine and stress effects on model-free and model-based learning mechanisms is extremely intriguing, especially because stress is a critical risk factor for the development of addiction as well as for relapse to drug seeking after abstinence (Piazza and Le Moal, 1996, Sinha, 2001). Hence, from an ethological view, the strategy to reduce complex processing towards more primitive behaviors under stressful conditions may have huge impact in order to aid survival but, in extreme cases, it may be detrimental and participate to development of mental disorders, such as drug addiction, by preventing flexible adaptation. Further studies may elucidate in which extent cocaine exposure and stressful conditions participate together to the shift between model-based to model-free behavioral control.

Acknowledgments

The work of writing it was supported by funding from the intramural programs of NIDA.

References

- Alexander JK, Hillier A, Smith RM, Tivarus ME, Beversdorf DQ. Beta-adrenergic modulation of cognitive flexibility during stress. *J Cogn Neurosci*. 2007; 19:468–478. [PubMed: 17335395]
- Balleine BW, Delgado MR, Hikosaka O. The role of the dorsal striatum in reward and decision-making. *J Neurosci*. 2007; 27:8161–8165. [PubMed: 17670959]
- Balleine BW, Dickinson A. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology*. 1998; 37:407–419. [PubMed: 9704982]
- Balleine BW, O'Doherty JP. Human and rodent homologues in action control: corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacol*. 2010; 35:48–69.
- Berridge KC, Robinson TE, Aldridge JW. Dissecting components of reward: 'liking', 'wanting', and learning. *Curr Opin Pharmacol*. 2009; 9:65–73. [PubMed: 19162544]
- Brogden WJ. Sensory pre-conditioning. *J Exp Psychol*. 1939; 25:323.
- 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. [PubMed: 18563088]
- Burke KA, Takahashi YK, Correll J, Brown PL, Schoenbaum G. Orbitofrontal inactivation impairs reversal of Pavlovian learning by interfering with 'disinhibition' of responding for previously unrewarded cues. *Eur J Neurosci*. 2009; 30:1941–1946. [PubMed: 19912335]
- Cardinal RN, Parkinson JA, Hall J, Everitt BJ. Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci Biobehav R*. 2002; 26:321–352.
- Coffey SF, Gudleski GD, Saladin ME, Brady KT. Impulsivity and rapid discounting of delayed hypothetical rewards in cocaine-dependent individuals. *Exp Clin Psychopharmacol*. 2003; 11:18–25. [PubMed: 12622340]
- Colwill RM, Rescorla RA. The Role of Response-Reinforcer Associations Increases Throughout Extended Instrumental Training. *Anim Learn Behav*. 1988; 16:105–111.
- Colwill, RMaRAR. The Role of Response-Reinforcer Associations Increases Throughout Extended Instrumental Training. *Animal Learning & Behavior*. 1988; 16:105–111.
- Corbit LH, Janak PH. Inactivation of the lateral but not medial dorsal striatum eliminates the excitatory impact of Pavlovian stimuli on instrumental responding. *J Neurosci*. 2007; 27:13977–13981. [PubMed: 18094235]
- 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. [PubMed: 8727405]
- Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci*. 2005; 8:1704–1711. [PubMed: 16286932]

