Psychopharmacology
https://doi.org/10.1007/s00213-018-5062-x

REVIEW

Translational tests involving non-reward: methodological considerations

Benjamin U. Phillips[1] · Laura Lopez-Cruz[1] · Lisa M. Saksida[1,2,3] · Timothy J. Bussey[1,2,3]

Received: 27 August 2018 / Accepted: 2 October 2018

Abstract

This review is concerned with methods for assessing the processing of unrewarded responses in experimental animals and the mechanisms underlying performance of these tasks. A number of clinical populations, including Parkinson's disease, depression, compulsive disorders, and schizophrenia demonstrate either abnormal processing or learning from non-rewarded responses in laboratory-based reinforcement learning tasks. These effects are hypothesized to result from disturbances in modulatory neurotransmitter systems, including dopamine and serotonin. Parallel work in experimental animals has revealed consistent behavioral patterns associated with non-reward and, consistent with the human literature, modulatory roles for specific neurotransmitters. Classical tests involving an important reward omission component include appetitive extinction, ratio schedules of responding, reversal learning, and delay and probability discounting procedures. In addition, innovative behavioral tests have recently been developed leverage probabilistic feedback to specifically assay accommodation of, and learning from, non-rewarded responses. These procedures will be described and reviewed with discussion of the behavioral and neural determinants of performance. A final section focusses specifically on the benefits of trial-by-trial analysis of responding during such tasks, and the implications of such analyses for the translation of findings to clinical studies.

Keywords Extinction · Operant · Schedules of reinforcement · Learning · Dopamine · Serotonin · Touchscreen

Deficits in reward omission processing in clinical populations

Appropriate processing of positive and negative outcomes is a requirement for adaptive reinforcement learning, broadly defined as the ability to select actions in a manner that maximizes positive outcomes (wins) and minimizes negative outcomes (losses) (Sutton and Barto 1998). Abnormal sensitivity to wins and losses is a central feature of a number of psychopathological conditions that are characterized by a failure to learn about and respond adequately to outcomes or shifts in environmental contingencies. Importantly, these changes in sensitivity to wins and losses are thought to result in significant consequences for functional outcomes in these conditions, including enduring reductions in quality of life in both major depressive disorder (MDD) (Victoria et al. 2018) and schizophrenia (Mueser et al. 1991; Dowd et al. 2016). Despite this, these deficits in function have not been fully characterized in all psychopathologies and much is still unknown regarding the mechanistic basis of these disruptions. Reward processing can be assessed in clinical populations via a number of procedures. Amongst these, reversal learning, which requires subjects to accommodate a contingency shift (a change in outcomes associated with specific responses), and probabilistic reinforcement learning, which requires subjects to maximize reward in the face of a set of uncertain outcomes, are highly prevalent and have revealed distinctive patterns of performance impairments across distinct clinical populations, with patient groups typically failing to accommodate shifts in task contingencies or adapt appropriately

This article belongs to a Special Issue on Psychoparmacology of Extinction.

✉ Benjamin U. Phillips
bp342@cam.ac.uk

[1] Department of Psychology and MRC/Wellcome Trust Behavioural and Clinical Neuroscience Institute, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK

[2] Molecular Medicine Research Group, Robarts Research Institute & Department of Physiology and Pharmacology, Schulich School of Medicine & Dentistry, Western University, London, ON, Canada

[3] The Brain and Mind Institute, Western University, London, ON, Canada

Published online: 10 October 2018

to wins and losses (Elliott et al. 1997; Frank et al. 2004; Dowd et al. 2016). In recent years, a number of closely related procedures have been developed and validated for use in experimental animals, thus providing a methodological framework for translational study of reinforcement-related processes.

As characterization of clinical populations on reinforcement learning tasks has increased, it has become increasingly evident that balancing learning from reward and non-reward and reacting with appropriate behavioral responses is a common disturbance across disorders. Moreover, many common procedures that are used to characterize clinical populations comprise responses that are non-rewarded, either inevitably due to task design (e.g., ratio schedules, stochastically reinforced tasks) or because subjects typically exhibit sub-optimal patterns of responding (e.g., discounting procedures, reversal learning). For example, progressive ratio schedules of reinforcement, which are traditionally used almost exclusively in experimental animals, have recently been used to assess effortful motivation in schizophrenia (Strauss et al. 2016; Bismark et al. 2018) and anorexia nervosa patients (Schebendach et al. 2007).

In a similar vein, cognitive flexibility has been assessed in a number of clinical populations using reversal learning procedures. In reversal learning, subjects must accommodate a shift in task contingencies whereby a previously rewarded response becomes non-rewarded and vice versa. Patients often display "perseverative" deficits, in which responses are persistently emitted at a previously rewarded but newly non-rewarded response option (Miller 1990; Chamberlain and Sahakian 2006; Cools et al. 2007; Murray et al. 2008). These deficits may be at least partially mediated by a failure to appropriately process reward omission and reflect the central nature of reinforcement-related executive dysfunction across a broad range of psychopathologies. In perhaps the clearest clinical example, MDD, which is known to possess a serotonin (5-hydroxytryptamine; 5-HT) dysregulation component, is characterized by both hyposensitivity to rewarding outcomes and hypersensitivity to omission of reward. This is as measured by non-optimal response switching in probabilistic learning tasks in which MDD patients tend to switch response strategy following a loss, even if their previous strategy was optimal (Murphy et al. 2003). This abnormal responsivity has been shown to correlate with reduced activity in the dorsomedial and ventrolateral prefrontal cortices, and increased activity in the amygdala in unmedicated MDD patients (Taylor Tavares et al. 2008). Moreover, the identified pattern of deficits is consistent with both anhedonia and hyperactive emotional responsivity to negative outcomes and predicts disease outcomes (Vrieze et al. 2013), suggesting that imbalances in reward processing in MDD may play a central role in the maintenance of low mood in this disorder. In addition to 5-HT, other neurotransmitters are implicated in this set of behavioral functions. For example, a distinguishable set of deficits related to processing non-reward has been reported in Parkinson's disease, in which patients performing probabilistic discrimination tasks demonstrate heightened learning from reward omission as compared to rewarding outcomes (Frank et al. 2004). This abnormality is dependent on treatment regimen, as patients taking levodopa show heightened sensitivity to reward and generally intact learning from reward omission, a set of findings in general agreement with a model that emphasizes the importance of dopamine (DA) levels in reinforcement learning feedback sensitivity.

Choice discounting of a preferred reward has also been assessed in clinical patients using probabilistic discounting and delay discounting. In typical discounting procedures, an increasing cost, such as an escalating delay, or reduction in probability, is systematically imposed on access to a preferred reward. Subject choices typically shift from selecting the preferred reward to a less preferred but free reward as the cost increases. Pathological gamblers have been characterized on both probability (Miedl et al. 2012) and delay discounting (Madden et al. 2011; Wiehler et al. 2015), with the results tending to indicate that patients suffering from this condition both select more risky choices and discount future rewards more steeply. Moreover, a large number of clinical conditions, including Parkinson's disease (Housden et al. 2010), frontotemporal dementia (Bertoux et al. 2015), and major depressive disorder (Pulcu et al. 2014), are also characterized by reduced tolerance of delayed reward. Imaging studies suggest that brain regions including the lateral prefrontal cortex, posterior parietal cortex (McClure et al. 2004), and inferior frontal gyrus (Lin et al. 2015) are involved in the valuation of delayed rewards. Additionally, a number of regions, including the ventral anterior cingulate cortex (vACC) (Kruse et al. 2017) and medial orbitofrontal cortex (mOFC) (Finger et al. 2008), have been implicated in human appetitive extinction learning, suggesting that a broad network of structures are involved in processing reward omission in humans.

Taken as a whole, the extant body of clinical evidence suggests that deficits in translationally viable tasks comprising a substantial reward omission component are present and distinguishable across a number of psychopathological conditions. Thus, this research area represents a valuable opportunity for parallel study of these processes in experimental animals. Moreover, whilst such patient deficits are well characterized and the available evidence suggests that they play a central role in the development and maintenance of psychopathology, the systems that mediate abnormal responsivity to reward omission are not yet fully understood and specific targets for novel treatments remain largely elusive. To facilitate understanding of the psychological and neural mechanisms that govern reward omission processing, it is critical to carry out preclinical studies in experimental animals, ideally using procedures that measure either identical or closely related psychological processes. As feedback integration and reward omission processing are clearly disturbed in numerous

psychopathological conditions, the aim of this review is to consider the preclinical application of operant methods that involve an explicit reward omission component, including single-contingency appetitive extinction (where only one response option is available), ratio schedules, reinforcement learning tasks, and discounting procedures. Both procedural considerations and the neural systems implicated in the performance of these tasks are discussed in the context of facilitating understanding of the behavioral methodology. We suggest that applying multiple tasks characterized by reward omission to the same experimental question can greatly facilitate understanding of the change in behavioral state resulting from a defined manipulation. Additionally, this review aims to examine the potential for cross-species translation of results obtained on these tasks (in the context of trial-by-trial analysis of performance (Daw 2011), and touchscreen operant testing (Bussey et al. 2008) is considered.

