Journal Pre-proof

Adaptive behaviour under conflict: deconstructing extinction, reversal, and active avoidance learning

Elizabeth E Manning, Laura A Bradfield, Mihaela D Iordanova

PII: S0149-7634(20)30588-1

DOI: https://doi.org/10.1016/j.neubiorev.2020.09.030

Reference: NBR 3922

To appear in: *Neuroscience and Biobehavioral Reviews*

Received Date: 16 April 2020

Revised Date: 28 September 2020

Accepted Date: 30 September 2020

Please cite this article as: Manning EE, Bradfield LA, Iordanova MD, Adaptive behaviour under conflict: deconstructing extinction, reversal, and active avoidance learning, *Neuroscience and Biobehavioral Reviews* (2020), doi: https://doi.org/10.1016/j.neubiorev.2020.09.030

Adaptive behaviour under conflict: deconstructing extinction, reversal, and active avoidance learning

Elizabeth E Manning*#[1-2], Laura A Bradfield*#[3-4], Mihaela D Iordanova*#[5]

Affiliations

[1] Department of Psychiatry, University of Pittsburgh, Suite 223, 450 Technology Drive, Pittsburgh, PA, 15224, USA

[2] School of Biomedical Sciences and Pharmacy, University of Newcastle, MS306, University Drive, Callaghan, NSW 2308, Australia

[3] Centre for Neuroscience and Regenerative Medicine, University of Technology Sydney (St. Vincent's Campus), 405 Liverpool St, Darlinghurst, NSW 2010, Australia

[4] St. Vincent's Centre for Applied Medical Research, St. Vincent's Hospital Sydney Limited, 405 Liverpool St, Darlinghurst, NSW 2010, Australia

[5] Department of Psychology/Centre for Studies in Behavioural Neurobiology, Concordia University, Montreal, Canada

*all authors contributed equally

Corresponding author information

corresponding authors: EEM: lizzie.e.manning@gmail.com, LAB: Laura.Bradfield@uts.edu.au, MDI: mihaela.iordanova@concordia.ca

Highlights

- Changing associations between events in the environment cause behavioural conflict
- The prefrontal cortex and amygdala contribute to adaptive behaviour during conflict
- Conflict exists in popular extinction, reversal learning and active avoidance tasks
- Deconstructing behaviour and neural control reveals common features of conflict

Abstract

In complex environments, organisms must respond adaptively to situations despite conflicting information. Under natural (i.e. non-laboratory) circumstances, it is rare that cues or responses are consistently paired with a single outcome. Inconsistent pairings are more common, as are situations where cues and responses are associated with multiple outcomes. Such inconsistency creates conflict, and a response that is adaptive in one scenario may not be adaptive in another. Learning to adjust responses accordingly is important for species to survive and prosper. Here we review the behavioural and brain mechanisms of responding under conflict by focusing on three popular behavioural procedures: extinction, reversal learning, and active avoidance. Extinction involves adapting from reinforcement to non-reinforcement, reversal learning involves swapping the reinforcement of cues or responses, and active avoidance involves performing a response to avoid an aversive outcome, which may conflict with other defensive strategies. We note that each of these phenomena relies on somewhat overlapping neural circuits, suggesting that such circuits may be critical for the general ability to respond appropriately under conflict.

Keywords. Conflict, Extinction, Reversal, Active avoidance, Prefrontal cortex, Amygdala

1 Introduction

The ability to behave in a flexible manner is fundamental to adaptive functioning. Flexible behaviour requires not only that we learn the relationships between events in the environment, but that we update these relationships in the face of changing contingencies. Often, such updating presents as a conflict between previously-established and current knowledge of the world. For example, a stimulus such as a raspberry bush signals the availability of a reward (raspberries) until the berries are depleted in which case the same bush becomes a signal for the absence of reward. In this example, behaviour is modified so that resources are not spent in pursuit of unavailable rewards, and the associative change from a positive to a negative contingency is known as extinction. Beyond this simple form of associative conflict, animals can be faced with situations in which contingencies reverse across pairs or classes of stimuli. In keeping with the example above, the depletion of berries in one place may coincide with the availability of reward, same or different, in another place that was previously non-reinforced. An animal that learns to visit the new berry bush has effectively undergone reversal learning. A third kind of conflict might arise if the animal's return to the berry bush is associated with the risk of predation which, when signalled, may induce a shift in behaviour to avoid harm. In the laboratory this behaviour is referred to as active avoidance.

These three procedures: extinction, reversal learning, and active avoidance, are common laboratory-tested behaviours, and as such, much has been discovered about their behavioural and neural mechanisms. Such procedures are often complex, and the specific parameters employed (e.g. cues, responses, outcomes, types of training) can vary considerably from study to study. Here we have attempted to deconstruct each phenomenon into their often-overlooked subcomponents to better understand their underlying behavioural principles as well as to provide insight into their underlying neural substrates. We hope to highlight how considering these processes individually can enhance interpretation of

experimental findings and illuminate the neural mechanisms associated with particular types of conflict versus those associated with responding under conflict more generally.

2. Extinction and Overexpectation

2.1 Extinction.

One simple but particularly common and widely studied form of associative conflict is captured in extinction. In experimental extinction, a stimulus that has previously been established as a signal for the *presence* of an outcome is now presented in the *absence* of that outcome (S-O to S-Ø; Table 1). As a result of extinction training, an organism must update its behavioural repertoire from exhibiting high levels of responding to the conditioned stimulus to supressing this responding in the presence of the same stimulus.

The conflict an animal is presented with when trained in extinction is not eliminated once extinction has been acquired. It is generally believed that extinction training does not erase the original association. Rather, two conflicting or opposing associations (S-O; S-Ø) develop and they compete for behavioural expression. Under varying test conditions, the relative strength of these associations contributes to behavioural responding. For example, one way to manipulate the relative strength of acquisition and extinction is through the number of training trials, which reveals that stronger acquisition competes with extinction more effectively than weaker acquisition, and stronger extinction competes more effectively with acquisition compared to weaker extinction (Chan & Harris, 2017; Harris & Andrew, 2017; Kalish, 1954; Rosas, García-Gutiérrez, & Callehas-Aguilera, 2007; Wagner, 1961). Another way to manipulate the relative strengths of the acquisition and extinction memories is through retrieval cues. This is best exemplified in the renewal effect. In renewal, a change in the context from extinction training to test results in high levels of behavioural responding. This is believed to be due to minimal contextual control over the acquisition memory compared to that of extinction (Bouton, 1994; 2004; Harris, Jones, Bailey, & Westbrook, 2000). As a result, any change in the extinction context tips the associative scales in favour of the acquisition memory. A similar argument has been made to account for the

spontaneous recovery of behavioural responding, as the passage of time following extinction training is also thought to constitute a change in context (e.g., Bouton, 1993; 1994).

2.2 Overexpectation.

Another, much less often investigated, way to reduce previously-established behavioural responding is through overexpectation. In overexpectation two individually trained good predictors of the same outcome are presented together in compound, yielding an inflated expectation of double the outcome. This inflated expectation is reinforced by a single outcome, which drives a reduction in the associative strength of each stimulus with the outcome and results in a reduced level of responding on test. Overexpectation, like extinction, generates two opposing associations (S-O; S-Ø or S-reducedO, depending on the design) which compete for behavioural expression. In other words, overexpectation provides conditions for learning that are in conflict with what was learned during acquisition. And again, like extinction, the strength of the overexpectation memory relative to that of acquisition depends on the context where retrieval takes place. Testing outside of the context where overexpectation training took place, be this a physical context or one determined by the passage of time, restores behavioural responding (Rescorla, 2006; 2007).

2.3 Common neural mechanisms in extinction and overexpectation.

The similarity between extinction and overexpectation outlined above suggests that both forms of learning may be reliant on common neurobiological structures. Investigations into the brain mechanisms of overexpectation are sparse, but evidence for common neural ground between both tasks exists. Behavioural electrophysiological recording during learning in extinction and overexpectation revealed that single units in the central nucleus of the amygdala (CeA) track reduction in outcome expectation in both designs and do so in a correlated fashion. That is, the greater the reduction in neural firing to a stimulus undergoing extinction, the greater the reduction in neural firing to a stimulus undergoing overexpectation

(Iordanova, Deroche, Esber, & Schoenbaum, 2016). These changes in neural firing were also linked to changes in conditioned responding. Causal examination confirms that inactivation of the CeA disrupts overexpectation learning (Haney, Calu, Takahashi, Hughes, & Schoenbaum, 2010) and data from Mihaela Iordanova's lab using the Daun02 inactivation procedure (Koya, Margetts-Smith, & Hope, 2016) provide evidence for the causal contribution of neuronal ensembles in the CeA in extinction learning (Lay, Koya, Esber, & Iordanova, 2020b). Studies in fear have also implicated the CeA in extinction using recording methods (Duvarci, Popa, & Paré, 2011) whereas disruptive methods reveal a role of the CeA in fear expression (Amorapanth, LeDoux, & Nader, 2000; Campeau & Davis, 1995; Kim & Davis, 1993). To our knowledge, there is no current evidence for the role of the CeA in overexpectation of fear.