- Dayan P, Daw ND. Decision theory, reinforcement learning, and the brain. *Cogn Affect Behav Neurosci*. 2008; 8:429–453. [PubMed: 19033240]
- de Wit H, Stewart J. Reinstatement of cocaine-reinforced responding in the rat. *Psychopharmacology (Berl)*. 1981; 75:134–143. [PubMed: 6798603]
- de Wit S, Corlett PR, Aitken MR, Dickinson A, Fletcher PC. Differential engagement of the ventromedial prefrontal cortex by goal-directed and habitual behavior toward food pictures in humans. *J Neurosci*. 2009; 29:11330–11338. [PubMed: 19741139]
- Di Ciano P. Facilitated acquisition but not persistence of responding for a cocaine-paired conditioned reinforcer following sensitization with cocaine. *Neuropsychopharmacol*. 2008; 33:1426–1431.
- Diamond DM, Campbell AM, Park CR, Halonen J, Zoladz PR. The temporal dynamics model of emotional memory processing: a synthesis on the neurobiological basis of stress-induced amnesia, flashbulb and traumatic memories, and the Yerkes-Dodson law. *Neural Plast*. 2007; 2007:60803. [PubMed: 17641736]
- Dickinson, AB.; B, W. The role of learning in motivation. In: Gallistei, C., editor. *Stevens' Handbook of experimental psychology*. New York, NY: Wiley; 2002.
- Doll BB, Simon DA, Daw ND. The ubiquity of model-based reinforcement learning. *Curr Opin Neurobiol*. 2012
- Ersche KD, Roiser JP, Robbins TW, Sahakian BJ. Chronic cocaine but not chronic amphetamine use is associated with perseverative responding in humans. *Psychopharmacology (Berl)*. 2008; 197:421–431. [PubMed: 18214445]
- Everitt BJ, Robbins TW. Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nat Neurosci*. 2005; 8:1481–1489. [PubMed: 16251991]
- Gallagher M, McMahon RW, Schoenbaum G. Orbitofrontal cortex and representation of incentive value in associative learning. *J Neurosci*. 1999; 19:6610–6614. [PubMed: 10414988]
- Ghods-Sharifi S, Haluk DM, Floresco SB. Differential effects of inactivation of the orbitofrontal cortex on strategy set-shifting and reversal learning. *Neurobiol Learn Mem*. 2008; 89:567–573. [PubMed: 18054257]
- Gottfried JA, O'Doherty J, Dolan RJ. Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science*. 2003; 301:1104–1107. [PubMed: 12934011]
- 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. [PubMed: 16899731]
- Hampton AN, Bossaerts P, O'Doherty JP. Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proc Natl Acad Sci U S A*. 2008; 105:6741–6746. [PubMed: 18427116]
- Harmer CJ, Phillips GD. Enhanced appetitive conditioning following repeated pretreatment with d-amphetamine. *Behav Pharmacol*. 1998; 9:299–308. [PubMed: 10065918]
- Harmer CJ, Phillips GD. Enhanced conditioned inhibition following repeated pretreatment with d-amphetamine. *Psychopharmacology (Berl)*. 1999; 142:120–131. [PubMed: 10102763]
- Heil SH, Johnson MW, Higgins ST, Bickel WK. Delay discounting in currently using and currently abstinent cocaine-dependent outpatients and non-drug-using matched controls. *Addict Behav*. 2006; 31:1290–1294. [PubMed: 16236455]
- Holland PC, Straub JJ. Differential effects of two ways of devaluing the unconditioned stimulus after Pavlovian appetitive conditioning. *J Exp Psychol Anim Behav Process*. 1979; 5:65–78. [PubMed: 528879]
- Holman EW. Some conditions for the dissociation of consummatory and instrumental behavior in rats. *Learn Motiv*. 1975; 6:358–366.
- 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. [PubMed: 15329401]
- Jentsch JD, Olsson P, De La Garza R 2nd, Taylor JR. Impairments of reversal learning and response perseveration after repeated, intermittent cocaine administrations to monkeys. *Neuropsychopharmacology*. 2002; 26:183–190. [PubMed: 11790514]