Behavioral procedures involving reward omission

Single contingency procedures: extinction and ratio schedules

Single contingency procedures—as opposed to choice procedures, considered below—involve only one possible response. To assess responding in the face of non-reward in such procedures, reward is omitted following a response. Two widely used approaches of this type are appetitive instrumental extinction schedules, and ratio schedules. In an appetitive extinction procedure, a previously rewarded response becomes abruptly non-rewarded and extinction learning is indicated by discontinued responding. Persistent responding, relative to controls, in the absence of reward is taken to indicate an extinction impairment (Balleine and Dickinson 2000). In addition to the number of responses emitted on extinction sessions, the rate of responding can be measured to generate an additional index of response to non-reward. These schedules have been used to characterize extinction learning in a number of diverse rodent models, including models of fragile × syndrome (Sidorov et al. 2014), NMDA receptor subunit dysfunction (Brigman et al. 2008), and deletion of postsynaptic density protein 95 (Horner et al. 2017).

A number of methodological considerations should be taken into account when applying extinction learning procedures in experimental animals. For example, multiple distinct mechanisms potentially contribute to performance on appetitive extinction schedules, including response suppression, instrumental learning, Pavlovian processes, and detection of non-rewarded responses (Bouton 2004). The reduction in responding under appetitive extinction schedules following reinforcer devaluation, an experimentally induced degradation of reinforcer value, can also depend on reinforcer properties as demonstrated elegantly by Adams and Dickinson (1981). In this study, a food reinforcer was paired with an injection of lithium chloride, thus inducing a conditioned food aversion, and was used in combination with instrumental extinction to reveal the mechanisms of reinforcer relationship to extinction of responding. It was shown that conditioned aversion to the reinforcer attenuated responding at extinction relative to controls, suggesting that reinforcement is represented in the associative structure even when not present within the schedule structure (Adams and Dickinson 1981). Thus, reinforcer valuation effects play a contributory role in extinction schedule performance, even when the reinforcer itself is absent. Thus, researchers should carefully consider the choice of reward used in studies investigating extinction learning under an intended experimental manipulation, as differences in reward processing during acquisition may spuriously affect extinction schedule performance.

Another methodological consideration in the context of extinction procedures is that previous instrumental experience can modulate performance at extinction. For example, animals previously exposed to partially reinforced schedules exhibit enhanced resistance to extinction as compared to animals exposed to a continuously reinforced schedule, an effect termed the partial reinforcement extinction effect (PREE) (Weiner et al. 1985; Bouton et al. 2014). This effect has been shown to depend on dopamine (DA), as potentiation of DAergic neurotransmission with d-amphetamine administration further increases PREE (Weiner et al. 1985). Conversely, DA D2 receptor blockade exerts effects that are partially overlapping with, but not identical to, extinction in animals working on reinforced schedules (Wise et al. 1978; Salamone 1986). Taken together, this evidence suggests that response to reward omission on appetitive instrumental extinction schedules is dependent on exposure to previous schedule, previously encoded incentive value of the reward and DA dynamics. In particular, these DA-dependent effects on extinction responding represent an important consideration in the context of characterization of instrumental extinction learning in genetically modified models with a reasonable probability of DA dysregulation (e.g., mouse models of Parkinson's disease or schizophrenia).

In contrast with extinction schedules, ratio schedules require emission of a set number of responses in order to gain access to a reward (Sidman and Stebbins 1954; Hodos 1961). Contingencies in ratio schedule testing can be arranged in a number of ways and are typically designed to probe aspects of reward-related behavior. In progressive ratio (PR) schedules, the number of responses (i.e., lever presses, nose-spokes or touches) required to obtain a single reinforcer increases progressively during the session according to a defined ramp. The final ratio completed in a session (referred to as the "breakpoint") is frequently interpreted as an operational measure of reward value (Hodos 1961) or animal effort capacity (Aberman et al. 1998). However, PR schedules may also probe

parallel psychological processes including extinction learning, reward expectancy, and tolerance of unrewarded delays (Ward et al. 2011). In particular, as challenging PR schedules are characterized by eventual large runs of non-rewarded responses at high work requirements, the results may reflect a substantial appetitive extinction component, rather than motivation per se (Ward et al. 2011). Researchers may carry out an extinction schedule control in conjunction with their intended manipulation on PR to assess this possibility.

Moreover, there is substantial evidence that PR and extinction do not assess fully overlapping processes. For example, modulating the DA system affects effort in isolation in other tasks, with systemic administration of DA antagonists and agonists resulting in bi-directional modulation of performance. For example, DA antagonists (i.e., raclopride, haloperidol, and eticlopride) decrease lever presses for a preferred reinforcer (i.e., sucrose pellets) in PR schedules and also increase consumption of freely available chow in fixed ratio (FR) or PR/free chow concurrent choice, suggesting that a number of the effects observed following DAergic manipulations on PR may be attributable to effort processes rather than reward omission processing alone (Farrar et al. 2010; Nunes et al. 2013; Randall et al. 2012; Heath et al. 2015). In addition, there are experimental dissociations between ratio and extinction schedule performance. For example, heterozygous and homozygous DA transporter knockout (DAT-KO) mice show neither higher breakpoints in PR nor higher responses on an FR schedule compared with their wild-type (WT) counterparts (Hironaka et al. 2004). However, during extinction, homozygous DAT-KO mice were more resistant to extinction than WT and heterozygous DAT-KO mice (Hironaka et al. 2004).

Thus, whilst the potential effects of reward omission and consequent extinction processes should be carefully considered in the context of the performance of PR schedules, there is also a large body of experimental evidence suggesting that PR depends on other processes including effort allocation. Testing both PR and extinction procedures may help disentangle the contribution of different psychological processes to task performance. Further approaches to delineating the contribution of different processes to PR performance, including extinction learning and effort, are discussed in the third section of this review. Overall, instrumental extinction performance comprises a large reward omission processing component, but performance also depends on prior reinforcer valuation processes and DA-dependent processes, including previous instrumental contingency experience. Whilst PR performance may depend partially on extinction learning, there is also a large body of evidence suggesting the involvement of effort processes in the determination of responding.

In addition to studies that have focused on general neuromodulatory control of reward omission processing in single contingency procedures, a large body of previous literature has studied the neural circuitry involved in this set of behaviors, especially with respect to instrumental habits. Given that habits are operationally defined as outcome insensitivity (Balleine and Dickinson 1998; Robbins and Costa 2017), these investigations are of obvious relevance to the systems involved in processing reward omission. This body of work has revealed the involvement of a broad network of cortical and sub-cortical structures in the support of habitual responding. Within these broad networks, one particularly influential model of the transition from initial goal-directed to eventual habitual responding at the neural level suggests that initial goal-directed responding is dorsomedial striatum (DMS) dependent but later shifts to dorsolateral striatum (DLS) dependency as habitual processes come to control responding (Corbit et al. 2012). Some studies in humans support this view that this region is critical for control of habitual behavior (Tricomi et al. 2009), whilst also implicating a reduction in activity of the ventromedial prefrontal cortex (vmPFC) in habitual responding (de Wit et al. 2009). With respect to the precise striatal micro-circuitry mediating habits in the DLS, a recent study demonstrated that the activity of fast-spiking interneurons located in this region is closely linked with food-reinforced habitual behavior (O'Hare et al. 2017). In this study, it was shown that silencing this population of interneurons via chemogenetics blocks the behavioral expression of a previously acquired habitual response. In other words, neuronal inhibition targeted at this specific population of DLS interneurons caused mice that had acquired an instrumental habit to behave in a goal-directed manner, perhaps suggesting that these neurons are specifically involved in determining responsivity to rewarding outcomes. Other midbrain systems are also known to encode both habitual responding and reward omission processes. For example, building on the highly influential finding that DA-releasing neurons in the ventral tegmental area (VTA) and substantia nigra encode "reward prediction errors" (i.e., the discrepancy between expected and actual outcomes) (Schultz et al. 1997), Tobler and colleagues later demonstrated that these cells are also predictive of reward omission and behave in a manner consistent with a negative reward prediction error (Tobler et al. 2003). Moreover, a recent study by Verharen and colleagues utilizing fiber photometry and chemogenetics suggested that processing of DA "dips" in the nucleus accumbens encodes learning from losses, further underlining the involvement of this neurotransmitter in reward omission processing (Verharen et al. 2018).

