Review

A Primer on Foraging and the Explore/Exploit Trade-Off for Psychiatry Research

MA Addicott*,[1], JM Pearson[2], MM Sweitzer[1], DL Barack[3] and ML Platt[4]

[1]Department of Psychiatry and Behavioral Sciences, Duke University, Durham, NC, USA; [2]Center for Cognitive Neuroscience, Duke University, Durham, NC, USA; [3]Department of Philosophy and Neuroscience, Columbia University, New York, NY, USA; [4]Departments of Psychology, Neuroscience, and Marketing, University of Pennsylvania, Philadelphia, PA, USA

Foraging is a fundamental behavior, and many types of animals appear to have solved foraging problems using a shared set of mechanisms. Perhaps the most common foraging problem is the choice between exploiting a familiar option for a known reward and exploring unfamiliar options for unknown rewards—the so-called explore/exploit trade-off. This trade-off has been studied extensively in behavioral ecology and computational neuroscience, but is relatively new to the field of psychiatry. Explore/exploit paradigms can offer psychiatry research a new approach to studying motivation, outcome valuation, and effort-related processes, which are disrupted in many mental and emotional disorders. In addition, the explore/exploit trade-off encompasses elements of risk-taking and impulsivity—common behaviors in psychiatric disorders—and provides a novel framework for understanding these behaviors within an ecological context. Here we explain relevant concepts and some common paradigms used to measure explore/exploit decisions in the laboratory, review clinically relevant research on the neurobiology and neuroanatomy of explore/exploit decision making, and discuss how computational psychiatry can benefit from foraging theory.

The explore/exploit trade-off is relatively new to psychiatry but already has a rich history in behavioral ecology and computational neuroscience research. All organisms that search for food or other resources make explore/exploit decisions; thus, explore/exploit paradigms are excellent tools for translational research. There is also strong experimental support for the underlying neurobiology and neuroanatomy that regulates these decisions. In contrast to many psychiatry measures that rely on overall task summations of risky or impulsive behaviors, advantageous explore/exploit decision-making relies on trial-by-trial updates of reward value estimates and flexible, adaptive behavioral changes in response to environmental uncertainties (Sutton and Barto, 1998). Furthermore, mathematical modeling of these explore/exploit decisions provides quantitative assessment of underlying behavioral mechanisms. For these reasons, we believe foraging, in general, and the explore/exploit trade-off, in particular, can provide a powerful new framework for understanding how disrupted decision-making mechanisms contribute to psychiatric disorders (for review, see Barack and Platt, 2016; Pearson *et al*, 2014; Stephens *et al*, 2007; Stephens and Krebs, 1986). In this review, we provide a beginner's guide to the explore/exploit trade-off in four parts.

First, we explain the concepts behind the explore/exploit trade-off. Second, we describe the paradigms and parameters used to measure explore/exploit decisions. Third, we review recent research on the neurobiology and neuroanatomy of explore/exploit decision making, followed by a review of its application in psychiatric research. Finally, we discuss future directions and how computational psychiatry can benefit from foraging theory.

WHAT IS THE EXPLORE/EXPLOIT TRADE-OFF?

Imagine lunch time has arrived and you must make a decision about what to eat. You can go to the nearby deli and order your usual sandwich, or you could try the new restaurant that just opened next door. What should you do?

Foraging is easily understood as the search for food, but it also encompasses a broad range of behaviors that support survival. All mobile animals must forage for resources—such as food, shelter, and mates—in the face of environmental uncertainty and limited abundance. One important problem faced in foraging is the explore/exploit trade-off, which is the decision between choosing a familiar option with a known reward value or choosing an unfamiliar option with an unknown or uncertain reward value. This unfamiliar option may be more or less valuable than your familiar option, but either way there are time and energy costs that must be paid for this information. Exploring a new restaurant means spending time and money on food before you know how much you like it. For other animals, exploring new territory

*Correspondence: Dr MA Addicott, Psychiatry and Behavioral Sciences, Duke University, 402 Trent Dr, Durham, NC 27705, USA, Tel: +919 684 5237, Fax: +919 681 0016, E-mail: merideth.addicott@duke.edu

means spending time and energy looking for food that might not be found there.

This raises a fundamental question: when is the right time to assume this risk and explore? Efficient performance (ie, maximizing the rate of rewards obtained and minimizing the costs—such as time, effort, or money—expended to obtain them) is a careful and deliberate balance between exploration and exploitation. Exploitation maximizes rewards in the near-term, while the information obtained during exploration can later be used to maximize rewards in the long-term (Barack and Gold, 2016). In an uncertain and changing environment, where values of all potential options are unknown and/or the values of these options change over time, one must adapt by flexibly alternating between exploration and exploitation in order to maintain efficient performance over time and to keep track of the state of the environment.

This duality raises a question of whether explore/exploit decisions are qualitatively different. Exploratory decisions could be separate processes in which automatic, exploitative decisions are actively suppressed in order to consider other possibilities. Alternatively, explore/exploit decisions could be better described as extreme ends of a continuous scale (Berger-Tal et al, 2014; Cohen et al, 2007), too much exploitation could promote habit formation (Beeler et al, 2014), and too much exploration may result in an individual who is 'jack of all trades, but master of none'. When explore/exploit decisions are balanced, the uncertainty of exploration can be reduced by exploiting information or past experience with similar options; for instance, patronizing a new restaurant that serves a particular cuisine that you have enjoyed in the past. There may be externally or internally driven biases towards the exploratory or exploitative end of the continuum. For instance, in the summer months when food is abundant, a forager should spend more time exploring potential food sources; this information will then be exploited in the winter months when food is scarce. In addition, for many species, adolescence is marked by a period of increased risky, exploratory behavior (eg, Laviola et al, 2003); knowledge gained from these formative experiences will be exploited later in adulthood (also see Mata et al, 2013). Personality traits may also bias an individual's decisions and may even influence their career choice (eg, Laureiro-Martinez et al, 2014), and society benefits from this diversity. Extreme biases in explore/exploit decisions may be advantageous if these behaviors are adaptations to the environment. Otherwise, extreme biases are (most likely) disadvantageous, and may be a symptom of an underlying psychiatric disorder, such as addiction. See Figure 1.