Another common neural target between extinction and overexpectation is the basolateral amygdala (BLA, (Herry et al., 2008; Laurent & Westbrook, 2010; Sengupta, Winters, Bagley, & McNally, 2016). Evidence for the role for the BLA in both paradigms comes from fear learning. Inactivation or blockade of the NR2B subunit of the NMDA receptor in the BLA prior to extinction learning disrupts the reduction in responding normally seen on test (Amano, Duvarci, Popa, & Paré, 2011; Herry et al., 2008; Laurent & Westbrook, 2010; Lingawi, Westbrook, & Laurent, 2017; Livneh & Paz, 2012; Sierra-Mercado, Padilla-Coreano, & Quirk, 2011). Activation of BLA glutamatergic neurons using the excitatory hM3Dq Designer Receptors Exclusively Activated by Designer Drugs (DREADDs) during overexpectation learning prevented the reduction in conditioned responding (Sengupta et al., 2016). Examinations into the role of the BLA in extinction and overexpectation of reward learning reveal less consistency between correlational and causal method. Single unit recordings implicate the BLA in tracking outcome omission during extinction (Toyomitsu, Nishijo, Uwano, Kuratsu, & Ono, 2002; Tye, Cone, Schairer, & Janak, 2010), and in changes in reward expectancy in overexpectation (Lucantonio et al., 2015). Yet, disruption of BLA

function prior to overexpectation or extinction learning did not disrupt either effect (Haney et al., 2010; Lindgren, Gallagher, & Holland, 2003).

2.4 Dissociable neural mechanisms of extinction and overexpectation.

Despite these commonalities, dissociable neural processing of extinction and overexpectation does exist. In the context of reward learning, pharmacological inactivation of the infralimbic cortex (IL) prior to extinction training disrupts extinction recall the next day (Lay, Nicolosi, Usypchuk, Esber, & Iordanova, 2019) and optogenetic activation of IL neurons enhance extinction recall following renewal and spontaneous recovery (Villaruel, Lacroix, Sanio, Sparks, Chapman, & Chaudhri, 2018). Inactivation of the IL prior to overexpectation training, however, leaves this effect intact (Lay et al., 2019). This dissociation has also been reported in fear (Lay, Pitaru, Boulianne, Esber, & Iordanova, 2020a) with considerable evidence specifically linking the IL to fear extinction (Burgos-Robles, Vidal-Gonzalez, Santini, & Quirk, 2007; Do-Monte, Manzano-Nieves, Quiñones-Laracuente, Ramos-Medina, & Quirk, 2015; Laurent & Westbrook, 2009; Sierra-Mercado et al., 2011). Single-unit electrophysiological recordings are consistent with the causal evidence. Increase in neural firing in the IL is seen to cues that have undergone extinction training, and the greater the firing the lower the fear (Milad & Quirk, 2002). Not surprisingly, disruption of IL input to the BLA disrupts extinction of fear (Bloodgood, Sugam, Holmes, & Kash, 2018).

Another brain area implicated in reduction of conditioned responses as a result of reduction in outcome expectation is the ventrolateral orbitofrontal cortex (vlOFC). VLOFC neurons track reward expectation during overexpectation (Takahashi et al., 2013; 2009) and inactivation of the vlOFC prior to appetitive overexpectation training disrupts the reduction in response normally seen on test (Haney et al., 2010). The role of the vlOFC in extinction is less clear. Inactivation of the vlOFC does not disrupt appetitive extinction (Burke, Takahashi, Correll, Brown, & Schoenbaum, 2009; Clarke, Robbins, & Roberts, 2008). However, reversal

studies (i.e., A→ O; B→ Ø in phase 1 followed by A→ Ø; B→ O in phase 2; see below) implicate the OFC in learning from negative (but also positive) feedback (Clarke et al., 2008). That is, in retardation in learning the A→ Ø association in phase 2. Investigations into extinction and overexpectation in fear may provide some insight. Chemogenetic inactivation of the vlOFC with hM4Di DREADDs (Zimmermann, Li, Rainnie, Ressler, & Gourley, 2018) or NMDA receptor *activation* (Y.-H. Chang, Liu, & Chang, 2018) both lead to a disruption in the loss of the conditioned response. Further data from Lay et al (2020a) confirm that vlOFC inaction prior to extinction results in a mild disruptive effect compared to an extinction group trained with a functional vlOFC. Importantly, when compared to non-extinguished controls, vlOFC-inactivated rats show a considerable extinction effect, suggesting that the vlOFC is not necessary to learn from extinction per se. In contrast to extinction, Lay et al. (2020a) report a catastrophic effect of vlOFC inactivation on overexpectation of fear. In the absence of a functional vlOFC during overexpectation learning, conditioned responding on test remained as high as that of controls that had not received overexpectation training at all.

2.5 Behavioural analyses of the extinction and overexpectation.

The key to understanding the specific involvement of the above-mentioned neural substrates in extinction and overexpectation lies in uncovering the behavioural processes that underscore each paradigm. Discovering common neural ground between these two designs offers insight into brain function by pointing to the common behavioural process that may underlie them. For example, reduction in outcome expectancies or changes in attention are plausible processes that account for extinction and overexpectation and the CeA and BLA have both been linked to those processes (Holland & Gallagher, 1993; Holland & Maddux, 2010; Iordanova et al., 2016; Roesch, Calu, Esber, & Schoenbaum, 2010). Alternatively, the dissociation of the IL and vlOFC in extinction and overexpectation could be understood in terms of differences in the underlying behavioural processes. The study of the behavioural processes that underlie experimental extinction has a long-standing tradition. Important and

extensive insight into these processes have been provided in influential reviews by Mark Bouton (Bouton, 1994; 2004), Andrew Delamater, Fred Westbrook (Delamater, 2004; 2012; Delamater & Westbrook, 2014) and Robert Rescorla (Rescorla, 2001).

One particularly exciting piece of evidence that has emerged is the finding that the associative process that underscores the inhibitory effect of extinction training on conditioned responding depends on the strength of the conditioning memory (Delamater, 1996; 2012; Delamater & Oakeshott, 2007; Rescorla, 1996). Delamater and colleagues trained animals to acquire two Pavlovian (S1-O1 and S2-O2) and two instrumental (R-O1 and R2-O2) associations. Subsequently, one of the Pavlovian associations was extinguished (S1-Ø), but not the other. To examine what was learned during extinction of S1, Delamater and colleagues used a Pavlovian-to-Instrumental transfer test during which presentation of the Pavlovian stimuli (S1 and S2) were presented while the animals engage in instrumental responding. Under normal conditions, i.e., when the Pavlovian stimuli have not undergone extinction, each of the stimuli potentiates the instrumental response that yields a congruent outcome. That is, S1 potentiates R1 and S2 potentiates R2 because they are associated with O1 and O2, respectively. Importantly, successful extinction of the S1-O1 association evidenced by reduced magazine approach reduced responding to R1 in the presence of S1 only when the initial acquisition of the Pavlovian S1-O1 association was limited in terms of training trials/days. When acquisition of the S1-O1 association was extensive, extinction of S1 did not prevent S1 from invigorating responding to R1.

These findings are compelling. They suggest that extensive training of Pavlovian associations leaves the S-O association intact following extinction (which presumably establishes a separate inhibitory S-R association), whereas limited training renders those associations subject to disruption. It is worth emphasizing that using the same paradigm but varying training duration can lead to qualitative differences in learning, that is, in the underlying associative architecture that supports behaviour.

These data are worth relating to the brain function mechanisms of extinction and overexpectation discussed earlier. Specifically, Lay et al. (2019) conducted extensive appetitive Pavlovian training before extinction, which according to the data reported by Delamater and colleagues would spare S-O associations and establish inhibitory S-R associations. Therefore, the disruption in extinction following IL inactivation reported by Lay et al (2019) may be due to modulation in the S-R association, leaving the downregulation of S-O associations to the OFC (Takahashi et al., 2009; 2013). It remains unknown whether the IL or the OFC are involved in appetitive extinction following limited acquisition training, but this examination would be of considerable interest. The distinction of limited versus extensive training is less clear in fear studies, in which conditioning often consists of few trials (e.g. under 10). Perhaps conditioning with a salient fear-eliciting outcome speeds the asymptote of learning essentially mimicking extensive training with rewarding outcomes, resulting in similar neural modulation of extinction and overexpectation in fear and reward. This of course would require experimental analysis.

Alternatively, the dissociation in the role of the IL and OFC in learning from extinction and overexpectation may be due to their procedural differences. Specifically, the role of the IL in extinction may be specific to the case of outcome omission. On the other hand, the role of the OFC in overexpectation may be in generating novel predictions as is the case when two individually trained cues sum their associative strengths to generated an inflated expectation of the outcome. Whatever the case may be, uncovering the functional role of specific brain structures (or pathways) requires and in-depth analysis of the behavioural processes that underlie behavioural tasks.

3 Reversal learning

Reversal learning, like extinction learning, involves conflict. Reversal learning broadly refers to any situation in which two or more cues or responses are associated with a particular outcome, or no outcome, and these cue-outcome (or response-outcome) relationships are

switched. There are a diverse variety of reversal tasks, each of which invoke different psychological processes and rely on distinct neural substrates. For instance, tasks can vary with regards to their structure, the type of reinforcement, stimuli that are present, and the responses required. Despite this diversity, effects on reversal following neural manipulations or disease states are often talked about broadly, in terms of behavioural flexibility. Here we aim to deconstruct the psychological phenomena underlying different types of reversal learning in order to hone in on the specific behavioural mechanisms (e.g. inhibition, attention) that underlie behavioural flexibility captured across these different reversal designs. As mentioned earlier, not only would this behaviour-centric approach provide more information with regard to brain function it also limits possible misattribution of the function of the particular neural structure, circuit, or ensemble to a general flexibility process that may be hard to define conceptually.

3.1 Reward/No Reward Reversal task

Probably the most common type of reversal task is that which we will refer to as the 'reward/no reward' task (e.g. Schoenbaum, Chiba and Gallagher, 1999; Schoenbaum *et al.*, 2002; Chudasama and Robbins, 2003; Rudebeck and Murray, 2011; Bell *et al.*, 2019). In this task, the rodent (or primate) is first presented with a stimulus (S1, e.g. a tone) paired with a rewarding outcome (O1, e.g. a grain pellet), whereas an alternative stimulus (S2, e.g. a light) is paired with nothing (i.e. S1-O1, S2-Ø, design in Table 1). During this initial phase, the animal learns to perform the response during presentations of S1 but not S2 (e.g. to enter the food receptacle during tone but not light presentations). These contingencies are then reversed (S1-Ø, S2-O1) such that the animal learns to perform the response during the now-rewarded stimulus (S2) and to reduce responding during the now non-rewarded stimulus (S1). Instrumental versions of this task are also common in which stimuli are substituted for responses (e.g. a left and a right lever press).