- Jentsch JD, Taylor JR. Impulsivity resulting from frontostriatal dysfunction in drug abuse: implications for the control of behavior by reward-related stimuli. *Psychopharmacology (Berl)*. 1999; 146:373–390. [PubMed: 10550488]
- Jog MS, Kubota Y, Connolly CI, Hillegaart V, Graybiel AM. Building neural representations of habits. *Science*. 1999; 286:1745–1749. [PubMed: 10576743]
- Jones JL, Esber GR, McDannald MA, Gruber AJ, Hernandez A, Mirens A, Schoenbaum G. Orbitofrontal cortex supports behavior and learning using inferred but not cached values. *Science*. 2012; 338:953–956. [PubMed: 23162000]
- Killcross S, Coutureau E. Coordination of actions and habits in the medial prefrontal cortex of rats. *Cereb Cortex*. 2003; 13:400–408. [PubMed: 12631569]
- Kirby KN, Petry NM. Heroin and cocaine abusers have higher discount rates for delayed rewards than alcoholics or non-drug-using controls. *Addiction*. 2004; 99:461–471. [PubMed: 15049746]
- Kremer EF. The Rescorla-Wagner model: losses in associative strength in compound conditioned stimuli. *J Exp Psychol Anim Behav Process*. 1978; 4:22–36. [PubMed: 621465]
- Lucantonio F, Stalnaker TA, Shaham Y, Niv Y, Schoenbaum G. The impact of orbitofrontal dysfunction on cocaine addiction. *Nat Neurosci*. 2012; 15:358–366. [PubMed: 22267164]
- 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. [PubMed: 21325538]
- McDannald MA, Takahashi YK, Lopatina N, Pietras BW, Jones JL, Schoenbaum G. Model-based learning and the contribution of the orbitofrontal cortex to the model-free world. *Eur J Neurosci*. 2012; 35:991–996. [PubMed: 22487030]
- Montague PR, Dayan P, Sejnowski TJ. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J Neurosci*. 1996; 16:1936–1947. [PubMed: 8774460]
- Nelson A, Killcross S. Amphetamine exposure enhances habit formation. *J Neurosci*. 2006; 26:3805–3812. [PubMed: 16597734]
- Niv Y, Daw ND, Dayan P. Choice values. *Nat Neurosci*. 2006; 9:987–988. [PubMed: 16871163]
- O'Doherty J, Rolls ET, Francis S, Bowtell R, McGlone F, Kobal G, Renner B, Ahne G. Sensory-specific satiety-related olfactory activation of the human orbitofrontal cortex. *Neuroreport*. 2000; 11:893–897. [PubMed: 10757540]
- O'Doherty JP. Contributions of the ventromedial prefrontal cortex to goal-directed action selection. *Ann N Y Acad Sci*. 2011; 1239:118–129. [PubMed: 22145881]
- 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. [PubMed: 12718865]
- Ostlund SB, Balleine BW. Orbitofrontal cortex mediates outcome encoding in Pavlovian but not instrumental conditioning. *J Neurosci*. 2007; 27:4819–4825. [PubMed: 17475789]
- Overmier JB, Bull JA 3rd, Trapold MA. Discriminative cue properties of different fears and their role in response selection in dogs. *J Comp Physiol Psychol*. 1971; 76:478–482. [PubMed: 5127953]
- Owen AM. Cognitive planning in humans: neuropsychological, neuroanatomical and neuropharmacological perspectives. *Prog Neurobiol*. 1997; 53:431–450. [PubMed: 9421831]
- Piazza PV, Le Moal ML. Pathophysiological basis of vulnerability to drug abuse: role of an interaction between stress, glucocorticoids, and dopaminergic neurons. *Annu Rev Pharmacol Toxicol*. 1996; 36:359–378. [PubMed: 8725394]
- Pickens CL, Saddoris MP, Gallagher M, Holland PC. Orbitofrontal lesions impair use of cue-outcome associations in a devaluation task. *Behav Neurosci*. 2005; 119:317–322. [PubMed: 15727536]
- Pickens CL, Saddoris MP, Setlow B, Gallagher M, Holland PC, Schoenbaum G. Different roles for orbitofrontal cortex and basolateral amygdala in a reinforcer devaluation task. *J Neurosci*. 2003; 23:11078–11084. [PubMed: 14657165]
- Quirarte GL, de la Teja IS, Casillas M, Serafin N, Prado-Alcala RA, Roozendaal B. Corticosterone infused into the dorsal striatum selectively enhances memory consolidation of cued water-maze training. *Learn Mem*. 2009; 16:586–589. [PubMed: 19794182]
- Redish AD. Addiction as a computational process gone awry. *Science*. 2004; 306:1944–1947. [PubMed: 15591205]