The explore/exploit trade-off is a broad problem faced by foragers, and this trade-off can be influenced by solutions to more specific problems, such as prey selection, time horizon, and patch leaving. For example, patch leaving refers to the decision to leave one patch of food in search of another patch, given that sources of food clump together in unevenly distributed patches and the value of a patch decreases as the forager consumes the food there (for review, see Stephens and Krebs, 1986). Importantly, optimal solutions to these foraging problems can be mathematically predicted, and extensive research has shown that animals conform to these predictions (with some exceptions, see Constantino and
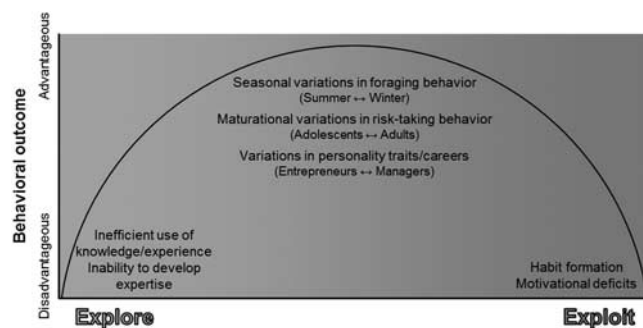

Figure 1 Conceptual overview of the explore/exploit trade-off. Decisions may vary along a continuum between exploration and exploitation with the most advantageous behaviors occurring around a point of balance between the two. Around this balance, there may be slight externally or internally driven biases towards one decision vs the other. Extreme exploitative or exploratory decisions that are not adaptations to the environment may be disadvantageous, leading to either inefficiency and lack of expertise (ie, overly exploratory) or habit formation and motivational deficiencies (ie, overly exploitative).

Daw, 2015; Hayden et al, 2011; Charnov, 1976; Stephens and Krebs, 1986; Stephens and Dunlap, 2011). Therefore, it is our opinion that foraging models tap into a deep neurobehavioral decision-making schemata pertinent to the health and fitness of an individual.

HOW ARE EXPLORE/EXPLOIT DECISIONS MEASURED IN THE LAB?

There are three behavioral paradigms that have been widely used in research on explore/exploit decisions. The most common is the n-armed bandit task, based on the n-armed bandit learning problem. The n-armed bandit is analogous to a slot machine (a.k.a. one-armed bandit) with n levers (Gittins and Jones, 1974). For instance, a 4-armed bandit task (Daw et al, 2006), presents four options (ie, slot machines) and the player is free to select any one of them. After an option is selected, the reward value (ie, points) for that option on that trial is shown briefly and the next trial begins. Option values are non-stationary, that is, the values of each option change gradually and independently of one another. Players learn the current value of an option by selecting it, and they must continually balance exploiting the option with the highest expected value with exploring lesser-known options in order to track their relative value and ensure exploitation of the best option. Option values are predetermined by an algorithm in which they drift around a specified mean using a fixed standard deviation for step size. The underlying option values across the trials are predetermined and identical for each player. See Figure 2.

A variant of the bandit task is the 2-armed 'leap frog' task (Knox et al, 2011). Here, there are two options, one option always has higher rewards than the other, and the value of each option is revealed after its selection. There is a fixed probability on each trial that the option with the smaller reward can increase in value, thereby becoming the better option. Since the relative values of the two options change over time, the player chooses between selecting the option with the highest known reward and sampling the alternative to see whether its value is now the greater of the two.
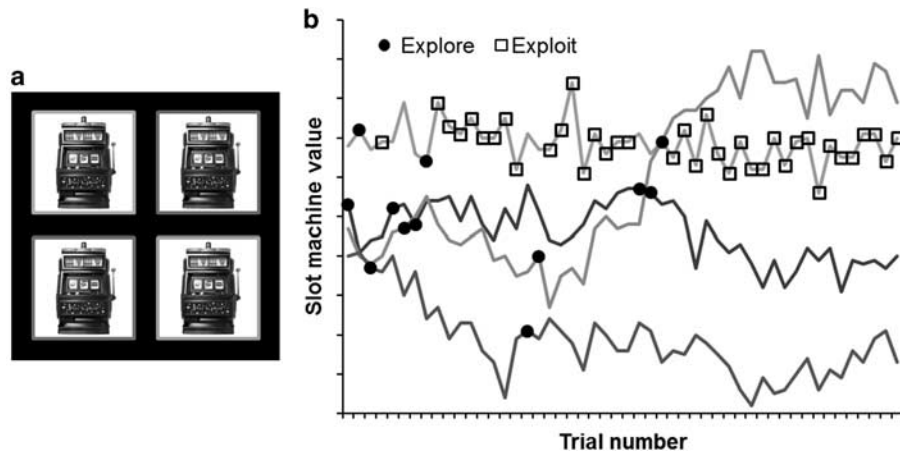
Figure 2 Design of the 4-armed bandit task. (a) In each trial, four options (ie, slot machines) are presented. The player selects one option, then the rewards (ie, points) paid off for that option on that trial are shown. (b) Example of the latent value structure of the options across trials. The value of each option changes gradually and independently of the other options. Based on the decision rule described in Daw et al (2006), exploratory choices made by a player are marked with closed circles and exploitative choices are marked with open squares.

A third task that has been used frequently is the clock task (Moustafa et al, 2008). Here, players are shown a clock face with a hand that makes a clockwise rotation over a 5 s window. In each trial, players must choose when to stop the hand to obtain a reward of unknown value, which is revealed after the choice has been made. The exploration of reward values involves selecting different time points within the 5 s rotation. Players are aware that the reward values available for different time points are fixed across the 50 trials, but they must learn experientially how reward probability and magnitude vary as a function of time. However, unlike the non-stationary option values of the bandit task, the option values in the clock task are fixed across trials; thus, the clock task cannot capture the transition from exploitative to exploratory choices precipitated by trial-to-trial changes in option value. One limitation of the clock task is that choices are classified as exploratory or exploitative based on the difference between previous trial and current trial response times, with larger differences inferred as exploratory decisions. Unfortunately, this means that exploratory decisions may be more likely to arise from decision noise than in other types of tasks.