Choice procedures: reversal learning and probabilistic reinforcement learning procedures

Reversal learning procedures typically require subjects to learn to discriminate between rewarded and non-rewarded responses, before the contingencies are switched so that the previously optimal selection becomes non-optimal and vice versa. At the point of reversal, subjects will typically continue to respond at the previous location for an extended period, a

pattern of responding termed perseverative. Whilst reversal learning is designed to assess cognitive flexibility, this term describes a complex construct with multiple constituent contributory psychological components, including response inhibition, attentional processing, reinforcement learning, and reward sensitivity (Nilsson et al. 2015). Critically, in deterministic reversal learning, the previous always-rewarded stimulus (S+) essentially shifts from maintenance under a continuous reinforcement schedule to an extinction schedule, thus raising the possibility that reversal learning performance in the perseverative phase can be attributed to reward omission sensitivity. However, experimental evidence suggests that reversal learning and extinction learning, though perhaps overlapping, are dissociable processes. For example, as pointed out by Horner and colleagues, there is evidence dissociating performance of these procedures (Horner et al. 2017). For example, whilst Grin2a$^{-/-}$ mice are impaired on reversal learning but not extinction (Brigman et al. 2008), Gria1$^{-/-}$ mice demonstrate the opposite pattern of impairments (Barkus et al. 2012). This suggests that, whilst reversal learning deficits can be attributable to loss of sensitivity to extinction, other mechanisms can also mediate impairments.

To perform reversal learning optimally, it is plainly advantageous to not only learn from the absence of reward following a response at the previous S+, but also from reward at the newly rewarded S+. Thus, whilst non-reward processing is an important factor in reversal learning, perhaps particularly in the early stages of a reversal, learning from positive feedback will likely come to predominate toward the end of reversal learning (Phillips et al. 2018). This model of performance at different stages of reversal is consistent with the density of outcomes that is experienced at these stages whereby early reversal is characterized by low performance levels and higher losses and late reversal by high performance levels and higher wins. Such patterns of learning at different phases of both discrimination and reversal learning may also reflect the relative contributions of goal-directed and habitual processes to choice behavior (Graybeal et al. 2011; Brigman et al. 2013; DePoy et al. 2013; Izquierdo and Jentsch 2012; Bergstrom et al. 2018). A number of behavioral probes and manipulations are available to test the balance of these mechanisms, with some results obtained using these procedures suggesting that a number of manipulations that affect reversal learning performance exert their effects by altering sensitivity to reward omission. Methodological approaches for dissociating these contributions will be considered below in the context of studies focused on the neural basis of reversal learning.

Much research has been directed toward the role of 5-HT in the performance of reversal learning tasks. Landmark studies in the marmoset demonstrate that orbitofrontal cortex 5-HT depletions impair reversal learning by increasing the number of perseverative responses emitted in the early phases of reversal (Clarke et al. 2004). This deficit is highly specific, as performance of an extradimensional shifting task, in which subjects have to adapt to discriminating another pair of stimuli based on a different rule, was unimpaired (Clarke et al. 2005). Additionally, the deficit was further investigated using stimulus replacement procedures, in which either the previously rewarded or non-rewarded stimulus is replaced with a novel stimulus. These procedures are designed to determine whether a reversal learning deficit is attributable to stimulus perseveration or learned avoidance of the previous S−. The observed pattern of responding under these manipulations revealed a specific failure to disengage from responding at the previously rewarded stimulus, perhaps suggesting a specific non-reward omission processing deficit in orbitofrontal cortex (OFC) 5-HT depleted subjects (Clarke et al. 2007).

A large body of additional evidence supports a key function for 5-HT in reversal learning, at least partially through influencing the processing of non-rewarded responses. In rodents, treatment with selective serotonin reuptake inhibitors (SSRI) and transgenic SERT (5-HT transporter) inactivation improves performance on reversal learning tasks (Bari et al. 2010; Brigman et al. 2010). At the level of neural circuitry, a recent study combined SERT-cre transgenic mice with fiber photometry to image the activity of dorsal raphe nucleus (DRN) neurons during a reversal learning task (Matias et al. 2017). The results suggest that DRN 5-HT neurons encode trial-by-trial prediction errors in a positive/negative outcome-independent fashion (termed "unsigned" prediction errors). Thus, the authors suggest that DRN 5-HT may have non-specific involvement in processing worse-than-expected rewards by modulating attention or general learning capacity, providing a potential neural substrate for reward omission in the 5-HT system.

In addition to subcortical 5-HT systems, other circuitry has also been shown to be involved in encoding multi-contingency task outcomes during procedures characterized by frequent reward omission. For example, a recent study, utilizing an attentional set-shifting procedure in which rats had to flexibly adapt to discriminate between different reward-predictive features, found that a set of dorsomedial prefrontal cortex (dmPFC) neurons reliably predicted task outcome not only prior to outcome presentation but also post-trial outcome encoding (Del Arco et al. 2017). The role of the PFC is further emphasized by multi-unit recordings carried out in rats performing a gambling task, in which subjects choose between multiple options with different probabilities of more or less desirable outcomes, demonstrating that poor performance in a model of maternal separation was correlated with a loss of synchrony between the anterior cingulate cortex and amygdala (Cao et al. 2016). More broadly, studies in animals carrying out reversal learning procedures have implicated a broad network of structures in encoding outcomes on a trial-by-trial basis including the dorsal raphe nucleus (Barlow et al. 2015; Matias et al. 2017), striatal regions (Klanker et al. 2015), and frontal cortices (Marquardt et al. 2017), suggesting

that the neural encoding of outcome signaling is broadly represented in both cortical and sub-cortical circuitries.

In similar approaches, a number of studies have sought to determine the 5-HT subreceptor-specific mechanisms involved in reversal learning performance. The available evidence indicates a central role for 5-HT2C receptors in mediating the influence of 5-HT in processing non-rewarded responses in reversal learning tasks. Specifically, a number of studies have demonstrated that antagonism of the 5-HT2C receptor facilitates reversal learning (Boulougouris et al. 2008), and this effect has been localized both temporally, to the early phases of reversal, and anatomically, to the lateral orbitofrontal cortex (Alsiö et al. 2015). Again, specific effects on early reversal performance are perhaps indicative of alterations in the processing of reward omission. To formally assess the contribution of positive and negative feedback learning to discrimination and reversal performance, a recently developed version of a stimulus-based visual discrimination task leverages a third probabilistically reinforced stimulus to assess learning from positive and negative feedback (Nilsson et al. 2015; Phillips et al. 2018). In the valence-probe visual discrimination (VPVD) task, deterministically reinforced S+ and S− stimuli are occasionally presented in conjunction with an S50 which is reinforced on 50% of the trials. By comparing performance on S+ vs S50 and S50 vs S− trials, it is possible to assess to which stimulus more learning has accrued, thus determining whether the subject has a bias toward learning from reward or omission of reward. The partially reinforced discrimination trials can be presented either during initial discrimination or following a reversal, enabling the contribution of positive and negative feedback learning to these distinct probe trials to be assessed. The task is designed to assess learning from wins and losses in a similar way to multiple procedures available for application in humans. Notable examples include the probabilistic and transitive selection tasks used to characterize reinforcement learning biases in Parkinson's disease (Frank et al. 2004), Schizophrenia (Waltz et al. 2007), and depression patients (Chase et al. 2010). Additional related tasks used in humans include probabilistic reversal learning tasks that are designed to assess immediate responses to positive and negative feedback (Evers et al. 2005) and reinforcement learning tasks intended to dissociate model based from model free learning (Gläscher et al. 2010). A common structural theme amongst these procedures is that subjects are required to emit responses at (often visually complex) stimuli that are each associated with a probability of reward. Thus, in adopting the same basic features, VPVD may represent a viable translational tool for assessing similar biases in reinforcement learning function in rodent models.

In addition to this potential for translational study, VPVD has already been used to evaluate a potential serotonergic antidepressant in mice (Phillips et al. 2018). In this study, it was found that the 5-HT2C receptor antagonist, SB 242084, impaired discrimination learning of the standard S+ > S− trials. This effect was particularly pronounced in the late sessions, where the performance level of vehicle-treated control animals is high. In terms of the impact of positive and negative feedback on learning as assayed by the additional trial types, 5-HT2C antagonism appeared to shift the balance toward learning from negative feedback. Subsequent experiments employing a spatial probabilistic reversal learning procedure provided further support for this finding, as the same manipulation resulted in a reduction in "win-stay" proportions, an operational index of positive feedback sensitivity.

In addition to deterministic discrimination and reversal tasks, the role of non-rewarded responses in performance of tasks assaying cognitive flexibility and reinforcement learning is also demonstrated by probabilistic reinforcement learning tasks in humans (Elliott et al. 1997) and experimental animals (Bari et al. 2010) (Fig. 1). In these procedures, responses are stochastically reinforced so that optimal choices are occasionally non-reinforced and vice versa. Contingencies are typically arranged so that overall, approximately 20% of trials are spuriously reinforced. Thus, a requirement for optimal responding in such tasks is that subjects continue to response at the highly reinforced stimulus even following spurious reward omission. Sensitivity to reward and reward omission is measured by analyzing the proportion of trials on which a subject persists at the same stimulus following a rewarded response (win-stay) or responds at the alternative stimulus following a non-rewarded response (lose-shift).