For these and all reversal tasks, the reversal phase is typically only completed if the manipulation of interest (neural manipulation, recording, psychiatric disorder) did not affect initial learning. This is because a manipulation that affects initial learning likely affects learning generally (or reward processing, or some other general process), making any specific effects on reversal impossible to disentangle from general learning effects. If a deficit/facilitation is found to be specific to the reversal phase however, the question remains as to what specific underlying psychological phenomena is being targeted.

One type of learning that underlies reversal and could be affected by manipulations is extinction (see Table 1). Indeed, if the training history of S1 is taken alone it is identical to that of extinction (i.e. S1-O1, S1-Ø), with the only difference that reversal additionally features the presence of the oppositely-reinforced S2. Therefore, studies have asked whether the presence of S2 completely alters the learning process to S1 so that it does not undergo extinction learning. Performance to S1 during reversal is certainly consistent with extinction learning. That is, 'perseverative errors', defined as the continued responding during the previously-rewarded stimulus (S1) during reversal (Rayburn-Reeves, Molet and Zentall, 2011; Butts, Floresco and Phillips, 2013), typically decrease in frequency in a negatively accelerated fashion which mirrors the reduction in responding observed in extinction (Rescorla, 2002). Whether this *performance* reduction is also reflective of extinction *learning*, however, is not immediately apparent.

To investigate this question further, Rescorla (2007) examined whether the phenomenon of spontaneous recovery (as described earlier, the observation that extinguished responding will return after a period of time has passed) was also present after reversal learning. Spontaneous recovery is thought to indicate that the original (S1-O1) contingency survives extinction but is temporarily inhibited by a new S1-Ø1 association formed during extinction. Rescorla produced evidence that, just as it does after extinction, responding to S1 spontaneously recovers when animals are tested 1 week after the end of reversal training. This suggests that, the initial S1-O1 contingency survives reversal just as it does in

extinction, and likewise points to the formation of a new S1-Ø1 contingency at the reversal stage, as also occurs in extinction. The implication of these observed parallels, of course, is that any manipulation that affects extinction learning might also affect reversal learning in this type of task.

Just as the training history of S1 during reversal follows that of extinction, if the training history of S2 is taken alone then its trajectory is identical to latent inhibition (initially S2-Ø, then S2-O1 during reversal, design is in Table 1). Latent inhibition occurs when a stimulus is initially presented with no reward, which 'latently inhibits' (i.e. reduces) the propensity of an animal to later associate that cue with reward (Lubow and Moore, 1959). In several findings that are analogous to spontaneous recovery from latent inhibition, there have been demonstrations that the inhibitory stimulus later paired with excitation will return to its initial inhibitory status following the passage of time (De la Casa and Lubow, 2002; Sissons and Miller, 2009) (but see (Rescorla, 2005)). Within the context of a reward/no reward reversal task, however, the S2 stimulus was found to *increase* its responding over time, suggesting that it wasn't latently inhibited (Rescorla, 2007). This finding suggests that the content of what is learned about S2 during reversal is different to latent inhibition. However, it is also possible that the presence of the non-reinforced S1 on the same test as S2 acted as a retrieval cue for the S1-Ø, S2-O1 (reversal) contingencies, which then excited responding to S2. Distinguishing between these possibilities can only be achieved through future testing, and one way to do so might be to test S1 and S2 in separate sessions such that S1 could not act as a retrieval cue for S2. Establishing that the trajectory of S2 during reversal is a separate phenomenon to that of latent inhibition, as is suggested by Rescorla's study, would imply that any manipulation that affects latent inhibition would not necessarily affect reversal learning.

3.2 Probabilistic Reversal Tasks

A similar but distinct task is the probabilistic reversal task (Cools *et al.*, 2002; Dalton, Phillips and Floresco, 2014). In these tasks, one stimulus (or response) is associated with an outcome say 70% of the time, and another is also associated with that outcome but with a lower probability, say 30% of the time (see Table 1). These probabilities are then later reversed such that the high probability stimulus becomes the low probability stimulus and vice versa.

Probabilistic tasks are inherently more complex than the reward/no reward reversal task, and likely engage a number of different attentional and learning processes even prior to reversal such that separating these from specific reversal effects is difficult. For instance, a stimulus that is probabilistically associated with an outcome 70% of the time fails to become a perfect predictor such that the outcome – or indeed lack of outcome – is still surprising on some trials. This might maintain attention to the stimulus and thus affect its tendency to associate with an outcome (i.e. its 'associability' Pearce and Hall, 1980; Esber and Haselgrove, 2011). In addition, model-based and model-free processes may be differentially engaged (Gläscher *et al.*, 2010), as well as different win/stay, lose/shift strategies (see Coutanche and Thompson-Schill, 2012; Brady and Floresco, 2015, for discussion). These differences could conceivably be ‘carried over’ to affect reversal performance if, say, differences in associability between two stimuli affect how quickly the reversed contingencies are attributed to each stimulus. Specifically, the more ‘associable’ stimulus should enter more readily into the new, reversed contingencies than the less associable stimulus. However, experimenters who employ this task often recognise the potential for confounds such as this, and run control experiments, and/or engage in deeper analysis of reversal performance to determine which particular aspect of probabilistic reversal might have been affected (e.g. Dalton, Phillips & Floresco, 2014; Groman et al., 2019; Harris et al., 2020).

3.3 Counterconditioning

Counterconditioning refers to the initial pairing of a stimulus with an outcome of a particular valence (appetitive or aversive) which is later paired with an outcome of the opposite valence (see Table 1). Although not typically considered to be a reversal task, it does involve a reversal of sorts as the animal must learn to 'reverse' the valence it associates with the stimulus.

In two experiments similar to that described for reward/no-reward reversal learning described above, Bouton and Peck (1992) investigated the propensity of an initial S-O association to survive counterconditioning using spontaneous recovery. Specifically, for these experiments a stimulus was initially paired with shock, then paired with food (Experiment 1) or vice versa (Experiment 2). If tested immediately after initial training, the animals responded in accordance with the most recent training experience (head-jerking in Experiment 1 where food-pairing was most recent, and freezing in Experiment 2 where shock-pairing was most recent). If, however, animals were tested 4 weeks later, they again responded in accordance with the initially-trained contingencies. This suggests that, like extinction, the initially-trained contingencies also survive counterconditioning, and are likely inhibited rather than erased.

3.4 Outcome identity reversal in an instrumental task

Bradfield et al., (2013; 2017) employed yet another kind of reversal task in an instrumental paradigm that involved the reversal of two response-outcome contingencies in which the outcomes were assumed to be of equivalent value but different sensory-specific characteristics (R1-O1, R2-O2 were reversed to become R1-O2, R2-O1, see Table 1). Specifically, animals were first trained to press a left lever for pellets and a right lever for sucrose (or vice versa, counterbalanced), which was reversed such that each lever now earned the opposite outcome. Contingency knowledge was probed after each phase using devaluation tests (see (Bradfield and Balleine, 2017) for details).

In this instance, and in contrast to the other reversal tasks described above, whether animals were tested 1 or 3 weeks after reversal training, the initial contingencies were not found to spontaneously recover (Bradfield and Balleine, 2017). That is, even after delays were inserted between training and test, animals continued to respond in accordance with the reversed contingencies. This is of interest, because it suggests that whatever learning process occurs during outcome identity reversal in an instrumental task, it is different to that which occurs in a reward/no reward reversal task and in counterconditioning. Indeed, taken alone, this finding might be interpreted to suggest that this type of reversal learning leads to an erasure rather than an inhibition of the initial contingencies. However, once the reversed contingencies had been explicitly extinguished in this same task, animals once again responded in accordance with the first-trained set of contingencies. This suggests that, just as the initial contingencies survive in both the reward/no reward and counterconditioning reversal tasks, they also survive the instrumental outcome-identity reversal task.

3.5 Common and distinct psychological processes that underlie reversal learning

The reversal tasks outlined here involve a number of different psychological processes such as stimulus-outcome learning, response-outcome learning, extinction, strategic thinking, and counterconditioning, among others. As such, the findings that result from the employment of such tasks in conjunction with some other manipulation (e.g. neural structure/circuit inactivation, neural recordings, psychiatric disease) must be interpreted carefully, to determine whether it is one of these processes being affected/measured, or whether it is a process that is specific to reversal. If a manipulation does broadly affect reversal learning across tasks, then it is less likely that it is doing so indirectly (e.g. through effects on extinction). Thus, the question remains as to what cognitive capacity is common across reversal tasks.

One commonality observed between all of the reversal tasks outlined above (with probabilistic reversal learning remaining to be tested) is that the initial learning survives the reversal process. Thus, one psychological process that might be necessary for accurate

performance on all reversal tasks is that of 'context' or 'state' modulation (Bouton, 1993; Wilson *et al.*, 2014; Bradfield & Hart, 2020), a concept that is similar to that of occasion setting (Holland, 1992; Bouton, 1993). That is, when the animal is faced with two sets of contingencies that compete with each other, as they do in reversal learning, they must parse these contingencies into two separate states or contexts. For example, for the reward/no reward reversal task, they might learn that "when State 1 is active, S1 leads to reward and S2 doesn't, but when State 2 is active, S2 leads to reward and S1 doesn't". To put this in the language of occasion setting, the internal state 1 'sets the occasion' for S1-O, S2-Ø, and internal state 2 'sets the occasion' for S1-Ø, S2-O1. An inability to parse contingencies into separate states in this manner could lead to deficits on any reversal task, regardless of its nature. Indeed, the orbitofrontal cortex is currently the primary structural candidate for state representation, and as referred to in more detail in the next section, its inactivation has been found to impair reversal learning across a number of different types of reversal task (Wilson *et al.*, 2014; Bradfield and Hart, 2020).