These tasks and others used to measure explore/exploit decisions (eg, Constantino and Daw, 2015; Costa et al, 2014; Wilson et al, 2014; Glass et al, 2011) all share several common features, including multiple options to choose from, an a priori unknown reward structure, the opportunity to select options other than that with highest immediate value (ie, exploration), and the need for experiential learning to make predictions about current and future reward values. Importantly, for each option there is a trade-off between (1) information gathering to reduce uncertainty and (2) opportunity cost. Ultimately, the explore/exploit trade-off is a problem of behavioral allocation—what to do right now— with the intended goal of efficient performance in the long-term. Analysis of data from explore/exploit paradigms relies on mathematical modeling of trial-by-trial changes, and this modeling is an important difference from other behavioral tasks used in psychiatry research. For example, the Iowa gambling task has four options to choose from, and each option has a different overall value, which is similar to a bandit task. However, the gambling task analysis averages the number of selections for each option, without regard to the changes in trial-by-trial values that affect the behavioral schedule (Bechara et al, 1997). Conversely, explore/exploit decisions are not necessarily identified with the selection of a certain fixed option; rather, they require an ongoing evaluation of reward values—meaning that an option considered exploitative at one moment in time might represent an exploratory option in the future. As a result, models from the reinforcement learning (RL) literature are used to fit players' choice behavior (Sutton and Barto, 1998; Rushworth and Behrens, 2008); the parameter values inferred for these models then become measures of individual differences.

A key parameter in these models is the 'temperature' or 'softmax gain,' which controls the premium placed on the option with the highest current value. Thus, higher 'gain' reflects a stronger tendency to choose the option with the highest previously experienced payoff (ie, exploitation), whereas lower 'gain' reflects a tendency to deviate from this tendency (ie, exploration). Based on this decision rule, a trial in which a player chooses the option with the highest expected value can be classified as 'exploitative', all other trials can be classified as 'exploratory' (Daw et al, 2006). A second key parameter is the learning rate, which determines how much prior beliefs determine choice, or the degree by which expectations are updated by the prediction error (ie, the difference between the expected and the actual outcome). A subject's learning rate should be balanced between two extremes of too much influence of prior beliefs or none at all. Each individual's 'gain' and learning rate (among other parameters) influences their explore/exploit tendencies.

WHAT DO WE KNOW ABOUT THE EXPLORE/EXPLOIT TRADE-OFF?

The last general review of explore/exploit research was conducted by Cohen et al in 2007, and many more studies have been published in the past 10 years. Here, we present an

update based on a select review of recent literature, with a focus on clinical applications of the explore/exploit trade-off.

Temporal Stability of the Environment

Not all exploratory decisions are information seeking, sometimes they are the result of random decision noise leading to exploration by chance. A study by Wilson et al quantified the contributions of these two strategies by modeling decisions in both a short and a long time horizon for decision making. After four forced-choice trials of a 2-armed bandit task, in which either equal or unequal information was given about the two options, participants made either one (short horizon) or six (long horizon) free-choice trials. The authors reported that participants were more information seeking and had higher decision noise with the longer horizon, suggesting that humans use both strategies to adapt their decision-making strategy to the temporal statistics of the environment (Wilson et al, 2014).

This evidence implies that agents can adjust information-seeking behavior according to the temporal stability of the environment. Theoretically, environmental factors like volatility can also affect the learning rate. In a stable environment in which knowledge of the distant past is relevant to the present, the learning rate should be small; conversely, in a rapidly changing, volatile environment, the learning rate should be larger. This hypothesis was tested by manipulating the environmental stability in a 2-armed bandit task (Behrens et al, 2007). In this version, the two options had different probabilities of reward. Players first experienced a stable environment in which one option always had a higher probability, followed by a volatile environment in which the options switched between high and low probability every 30–40 trials. Players displayed higher learning rates in more volatile environments (Behrens et al, 2007). The stability/volatility of the environment, in addition to how recently options have been sampled, also affects the decision to explore or exploit. Actors should be inclined to explore in volatile environments when options have not been sampled recently. In support of this, Knox et al reported that players in a leapfrog task made more exploratory decisions as the environmental volatility increased (Knox et al, 2011).

Conservation Across Species

One of the strengths of the explore/exploit trade-off is the conservation of behavior across species, making translational studies possible. Shared behavior across humans and nonhuman animals is consistent with deep homology in the underlying circuitry, making preclinical, translational studies possible (Pearson et al, 2014). Two studies have compared explore/exploit performance in humans and other species. In the first study, Pearson et al administered a 4-armed bandit task to macaques and humans. The two species performed the task comparably well, and both humans and macaques made exploratory decisions about 25% of the time. However, human behavior was best fit by a model suggesting they possessed more accurate estimates of the task parameters. For instance, humans possessed a longer memory window, which improved the value estimates of options not chosen recently (Pearson et al, 2009). In the

second study, Racey et al administered a stationary 8-armed bandit task to pigeons and humans. In their version of the task, rewards were administered at variable intervals and reward values assigned to each option were fixed within a session, but differed between sessions. Both humans and pigeons were sensitive to the change in values within each session and preferentially chose the option with the highest value, although humans learned more quickly about changes in value structure (Racey et al, 2011).

Neurobiology

There is extensive evidence that phasic midbrain dopamine (DA) encodes reward prediction errors (Schultz and Dickinson, 2000), which is a basic element of RL that can guide explore/exploit decisions. There is also evidence that tonic DA is involved in explore/exploit decisions in other ways, such as the regulation of effort expenditure in food-related foraging. Studies have shown that effort-based choice behavior is regulated by DA in the nucleus accumbens (Salamone et al, 2006, 2009). In particular, rats with DA depletion in the accumbens were more sensitive to work-related response costs and less likely to trade high levels of work for food (Salamone et al, 2001). Conversely, DA transporter knock-down mice, which have elevated extra-cellular DA and increased tonic DA-firing activity, were more exploratory and less sensitive to work-related response costs, even though their increased effort did not increase the likelihood of receiving food rewards (Beeler et al, 2010). Based on this work, DA has been proposed to modulate behavioral energy expenditure along two axes: (1) a conserve-expend axis that regulates activity levels and (2) an explore–exploit axes that regulates the coupling of activity to reward (Beeler et al, 2012). Increased tonic DA function is thought to promote energy expenditure and exploration while decreased tonic DA function favors energy conservation and exploitation; thus, DA interfaces between the internal and external environments and helps match behavioral energy expenditure to the external environmental energy economy (Beeler et al, 2012).

DA receptor subtypes in the prefrontal cortex (PFC) also influence behavioral components of explore/exploit decisions, such as working memory, risk preference, and behavioral flexibility (for review, see Floresco, 2013). Both too much and too little D1 receptor activity can impair working memory, which could reduce the learning rate component of explore/exploit decisions (ie, reduce the influence of prior beliefs on choice), and increased D1 and D2 receptor activity reduces perseverative errors and improves behavioral flexibility, perhaps by strengthening the signal indicating changes in reward contingencies (Floresco, 2013). Furthermore, infusions of selective D1 or D2 receptor antagonists in the medial PFC of rats have been shown to have distinct effects on risk preference: D1 antagonists decreased preference for the large-magnitude/high-risk option, perhaps due to increased sensitivity to negative feedback. Conversely, D2 antagonists increased preference for the large-magnitude/high-risk option (St. Onge et al, 2011). This suggests that DA receptors in the medial PFC help monitor changes in reward probabilities, which supports the behavioral-flexibility component of explore/exploit decisions.