A consistent finding using such procedures is that lose-shift proportions are substantially higher than win-shift proportions, indicating that experimental animals are in general highly sensitive to omission of reward on a trial-by-trial basis (Bari et al. 2010; Ineichen et al. 2012). These proportions are sensitive to a number of manipulations including modulation of 5-HT and glutamate systems (Bari et al. 2010; Rychlik et al. 2016). Specifically, 5-HT appears to be particularly implicated in encoding win-stay lose-shift proportions. In rats, the selective serotonin reuptake inhibitor (SSRI) citalopram exerts dose-dependent effects on accommodation of spurious reward omission (Bari et al. 2010). A recent study suggests that these effects may at least be partially attributable to activity at the 5-HT2C receptor, as systemic agonism of this receptor in mice recapitulates this reduction in lose-shift proportions (Phillips et al. 2018). Taken together, these data suggest that the 5-HT2C receptor is a likely target for the capacity of 5-HT to process reward omission in probabilistic tasks.

Delay and probability discounting procedures

Discounting refers to the reduction in value of a preferred option when it becomes associated with a cost such as delay, uncertainty, or effort (Cardinal et al. 2000; Bickel 2015). In perhaps the simplest version of delay discounting, an incrementally increasing delay is imposed between choice and

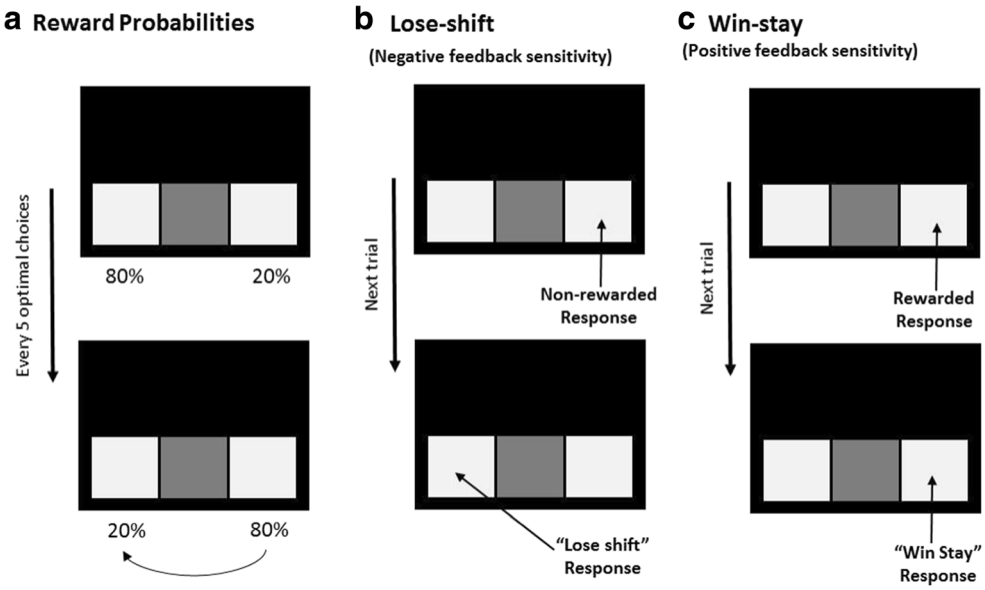

Fig. 1 Schematic of a typical rodent spatial probabilistic reversal learning procedure. Rewarding outcomes are stochastically delivered following responses at both spatial locations. Lose-shift is defined as a selection of the alternate spatial location following a non-rewarded trial and is taken as a measure of negative feedback sensitivity. Conversely, win-stay is defined as a selection at the same location following a rewarded trial and is taken as an index of positive feedback sensitivity

reward delivery. A characteristic discounting choice curve follows a hyperbolic function, reflecting the iterative devaluation of the large reward as a function of delay (Fig. 2). In a similar approach, decision-making under risk can be assessed by probabilistic discounting, in which the probability that a preferred large reward is delivered is systematically reduced across a session (Ghods-Sharifi et al. 2009; Abela and Chudasama 2013). For example, the probability of large reward delivery may begin at 100% and then reduce by 25% across subsequent trial blocks until the probability of large reward is small.

Despite superficial similarities, it is widely recognized that delayed and probabilistic reinforcement recruit distinct neural and psychological processes, as they display differential sensitivity to a number of manipulations (Cardinal 2006). Moreover, since probabilistic discounting is the only procedure amongst these examples in which the outcome is unpredictable, it can be methodologically leveraged to isolate processes associated with reward omission decision-making by comparing results obtained with this procedure with results obtained under delay discounting. In this vein, some studies have utilized delay and probabilistic discounting in parallel to reveal the specific neural mechanisms involved in reward uncertainty and reward delay, thus isolating mechanisms unique to reward omission within an otherwise consistent framework (Yates et al. 2015; Yates et al. 2018).

However, accurate interpretation of results acquired from discounting procedures requires a number of considerations. For example, a potential alternative explanation of delay discounting data is that the response becomes uncoupled from the rewarding outcome as the delay increases. This shift in contingencies at the large reward location could result in a reduction in associability between a response at the large reward response contingency and the delivery of the large reward itself. Thus, in the same sense that PR schedules could potentially be characterized as reflective of extinction processes (because the time between the initiation of a bout of responding and the reward deliver increases), delay discounting could also possess an extinction component. Experimental findings suggest that delay discounting does potentially comprise an extinction component. For example, rats may exhibit reduced preference for the large reward choice during the very first session in which delays are presented (Cardinal et al. 2000). However, it has been demonstrated that reward omission only partially resembles a typical discounting function in well-trained animals, suggesting that extinction learning likely cannot fully account for typical discounting-related preference reduction (Cardinal et al. 2000).

Another set of important considerations regarding the application of delay discounting procedures has been highlighted by analysis of the effects of DAergic manipulations on choice behavior. Specifically, the effects of d-amphetamine on choice behavior are equivocal, with studies reporting both decreased (Evenden and Ryan 1996; Helms et al. 2006) and increased (Krebs and Anderson 2012) large, delayed reward preference. Critically, the effects of this compound on choice

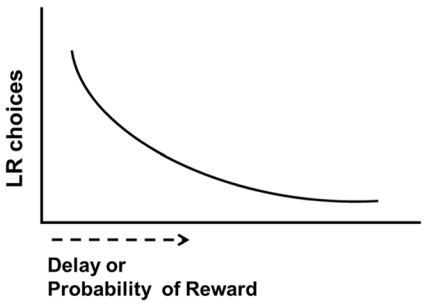

Fig. 2 Typical hyperbolic discounting curve observed as a result of discounting procedures. Raw number or percentage choices resulting in the large reward decrease as delay to large reward increases or probability of large reward decreases

behavior can be reversed by reversing the delay presentation order, perhaps suggesting that the effects of amphetamine under baseline conditions reflect a tendency to perseveratively repeat patterns of choices emitted early in the session (Tanno et al. 2014). Alternatively, these results may reflect the tendency of d-amphetamine to accentuate the effects of conditioned reinforcers (Taylor and Robbins 1984; Cardinal et al. 2000), or control timing perception, a set of cognitive processes in which DA is known to play a role (Meck 1983; Soares et al. 2016). Thus, discounting procedures undoubtedly recruit a large number of complex mechanisms and processes that could be examined by modification of discounting procedures to probe alternative psychological explanations. For example, researchers may consider reversing the order of presentation of trial bins on some probe sessions, carrying out sessions run in extinction and closely examine other parameters (e.g., the presence or absence of cues during task performance).

In comparison with delay discounting, the underlying neural and pharmacological mechanisms underlying probabilistic discounting are less well studied. However, a number of investigations have sought to determine the neuropharmacological basis of probabilistic discounting. The available evidence indicates an important role for DA, with activation of both D1 and D2 receptors increasing the proportion of choices emitted for a larger but riskier reward (St Onge and Floresco 2009). In addition to DA manipulations, the involvement of defined neural structures involved in probabilistic discounting has been studied in rodent models. These studies have demonstrated the involvement of the nucleus accumbens (Stopper and Floresco 2011), amygdala (Jenni et al. 2017), and OFC (Abela and Chudasama 2013). In one example, Abela and colleagues (Abela and Chudasama 2013) investigated the effects of OFC and ventral hippocampus lesions in rats performing touchscreen-based delay and probabilistic discounting. It was found that, compared to sham-lesioned controls, OFC-lesioned rats tended to show a reduced tolerance for risk but normal tolerance of delay. However, this pattern of performance was reversed in ventral hippocampus lesioned rats, which exhibited reduced tolerance of delay but normal tolerance of risk. These results suggest an important role for the OFC in encoding reward omission in this task framework, thus demonstrating the utility of comparison between delay and probabilistic discounting following defined experimental interventions.