3.6 The neural substrates of reversal learning

Several of the neural substrates of extinction learning outlined above have also been investigated for their role in reversal. For instance, sub-structures within the prefrontal cortex such as prelimbic (PL) and infralimbic (IL) cortices appear to have specific roles in certain types of reversal learning, whereas the vlOFC appears to play a broader role in regulating reversal learning across different tasks. The basolateral amygdala has also been found to regulate some forms of reversal. In addition, several other neural structures that are not particularly known to have any role in extinction learning have been found to regulate forms of reversal, including hippocampus, mediodorsal thalamus, and the parafascicular-controlled cholinergic interneurons in the striatum.

With regards the prefrontal cortical substructures, evidence for their role in reversal learning has been mixed. Boulougouris et al., (2007) and Ashwell and Ito (2014) both found no effect of PL lesions on reversal in a spatial discrimination task, and Floresco et al., (2008) similarly

found no effect of medial prefrontal cortical (primarily PL) inactivation on reversal learning in a cue-response reversal task. By contrast, when the task involved a shift in strategy (e.g. from responding on a left or right lever based on whether the cue light above it is illuminated to responding only on the left lever regardless of which cue light is illuminated), whether in the context of reversal or not, PL inactivation did appear to impair performance (Floresco, Block and Tse, 2008; Oualian and Gisquet-Verrier, 2010; Dalton *et al.*, 2016). These studies appear to suggest, therefore, that the PL does not regulate reversal learning *per se* but does regulate switching between strategies.

Several studies have also failed to find any clear role for the IL in reversal learning (Boulougouris, Dalley and Robbins, 2007; Dalton *et al.*, 2016), although Ashwell and Ito (2014) found a facilitation of reversal learning after lesions of the IL. Oualian and Gisquet-Verrier (2010) also found that the IL, like the PL, was involved in the choice of a new strategy after reversal learning had taken place. Thus, the evidence suggesting a role for the IL in reversal learning is mixed and depends on the type of task employed, suggesting that it relates to an underlying function that is not reversal specific, such as extinction learning which is IL-dependent.

The OFC (primarily vlOFC) has been heavily implicated in reversal and not initial learning across a diverse range of tasks. For example, Schoenbaum et al., (2003) found a role for vlOFC in reversal of an odor discrimination task, whereas Thorpe et al., (1983) found a role for vlOFC in reversal of a visual discrimination task, which is consistent with a number of other findings (O'Doherty *et al.*, 2001; Chudasama and Robbins, 2003; Izquierdo, Suda and Murray, 2004; Hervig *et al.*, 2019) . Moreover, Parkes et al., (2018) found that identity-based reversal learning of action-outcome contingencies depended on vlOFC, and several other studies have found a role for OFC on reversal in spatial tasks (Boulougouris, Dalley and Robbins, 2007; Klanker *et al.*, 2013). Common across these results was the finding that OFC manipulations and recordings detected effects specific to reversal and not initial learning. The fact that OFC has such a broad role across different types of reversal tasks suggests

that it might encompass a function that is specific to reversal, rather than a different type of learning such as extinction (see earlier). A recent study by Farovik et al. (2015) proposed a potential mechanism for OFC regulation of reversal learning via state representation, showing that in different contexts the same object (a ceramic pot) can create very different firing patterns depending on its association with a reward. They identified an apparent hierarchy with regards to the systematic structure of OFC neuronal population representation, demonstrating that events with opposite reward value (as occurs in reversal learning) are represented most distinctly from each other, objects associated with the same outcome but in different contexts represented next most distinctly, whereas different objects in different locations that were nonrewarded were encoded similarly. The authors suggested that the organisation of this structure was suggestive of OFC representing 'states' that were separated according to likelihood of receiving reward; partitioning of OFC responses according to object-reward associations was more definitive than any other partitioning by context, location, response, or object.

There has been some evidence for the role of other structures in reversal learning, such as the hippocampus (McDonald, Ko and Hong, 2002; López *et al.*, 2003), although the evidence suggests that this is only the case for tasks that have a strong spatial or contextual element (McDonald, Ko and Hong, 2002; Schoenbaum, S. Nugent, *et al.*, 2002). Likewise, there have been some studies linking basolateral amygdala function to reversal learning, possibly via its connections with OFC (Schoenbaum, 2003), however the basolateral amygdala is known to play such a central role in fundamental associative processes that any role it plays in reversal is likely linked to these.

Finally, several thalamic structures have been implicated in different types of reversal learning also, with mediodorsal thalamus implicated in reversal of instrumental contingencies (Parnaudeau *et al.*, 2015). Likewise, the parafascicular thalamic nucleus has been implicated in regulation of an instrumental identity-based reversal learning task (Bradfield *et al.*, 2013; Bradfield and Balleine, 2017), and has also been implicated in the regulation of

spatial reversal learning (Thompson, Kao and Yang, 1981; Brown, Baker and Ragozzino, 2010). This role for parafascicular thalamic nucleus is likely achieved through its inputs onto striatal cholinergic neurons (Brown, Baker and Ragozzino, 2010; Bradfield *et al.*, 2013), which have also been implicated in the regulation of a number of different reversal tasks (Ragozzino *et al.*, 2009; Okada *et al.*, 2014; Aoki *et al.*, 2018). Moreover, the function of striatal cholinergic interneurons is also thought to depend on the integrity of the lateral OFC (Stalnaker *et al.*, 2016), making this a likely circuit for regulating reversal performance.

4 Active Avoidance learning

When an individual faces threat, different aversive learning systems can compete with each other for control of behavioural output, creating conflict. This section focusses on active avoidance, which describes behaviours that reduce the occurrence of aversive outcomes when they are performed. This is in contrast to passive avoidance and punishment, which reduces the occurrence of aversive outcomes when specific behaviours are *not* performed. Active avoidance learning often conflicts with other behavioural responses when an individual faces threat, including Pavlovian defensive behaviours like freezing. For example, a rat in the wild might experience conflict between navigating to a safe location to avoid potential threat of attack or freezing to avoid being detected, and adaptive resolution of these scenarios is critical for survival.

In the laboratory, active avoidance is typically studied in rodents using a shuttle box or operant chamber, where specific behavioural responses (e.g. moving to the opposite side of the shuttle box or pressing a lever) are associated with omission of the aversive outcome (R-$Ø_{aversive}$). Experimental paradigms often include stimuli that predict the adverse outcome, and thus indicate that avoidance responses can be made, or stimuli that signal successful completion of the avoidance behaviour (i.e. safety). Debate around active avoidance learning has largely centred around whether avoidance responses are Pavlovian reactions to conditioned stimuli (S-$O_{aversive}$), or instrumental actions elicited with the aim of avoiding

potential threats (i.e. outcomes) in the presence of particular stimuli [S(R-Ø$_{aversive}$)]. Here we suggest that both processes contribute to active avoidance learning, and aim to dissect distinct stages of learning to provide a clearer examination of the associations underlying behaviour (Table 2). Elucidating these processes could, in turn, open new avenues for investigation of the associated neural circuitry.

4.1 What are the reinforcers of avoidance learning?

Debate about whether avoidance learning produces outcome dependent actions has focussed, in part, on identifying the 'reinforcer' of avoidance learning. The "two-factor" theory – a dominant theory in the field – suggests that avoidance learning occurs in two distinct phases (Mowrer & Lamoreaux, 1946). First, Pavlovian fear conditioning generates an association between a stimulus (S; e.g. tone) and the aversive outcome (e.g. shock; S-O$_{aversive}$), so that the stimulus itself comes to elicit fear. Next, the organism learns to perform (instrumental) avoidance responses to the stimulus, which cause omission of the aversive outcome and reduces fear to the stimulus (R-O$_{fear\ reduction}$). The stimulus can then be considered as a "warning signal" that indicates when avoidance responses can be made. During the instrumental phase, learning is thought to be supported through negative reinforcement, because behaviours are 'reinforced' by the reduction or removal of an aversive outcome.

Two-factor theory suggests that negative reinforcement of the instrumental avoidance responses is mediated by fear reduction resulting from removal or termination of the fear-evoking stimulus. There is some evidence to support this (Cain & LeDoux, 2007; Miller, 1948; Overmier & Bull, 1969; Rescorla & Solomon, 1967), however it has also been repeatedly demonstrated that acquisition of avoidance learning is associated with reduced expression of fear-associated behaviours (e.g. freezing) during the stimulus (Bravo-Rivera, Roman-Ortiz, Brignoni-Perez, Sotres-Bayon, & Quirk, 2014; Choi, Cain, & LeDoux, 2010; Fernando, Urcelay, Mar, Dickinson, & Robbins, 2014b). This reduction in fear would be

expected to decrease negative reinforcement and potentially extinction of avoidance responses, but this is not observed in well-trained animals. This suggests that a process other than fear reduction may be responsible for reinforcing avoidance responses, at least once avoidance responding has been established and fear is low.

What this alternative process might be is unclear, but there is evidence suggesting that safety signals could play a role. Safety signals are stimuli that are present when an avoidance response is made, such as a tone or contextual stimuli associated with a safe area in the testing environment. In contrast to warning signals, safety signals positively reinforce behaviour, and there is some evidence that they act as a positive reinforcer during avoidance learning (Fernando et al., 2014b). Interestingly, it has also been demonstrated that the presentation of a safety signal following avoidance responses can replace warning signals to promote avoidance learning, suggesting that under some conditions warning signals (and their termination) may act as a safety signal that positively reinforces avoidance responses (Bolles & Grossen, 1969; Bower, Starr, & Lazarovitz, 1965). Overall, while there is strong evidence that specific outcomes can reinforce instrumental avoidance responding, it remains unclear what conditions are associated with negative vs positive reinforcement (or some combination of both) as the driver of this instrumental avoidance learning.