Additional evidence for the role of DA in explore/exploit decisions comes from human genotyping for COMT (Blanco *et al*, 2015; Frank *et al*, 2009). The COMT (Catechol-O-methyltransferase) gene modulates DA levels in the PFC, with Met allele carriers having lower COMT enzyme activity and higher DA levels compared to Val/Val homozygotes. An early study found that individuals with lower COMT enzyme activity associated with the Met/Met genotype had greater exploration in the clock task than those with Met/Val or Val/Val genotypes (Frank *et al*, 2009). Although a later study did not replicate this result, it found that the COMT inhibitor tolcapone increased exploratory choices in Met/Met, but not Val/Val subjects (Kayser *et al*, 2015). A third study found no difference in the rate of exploration between Met/Met, Val/Met or Val/Val subjects in the leapfrog task, although Met carriers were more likely to be best fit by the Ideal Actor model (which reflexively updates beliefs and plans ahead to maximize long-term rewards) and Val/Val carriers were more likely to be best fit by Naive RL (which values options based only on the rewards experienced so far) (Blanco *et al*, 2015).

The locus coeruleus (LC) norepinephrine system has also been proposed to regulate explore/exploit trade-off (Aston-Jones and Cohen, 2005). The phasic LC mode (ie, activated due to presynaptic activity) is thought to optimize performance in the current task (ie, exploitation), while the tonic mode (ie, steady action potential firing at a constant frequency) is thought to facilitate the disengagement of attention from the current course of action and redirect it to processing of other actions (ie, exploration).

Under constant illumination, pupil diameter is correlated with LC activity and may be an indirect measure of tonic LC firing rate (for review, see Jepma and Nieuwenhuis, 2011). One study showed that baseline pupil diameters preceding exploratory choices in a 4-armed bandit task were larger than those preceding exploitative choices, and individual differences in baseline pupil diameter predicted exploratory choices (Jepma and Nieuwenhuis, 2011). However, the acute administration of reboxetine (a selective norepinephrine reuptake inhibitor) did not affect explore/exploit decisions on the 4-armed bandit task (Jepma *et al*, 2010).

A recent study investigated the role of acetylcholinergic (ACh) transmission on explore/exploit decisions using nicotinic ACh receptor (nAChR) $\beta2^*$ knockout mice. The $\beta2$ subunit influences DA activity in the ventral tegmental area and is involved in value-based decisions. Using a spatial version of a 3-armed bandit task and intra-cranial self-stimulation as reward, the $\beta2$ knockout mice made fewer exploratory choices than wild-type mice. This suggests a role for $\beta2^*$-nAChRs in translating expected uncertainty into motivational value and exploratory decision making (Naude *et al*, 2016).

Neuroanatomy

Functional neuroimaging studies have investigated the neuroanatomy that subserves explore/exploit decision making (Addicott *et al*, 2014; Boorman *et al*, 2009; Daw *et al*, 2006; Laureiro-Martinez *et al*, 2014). Compared to exploitative choices, exploratory choices activate the frontopolar cortex and the intraparietal sulcus (Addicott *et al*, 2014; Boorman *et al*, 2009; Daw *et al*, 2006; Laureiro-Martinez

et al, 2014). The frontopolar cortex is thought to subserve switching between behavioral options while maintaining other options in working memory (Boorman *et al*, 2009; Koechlin and Hyafil, 2007). The intraparietal sulcus is thought to subserve behavioral responses (ie, button-press actions), interface between frontal areas and motor output (Daw *et al*, 2006), support mental calculations (Dehaene *et al*, 2003) and support decision-making during uncertainty (Huettel *et al*, 2005). Two studies have also reported activation in the brain stem, possibly the LC (Addicott *et al*, 2014; Laureiro-Martinez *et al*, 2014). Given the spatial limitations of functional neuroimaging, it is uncertain whether this activation stems from the LC, but if it did, this would support the idea that the LC helps regulate exploratory decisions (Aston-Jones and Cohen, 2005).

Compared to exploratory choices, exploitative choices activate a lesser extent of brain regions, and exploitative activation has been inconsistent across studies. One study reported activation in the bilateral temporal lobes, including the middle and superior temporal gyri, planum temporale, and left angular gyrus (Addicott *et al*, 2014); while another study reported activation in the medial PFC, hippocampus, and middle temporal gyri (among other regions) (Laureiro-Martinez *et al*, 2014).

Given the significance of the frontopolar cortex to exploratory decision making, Kovach *et al* administered a 4-armed bandit task to patients with frontopolar lesions and to patients with control lesions. Unexpectedly, patients with frontopolar lesions were not grossly impaired in overall task performance or exploratory switching, although a model-based analysis of learning revealed a selective deficit in frontopolar lesion patients' ability to use the most recent trial outcomes to make decisions (Kovach *et al*, 2012). This suggests that the frontopolar cortex subserves the extrapolation of trends in reward value based on recent reward history.

Evidence from a lesion study in monkeys suggests that the frontopolar cortex is specialized for disengaging executive control from the current task in order to explore new opportunities (Mansouri *et al*, 2015). Another investigation of the role of the frontopolar cortex used anodal and cathodal transcranial direct current stimulation on this cortical region during a modified 3-armed bandit task (Beharelle *et al*, 2015). Compared to baseline performance, anodal (excitatory) stimulation increased the number of exploratory decisions and cathodal (inhibitory) stimulation decreased the number of exploratory choices. Furthermore, the estimated rewards of the highest-paying option became less influential in driving the more exploratory anodal group's choices, but had a stronger effect on the choices of the more exploitative cathodal group. The authors suggest that the increased exploration in the anodal stimulation group reflected an increased responsiveness to previous lower-than-expected outcomes of exploitative choices, whereas the increased exploitation in the cathodal group related to a weaker influence of recent prediction errors and a stronger focus on the current monetary reward of the highest-paying option (Beharelle *et al*, 2015).