From a methodological perspective, the test species is an important consideration in the context of the application of discounting tasks. There are few published studies that have successfully demonstrated the ability of mice to distinguish between large and small rewards in the context of discounting procedures. However, some authors have reported successful quantity discrimination in the context of "adjusting-amounts" procedures in which the delays are held constant and the amount of reward delivered parametrically varied (Mitchell 2014). In our laboratory, we have found that mice are capable of discriminating different quantities of a palatable liquid reward (strawberry milkshake) to a high level and that this preference supports robust delay and effort discounting (Phillips et al. unpublished data; Lopez-Cruz et al. unpublished data). Moreover, in laboratories employing variants of these tasks with a range of apparatus, rewards, and species, it is important to confirm that subjects are able to both discriminate reward quantities under baseline conditions (i.e., no delay or 100% reward probability) and that this preference is not disrupted by the experimental manipulation.

More broadly, some studies have demonstrated that rodents behaviorally express preferences for rewards of different sweetness/flavor that are of the same quantity (Pardo et al. 2015) and that such preferences may be subject to a high degree of variability between individuals (Sweis et al. 2018). Whether the same processes would be recruited for the performance of discounting tasks leveraging preferences for the same quantity of differentially preferred rewards is an interesting open question.

Trial-by-trial analysis and cross-species translation

Assessing responses to non-reward in experimental animals arguably necessitates trial-by-trial approaches to data analysis. This is because the usual whole-session measures fail to capture the animal's immediate responses to changes in the patterns of outcomes. For ratio tasks, one approach to trial-by-trial analysis is to calculate a rate of responding for each trial, the form of which can then be subsequently described by mathematical equations (Killeen 1994; Bradshaw and Killeen 2012). In PR for example, it is typical for rodents to first emit responses rapidly, before declining toward inactivity as the number of trials increases. This decay is known to follow an exponential function. A simple approach is to fit this in a trial-by-trial manner with an equation that captures this decay (Simpson et al. 2011; Phillips et al. 2017). From there, coefficients that represent the predicted peak response rate and decay rate can be extracted and tested for between-group significance. From these, decay rate is hypothesized to reflect reinforcer control over behavior and extinction sensitivity (Ward et al. 2011), whereas peak response rate may reflect initial motivation for the reward or differences in motoric capacity.

For reinforcement learning procedures, a set of recent advances have utilized trial-by-trial modeling to reveal hidden parameters that drive task performance (Daw 2011). These models tend to be based on prediction error learning algorithms that update choice values based on the discrepancy between expected and actual outcomes. From such models, it is possible to extract values for a number of different parameters, including but not limited to learning rates, which can be separated by wins and losses, "stickiness" (the tendency to

repeat choices), and inverse temperature (a measure reflective of the sharpness of choice, i.e., the degree to which a subject chooses either in accordance with the perceived value of the available responses). These approaches have already been applied in human (Gläscher et al. 2010), non-human primate (Clarke et al. 2014), and rodent (Verharen et al. 2018) studies, and have been used to investigate the roles of both DA (den Ouden et al. 2013; Eisenegger et al. 2014) and 5-HT (den Ouden et al. 2013; Iigaya et al. 2018) on performance. For example, this class of approaches has shown that rats treated with a stimulant that potentiates DAergic neurotransmission, methamphetamine, exhibit impaired learning from losses (reward omission) on a reversal learning task (Verharen et al. 2018). Additionally, in the general context of decision-making under risk, it has been demonstrated that D2Rs play an important role in encoding trial-by-trial choices when reward omission is possible (Zalocusky et al. 2016). Thus, reinforcement learning modeling is a promising novel direction for investigation of the learning mechanisms related to reward omission across species. However, ensuring equivalent task design as far as possible is an important requirement to fully realize the potential of such approaches so that the extracted parameters can be compared as far as is reasonably possible.

More broadly, in translational terms, it is beneficial to employ tasks with a high degree of translational potential across multiple species, not only to ensure that findings in animals are as translational as possible but also to facilitate back-translation of human findings to preclinical studies. One approach directed toward this end is the operant touchscreen testing system, which has already been used to directly assess the performance of both rodents and humans with comparable genetic mutations on the same cognitive tasks (Nithianantharajah et al. 2013; Nithianantharajah et al. 2015). This approach is advantageous not only because tasks can be applied in very similar ways in multiple species but also because a battery of tasks can be applied using the same stimuli, responses, and reinforcers within the same species (Bussey et al. 2008). Indeed, all the tasks described in this review are available in the touchscreen apparatus, and the internal consistency of this approach may allow for better comparison of the effects of reward omission across multiple tasks. It would in theory be possible to apply a large set of the preclinical tasks described here in the same cohort of experimental animals, thus facilitating more precise comparison of results and a better understanding of the mechanisms involved in different aspects of behavioral responses to reward omission.

Conclusions

Clinical data suggest that the processing of the omission of reward is a dysregulated process in a number of neuropsychiatric and neurodegenerative conditions including MDD, Parkinson's disease, obsessive compulsive disorder (OCD), and addiction. A clear example is MDD, in which oversensitivity to reward omission manifests as impairments in reinforcement learning in laboratory tasks and may constitute a central pathological process in the maintenance of low mood in this disorder (Elliott et al. 1997). Failures in appropriate reward omission processing can be hypothesized to result from dysregulation of a number of neural systems. For example, both DA and 5-HT appear to play a particularly important mechanistic role in this domain in both health and disease.

A large number of procedures at the preclinical level feature omission of rewarding outcomes, though the focus and interpretation of many such studies often hinge on response to reward. A good deal of progress has been made in elucidating the neural basis of processing reward omission using such tasks. Some conclusions regarding behavioral effects across preclinical tasks can be drawn. For example, reward omission tends to promote decreased vigor in single contingency tasks whilst tending to promote choice switching in multiple contingency tasks. These baseline tendencies can be affected by manipulation of the neuromodulatory systems involved in encoding these responses, specifically DA and 5-HT. Importantly, it is known that whilst these systems may possess distinct behavioral functions, they closely interact in the context of encoding wins and losses (Daw et al. 2002).

Given the clear importance of reinforcement learning deficits from a clinical perspective, we suggest that researchers continue to develop preclinical procedures specifically designed to assess processing of, and response to, reward omission, and that trial-by-trial analytical techniques are applied to maximize translational potential. Moreover, we suggest the application of multiple types of preclinical procedures to study this set of clinically relevant domains, as patterns of responding are divergent across tasks and afford an unprecedented opportunity to dissect the precise neural circuitry involved in encoding reward omission responses.

Compliance with ethical standards

Conflict of interest TJB and LMS consult for Campden Instruments, Ltd. BUP and LLC disclose no interests.

References

Abela AR, Chudasama Y (2013) Dissociable contributions of the ventral hippocampus and orbitofrontal cortex to decision-making with a delayed or uncertain outcome. Eur J Neurosci 37:640–647. https://doi.org/10.1111/ejn.12071

Aberman JE, Ward SJ, Salamone JD (1998) Effects of dopamine antagonists and accumbens dopamine depletions on time-constrained progressive-ratio performance. Pharmacol Biochem Behav 61:341–348. https://doi.org/10.1016/S0091-3057(98)00112-9

Adams CD, Dickinson A (1981) Instrumental responding following reinforcer devaluation. Q J Exp Psychol B 33:109–121. https://doi.org/10.1080/14640748108400816

Alsiö J, Nilsson SRO, Gastambide F, Wang RAH, Dam SA, Mar AC, Tricklebank M, Robbins TW (2015) The role of 5-HT2C receptors in touchscreen visual reversal learning in the rat: a cross-site study. Psychopharmacology 232:4017–4031. https://doi.org/10.1007/s00213-015-3963-5

Balleine BW, Dickinson A (1998) Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. Neuropharmacology 37:407–419. https://doi.org/10.1016/S0028-3908(98)00033-1

Balleine BW, Dickinson A (2000) The effect of lesions of the insular cortex on instrumental conditioning: evidence for a role in incentive memory. J Neurosci 20:8954–8964

Bari A, Theobald DE, Caprioli D, Mar AC, Aidoo-Micah A, Dalley JW, Robbins TW (2010) Serotonin modulates sensitivity to reward and negative feedback in a probabilistic reversal learning task in rats. Neuropsychopharmacology 35:1290–1301. https://doi.org/10.1038/npp.2009.233

Barkus C, Feyder M, Graybeal C, Wright T, Wiedholz L, Izquierdo A, Kiselycznyk C, Schmitt W, Sanderson DJ, Rawlins JNP, Saksida LM, Bussey TJ, Sprengel R, Bannerman D, Holmes A (2012) Do GluA1 knockout mice exhibit behavioral abnormalities relevant to the negative or cognitive symptoms of schizophrenia and schizoaffective disorder? Neuropharmacology 62:1263–1272. https://doi.org/10.1016/j.neuropharm.2011.06.005

Barlow RL, Alsiö J, Jupp B, Rabinovich R, Shrestha S, Roberts AC, Robbins TW, Dalley JW (2015) Markers of serotonergic function in the orbitofrontal cortex and dorsal raphé nucleus predict individual variation in spatial-discrimination serial reversal learning. Neuropsychopharmacology 40:1619–1630. https://doi.org/10.1038/npp.2014.335