- Redish AD, Jensen S, Johnson A. A unified framework for addiction: Vulnerabilities in the decision process. *Behav Brain Sci.* 2008; 31 415–+.
- Robbins TW, Everitt BJ. Drug addiction: bad habits add up. *Nature.* 1999; 398:567–570. [PubMed: 10217139]
- Roesch MR, Takahashi Y, Gugs N, Bissonette GB, Schoenbaum G. Previous cocaine exposure makes rats hypersensitive to both delay and reward magnitude. *J Neurosci.* 2007a; 27:245–250. [PubMed: 17202492]
- Roesch MR, Takahashi Y, Gugs N, Bissonette GB, Schoenbaum G. Previous cocaine exposure makes rats hypersensitive to both delay and reward magnitude. *J Neurosci.* 2007b; 27:245–250. [PubMed: 17202492]
- Rudebeck PH, Murray EA. Amygdala and orbitofrontal cortex lesions differentially influence choices during object reversal learning. *J Neurosci.* 2008; 28:8338–8343. [PubMed: 18701696]
- Rushworth MF, Noonan MP, Boorman ED, Walton ME, Behrens TE. Frontal cortex and reward-guided learning and decision-making. *Neuron.* 2011; 70:1054–1069. [PubMed: 21689594]
- Ryguła R, Walker SC, Clarke HF, Robbins TW, Roberts AC. Differential contributions of the primate ventrolateral prefrontal and orbitofrontal cortex to serial reversal learning. *J Neurosci.* 2010; 30:14552–14559. [PubMed: 20980613]
- Schoenbaum G, Chiba AA, Gallagher M. Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. *Nat Neurosci.* 1998; 1:155–159. [PubMed: 10195132]
- Schoenbaum G, Chiba AA, Gallagher M. Neural encoding in orbitofrontal cortex and basolateral amygdala during olfactory discrimination learning. *J Neurosci.* 1999; 19:1876–1884. [PubMed: 10024371]
- Schoenbaum G, Setlow B. Cocaine makes actions insensitive to outcomes but not extinction: implications for altered orbitofrontal-amygdalar function. *Cereb Cortex.* 2005; 15:1162–1169. [PubMed: 15563719]
- Schoenbaum G, Setlow B, Nugent SL, Saddoris MP, Gallagher M. Lesions of orbitofrontal cortex and basolateral amygdala complex disrupt acquisition of odor-guided discriminations and reversals. *Learn Mem.* 2003; 10:129–140. [PubMed: 12663751]
- Schoenbaum G, Shaham Y. The role of orbitofrontal cortex in drug addiction: a review of preclinical studies. *Biol Psychiatry.* 2008; 63:256–262. [PubMed: 17719014]
- Schwabe L, Oitzl MS, Philippson C, Richter S, Bohringer A, Wippich W, Schachinger H. Stress modulates the use of spatial versus stimulus-response learning strategies in humans. *Learn Mem.* 2007; 14:109–116. [PubMed: 17272656]
- Seymour B, O'Doherty JP, Dayan P, Koltzenburg M, Jones AK, Dolan RJ, Friston KJ, Frackowiak RS. Temporal difference models describe higher-order learning in humans. *Nature.* 2004; 429:664–667. [PubMed: 15190354]
- Shansky RM, Lipps J. Stress-induced cognitive dysfunction: hormone-neurotransmitter interactions in the prefrontal cortex. *Front Hum Neurosci.* 2013; 7:123. [PubMed: 23576971]
- Shiflett MW. The effects of amphetamine exposure on outcome-selective Pavlovian-instrumental transfer in rats. *Psychopharmacology (Berl).* 2012; 223:361–370. [PubMed: 22562522]
- Shiflett MW, Brown RA, Balleine BW. Acquisition and performance of goal-directed instrumental actions depends on ERK signaling in distinct regions of dorsal striatum in rats. *J Neurosci.* 2010; 30:2951–2959. [PubMed: 20181592]
- Simon NW, Mendez IA, Setlow B. Cocaine exposure causes long-term increases in impulsive choice. *Behav Neurosci.* 2007; 121:543–549. [PubMed: 17592945]
- Sinha R. How does stress increase risk of drug abuse and relapse? *Psychopharmacology (Berl).* 2001; 158:343–359. [PubMed: 11797055]
- Stalnaker TA, Roesch MR, Franz TM, Burke KA, Schoenbaum G. Abnormal associative encoding in orbitofrontal neurons in cocaine-experienced rats during decision-making. *Eur J Neurosci.* 2006; 24:2643–2653. [PubMed: 17100852]
- Stalnaker TA, Roesch MR, Franz TM, Calu DJ, Singh T, Schoenbaum G. Cocaine-induced decision-making deficits are mediated by miscoding in basolateral amygdala. *Nat Neurosci.* 2007; 10:949–951. [PubMed: 17603478]