4.2 Outcome dependence across different phases of avoidance training

An important consideration for assessment of the outcome dependency of active avoidance is the amount and type of training a subject has received, as there is growing evidence that the associative structure of avoidance responses changes across training. Often laboratory avoidance paradigms expose subjects to hundreds of avoidance trials in order to reach asymptotic performance before some experimental manipulation is performed. In appetitive instrumental learning, this type of over-training is known to cause a transfer from outcome-dependent goal-directed actions to stimulus-dependent habits (Dickinson, Balleine, Watt, Gonzalez, & Boakes, 1995). Similar mechanisms have been proposed for avoidance

learning, whereby avoidance responses are dependent on avoiding the aversive outcome early in training, but become increasingly dependent on the stimulus that predicts the aversive outcome as training continues. It is possible that such 'habitual' transfer is what supports maintenance of avoidance responding when exposure to aversive outcomes becomes rare, and fear is reduced, which might be expected to reduce motivation to make instrumental avoidance responses as described above (LeDoux, Moscarello, Sears, & Campese, 2017).

However the idea that avoidance responses become habitual with overtraining has not been demonstrated experimentally. In the appetitive learning literature, outcome dependence of responding is tested using procedures such as outcome devaluation (or revaluation), to determine whether responding is sensitive to changes in the value of the outcome (Balleine & Dickinson, 1998), with habitual responses being insensitive to these manipulations. In rats, attenuation of avoidance responding following revaluation of the aversive outcome has been demonstrated in two distinct paradigms (Fernando, Urcelay, Mar, Dickinson, & Robbins, 2014a; Hendersen & Graham, 1979), and in healthy human subjects avoidance responses have been shown to be sensitive to outcome devaluation (Gillan et al., 2014; Patterson, Craske, & Knowlton, 2019). Thus, these studies provide support that avoidance is outcome-, not stimulus- dependent, which supports the notion that their elicitation was goal-directed rather than habitual, at least at the time of testing. To our knowledge, however, there has been no similar investigation of outcome revaluation in avoidance learning after extended training, thus whether it does indeed become habitual with additional training remains an open question.

4.3 Avoidance responding along the threat imminence continuum

An alternative to the two-factor theory is the proposal that all avoidance responses are stimulus-dependent reactions (Bolles, 1970). Specifically, Bolles (1970) suggested that inflexible, pre-wired defensive reactions are the most efficient strategy for responding under

threat. In addition, he proposed that under certain experimental conditions, inflexible reactions could meet the avoidance response criteria, and be incorrectly interpreted as an outcome-dependent avoidance response. For example, in shuttle box avoidance, a stimulus-dependent flight reaction to a warning signal might result in the subject moving to the safe compartment, which could be misconstrued as an avoidance response that was elicited to avoid the footshock outcome.

This account is clearly at odds with the two-factor theory outlined above, which states a clear role for outcome-dependent avoidance responses. However, Fanselow and colleagues helped to reconcile these contradictory hypotheses by introducing the notion that defensive responses are determined by the current levels of threat imminence (Fanselow, 1997; Fanselow & Lester, 1988). They proposed that at high levels of threat imminence, only pre-wired species-specific reactions can be selected (in a stimulus-dependent manner) to ensure a rapid response, similar to what Bolles described, whereas at low threat imminence, behaviour is more flexible. Others have extended this framework to suggest that that flexible outcome-dependent avoidance learning may occur when threat imminence is lower (Cain, 2019; Campese et al., 2016; Mobbs, Headley, Ding, & Dayan, 2020), and where slower trial-and-error learning does not risk severe outcomes like injury or death.

It is likely that some laboratory avoidance research has been conducted at high threat imminence, where the responses measured by experimenters may have reflected stimulus-dependent reactions in the manner proposed by Bolles (Cain, 2019). Working from this hypothesis, Cain and colleagues predicted that reduction of threat imminence would promote avoidance learning over inflexible stimulus-dependent responses. Indeed, they found that introducing a longer warning signal in a two-way shuttle box task, which decreases the imminence of the threatening outcome, led to more rapid acquisition and improved learning rates for avoidance responses (Laughlin, Moloney, Samels, Sears, & Cain, 2020). This result clearly supports the threat imminence account and provides a useful framework for examining how outcome- and stimulus-dependent responses contribute to active avoidance

depending on the level of threat, using outcome revaluation procedures described previously.

4.4 Avoidance extinction learning

Earlier we outlined how extinction learning is associated with conflict between previously learned and new contingencies. In the case of active avoidance, unique challenges are faced to elicit extinction learning. Active avoidance is often considered to be highly resistant to extinction, and there are indeed examples where subjects exposed to hundreds of extinction trials continue to make avoidance responses (Solomon, Kamin, & Wynne, 1953). However, it is important to carefully consider the learning processes associated with the specific avoidance extinction procedures used in these experiments, as this may provide further explanation for the weak extinction effects that have been observed. Similar to extinction protocols described earlier, avoidance extinction training typically involves continued presentation of the warning stimulus, but the aversive outcome (e.g. shock) is now removed. However, in well-trained animals prior to extinction training, the shock is rarely encountered due to high levels of avoidance responding, and therefore the change in contingency during extinction may not be detected. Alternatively, performing the avoidance response may protect the stimulus from extinction by acting as a conditioned inhibitor, since the stimulus is associated with shock *except* when its accompanied by an avoidance response (Rescorla, 1968, 2003). Therefore, during extinction the absence of shocks can be attributed to the presence of the avoidance response, thereby 'protecting' the stimulus from the extinction effects of nonreinforcement. A solution to both scenarios is to block the avoidance response during extinction, which ensures that the change in contingency (removal of the aversive outcome) is encountered by the subject, and prevents conditioned inhibitory effects of the avoidance response that may protect the stimulus from extinction. Response prevention or “flooding” procedures have been developed to improve the efficiency of avoidance extinction (Baum, 1970; Bravo-Rivera, Roman-Ortiz, Montesinos-

Cartagena, & Quirk, 2015; Rodriguez-Romaguera, Greenberg, Rasmussen, & Quirk, 2016), however given the fairly limited research on the associative processes underlying avoidance extinction (with and without response prevention procedures), more detailed investigation is warranted.

4.5 Neural substrates resolving conflict associated with avoidance learning

The neural substrates of avoidance learning overlap with many of the same brain regions that help resolve conflict in both extinction and reversal learning. This appears to be particularly true for the neural substrates of fear extinction, and this is potentially a consequence of active avoidance recruiting extinction processes as outlined above (i.e. in accordance with two-factor theory). For instance, the CeA appears to play a role in both extinction and avoidance learning. Its role in both paradigms appears to relate back to its central role in the expression of conditioned fear, because reduction of neuronal firing in CeA tracks with suppression of conditioned fear responses across extinction (Duvarci, Popa, & Paré, 2011), and likewise, in active avoidance, the suppression of CeA activity is thought to reduce competing fear responses so that avoidance responses can be elicited (Choi et al., 2010; Lazaro-Munoz, LeDoux, & Cain, 2010).

Another structure that plays a central role in fear learning and extinction is BLA. Therefore, as one might expect, BLA activity has been consistently demonstrated to be necessary for the active avoidance learning across a variety of paradigms (Choi et al., 2010; Choi & Kim, 2010; Poremba & Gabriel, 1999). However this is dependent on how much training subjects have received, and expression of avoidance responses becomes BLA-independent following extensive training (Lazaro-Munoz et al., 2010; Poremba & Gabriel, 1999). This finding is thought to reflect BLA involvement in outcome-dependent avoidance responses, and suggests that neural control shifts away from BLA following overtraining when S-R habitual control is believed to have taken over (LeDoux et al., 2017). The involvement of other brain regions in the control of avoidance responding following overtraining has not been explored,

however examination of areas involved in the shift between goal-directed and habitual appetitive behaviour has been proposed as a useful starting point (Cain, 2019) [e.g. dorsolateral striatum (Yin, Knowlton, & Balleine, 2006)].

Similar to BLA, dopamine release in the nucleus accumbens core (NAcC) controls avoidance learning in a manner that is dependent on training duration. Upon acquisition of avoidance responding, NAcC dopamine increases at the onset of the warning signal during trials where the subject successfully avoids the aversive outcome, whereas this is absent during escape trials, and decreased dopamine is observed during warning signals that predict inescapable shock (Oleson, Gentry, Chioma, & Cheer, 2012). The increase in dopamine observed to warning signals during successful avoidance trials mimics the increase in dopamine release patterns observed to stimuli associated with rewards, suggesting that the mesolimbic dopamine system may encode successful avoidance similarly to reward. Infusion of a D1 receptor antagonist into the NAcC impairs avoidance responding when the behaviour has just been acquired, however in well-trained animals NAcC dopamine no longer controls avoidance responding, which may reflect a shift to control by nigrostriatal dopamine systems that are implicated in appetitive habits (Wenzel et al., 2018). There is some evidence that safety signals that follow a successful avoidance response may also recruit mesolimbic dopamine circuitry during avoidance learning. Fernando (2014c) found a critical role of dopamine signalling in the NAc shell (NAcS) for the enhancement of avoidance responding elicited by addition of safety signals (Fernando, Urcelay, Mar, Dickinson, & Robbins, 2014).