Bergstrom HC, Lipkin AM, Lieberman AG, Pinard CR, Gunduz-Cinar O, Brockway ET, Taylor WW, Nonaka M, Bukalo O, Wills TA, Rubio FJ, Li X, Pickens CL, Winder DG, Holmes A (2018) Dorsolateral striatum engagement interferes with early discrimination learning. Cell Rep 23:2264–2272. https://doi.org/10.1016/j.celrep.2018.04.081

Bertoux M, de Souza LC, Zamith P, Dubois B, Bourgeois-Gironde S (2015) Discounting of future rewards in behavioural variant frontotemporal dementia and Alzheimer's disease. Neuropsychology 29:933–939. https://doi.org/10.1037/neu0000197

Bickel WK (2015) Discounting of delayed rewards as an endophenotype. Biol Psychiatry 77:846–847. https://doi.org/10.1016/j.biopsych.2015.03.003

Bismark AW, Thomas ML, Tarasenko M, Shiluk AL, Rackelmann SY, Young JW, Light GA (2018) Relationship between effortful motivation and neurocognition in schizophrenia. Schizophr Res 193:69–76. https://doi.org/10.1016/j.schres.2017.06.042

Boulougouris V, Glennon JC, Robbins TW (2008) Dissociable effects of selective 5-HT2A and 5-HT2C receptor antagonists on serial spatial reversal learning in rats. Neuropsychopharmacology 33:2007–2019. https://doi.org/10.1038/sj.npp.1301584

Bouton ME (2004) Context and behavioral processes in extinction. Learn Mem 11:485–494. https://doi.org/10.1101/lm.78804

Bouton ME, Woods AM, Todd TP (2014) Separation of time-based and trial-based accounts of the partial reinforcement extinction effect. Behav Process 101:23–31. https://doi.org/10.1016/j.beproc.2013.08.006

Bradshaw CM, Killeen PR (2012) A theory of behaviour on progressive ratio schedules, with applications in behavioural pharmacology. Psychopharmacology 222:549–564. https://doi.org/10.1007/s00213-012-2771-4

Brigman JL, Feyder M, Saksida LM, Bussey TJ, Mishina M, Holmes A (2008) Impaired discrimination learning in mice lacking the NMDA receptor NR2A subunit. Learn Mem 15:50–54. https://doi.org/10.1101/lm.777308

Brigman JL, Mathur P, Harvey-White J, Izquierdo A, Saksida LM, Bussey TJ, Fox S, Deneris E, Murphy DL, Holmes A (2010) Pharmacological or genetic inactivation of the serotonin transporter improves reversal learning in mice. Cereb Cortex 20:1955–1963. https://doi.org/10.1093/cercor/bhp266

Brigman JL, Daut RA, Wright T, Gunduz-Cinar O, Graybeal C, Davis MI, Jiang Z, Saksida LM, Jinde S, Pease M, Bussey TJ, Lovinger DM, Nakazawa K, Holmes A (2013) GluN2B in corticostriatal circuits governs choice learning and choice shifting. Nat Neurosci 16:1101–1110. https://doi.org/10.1038/nn.3457

Bussey TJ, Padain TL, Skillings EA, Winters BD, Morton AJ, Saksida LM (2008) The touchscreen cognitive testing method for rodents: how to get the best out of your rat. Learn Mem 15:516–523. https://doi.org/10.1101/lm.987808

Cao B, Wang J, Zhang X, Yang X, Poon DCH, Jelfs B, Chan RHM, Wu JCY, Li Y (2016) Impairment of decision making and disruption of synchrony between basolateral amygdala and anterior cingulate cortex in the maternally separated rat. Neurobiol Learn Mem 136:74–85. https://doi.org/10.1016/j.nlm.2016.09.015

Cardinal RN (2006) Neural systems implicated in delayed and probabilistic reinforcement. Neural Netw 19:1277–1301. https://doi.org/10.1016/j.neunet.2006.03.004

Cardinal RN, Robbins TW, Everitt BJ (2000) The effects of d-amphetamine, chlordiazepoxide, alpha-flupenthixol and behavioural manipulations on choice of signalled and unsignalled delayed reinforcement in rats. Psychopharmacology 152:362–375

Chamberlain SR, Sahakian BJ (2006) The neuropsychology of mood disorders. Curr Psychiatry Rep 8:458–463

Chase HW, Frank MJ, Michael A, Bullmore ET, Sahakian BJ, Robbins TW (2010) Approach and avoidance learning in patients with major depression and healthy controls: relation to anhedonia. Psychol Med 40:433–440. https://doi.org/10.1017/S0033291709990468

Clarke HF, Dalley JW, Crofts HS et al (2004) Cognitive inflexibility after prefrontal serotonin depletion. Science 304:878–880. https://doi.org/10.1126/science.1094987

Clarke HF, Walker SC, Crofts HS et al (2005) Prefrontal serotonin depletion affects reversal learning but not attentional set shifting. J Neurosci 25:532–538. https://doi.org/10.1523/JNEUROSCI.3690-04.2005

Clarke HF, Walker SC, Dalley JW, Robbins T, Roberts A (2007) Cognitive inflexibility after prefrontal serotonin depletion is behaviorally and neurochemically specific. Cereb Cortex 17:18–27. https://doi.org/10.1093/cercor/bhj120

Clarke HF, Cardinal RN, Rygula R, Hong YT, Fryer TD, Sawiak SJ, Ferrari V, Cockcroft G, Aigbirhio FI, Robbins TW, Roberts AC (2014) Orbitofrontal dopamine depletion upregulates caudate dopamine and alters behavior via changes in reinforcement sensitivity. J Neurosci 34:7663–7676. https://doi.org/10.1523/JNEUROSCI.0718-14.2014

Cools R, Lewis SJG, Clark L, Barker RA, Robbins TW (2007) L-DOPA disrupts activity in the nucleus accumbens during reversal learning in Parkinson's disease. Neuropsychopharmacology 32:180–189. https://doi.org/10.1038/sj.npp.1301153

Corbit LH, Nie H, Janak PH (2012) Habitual alcohol seeking: time course and the contribution of subregions of the dorsal striatum. Biol Psychiatry 72:389–395. https://doi.org/10.1016/j.biopsych.2012.02.024

Daw ND (2011) Trial-by-trial data analysis using computational models. In: Decision making, affect, and learning. Oxford University Press, Oxford, pp 3–38

Daw ND, Kakade S, Dayan P (2002) Opponent interactions between serotonin and dopamine. Neural Netw 15:603–616. https://doi.org/10.1016/S0893-6080(02)00052-7

De Wit S, Corlett PR, Aitken MR et al (2009) Differential engagement of the ventromedial prefrontal cortex by goal-directed and habitual behavior toward food pictures in humans. J Neurosci 29:11330–11338. https://doi.org/10.1523/JNEUROSCI.1639-09.2009

Del Arco A, Park J, Wood J et al (2017) Adaptive encoding of outcome prediction by prefrontal cortex ensembles supports behavioral flexibility. J Neurosci 37:8363–8373. https://doi.org/10.1523/JNEUROSCI.0450-17.2017

Den Ouden HEM, Daw ND, Fernandez G et al (2013) Dissociable effects of dopamine and serotonin on reversal learning. Neuron 80:1090–1100. https://doi.org/10.1016/j.neuron.2013.08.030

DePoy L, Daut R, Brigman JL, MacPherson K, Crowley N, Gunduz-Cinar O, Pickens CL, Cinar R, Saksida LM, Kunos G, Lovinger DM, Bussey TJ, Camp MC, Holmes A (2013) Chronic alcohol produces neuroadaptations to prime dorsal striatal learning. Proc Natl Acad Sci U S A 110:14783–14788. https://doi.org/10.1073/pnas.1308198110

Dowd EC, Frank MJ, Collins A, Gold JM, Barch DM (2016) Probabilistic reinforcement learning in patients with schizophrenia: relationships to anhedonia and Avolition. Biol Psychiatry Cogn Neurosci Neuroimaging 1:460–473. https://doi.org/10.1016/j.bpsc.2016.05.005

Eisenegger C, Naef M, Linssen A, Clark L, Gandamaneni PK, Müller U, Robbins TW (2014) Role of dopamine D2 receptors in human reinforcement learning. Neuropsychopharmacology 39:2366–2375. https://doi.org/10.1038/npp.2014.84

Elliott R, Sahakian BJ, Herrod JJ, Robbins TW, Paykel ES (1997) Abnormal response to negative feedback in unipolar depression: evidence for a diagnosis specific impairment. J Neurol Neurosurg Psychiatry 63:74–82

Evenden JL, Ryan CN (1996) The pharmacology of impulsive behaviour in rats: the effects of drugs on response choice with varying delays of reinforcement. Psychopharmacology 128:161–170. https://doi.org/10.1007/s002130050121