- Steinhauser M, Maier M, Hubner R. Cognitive control under stress: how stress affects strategies of task-set reconfiguration. *Psychol Sci.* 2007; 18:540–545. [PubMed: 17576268]
- Takahashi Y, Roesch MR, Stalnaker TA, Schoenbaum G. Cocaine exposure shifts the balance of associative encoding from ventral to dorsolateral striatum. *Front Integr Neurosci.* 2007; 1
- 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. [PubMed: 19409271]
- Taylor JR, Horger BA. Enhanced responding for conditioned reward produced by intra-accumbens amphetamine is potentiated after cocaine sensitization. *Psychopharmacology (Berl).* 1999; 142:31–40. [PubMed: 10102780]
- Taylor JR, Jentsch JD. Repeated intermittent administration of psychomotor stimulant drugs alters the acquisition of Pavlovian approach behavior in rats: differential effects of cocaine, d-amphetamine and 3,4- methylenedioxymethamphetamine ("Ecstasy"). *Biol Psychiatry.* 2001; 50:137–143. [PubMed: 11526995]
- Thorpe SJ, Rolls ET, Maddison S. The orbitofrontal cortex: neuronal activity in the behaving monkey. *Exp Brain Res.* 1983; 49:93–115. [PubMed: 6861938]
- Tremblay L, Schultz W. Relative reward preference in primate orbitofrontal cortex. *Nature.* 1999; 398:704–708. [PubMed: 10227292]
- Tricomi E, Balleine BW, O'Doherty JP. A specific role for posterior dorsolateral striatum in human habit learning. *Eur J Neurosci.* 2009a; 29:2225–2232. [PubMed: 19490086]
- Tricomi E, Balleine BW, O'Doherty JP. A specific role for posterior dorsolateral striatum in human habit learning. *Eur J Neurosci.* 2009b; 29:2225–2232. [PubMed: 19490086]
- Valentin VV, Dickinson A, O'Doherty JP. Determining the neural substrates of goal-directed learning in the human brain. *J Neurosci.* 2007; 27:4019–4026. [PubMed: 17428979]
- van der Meer M, Kurth-Nelson Z, Redish AD. Information processing in decision-making systems. *Neuroscientist.* 2012; 18:342–359. [PubMed: 22492194]
- Volkow ND, Fowler JS. Addiction, a disease of compulsion and drive: involvement of the orbitofrontal cortex. *Cereb Cortex.* 2000; 10:318–325. [PubMed: 10731226]
- Wunderlich K, Dayan P, Dolan RJ. Mapping value based planning and extensively trained choice in the human brain. *Nat Neurosci.* 2012; 15:786–791. [PubMed: 22406551]
- Wyvell CL, Berridge KC. Incentive sensitization by previous amphetamine exposure: increased cue-triggered "wanting" for sucrose reward. *J Neurosci.* 2001; 21:7831–7840. [PubMed: 11567074]
- Yin HH, Knowlton BJ, Balleine BW. Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. *Eur J Neurosci.* 2004; 19:181–189. [PubMed: 14750976]
- Yin HH, Ostlund SB, Knowlton BJ, Balleine BW. The role of the dorsomedial striatum in instrumental conditioning. *Eur J Neurosci.* 2005; 22:513–523. [PubMed: 16045504]
- Zapata A, Minney VL, Shippenberg TS. Shift from goal-directed to habitual cocaine seeking after prolonged experience in rats. *J Neurosci.* 2010; 30:15457–15463. [PubMed: 21084602]

Figure 1. Effect of amphetamine on changes in lever pressing as a result of reinforcer devaluation in rats

Rats were sensitized with amphetamine for 7 consecutive days and then trained to press a lever for a food reward. Subsequently, the food was devalued by pairing it with LiCl-induced illness and lever presses were assessed in a final probe session. Shown is the number of lever presses per minute as a proportion of baseline after devaluation, in the extinction probe test. Filled bars: 'non-devalued' groups, open bars: 'devalued' groups. During the devaluation test, amphetamine-sensitized rats failed to change lever pressing as a result of prior devaluation of the food reward. (** $p < 0.01$). Adapted from ref. (Nelson and Killcross, 2006).