Another neural substrate that is common to active avoidance, extinction, and reversal learning is the prefrontal cortex. For example, IL activity is necessary for retrieval of avoidance extinction learning (Bravo-Rivera et al., 2014), similar to its role in appetitive and fear extinction described above. Moreover, recruitment of the IL has been shown to play an important role in suppressing both fear and CeA activity (cFos), which is necessary for the acquisition and expression of active avoidance (Moscarello & LeDoux, 2013). In the PL,

pharmacological inactivation impairs active avoidance by delaying avoidance responses that typically have a very short latency in well trained animals (Bravo-Rivera et al., 2014; Diehl et al., 2018). Interestingly, *in vivo* electrophysiology demonstrated that inhibitory PL neural responses were uniquely associated with tone onset in avoidance trained animals, but not fear trained or naïve rats, and these inhibitory responses were observed even when rats failed to make an avoidance response (Diehl et al., 2018). These findings, along with causal optogenetics manipulations, suggested that PL neural inhibition may be involved in encoding discriminatory stimuli (e.g. warning signal/tone) that signal when a threat can be avoided, which may relate to the proposed role of PL in switching between strategies during reversal learning that was described earlier.

Surprisingly, and in contrast to reversal and extinction learning, relatively little attention has been paid to the role of the OFC during avoidance learning. One study did examine the role of the lateral OFC (lOFC) during retrieval of avoidance extinction using a platform avoidance paradigm, where rats can move onto a safe platform to a avoid a shock during tone presentations. Following acquisition of the task, extinction training with response prevention was performed for three days, by presenting tones without shocks while access to the platform was physically blocked. Avoidance extinction was heterogenous using this procedure, with approximately a quarter of rats showing poor extinction associated with high freezing across the 3 days of extinction and persistent avoidance responding on test day. In contrast, approximately half of the rats showed good extinction, associated with low freezing during extinction training and low avoidance responding at test. Based on these findings, freezing during extinction was subsequently used to predict whether rats were successfully extinguishing, to determine how lOFC manipulations affected retrieval of extinction learning. Interestingly, inactivation of the lOFC prior to the retrieval test had opposing effects on these subgroups; extinction retrieval was impaired in rats that were expected to show low avoidance at test (i.e. that had shown low freezing during extinction), whereas extinction retrieval was enhanced in rats that were expected to show persistent avoidance (Rodriguez-

Romaguera et al., 2016). These findings indicate that lOFC may be responsible for encoding state-dependent value of the outcome of avoidance responses (high value in extinction resistant rats, and low value in extinction sensitive rats) just as it seems to do in reversal learning.

More work is needed to tease out the neural mechanisms supporting outcome-dependent avoidance behaviour, and to understand if and how these shift between acquisition, expression, overtraining and extinction phases. Nonetheless, the studies to date demonstrate two interesting findings regarding outcome-dependent neural control of avoidance. Firstly, the BLA is critical for the acquisition and expression of avoidance, but is no longer necessary to express over-trained behaviour, which may reflect transfer to habitual S-R control (Lazaro-Munoz et al., 2010; Poremba & Gabriel, 1999). Second, the lOFC appears to encode state-dependent value of the outcome of avoidance responses, including revaluation across extinction training (Rodriguez-Romaguera et al., 2016). In appetitive instrumental learning, the lOFC-BLA circuit encodes changes in state-dependent reward value (Malvaez, Shieh, Murphy, Greenfield, & Wassum, 2019) and is necessary for behavioural sensitivity to reward contingency degradation (Zimmermann, Yamin, Rainnie, Ressler, & Gourley, 2017), however the role of this circuit has not been characterized during avoidance learning. The findings outlined above thus suggest that lOFC-BLA circuitry may support outcome-dependent avoidance behaviour, and may provide an important point of integration between conflicting behavioural responses including S-O dependent reactions like freezing, S-R habits, and avoidance extinction.

5 Conclusions

Extinction, reversal learning, and active avoidance are three laboratory paradigms that have been recruited to examine the behavioural and brain mechanisms of different aspects of responding under conflict. Here, we have investigated both the unique and overlapping behavioural mechanisms by deconstructing each paradigm into its component parts and we

have then compared the neural substrates that underlie each finding that, once again, there are some unique mechanisms as well as some overlap. Where there is overlap at a behavioural level, this likely derives from the same psychological process being employed across paradigms. For instance, both reversal and avoidance learning involve a degree of extinction, and all three paradigms require animals to parse conflicting contingencies using state or context modulation. It is no surprise, then, that similar circuitry also appears to underlie learning and performance in each procedure, with prefrontal cortical structures (e.g. PL, IL, and OFC) as well as nuclei within the amygdala being particularly broadly implicated across paradigms, suggesting that these structures might be particularly important for adaptive responding under conflict.

In future assessments of the functions of these brain regions, it may be advisable to take a holistic approach. For instance, if examining the role of a brain region in extinction, it might be worth also testing overexpectation to determine if the region of interest is specifically encoding the absence of the outcome, or more generally encoding the reduction in outcome expectation or inhibitory learning that results from both procedures. Indeed, extending this work to conditioned inhibition in which a cue has a specific role in signalling outcome omission would be particularly telling, as it eliminates the role of conflict between acquisition and extinction. Researchers investigating reversal learning may likewise want to investigate whether their manipulation also affects extinction, as well as the associability of the stimulus/response such that any effects specific to the 'reversal' phase are not a result of these processes. With regards to active avoidance, assuming that fundamental differences in instrumental responding and/or Pavlovian fear conditioning have been ruled out, researchers may wish to employ some kind of outcome evaluation procedure to determine whether the avoidance behaviour is elicited by consideration of the adverse outcome, or automatically as a response to surrounding stimuli. Adding these levels of analysis to experiments would help pin down the precise behavioural process that a particular manipulation may be targeting. This would provide deeper insight into the function of a

particular brain region, and could in turn serve as the backbone to uncovering the behavioural dysfunction that might be observed in specific models of psychiatric disease.

Acknowledgements

Competing interests

The authors declare no competing interests.

Funding sources

This work was supported by the Australian Research Council grant DP200102445 to L.A.B; the Canada Research Chair program to MDI; Brain and Behaviour Research Foundation Young Investigator grant to MDI; Canadian Institutes of Health Research Project Grant to MDI.

References

Amano, T., Duvarci, S., Popa, D., & Paré, D. (2011). The fear circuit revisited: contributions of the basal amygdala nuclei to conditioned fear. *The Journal of Neuroscience: the Official Journal of the Society for Neuroscience*, *31*(43), 15481–15489. http://doi.org/10.1523/JNEUROSCI.3410-11.2011

Amorapanth, P., LeDoux, J. E., & Nader, K. (2000). Different lateral amygdala outputs mediate reactions and actions elicited by a fear-arousing stimulus. *Nature Neuroscience*, *3*(1), 74–79. http://doi.org/10.1038/71145

Aoki, S. et al. (2018) 'Cholinergic interneurons in the rat striatum modulate substitution of habits', European Journal of Neuroscience. doi: 10.1111/ejn.13820.

Ashwell, R. and Ito, R. (2014) 'Excitotoxic lesions of the infralimbic, but not prelimbic cortex facilitate reversal of appetitive discriminative context conditioning: the role of the infralimbic cortex in context generalization', Frontiers in Behavioral Neuroscience, 8. doi: 10.3389/fnbeh.2014.00063.

Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. Neuropharmacology, 37(4-5), 407-419. doi:10.1016/s0028-3908(98)00033-1

Baum, M. (1970). Extinction of avoidance responding through response prevention (flooding). Psychological Bulletin, 74(4), 276-284. doi:10.1037/h0029789

Bell, T. et al. (2019) 'Regional Striatal Cholinergic Involvement in Human Behavioral Flexibility', The Journal of Neuroscience. doi: 10.1523/jneurosci.2110-18.2019.

Bloodgood, D. W., Sugam, J. A., Holmes, A., & Kash, T. L. (2018). Fear extinction requires infralimbic cortex projections to the basolateral amygdala. *Translational Psychiatry*, *8*(1), 60–11. http://doi.org/10.1038/s41398-018-0106-x

Bolles, R. C. (1970). Species-specific defense reactions and avoidance learning. Psychol Rev, 77(1), 32-48. doi:10.1037/h0028589

Bolles, R. C., & Grossen, N. E. (1969). Effects of an informational stimulus on the acquisition of avoidance behavior in rats. Journal of Comparative and Physiological Psychology, 68(1, Pt.1), 90-99. doi:10.1037/h0027677

Boulougouris, V., Dalley, J. W. and Robbins, T. W. (2007) 'Effects of orbitofrontal, infralimbic and prelimbic cortical lesions on serial spatial reversal learning in the rat', Behavioural Brain Research. doi: 10.1016/j.bbr.2007.02.005.

Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin*, *114*(1), 80–99.

Bouton, M. E. (1994). Conditioning, remembering, and forgetting. *Journal of Experimental Psychology: Animal Behavior Processes*, 20 (3), 219-231.

Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning & Memory*.

Bouton, M. E. and Peck, C. A. (1992) 'Spontaneous recovery in cross-motivational transfer (counterconditioning)', Animal Learning & Behavior. doi: 10.3758/BF03197954.

Bower, G., Starr, R., & Lazarovitz, L. (1965). Amount of response-produced change in the CS and avoidance learning. Journal of Comparative and Physiological Psychology, 59(1), 13-17. doi:10.1037/h0021623

Bradfield, L. A. and Balleine, B. W. (2017) 'Thalamic Control of Dorsomedial Striatum Regulates Internal State to Guide Goal-Directed Action Selection', The Journal of Neuroscience, 37(13), pp. 3721–3733. doi: 10.1523/JNEUROSCI.3860-16.2017.

Bradfield, L. A. and Hart, G. (2020) Rodent medial and lateral orbitofrontal cortices represent unique components of cognitive maps of task space. doi: https://doi.org/10.1016/j.neurobiorev.2019.11.009.