Evers EAT, Cools R, Clark L, van der Veen FM, Jolles J, Sahakian BJ, Robbins TW (2005) Serotonergic modulation of prefrontal cortex during negative feedback in probabilistic reversal learning. Neuropsychopharmacology 30:1138–1147. https://doi.org/10.1038/sj.npp.1300663

Farrar AM, Segovia KN, Randall PA, Nunes EJ, Collins LE, Stopper CM, Port RG, Hockemeyer J, Müller CE, Correa M, Salamone JD (2010) Nucleus accumbens and effort-related functions: behavioral and neural markers of the interactions between adenosine A2A and dopamine D2 receptors. Neuroscience 166(4):1056–1067

Finger EC, Mitchell DGV, Jones M, Blair RJR (2008) Dissociable roles of medial orbitofrontal cortex in human operant extinction learning. Neuroimage 43:748–755. https://doi.org/10.1016/j.neuroimage.2008.08.021

Frank MJ, Seeberger LC, O'reilly RC (2004) By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 306:1940–1943. https://doi.org/10.1126/science.1102941

Ghods-Sharifi S, St Onge JR, Floresco SB (2009) Fundamental contribution by the basolateral amygdala to different forms of decision making. J Neurosci 29:5251–5259. https://doi.org/10.1523/JNEUROSCI.0315-09.2009

Gläscher J, Daw N, Dayan P, O'Doherty JP (2010) States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. Neuron 66:585–595. https://doi.org/10.1016/j.neuron.2010.04.016

Graybeal C, Feyder M, Schulman E, Saksida LM, Bussey TJ, Brigman JL, Holmes A (2011) Paradoxical reversal learning enhancement by stress or prefrontal cortical damage: rescue with BDNF. Nat Neurosci 14:1507–1509. https://doi.org/10.1038/nn.2954

Heath CJ, Bussey TJ, Saksida LM (2015) Motivational assessment of mice using the touchscreen operant testing system: effects of dopaminergic drugs. Psychopharmacology 232(21–22):4043–4057

Helms CM, Reeves JM, Mitchell SH (2006) Impact of strain and D-amphetamine on impulsivity (delay discounting) in inbred mice. Psychopharmacology 188:144–151. https://doi.org/10.1007/s00213-006-0478-0

Hironaka N, Ikeda K, Sora I, Uhl GR, Niki H (2004) Food-reinforced operant behavior in dopamine transporter knockout mice: enhanced resistance to extinction. Ann N Y Acad Sci 1025(1):140–145

Hodos W (1961) Progressive ratio as a measure of reward strength. Science 134(3483):943–944

Horner AE, McLaughlin CL, Afinowi NO et al (2017) Enhanced cognition and dysregulated hippocampal synaptic physiology in mice with a heterozygous deletion of PSD-95. Eur J Neurosci 47:164–176. https://doi.org/10.1111/ejn.13792

Housden CR, O'Sullivan SS, Joyce EM et al (2010) Intact reward learning but elevated delay discounting in Parkinson's disease patients with impulsive-compulsive spectrum behaviors. Neuropsychopharmacology 35:2155–2164. https://doi.org/10.1038/npp.2010.84

Iigaya K, Fonseca MS, Murakami M, Mainen ZF, Dayan P (2018) An effect of serotonergic stimulation on learning rates for rewards apparent after long intertrial intervals. Nat Commun 9:2477. https://doi.org/10.1038/s41467-018-04840-2

Ineichen C, Sigrist H, Spinelli S, Lesch KP, Sautter E, Seifritz E, Pryce CR (2012) Establishing a probabilistic reversal learning test in mice: evidence for the processes mediating reward-stay and punishment-shift behaviour and for their modulation by serotonin. Neuropharmacology 63:1012–1021. https://doi.org/10.1016/j.neuropharm.2012.07.025

Izquierdo A, Jentsch JD (2012) Reversal learning as a measure of impulsive and compulsive behavior in addictions. Psychopharmacology 219:607–620. https://doi.org/10.1007/s00213-011-2579-7

Jenni NL, Larkin JD, Floresco SB (2017) Prefrontal dopamine D1 and D2 receptors regulate dissociable aspects of decision making via distinct ventral striatal and amygdalar circuits. J Neurosci 37:6200–6213. https://doi.org/10.1523/JNEUROSCI.0030-17.2017

Killeen PR (1994) Mathematical principles of reinforcement. Behav Brain Sci 17:105. https://doi.org/10.1017/S0140525X00033628

Klanker M, Sandberg T, Joosten R, Willuhn I, Feenstra M, Denys D (2015) Phasic dopamine release induced by positive feedback predicts individual differences in reversal learning. Neurobiol Learn Mem 125:135–145. https://doi.org/10.1016/j.nlm.2015.08.011

Krebs CA, Anderson KG (2012) Preference reversals and effects of D-amphetamine on delay discounting in rats. Behav Pharmacol 23:228–240. https://doi.org/10.1097/FBP.0b013e32835342ed

Kruse O, Tapia León I, Stark R, Klucken T (2017) Neural correlates of appetitive extinction in humans. Soc Cogn Affect Neurosci 12:106–115. https://doi.org/10.1093/scan/nsw157

Lin X, Zhou H, Dong G, Du X (2015) Impaired risk evaluation in people with internet gaming disorder: fMRI evidence from a probability discounting task. Prog Neuro-Psychopharmacol Biol Psychiatry 56:142–148. https://doi.org/10.1016/j.pnpbp.2014.08.016

Madden GJ, Francisco MT, Brewer AT, Stein JS (2011) Delay discounting and gambling. Behav Process 87:43–49. https://doi.org/10.1016/j.beproc.2011.01.012

Marquardt K, Sigdel R, Brigman JL (2017) Touch-screen visual reversal learning is mediated by value encoding and signal propagation in the orbitofrontal cortex. Neurobiol Learn Mem 139:179–188. https://doi.org/10.1016/j.nlm.2017.01.006

Matias S, Lottem E, Dugué GP, Mainen ZF (2017) Activity patterns of serotonin neurons underlying cognitive flexibility. elife. https://doi.org/10.7554/eLife.20552

McClure SM, Laibson DI, Loewenstein G, Cohen JD (2004) Separate neural systems value immediate and delayed monetary rewards. Science 306:503–507. https://doi.org/10.1126/science.1100907

Meck WH (1983) Selective adjustment of the speed of internal clock and memory processes. J Exp Psychol Anim Behav Process 9:171–201

Miedl SF, Peters J, Büchel C (2012) Altered neural reward representations in pathological gamblers revealed by delay and probability discounting. Arch Gen Psychiatry 69:177–186. https://doi.org/10.1001/archgenpsychiatry.2011.1552

Miller L (1990) Neuropsychodynamics of alcoholism and addiction: personality, psychopathology, and cognitive style. J Subst Abus Treat 7:31–49

Mitchell SH (2014) Assessing delay discounting in mice. Curr Protoc Neurosci 66:Unit 8.30. https://doi.org/10.1002/0471142301.ns0830s66

Mueser KT, Douglas MS, Bellack AS, Morrison RL (1991) Assessment of enduring deficit and negative symptom subtypes in schizophrenia. Schizophr Bull 17:565–582

Murphy FC, Michael A, Robbins TW, Sahakian BJ (2003) Neuropsychological impairment in patients with major depressive disorder: the effects of feedback on task performance. Psychol Med 33:455–467. https://doi.org/10.1017/S0033291702007018

Murray GK, Cheng F, Clark L, Barnett JH, Blackwell AD, Fletcher PC, Robbins TW, Bullmore ET, Jones PB (2008) Reinforcement and reversal learning in first-episode psychosis. Schizophr Bull 34:848–855. https://doi.org/10.1093/schbul/sbn078

Nilsson SRO, Alsiö J, Somerville EM, Clifton PG (2015) The rat's not for turning: dissociating the psychological components of cognitive inflexibility. Neurosci Biobehav Rev 56:1–14. https://doi.org/10.1016/j.neubiorev.2015.06.015

Nithianantharajah J, Komiyama NH, McKechanie A, Johnstone M, Blackwood DH, Clair DS, Emes RD, van de Lagemaat LN, Saksida LM, Bussey TJ, Grant SGN (2013) Synaptic scaffold evolution generated components of vertebrate cognitive complexity. Nat Neurosci 16:16–24. https://doi.org/10.1038/nn.3276

Nithianantharajah J, McKechanie AG, Stewart TJ et al (2015) Bridging the translational divide: identical cognitive touchscreen testing in mice and humans carrying mutations in a disease-relevant homologous gene. Sci Rep 5:14613. https://doi.org/10.1038/srep14613

Nunes EJ, Randall PA, Hart EE, Freeland C, Yohn SE, Baqi Y, Muller CE, Lopez-Cruz L, Correa M, Salamone JD (2013) Effort-related motivational effects of the VMAT-2 inhibitor tetrabenazine: implications for animal models of the motivational symptoms of depression. J Neurosci 33(49):19120–19130