The frontopolar cortex shares a neuroanatomical link with the posterior cingulate cortex (PCC), and these two regions are often activated or deactivated together while subjects perform tasks (reviewed in Mansouri *et al*, 2015). Likewise,

1936

the PCC is also implicated in altering behavior in response to unexpected changes in reward. In an electrophysiological study conducted in macaques performing a 4-armed bandit task, firing rates of PCC neurons signaled single-trial reward outcomes and also predicted the probability of shifting between explore/exploit decisions. PCC neurons were also sensitive to reward, risk, and switching options (Pearson et al, 2009). This suggests that increased activity in the PCC reflects a change in either environmental structure or internal state and promotes flexibility, exploration, and renewed learning (Pearson et al, 2011).

Psychiatry Research

There is a small yet growing literature investigating the relationship between addiction and explore/exploit decisions. One study administered a 6-armed bandit task to tobacco smokers and nonsmokers. Smokers made fewer exploratory choices in the first 300 trials and had a higher learning rate, indicating that smokers were more sensitive to the most recent value of each option (Addicott et al, 2012). A later study revealed a relationship between smoking dependence motives and brain activation while smokers performed the 6-armed bandit task. After controlling for nicotine tolerance, there was a relationship between automaticity (ie, habitual smoking) and exploratory brain activation in the bilateral postcentral and supramarginal gyri of the parietal cortices. This suggests that as smoking becomes more automatic, more cognitive effort is necessary for exploratory decision making (Addicott et al, 2014). Harle et al administered a 2-armed bandit task with probabilistic rewards to methamphetamine-dependent participants and healthy controls (Harle et al, 2015). Although both groups showed similar overall performance based on earned points, methamphetamine-dependent participants were less likely to use a learning-supported strategy (ie, using estimated reward values) and instead simply paid attention to the previous trial outcome. Although explore/exploit decisions were not modeled, this result is consistent with research suggesting that methamphetamine-dependent individuals are impaired in learning and updating their knowledge of the environment and generally have difficulties 'seeing the big picture' (Harle et al, 2015). Most recently, Morris et al (2016) administered the clock task to participants with alcohol use disorder, binge-eating disorder, and healthy controls. The participants with alcohol use disorder displayed more repetitive or exploitative decisions rather than strategic exploratory decisions, but the participants with binge-eating disorders did not differ from healthy controls (Morris et al, 2016).

Addictive drugs produce reinforcing effects via rapid, transient increases in DA that mimic phasic DA signals. Chronic exposure to these drugs diminishes DA function (Thiruchselvam et al, 2016) leading to changes in motivation, and the subsequent hypodopaminergic state results in decreased sensitivity to natural rewards and anhedonia. Drug use is then perpetuated as a compensatory mechanism (Volkow et al, 2010; Epstein and Silbersweig, 2015). The hypothesis by Beeler et al (2012) suggests that the acute use of a dopaminergic drug would promote exploration and energy expenditure while chronic use would promote exploitation and energy conservation. The findings of greater exploitation in drug dependent samples by Morris et al (2016) and Addicott et al (2012) appear to support this hypothesis.

Other studies have investigated the effects of specific psychiatric symptoms on explore/exploit decisions, including anhedonia, anxiety, and depression. Symptom scores for anhedonia negatively correlated with the extent of exploratory choices made during a clock task among participants with schizophrenia, suggesting a possible deficit in learning to pursue actions with high reward probability or a preference for maintaining the status quo (Strauss et al, 2011). High trait anxiety was associated with a reduced ability to update outcome expectations between stable and volatile environments in an aversive 2-armed bandit task, suggesting a specific deficit in adjusting learning rate to changes in environmental volatility (Browning et al, 2015). Participants with depressive symptoms deviated from the optimal strategy by exploring when they should have been exploiting and exploring more frequently on exploit-optimal trials on the leapfrog task, possibly due to reduced working memory capacity or decreased sensitivity to changing reward contingencies (Blanco et al, 2013). Finally, among individuals with a tendency towards mood instability, a large unexpected gain or loss influenced their subsequent preference for familiar and unfamiliar options in a 3-armed bandit task, suggesting a biased perception of the subjective value of reward (Eldar and Niv, 2015). Although different paradigms were used, these results demonstrate how explore/exploit decisions and their component processes (eg, learning rate) can be affected by emotional dysregulation.

CONCLUSION

The circuitry mediating explore/exploit decisions evolved to help us survive in the natural world. These decisions represent adaptive behavioral responses to changes in the environment and can be quantified using mathematical modeling. An important goal of computational psychiatry, ie, the application of computational neuroscience to the study of mental disorders, is to identify component processes of transdiagnostic behavioral endophenotypes (Wang and Krystal, 2014; Friston et al, 2014; Adams et al, 2016). In accordance with this goal, the explore/exploit trade-off provides value to psychiatry research by measuring component processes of the positive valence and cognitive systems within the Research Domain Criteria (RDoC). The RDoC is an initiative sponsored by the National Institutes of Mental Health to understand dimensions of psychological function rather than diagnoses of heterogeneous disorders (NIH, 2016). The explore/exploit trade-off encompasses the component processes of cognitive control, working memory, effort valuation, and action selection. Dysregulations in action selection could lead to habitual behaviors, such as drug addiction, which are pathological expressions of processes than normally subserve adaptive goals. Addiction research, in particular, could benefit from formal quantitative measures of action selection, since addicts go to great lengths to obtain their preferred drug, and overcome many obstacles and constraints, both behavioral and economic, to do so at the expense of other (perhaps less certain) non-drug rewards (Salamone et al, 2006, 2007, 2009). Research on

explore/exploit decisions can also inform the neural basis for anhedonia, motivational deficits, and apathy shown in depression and other psychiatric disorders (Salamone et al, 2006, 2007, 2009). Furthermore, studying explore/exploit decisions in the context of foraging behavior combines computational psychiatry with evolutionary psychiatry, which considers the evolutionary function of emotions and behaviors (Nesse, 1984).

Previously, Adams et al (2012), argued for a 'bottom-up' neurobiological decision-making schemata conserved across species. The authors describe how problems, such as patch leaving, are solved similarly across different animal taxonomies. This suggests that natural selection has favored simple, repeated design patterns capable of being implemented by many biological configurations (Adams et al, 2012). This argument supports the idea that these decision-making computations and the circuits that embed them are applied to other behavioral-flexibility challenges. Thus, many behavioral impairments, such as addiction or problem gambling, may result from dysfunction in these basic, common foraging mechanisms. For example, patch leaving choices made by humans and other animals are consistent with outcomes predicted by the marginal value theorem (Charnov, 1976; Constantino and Daw, 2015), and neurons in the dorsal anterior cingulate encode the relative value of leaving a diminishing-value patch for a new one (Hayden et al, 2011). Importantly, research with patch leaving paradigms suggests that some cognitive fallacies, like the belief in 'winning streaks', may be due to instinctive psychological expectations that evolved while foraging in patchy environments (Wilke and Barrett, 2009). Altogether, foraging theory can provide new ways of understanding healthy decision making and open new avenues for investigating abnormal psychiatry.