Bradfield, L. A., Hart, G. and Bradfield at laurabradfield, L. (2020) 'Medial and lateral orbitofrontal cortices represent unique components of cognitive maps of task space', Neuroscience and Biobehavioral Reviews. Elsevier, 108(June 2019), pp. 287–294. doi: 10.1016/j.neubiorev.2019.11.009.

Bradfield, L. et al. (2013) 'The Thalamostriatal Pathway and Cholinergic Control of Goal-Directed Action: Interlacing New with Existing Learning in the Striatum', Neuron. Elsevier Inc., 79(1), pp. 153–166. doi: 10.1016/j.neuron.2013.04.039.

Brady, A. M. and Floresco, S. B. (2015) 'Operant Procedures for Assessing Behavioral Flexibility in Rats', Journal of Visualized Experiments, (96), p. 52387. doi: 10.3791/52387.

Bravo-Rivera, C., Roman-Ortiz, C., Brignoni-Perez, E., Sotres-Bayon, F., & Quirk, G. J. (2014). Neural structures mediating expression and extinction of platform-mediated avoidance. J Neurosci, 34(29), 9736-9742. doi:10.1523/JNEUROSCI.0191-14.2014

Bravo-Rivera, C., Roman-Ortiz, C., Montesinos-Cartagena, M., & Quirk, G. J. (2015). Persistent active avoidance correlates with activity in prelimbic cortex and ventral striatum. Front Behav Neurosci, 9, 184. doi:10.3389/fnbeh.2015.00184

Brown, H. D., Baker, P. M. and Ragozzino, M. E. (2010) 'The parafascicular thalamic nucleus concomitantly influences behavioral flexibility and dorsomedial striatal acetylcholine output in rats', Journal of Neuroscience. doi: 10.1523/JNEUROSCI.2167-10.2010.

Burgos-Robles, A., Vidal-Gonzalez, I., Santini, E., & Quirk, G. J. (2007). Consolidation of fear extinction requires NMDA receptor-dependent bursting in the ventromedial prefrontal cortex. *Neuron*, *53*(6), 871–880. http://doi.org/10.1016/j.neuron.2007.02.021

Burke, K. A., Takahashi, Y. K., Correll, J., Brown, P. L., & Schoenbaum, G. (2009). Orbitofrontal inactivation impairs reversal of Pavlovian learning by interfering with "disinhibition" of responding for previously unrewarded cues. *The European Journal*

of Neuroscience, *30*(10), 1941–1946. http://doi.org/10.1111/j.1460-9568.2009.06992.x

Butts, K. A., Floresco, S. B. and Phillips, A. G. (2013) 'Acute stress impairs set-shifting but not reversal learning', Behavioural Brain Research. doi: 10.1016/j.bbr.2013.06.007.

Cain, C. K. (2019). Avoidance Problems Reconsidered. Curr Opin Behav Sci, 26, 9-17. doi:10.1016/j.cobeha.2018.09.002

Cain, C. K., & LeDoux, J. E. (2007). Escape from fear: A detailed behavioral analysis of two atypical responses reinforced by CS termination. Journal of Experimental Psychology: Animal Behavior Processes, 33(4), 451-463. doi:10.1037/0097-7403.33.4.451

Campeau, S., & Davis, M. (1995). Involvement of the central nucleus and basolateral complex of the amygdala in fear conditioning measured with fear-potentiated startle in rats trained concurrently with …. *The Journal of Neuroscience*.

Campese, V. D., Sears, R. M., Moscarello, J. M., Diaz-Mataix, L., Cain, C. K., & LeDoux, J. E. (2016). The Neural Foundations of Reaction and Action in Aversive Motivation. Curr Top Behav Neurosci, 27, 171-195. doi:10.1007/7854_2015_401

Chan, C. K. J., & Harris, J. A. (2017). Extinction of Pavlovian conditioning: The influence of trial number and reinforcement history. *Behavioural Processes*, *141*, 19–25. http://doi.org/10.1016/j.beproc.2017.04.017

Chang, Y.-H., Liu, S.-W., & Chang, C.-H. (2018). Pharmacological activation of the lateral orbitofrontal cortex on regulation of learned fear and extinction. *Neurobiology of Learning and Memory*, *148*, 30–37. http://doi.org/10.1016/j.nlm.2017.12.011

Choi, J. S., & Kim, J. J. (2010). Amygdala regulates risk of predation in rats foraging in a dynamic fear environment. Proc Natl Acad Sci U S A, 107(50), 21773-21777. doi:10.1073/pnas.1010079108

Choi, J. S., Cain, C. K., & LeDoux, J. E. (2010). The role of amygdala nuclei in the expression of auditory signaled two-way active avoidance in rats. Learn Mem, 17(3), 139-147. doi:10.1101/lm.1676610

Chudasama, Y. and Robbins, T. W. (2003) 'Dissociable Contributions of the Orbitofrontal and Infralimbic Cortex to Pavlovian Autoshaping and Discrimination Reversal Learning: Further Evidence for the Functional Heterogeneity of the Rodent Frontal Cortex', The Journal of Neuroscience, 23(25), pp. 8771–8780. doi: 10.1523/JNEUROSCI.23-25-08771.2003.

Clarke, H. F., Robbins, T. W., & Roberts, A. C. (2008). Lesions of the medial striatum in monkeys produce perseverative impairments during reversal learning similar to those produced by lesions of the orbitofrontal cortex. *The Journal of Neuroscience : the Official Journal of the Society for Neuroscience*, *28*(43), 10972–10982. http://doi.org/10.1523/JNEUROSCI.1521-08.2008

Cools, R. et al. (2002) 'Defining the Neural Mechanisms of Probabilistic Reversal Learning Using Event-Related Functional Magnetic Resonance Imaging', Journal of Neuroscience. doi: 10.1523/jneurosci.22-11-04563.2002.

Coutanche, M. N. and Thompson-Schill, S. L. (2012) 'Reversal Without Remapping: What We Can (and Cannot) Conclude About Learned Associations From Training-Induced Behavior Changes', Perspectives on Psychological Science. doi: 10.1177/1745691611434211.

D'Cruz, A. M. et al. (2011) 'Human reversal learning under conditions of certain versus uncertain outcomes', NeuroImage. doi: 10.1016/j.neuroimage.2011.01.068.

Dalton, G. L. et al. (2016) 'Multifaceted Contributions by Different Regions of the Orbitofrontal and Medial Prefrontal Cortex to Probabilistic Reversal Learning', The

Journal of Neuroscience, 36(6), pp. 1996–2006. doi: 10.1523/JNEUROSCI.3366-15.2016.

Dalton, G. L., Phillips, A. G. and Floresco, S. B. (2014) 'Preferential involvement by nucleus accumbens shell in mediating probabilistic learning and reversal shifts', Journal of Neuroscience. doi: 10.1523/JNEUROSCI.5058-13.2014.

De la Casa, L. G. and Lubow, R. E. (2002) 'An empirical analysis of the super-latent inhibition effect', Animal Learning & Behavior, 30(2), pp. 112–120. doi: 10.3758/BF03192913.

Delamater, A. R. (1996). Effects of several extinction treatments upon the integrity of Pavlovian stimulus-outcome associations. *Learning & Behavior*.

Delamater, A. R. (2004). Experimental extinction in Pavlovian conditioning: behavioural and neuroscience perspectives. *The Quarterly Journal of Experimental Psychology. B, Comparative and Physiological Psychology*, *57*(2), 97–132. http://doi.org/10.1080/02724990344000097

Delamater, A. R. (2012). Issues in the extinction of specific stimulus-outcome associations in Pavlovian conditioning. *Behavioural Processes*, *90*(1), 9–19. http://doi.org/10.1016/j.beproc.2012.03.006

Delamater, A. R., & Oakeshott, S. (2007). Learning about multiple attributes of reward in Pavlovian conditioning. *Annals of the New York Academy ….*

Delamater, A. R., & Westbrook, R. F. (2014). Psychological and neural mechanisms of experimental extinction: a selective review. *Neurobiology of Learning and Memory*, *108*, 38–51. http://doi.org/10.1016/j.nlm.2013.09.016

Dickinson, A., Balleine, B., Watt, A., Gonzalez, F., & Boakes, R. A. (1995). Motivational control after extended instrumental training. Animal Learning & Behavior, 23(2), 197-206. doi:10.3758/BF03199935

Diehl, M. M., Bravo-Rivera, C., Rodriguez-Romaguera, J., Pagan-Rivera, P. A., Burgos-Robles, A., Roman-Ortiz, C., & Quirk, G. J. (2018). Active avoidance requires inhibitory signaling in the rodent prelimbic prefrontal cortex. Elife, 7. doi:10.7554/eLife.34657

Do-Monte, F. H., Manzano-Nieves, G., Quiñones-Laracuente, K., Ramos-Medina, L., & Quirk, G. J. (2015). Revisiting the role of infralimbic cortex in fear extinction with optogenetics. *The Journal of Neuroscience : the Official Journal of the Society for Neuroscience*, *35*(8), 3607–3615. http://doi.org/10.1523/JNEUROSCI.3137-14.2015

Duvarci, S., Popa, D., & Paré, D. (2011). Central amygdala activity during fear conditioning. *The Journal of Neuroscience : the Official Journal of the Society for Neuroscience*, *31*(1), 289–294. http://doi.org/10.1523/JNEUROSCI.4985-10.2011

Esber, G. R. and Haselgrove, M. (2011) 'Reconciling the influence of predictiveness and uncertainty on stimulus salience: A model of attention in associative learning', Proceedings of the Royal Society B: Biological Sciences. doi: 10.1098/rspb.2011.0836.