O'Hare JK, Li H, Kim N et al (2017) Striatal fast-spiking interneurons selectively modulate circuit output and are required for habitual behavior. elife. https://doi.org/10.7554/eLife.26231

Pardo M, López-Cruz L, San Miguel N et al (2015) Selection of sucrose concentration depends on the effort required to obtain it: studies using tetrabenazine, D1, D2, and D3 receptor antagonists. Psychopharmacology 232:2377–2391. https://doi.org/10.1007/s00213-015-3872-7

Phillips BU, Heath CJ, Ossowska Z, Bussey TJ, Saksida LM (2017) Optimisation of cognitive performance in rodent operant (touchscreen) testing: evaluation and effects of reinforcer strength. Learn Behav 45:252–262. https://doi.org/10.3758/s13420-017-0260-7

Phillips BU, Dewan S, Nilsson SRO, Robbins TW, Heath CJ, Saksida LM, Bussey TJ, Alsiö J (2018) Selective effects of 5-HT2C receptor modulation on performance of a novel valence-probe visual discrimination task and probabilistic reversal learning in mice. Psychopharmacology 235:2101–2111. https://doi.org/10.1007/s00213-018-4907-7

Pulcu E, Trotter PD, Thomas EJ, McFarquhar M, Juhasz G, Sahakian BJ, Deakin JFW, Zahn R, Anderson IM, Elliott R (2014) Temporal discounting in major depressive disorder. Psychol Med 44:1825–1834. https://doi.org/10.1017/S0033291713002584

Randall PA, Pardo M, Nunes EJ, Cruz LL, Vemuri VK, Makriyannis A, Baqi Y, Müller CE, Correa M, Salamone JD, Beeler JA (2012) Dopaminergic modulation of effort-related choice behavior as assessed by a progressive ratio chow feeding choice task: pharmacological studies and the role of individual differences. PLoS ONE 7(10):e47934

Robbins TW, Costa RM (2017) Habits. Curr Biol 27:R1200–R1206. https://doi.org/10.1016/j.cub.2017.09.060

Rychlik M, Bollen E, Rygula R (2016) Ketamine decreases sensitivity of male rats to misleading negative feedback in a probabilistic reversal-learning task. Psychopharmacology 234:613–620. https://doi.org/10.1007/s00213-016-4497-1

Salamone JD (1986) Different effects of haloperidol and extinction on instrumental behaviours. Psychopharmacology 88:18–23. https://doi.org/10.1007/BF00310507

Schebendach JE, Klein DA, Foltin RW, Devlin MJ, Walsh BT (2007) Relative reinforcing value of exercise in inpatients with anorexia nervosa: model development and pilot data. Int J Eat Disord 40:446–453. https://doi.org/10.1002/eat.20392

Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275:1593–1599. https://doi.org/10.1126/science.275.5306.1593

Sidman M, Stebbins WC (1954) Satiation effects under fixed-ratio schedules of reinforcement. J Comp Physiol Psychol 47:114–116

Sidorov MS, Krueger DD, Taylor M, Gisin E, Osterweil EK, Bear MF (2014) Extinction of an instrumental response: a cognitive behavioral assay in Fmr1 knockout mice. Genes Brain Behav 13:451–458. https://doi.org/10.1111/gbb.12137

Simpson EH, Kellendonk C, Ward RD, Richards V, Lipatova O, Fairhurst S, Kandel ER, Balsam PD (2011) Pharmacologic rescue of motivational deficit in an animal model of the negative symptoms of schizophrenia. Biol Psychiatry 69:928–935. https://doi.org/10.1016/j.biopsych.2011.01.012

Soares S, Atallah BV, Paton JJ (2016) Midbrain dopamine neurons control judgment of time. Science 354:1273–1277. https://doi.org/10.1126/science.aah5234

St Onge JR, Floresco SB (2009) Dopaminergic modulation of risk-based decision making. Neuropsychopharmacology 34:681–697. https://doi.org/10.1038/npp.2008.121

Stopper CM, Floresco SB (2011) Contributions of the nucleus accumbens and its subregions to different aspects of risk-based decision making. Cogn Affect Behav Neurosci 11:97–112. https://doi.org/10.3758/s13415-010-0015-9

Strauss GP, Whearty KM, Morra LF, Sullivan SK, Ossenfort KL, Frost KH (2016) Avolition in schizophrenia is associated with reduced willingness to expend effort for reward on a progressive ratio task. Schizophr Res 170:198–204. https://doi.org/10.1016/j.schres.2015.12.006

Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. IEEE Trans Neural Netw 9:1054–1054. https://doi.org/10.1109/TNN.1998.712192

Sweis BM, Thomas MJ, Redish AD (2018) Mice learn to avoid regret. PLoS Biol 16:e2005853. https://doi.org/10.1371/journal.pbio.2005853

Tanno T, Maguire DR, Henson C, France CP (2014) Effects of amphetamine and methylphenidate on delay discounting in rats: interactions with order of delay presentation. Psychopharmacology 231:85–95. https://doi.org/10.1007/s00213-013-3209-3

Taylor JR, Robbins TW (1984) Enhanced behavioural control by conditioned reinforcers following microinjections of d-amphetamine into the nucleus accumbens. Psychopharmacology 84:405–412. https://doi.org/10.1007/BF00555222

Taylor Tavares JV, Clark L, Furey ML, Williams GB, Sahakian BJ, Drevets WC (2008) Neural basis of abnormal response to negative

feedback in unmedicated mood disorders. Neuroimage 42:1118–1126. https://doi.org/10.1016/j.neuroimage.2008.05.049

Tobler PN, Dickinson A, Schultz W (2003) Coding of predicted reward omission by dopamine neurons in a conditioned inhibition paradigm. J Neurosci 23:10402–10410

Tricomi E, Balleine BW, O'Doherty JP (2009) A specific role for posterior dorsolateral striatum in human habit learning. Eur J Neurosci 29:2225–2232. https://doi.org/10.1111/j.1460-9568.2009.06796.x

Verharen JPH, de Jong JW, Roelofs TJM, Huffels CFM, van Zessen R, Luijendijk MCM, Hamelink R, Willuhn I, den Ouden HEM, van der Plasse G, Adan RAH, Vanderschuren LJMJ (2018) A neuronal mechanism underlying decision-making deficits during hyperdopaminergic states. Nat Commun 9:731. https://doi.org/10.1038/s41467-018-03087-1

Victoria LW, Gunning FM, Bress JN, Jackson D, Alexopoulos GS (2018) Reward learning impairment and avoidance and rumination responses at the end of engage therapy of late-life depression. Int J Geriatr Psychiatry 33:948–955. https://doi.org/10.1002/gps.4877

Vrieze E, Pizzagalli DA, Demyttenaere K, Hompes T, Sienaert P, de Boer P, Schmidt M, Claes S (2013) Reduced reward learning predicts outcome in major depressive disorder. Biol Psychiatry 73:639–645. https://doi.org/10.1016/j.biopsych.2012.10.014

Waltz JA, Frank MJ, Robinson BM, Gold JM (2007) Selective reinforcement learning deficits in schizophrenia support predictions from computational models of striatal-cortical dysfunction. Biol Psychiatry 62:756–764. https://doi.org/10.1016/j.biopsych.2006.09.042

Ward RD, Simpson EH, Kandel ER, Balsam PD (2011) Modeling motivational deficits in mouse models of schizophrenia: behavior analysis as a guide for neuroscience. Behav Process 87:149–156. https://doi.org/10.1016/j.beproc.2011.02.004

Weiner I, Bercovitz H, Lubow RE, Feldon J (1985) The abolition of the partial reinforcement extinction effect (PREE) by amphetamine. Psychopharmacology 86:318–323. https://doi.org/10.1007/BF00432221

Wiehler A, Bromberg U, Peters J (2015) The role of prospection in steep temporal reward discounting in gambling addiction. Front Psychiatry 6:112. https://doi.org/10.3389/fpsyt.2015.00112

Wise RA, Spindler J, deWit H, Gerberg GJ (1978) Neuroleptic-induced "anhedonia" in rats: pimozide blocks reward quality of food. Science 201:262–264

Yates JR, Batten SR, Bardo MT, Beckmann JS (2015) Role of ionotropic glutamate receptors in delay and probability discounting in the rat. Psychopharmacology 232:1187–1196. https://doi.org/10.1007/s00213-014-3747-3

Yates JR, Prior NA, Chitwood MR, Day HA, Heidel JR, Hopkins SE, Muncie BT, Paradella-Bradley TA, Sestito AP, Vecchiola AN, Wells EE (2018) Effects of GluN2B-selective antagonists on delay and probability discounting in male rats: modulation by delay/probability presentation order. Exp Clin Psychopharmacol. https://doi.org/10.1037/pha0000216

Zalocusky KA, Ramakrishnan C, Lerner TN, Davidson TJ, Knutson B, Deisseroth K (2016) Nucleus accumbens D2R cells signal prior outcomes and control risky decision-making. Nature 531:642–646. https://doi.org/10.1038/nature17400