Figure 2. Effect of amphetamine on Pavlovian-instrumental transfer (PIT) performance
(A) Rats underwent separate Pavlovian and instrumental training for two types of pellet rewards (grain- and chocolate-flavored pellets). Subsequently, they were sensitized with amphetamine for 7 consecutive days and tested on an outcome-selective PIT task. During the test, one lever was inserted into the chamber and the two cues associated with each reward were presented in pseudorandom order. Responses were classified as a ‘Match’ if the cue and the instrumental response shared a common outcome, whereas responses were classified as a ‘Mismatch’ if the response and the cue did not share an outcome. Shown is the number of lever presses per minute as a proportion of baseline in the extinction probe test. Filled bars: ‘match’ responses, open bars: ‘mismatch’ responses. During the test, amphetamine-sensitized rats failed to manifest outcome-selective transfer, as shown by performing Match and Mismatch responses at a similar rate. (*p < 0.01). **(B)** Rats were separately trained to lever press for sucrose pellets and to associate sucrose with a Pavlovian cue (CS+: conditioned stimulus). After amphetamine sensitization, rats were tested for lever pressing under extinction conditions during intermittent presentations of the CS+. Shown is the number of lever pressing as a proportion of baseline in the extinction probe test. Filled bars: ‘CS+’ responses, open bars: ‘CS–’ responses. (CS– : cue never paired with the sucrose). During the test, amphetamine-sensitized rats engaged in greater increases in lever pressing during the cue presentation compared to control animals. (*p < 0.02). Adapted from ref. (Wyvell and Berridge, 2001, Shiflett, 2012).

Figure 3. Effect of cocaine on associative encoding in DS neurons during discrimination and reversal learning

Rats were trained to sample odors at a central port and then respond at a nearby fluid well. In each odor problem, one odor predicted sucrose and a second quinine. Rats had to learn to respond for sucrose and to inhibit responding to avoid quinine, and then the odor-outcome associations were reversed. (A) Shown are population responses to each odor cue for cue-selective DS neurons and all DS neurons (small inset) during pre- and post-criterion trials (required criterion: 90% correct responses). Gray shading in each histogram indicates the timing of odor sampling. Bar graph within each histogram shows averaged firing rates of each odor sampling phase. The neural population in cocaine-exposed rats exhibited stronger phasic firing to the odor cue that predicted the sucrose outcome. (B) Population of neurons which show nonselective (black), sucrose cue-selective (blue), and quinine cue-selective (red) firing. Cocaine treatment caused an increase in the responsiveness of DS neurons to the positive odor cue. (* $p < 0.05$, ** $p < 0.01$). Adapted from ref. (Takahashi et al., 2007).

Figure 4. Effect of cocaine on changes in conditioned responding as a result of reinforcer devaluation in rats

Rats were trained to associate a light cue with food. Subsequently, the food was devalued by pairing it with LiCl-induced illness and response to the cue was assessed in a final probe session. Shown is the mean percentage of time spent in the food cup during presentation of the food-predicting cue after devaluation, in the extinction probe test. Filled bars, 'devalued' groups; open bars, 'non-devalued' groups. During the devaluation test, cocaine-exposed rats fail to change conditioned responding as a result of previous devaluation of the food reward.

* $P < 0.05$. Adapted from ref. (Schoenbaum and Setlow, 2005).

Figure 5. Effect of cocaine on associative encoding in OFC neurons during discrimination and reversal learning

Rats were trained in the same task described in Figure 3. Shown is an index of odor selectivity for each sucrose- or quinine-expectant neuron, which is neuron that responds selectively to one of the two outcomes. Positive values indicate greater firing to the sucrose-predicting cue and negative values indicate greater firing to the quinine-predicting odor cue. The number of neurons that developed significant odor selectivity is shown by green bars for the sucrose-expectant population and by red bars for the quinine-expectant population. Neurons in cocaine-exposed animals failed to fire selectively for the odor that predicts their preferred outcome at the time of cue sampling. Adapted from ref. (Stalnaker et al., 2006).

Figure 6. The proposed model for model-based to model-free behavioral control transition in cocaine addiction

A summary of the proposal reviewed in the text for explaining the transition between model-based to model-free behavior control in cocaine addiction. At the top is the healthy state in which the two learning mechanisms, OFC-dependent model-based and DS-dependent model-free, dynamically interplay to facilitate appropriate behavior. Repeated cocaine exposure gradually alters OFC and DS structures and functions. In particular, due to these neuroadaptations, model-based representations appear to be impaired and model-free representations enhanced by cocaine exposure. As a result, behavior is less flexible or reflective of new goals and more driven by past experience or rigid. The green ovals at the

top figure represent OFC and DS regions in a healthy state. The red ovals at the bottom figure represent OFC and DS areas in an addicted state. Changes in font size of OFC and DS between the top and the bottom figures indicate the diminished model-based control by OFC and the enhanced model-free control by DS in an addicted brain. The red arrow on the right indicates the transition between model-based to model-free control to be mediated by cocaine exposure. OFC, orbitofrontal cortex; DS, dorsal striatum.