Traditional behavioral measures used in psychiatry research have shown that individuals with substance dependence are more prone to risk-taking (ie, spending time, effort or money on an uncertain outcome) and impulsivity (ie, preference for short-term rewards over long-term rewards) than healthy individuals (for review, see deWit, 2009; Verdejo-Garcia et al, 2008). Presumably, the behaviors measured with risk-taking or impulsivity paradigms are inherently disadvantageous and represent a lack of self-control. However, within the explore/exploit trade-off, these behaviors can be understood as advantageous in the short- and long-term depending on the environmental context, and the flexibility to adapt advantageously to changes in the environment is critical. For example, when acorns are plentiful and the weather is warm, squirrels will bury nuts to exploit them later rather than eating them all right away. However, in the cold winter months, squirrels must consume more calories to stay warm and a delayed meal could be deadly. Similarly, saving for retirement is not a practical use of resources for someone struggling to provide food for their family every day.

The decision between delaying a reward to maximize long-term survival or enjoying it immediately is not necessarily a contest between a 'top-down' prefrontal control system and a 'bottom-up' striatal impulsive system, where delay discounting is always disadvantageous. Rather, it is an interplay of environmental pressures, biological needs, and brain chemistry that serves to flexibly adapt behavior to support

survival. Risk-taking and impulsivity can be advantageous if they help the individual adapt to changes in environmental uncertainty. Self-report questionnaire data may lead to the conclusion that risk attitudes are stable personality traits, but risk-taking behavior has been shown to be dynamically modulated by the environmental context and the past action-outcome sequence of events. For example, in a foraging game players became less risk averse as the remaining opportunities to earn points decreased (as the horizon for decision making was approached) (Kolling et al, 2014). Furthermore, another study showed that reward-based decision making under risk is better explained by homeostatic principles (ie, a combination of avoiding 'starvation' and the expected value of the outcome) than by standard economic models, which underscores the importance of factoring survival instincts into human decision making (Korn and Bach, 2015; but also see Kacelnik and El Mouden, 2013). The explore/exploit trade-off provides a neutral framework for investigating these behavioral components, and future research on risk-taking and impulsivity should take into account the stability of the individual's environment.

Although the findings reviewed above support the utility of explore/exploit trade-off to investigate the underlying neural mechanisms of reward-based decision making, interpreting the existing evidence is difficult due to the heterogeneity of methods and concepts across studies. Research in computational neuroscience has developed a number of measures and behavioral models to assess explore/exploit decision making, but this work is not readily translatable into psychiatry research. In particular, more replication of results is needed. A current limitation to replication is the variety of explore/exploit measures. We described the three most common paradigms developed to study explore/exploit decisions: the bandit task, the leapfrog task, and the clock task; but many more have been reported (eg, Constantino and Daw, 2015; Costa et al, 2014; Glass et al, 2011; Wilson et al, 2014). Even within measures, changes in the number of options (eg, 2- or 4-armed bandit tasks), reward values, and the predictability of these values can affect behavioral outcomes. In order to move forward in psychiatry research, there needs to be a standardized measure of explore/exploit decisions that is a reproducible and valid measure of real-world decision making, in addition to a behavioral model, or set of models, that can provide a mathematical characterization of parameters in normal and abnormal psychiatric populations. Ultimately, clinical cut-offs for extreme ends of the behavioral spectrum (ie, too much exploratory or too much exploitative choices) or a behavioral flexibility impairment could be defined and provide a new context for understanding psychiatric disorders.

Explore/exploit paradigms have already shown potential for distinguishing between people with and without psychiatric disorders. In particular, several studies suggest that addictive disorders are associated with reduced exploratory and increased exploitative decisions. This would indicate a preference in behavioral planning for the near-term and a preference for familiar, expected rewards over unfamiliar and/or unknown rewards. Furthermore, studies linking addiction to low levels of DA function (Volkow et al, 2004) are consistent with the proposal that decreased tonic DA function favors energy conservation and exploitation (Beeler et al, 2012). Future work could investigate this by comparing

the ratio of explore/exploit decisions to DA receptor availability among substance addicted individuals.

Most importantly, more research is needed comparing the sensitivity of explore/exploit paradigms for detecting group differences, compared to traditional behavioral measures used in psychiatry research. In particular, explore/exploit and other foraging paradigms need to be tested against existing behavioral tasks to determine the best predictor of group differences, symptom severity, neural dysregulation, and therapeutic treatment outcomes. For example, Addicott et al reported that individuals who gambled frequently made more exploratory choices on a 4-armed bandit task, although there were no differences across groups on other assessments of risky or impulsive behavior (eg, balloon analog risk task) (Addicott et al, 2015). It is our belief that foraging paradigms can provide new insights into psychiatric disorders, but first these paradigms must be incorporated into new and ongoing studies in a diverse field of research. We hope that this primer on the explore/exploit trade-off will encourage more psychiatry researchers to add explore/exploit or other foraging paradigms to their battery of behavioral tasks.

FUNDING AND DISCLOSURE

REFERENCES

Adams RA, Huys QJM, Roiser JP (2016). Computational psychiatry: towards a mathematically informed understanding of mental illness. *J Neurol* 87: 53–63.

Adams GK, Watson KK, Pearson JM, Platt ML (2012). Neuroethology of decision-making. *Curr Opin Neurobiol* 22: 982–989.

Addicott MA, Baranger DAA, Kozink RV, Smoski MJ, Dichter GS, McClernon FJ (2012). Smoking withdrawal is associated with increases in brain activation during decision making and reward anticipation: a preliminary study. *Psychopharmacology* 219: 563–573.

Addicott MA, Pearson JM, Froeliger B, Platt ML, McClernon FJ (2014). Smoking automaticity and tolerance moderate brain activation during explore-exploit behavior. *Psychiat Res* 224: 254–261.

Addicott MA, Pearson JM, Kaiser N, Platt ML, McClernon FJ (2015). Suboptimal foraging behavior: a new perspective on gambling. *Behav Neurosci* 129: 656–665.

Aston-Jones G, Cohen JD (2005). Adaptive gain and the role of the locus coeruleus-norepinephrine system in optimal performance. *J Comp Neurol* 493: 99–110.

Barack DL, Gold JI (2016). Temporal trade-offs in psychophysics. *Curr Opin Neurobiol* 37: 121–125.

Barack DL, Platt ML (2016). Neurocomputational nosology: malfunctions of models and mechanisms. *Front Psychol* 7.

Bechara A, Damasio H, Tranel D, Damasio AR (1997). Deciding advantageously before knowing the advantageous strategy. *Science* 275: 1293–1295.

Beeler JA, Cools R, Luciana M, Ostlund SB, Petzinger G (2014). A kinder, gentler dopamine. highlighting dopamine's role in behavioral flexibility. *Front Neurosci* 8: 1–2.

Beeler JA, Daw N, Frazier CRM, Zhuang XX (2010). Tonic dopamine modulates exploitation of reward learning. *Front Behav Neurosci* 4: 1–14.

Beeler JA, Frazier CR, Zhuang X (2012). Putting desire on a budget: dopamine and energy expenditure, reconciling reward and resources. *Front Integr Neurosci* 6: 1–22.

Beharelle AR, Polania R, Hare TA, Ruff CC (2015). Transcranial stimulation over frontopolar cortex elucidates the choice attributes and neural mechanisms used to resolve exploration-exploitation trade-offs. *J Neurosci* 35: 14544–14556.

Behrens TEJ, Woolrich MW, Walton ME, Rushworth MFS (2007). Learning the value of information in an uncertain world. *Nat Neurosci* 10: 1214–1221.

Berger-Tal O, Nathan J, Meron E, Saltz D (2014). The exploration-exploitation dilemma: a multidisciplinary framework. *PLoS ONE* 9: e95693.

Blanco NJ, Love BC, Cooper JA, McGeary JE, Knopik VS, Maddox WT (2015). A frontal dopamine system for reflective exploratory behavior. *Neurobiol Learn Mem* 123: 84–91.

Blanco NJ, Otto AR, Maddox WT, Beevers CG, Love BC (2013). The influence of depression symptoms on exploratory decision-making. *Cognition* 129: 563–568.

Boorman ED, Behrens TEJ, Woolrich MW, Rushworth MFS (2009). How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. *Neuron* 62: 733–743.

Browning M, Behrens TE, Jocham G, O'Reilly J, Bishop SJ (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nat Neurosci* 18: 590–596.

Charnov EL (1976). Optimal foraging, the marginal value theorem. *Theor Popul Biol* 9: 129–136.

Cohen JD, McClure SM, Yu AJ (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philos Trans R Soc Lond B Biol Sci* 362: 933–942.

Constantino SM, Daw ND (2015). Learning the opportunity cost of time in a patch-foraging task. *Cogn Affect Behav Neurosci* 15: 837–853.

Costa VD, Tran VL, Turchi J, Averbeck BB (2014). Dopamine modulates novelty seeking behavior during decision making. *Behav Neurosci* 128: 556–566.

Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006). Cortical substrates for exploratory decisions in humans. *Nature* 441: 876–879.

Dehaene S, Piazza M, Pinel P, Cohen L (2003). Three parietal circuits for number processing. *Cogn Neuropsychol* 20: 487–506.

de Wit H (2009). Impulsivity as a determinant and consequence of drug use: a review of underlying processes. *Addict Biol* 14: 22–31.

Eldar E, Niv Y (2015). Interaction between emotional state and learning underlies mood instability. *Nat Commun* 6: 1–9.

Epstein J, Silbersweig D (2015). The neuropsychiatric spectrum of motivational disorders. *J Neuropsychiatry and Clin Neurosci* 27: 7–18.

Floresco SB (2013). Prefrontal dopamine and behavioral flexibility: shifing from an 'inverted-U' toward a family of functions. *Front Neurosci* 7: 1–12.

Frank MJ, Doll BB, Oas-Terpstra J, Moreno F (2009). Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nat Neurosci* 12: 1062–U1145.

Friston KJ, Stephan KE, Montague R, Dolan RJ (2014). Computational psychiatry: the brain as a phantastic organ. *Lancet Psychiatry* 1: 148–158.

Gittins JC, Jones D. A dynamic allocation index for the sequential design of experiments. In: Gani J, Sarkadi K, Vincze I (eds). *Progress in Statistics*. North Holland Publishing Company: Amsterdam, 1974, pp 241-266.

Glass BD, Maddox WT, Bowen C, Savarie ZR, Matthews MD, Markman AB et al (2011). The effects of 24-hour sleep deprivation on the exploration-exploitation trade-off. *Biol Rhythm Res* 42: 99–110.

Harle KM, Zhang SN, Schiff M, Mackey S, Paulus MP, Yu AJ (2015). Altered statistical learning and decision-making in methamphetamine dependence: evidence from a two-armed bandit task. *Front Psychol* 6.

Hayden BY, Pearson JM, Platt ML (2011). Neuronal basis of sequential foraging decisions in a patchy environment. *Nat Neurosci* 14: 933–939.

Huettel SA, Song AW, McCarthy G (2005). Decisions under uncertainty: probabilistic context influences activation of prefrontal and parietal cortices. *J Neurosci* 25: 3304–3311.

Jepma M, Nieuwenhuis S (2011). Pupil diameter predicts changes in the exploration-exploitation tradeoff: evidence for the adaptive gain theory. *J Cogn Neurosci* 23: 1587–1596.

Jepma M, te Beek ET, Wagenmakers E-J, van Gerven JMA, Nieuwenhuis S (2010). The role of the noradrenergic system in the exploration-exploitation trade-off: a psychopharmacological study. *Front Hum Neurosci* 4: 1–13.

Kacelnik A, El Mouden C (2013). Triumphs and trials of the risk paradigm. *Anim Behav* 86: 1117–1129.

Kayser AS, Mitchell JM, Weinstein D, Frank MJ (2015). Dopamine, locus of control, and the exploration-exploitation tradeoff. *Neuropsychopharmacology* 40: 454–462.

Knox WB, Otto AR, Stone P, Love BC (2011). The nature of belief-directed exploratory choice in human decision-making. *Front Psychol* 2: 398.

Koechlin E, Hyafil A (2007). Anterior prefrontal function and the limits of human decision-making. *Science* 318: 594–598.

Kolling N, Wittmann M, Rushworth MFS (2014). Multiple neural mechanisms of decision making and their competition under changing risk pressure. *Neuron* 81: 1190–1202.

Korn CW, Bach DR (2015). Maintaining homeostasis by decision-making. *PLOS Comput Biol* 11: 1–19.

Kovach CK, Daw ND, Rudrauf D, Tranel D, O'Doherty JP, Adolphs R (2012). Anterior prefrontal cortex contributes to action selection through tracking of recent reward trends. *J Neurosci* 32: 8434–8442.

Laureiro-Martinez D, Canessa N, Brusoni S, Zollo M, Hare T, Alemanno F et al (2014). Frontopolar cortex and decision-making efficiency: comparing brain activity of experts with different professional background during an exploration-exploitation task. *Front Hum Neurosci* 7: 927.

Laviola G, Macri S, Morley-Fletcher S, Adriani W (2003). Risk-taking behavior in adolescent mice: psychobiological determinants and early epigenetic influence. *Neurosci Biobehav Rev* 27: 19–31.

Mansouri FA, Buckley MJ, Mahboubi M, Tanaka K (2015). Behavioral consequences of selective damage to frontal pole and posterior cingulate cortices. *Proc Natl Acad Sci USA* 112: E3940–E3949.

Mata R, Wilke A, Czienskowski U (2013). Foraging across the life span: is there a reduction in exploration with aging? *Front Neurosci* 7: 1–7.

Morris LS, Baek K, Kundu P, Harrison NA, Frank MJ, Voon V (2016). Biases in the explore-exploit tradeoff in addictions: the role of avoidance of uncertainty. *Neuropsychopharmacology* 41: 940–948.

Moustafa AA, Cohen MX, Sherman SJ, Frank MJ (2008). A role for dopamine in temporal decision making and reward maximization in Parkinsonism. *J Neurosci* 28: 12294–12304.

Naude J, Tolu S, Dongelmans M, Torquet N, Valverde S, Rodriguez G et al (2016). Nicotinic receptors in the ventral tegmental area promote uncertainty-seeking. *Nat Neurosci* 19: 471–-478.

Nesse RM (1984). An evolutionary perspective on psychiatry. *Compr Psychiatry* 25: 575–580.

NIH (2016). *Development and Definitions of the RDoC Domains and Constructs*. National Institutes of Health.

Pearson JM, Hayden BY, Raghavachari S, Platt ML (2009). Neurons in posterior cingulate cortex signal exploratory decisions in a dynamic multioption choice task. *Curr Biol* 19: 1532–1537.

Pearson JM, Heilbronner SR, Barack DL, Hayden BY, Platt ML (2011). Posterior cingulate cortex: adapting behavior to a changing world. *Trends Cogn Sci* 15: 143–151.

Pearson JM, Watson KK, Platt ML (2014). Decision making: the neuroethological turn. *Neuron* 82: 950–965.

Racey D, Young ME, Garlick D, Pham JN, Blaisdell AP (2011). Pigeon and human performance in a multi-armed bandit task in response to changes in variable interval schedules. *Learn Behav* 39: 245–258.

Rushworth MFS, Behrens TEJ (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nat Neurosci* 11: 389–397.

Salamone JD, Correa M, Farrar A, Mingote SM (2007). Effort-related functions of nucleus accumbens dopamine and associated forebrain circuits. *Psychopharmacology* 191: 461–482.

Salamone JD, Correa M, Farrar AM, Nunes EJ, Pardo M (2009). Dopamine, behavioral economics, and effort. *Front Behav Neurosci* 3: 1–12.

Salamone JD, Correa M, Mingote SM, Weber SM, Farrar AM (2006). Nucleus accumbens dopamine and the forebrain circuitry involved in behavioral activation and effort-related decision making: implications for understanding anergia and psychomotor slowing in depression. *Curr Psychiatry Rev* 2: 1–14.

Salamone JD, Wisniecki A, Carlson BB, Correa M (2001). Nucleus accumbens dopamine depletions make animals highly sensitive to high fixed ratio requirements but do not impair primary food reinforcement. *Neuroscience* 105: 863–870.

Schultz W, Dickinson A (2000). Neuronal coding of prediction errors. *Ann Rev Neurosci* 23: 473–500.

St. Onge JR, Abhari H, Floresco SB (2011). Dissociable contributions by prefrontal D1 and D2 receptors to risk-based decision making. *J Neurosci* 31: 8625–8633.

Stephens DW, Brown JS, Ydenberg RC (2007). *Foraging: Behavior and Ecology*. University of Chicago Press: Chicago, IL, USA.

Stephens DW, Dunlap AS (2011). Patch exploitation as choice: symmetric choice in an asymmetric situation? *Animal Behav* 81: 683–689.

Stephens DW, Krebs JR (1986). *Foraging Theory*. Princeton University Press: Princeton, NJ, USA.

Strauss GP, Frank MJ, Waltz JA, Kasanova Z, Herbener ES, Gold JM (2011). Deficits in positive reinforcement learning and uncertainty-driven exploration are associated with distinct aspects of negative symptoms in Schizophrenia. *Biol Psychiat* 69: 424–431.

Sutton RS, Barto AG (1998). *Reinforcement Learning: An Introduction*. MIT Press: Cambridge, MA, USA.

Thiruchselvam T, Malik S, Le Foll B (2016). A review of positron emission tomography studies exploring the dopaminergic system in substance use with a focus on tobacco as a co-variate. *Am J Drug Alcohol Abuse* 1–18.

Verdejo-Garcia A, Lawrence AJ, Clark L (2008). Impulsivity as a vulnerability marker for substance-use disorders: review of findings from high-risk research, problem gamblers and genetic association studies. *Neurosci Biobehav Rev* 32: 777–810.

Volkow ND, Fowler JS, Wang GJ, Swanson JM (2004). Dopamine in drug abuse and addiction: results from imaging studies and treatment implications. *Mol Psychiatr* 9: 557–569.

Volkow ND, Wang GJ, Fowler JS, Tomasi D, Telang F, Baler R (2010). Addiction: decreased reward sensitivity and increased expectation sensitivity conspire to overwhelm the brain's control circuit. *Bioessays* 32: 748–755.

Wang XJ, Krystal JH (2014). Computational psychiatry. *Neuron* 84: 638–654.

Wilke A, Barrett HC (2009). The hot hand phenomenon as a cognitive adaptation to clumped resources. *Evol Hum Behav* 30: 161–169.

Wilson RC, Geana A, White JM, Ludvig EA, Cohen JD (2014). Humans use directed and random exploration to solve the explore-exploit dilemma. *J Exp Psychol Gen* 143: 2074–2081.