Fanselow, M. S. (1997). Species-specific defense reactions: Retrospect and prospect. In Learning, motivation, and cognition: The functional behaviorism of Robert C. Bolles. (pp. 321-341). Washington, DC, US: American Psychological Association.

Fanselow, M. S., & Lester, L. S. (1988). A functional behavioristic approach to aversively motivated behavior: Predatory imminence as a determinant of the topography of defensive behavior. In Evolution and learning. (pp. 185-212). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Farovik, A., Place, R. J., McKenzie, S., Porter, B., Munro, C. E., & Eichenbaum, H. (2015). Orbitofrontal cortex encodes memories within value-based schemas and represents contexts that guide memory retrieval. *J Neurosci, 35*(21), 8333-8344. doi:10.1523/JNEUROSCI.0134-15.2015

Fernando, A. B., Urcelay, G. P., Mar, A. C., Dickinson, A., & Robbins, T. W. (2014b). Safety signals as instrumental reinforcers during free-operant avoidance. Learn Mem, 21(9), 488-497. doi:10.1101/lm.034603.114

Fernando, A. B., Urcelay, G. P., Mar, A., Dickinson, A., & Robbins, T. W. (2014a). Free-operant avoidance behavior by rats after reinforcer revaluation using opioid agonists and D-amphetamine. J Neurosci, 34(18), 6286-6293. doi:10.1523/JNEUROSCI.4146-13.2014

Fernando, A. B., Urcelay, G. P., Mar, A. C., Dickinson, T. A., & Robbins, T. W. (2014c). The role of the nucleus accumbens shell in the mediation of the reinforcing properties of a safety signal in free-operant avoidance: dopamine-dependent inhibitory effects of d-amphetamine. *Neuropsychopharmacology, 39*(6), 1420-1430. doi:10.1038/npp.2013.337

Floresco, S. B., Block, A. E. and Tse, M. T. L. (2008) 'Inactivation of the medial prefrontal cortex of the rat impairs strategy set-shifting, but not reversal learning, using a novel, automated procedure', Behavioural Brain Research. doi: 10.1016/j.bbr.2008.02.008.

Gillan, C. M., Morein-Zamir, S., Urcelay, G. P., Sule, A., Voon, V., Apergis-Schoute, A. M., . . . Robbins, T. W. (2014). Enhanced avoidance habits in obsessive-compulsive disorder. Biol Psychiatry, 75(8), 631-638. doi:10.1016/j.biopsych.2013.02.002

Gläscher, J. et al. (2010) 'States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning', Neuron. doi: 10.1016/j.neuron.2010.04.016.

Groman, S. M., Keistler, C., Keip, A. J., Hammarlund, E., DiLeone, R. J., Pittenger, C., . . . Taylor, J. R. (2019). Orbitofrontal Circuits Control Multiple Reinforcement-Learning Processes. *Neuron, 103*(4), 734-746 e733. doi:10.1016/j.neuron.2019.05.042

Haney, R. Z., Calu, D. J., Takahashi, Y. K., Hughes, B. W., & Schoenbaum, G. (2010). Inactivation of the central but not the basolateral nucleus of the amygdala disrupts learning in response to overexpectation of reward. *The Journal of Neuroscience : the Official Journal of the Society for Neuroscience*, *30*(8), 2911–2917. http://doi.org/10.1523/JNEUROSCI.0054-10.2010

Harris, C., Aguirre, C. G., Kolli, S., Das, K., Izquierdo, A., & Soltani, A. (2020). Unique features of stimulus-based probabilistic reversal learning. *bioRxiv*, 2020.2009.2024.310771. doi:10.1101/2020.09.24.310771

Harris, J. A., & Andrew, B. J. (2017). Time, trials, and extinction. *Journal of Experimental Psychology. Animal Learning and Cognition*, *43*(1), 15–29. http://doi.org/10.1037/xan0000125

Harris, J. A., Jones, M. L., Bailey, G. K., & Westbrook, R. F. (2000). Contextual control over conditioned responding in an extinction paradigm. *Journal of Experimental Psychology. Animal Behavior Processes*, *26*(2), 174–185.

Hendersen, R. W., & Graham, J. (1979). Avoidance of heat by rats: Effects of thermal context on rapidity of extinction. Learning and Motivation, 10(3), 351-363. doi:https://doi.org/10.1016/0023-9690(79)90038-9

Herry, C., Ciocchi, S., Senn, V., Demmou, L., Müller, C., & Lüthi, A. (2008). Switching on and off fear by distinct neuronal circuits. *Nature*, *454*(7204), 600–606. http://doi.org/10.1038/nature07166

Hervig, M. E. et al. (2019) 'Dissociable and Paradoxical Roles of Rat Medial and Lateral Orbitofrontal Cortex in Visual Serial Reversal Learning', Cerebral Cortex, (July), pp. 1–14. doi: 10.1093/cercor/bhz144.

Holland, P. C., & Gallagher, M. (1993). Amygdala central nucleus lesions disrupt increments, but not decrements, in conditioned stimulus processing. *Behavioral Neuroscience*, *107*(2), 246–253. http://doi.org/10.1037//0735-7044.107.2.246

Holland, P. C., & Maddux, J-M (2010) Brain Systems of attention in associative learning. In Attention and Associative Learning: From Brain to Behavior. Eds C. J. Michell & M. E. Le Pelley, Oxford University Press.

Iordanova, M. D., Deroche, M. L. D., Esber, G. R., & Schoenbaum, G. (2016). Neural correlates of two different types of extinction learning in the amygdala central nucleus. *Nature Communications*, *7*, 12330. http://doi.org/10.1038/ncomms12330

Izquierdo, A., Suda, R. K. and Murray, E. A. (2004) 'Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency', Journal of Neuroscience. doi: 10.1523/JNEUROSCI.1921-04.2004.

Kalish, H. I. (1954). Strength of fear as a function of the number of acquisition and extinction trials. *Journal of Experimental Psychology: …*, *47*(1), 1–9. http://doi.org/10.1037/h0053732

Kim, M., & Davis, M. (1993). Electrolytic lesions of the amygdala block acquisition and expression of fear-potentiated startle even with extensive training but do not prevent reacquisition. *Behavioral Neuroscience*, *107*(4), 580–595. http://doi.org/10.1037//0735-7044.107.4.580

Klanker, M. et al. (2013) 'Deep brain stimulation in the lateral orbitofrontal cortex impairs spatial reversal learning', Behavioural Brain Research. doi: 10.1016/j.bbr.2013.01.043.

Koya, E., Margetts-Smith, G., & Hope, B. T. (2016). Daun02 Inactivation of Behaviorally Activated Fos-Expressing Neuronal Ensembles. *Current Protocols in Neuroscience / Editorial Board, Jacqueline N. Crawley ... [Et Al.]*, *76*, 8.36.1–8.36.17. http://doi.org/10.1002/cpns.2

Laughlin, L., Moloney, D., Samels, S., Sears, R., & Cain, C. K. (2020). Reducing shock imminence eliminates poor avoidance in rats. bioRxiv, 2020.2001.2030.927152. doi:10.1101/2020.01.30.927152

Laurent, V., & Westbrook, R. F. (2009). Inactivation of the infralimbic but not the prelimbic cortex impairs consolidation and retrieval of fear extinction. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *16*(9), 520–529. http://doi.org/10.1101/lm.1474609

Laurent, V., & Westbrook, R. F. (2010). Role of the basolateral amygdala in the reinstatement and extinction of fear responses to a previously extinguished conditioned stimulus. *Learning & Memory (Cold Spring Harbor, N.Y.)*, *17*(2), 86–96. http://doi.org/10.1101/lm.1655010

Lay, B. P. P., Nicolosi, M., Usypchuk, A. A., Esber, G. R., & Iordanova, M. D. (2019). Corrigendum: Dissociation of Appetitive Overexpectation and Extinction in the Infralimbic Cortex. *Cerebral Cortex (New York, N.Y. : 1991)*, *29*(4), 1703–1703. http://doi.org/10.1093/cercor/bhy338

Lay, B. P. P., Pitaru, A., Boulianne, N., Esber, G. R., & Iordanova, M. D. (2020a). Different forms of fear extinction are supported by distinct cortical substrates. *bioRxiv*, 2020.01.27.921221. http://doi.org/10.1101/2020.01.27.921221

Lay, B. P. P., Koya, E., Esber, G. R., & Iordanova, M. D. (2020b). Distinct neuronal ensembles in the central nucleus of the amygdala regulate extinction learning *bioRxiv*.

Lazaro-Munoz, G., LeDoux, J. E., & Cain, C. K. (2010). Sidman instrumental avoidance initially depends on lateral and basal amygdala and is constrained by central amygdala-mediated Pavlovian processes. Biol Psychiatry, 67(12), 1120-1127. doi:10.1016/j.biopsych.2009.12.002

LeDoux, J. E., Moscarello, J., Sears, R., & Campese, V. (2017). The birth, death and resurrection of avoidance: a reconceptualization of a troubled paradigm. Mol Psychiatry, 22(1), 24-36. doi:10.1038/mp.2016.166

Lindgren, J. L., Gallagher, M., & Holland, P. C. (2003). Lesions of basolateral amygdala impair extinction of CS motivational value, but not of explicit conditioned responses,

in Pavlovian appetitive second-order conditioning. *The European Journal of Neuroscience*, *17*(1), 160–166. http://doi.org/10.1046/j.1460-9568.2003.02421.x

Lingawi, N. W., Westbrook, R. F., & Laurent, V. (2017). Extinction of relapsed fear does not require the basolateral amygdala. *Neurobiology of Learning and Memory*, *139*, 149–156. http://doi.org/10.1016/j.nlm.2017.01.005

Livneh, U., & Paz, R. (2012). Amygdala-prefrontal synchronization underlies resistance to extinction of aversive memories. *Neuron*, *75*(1), 133–142. http://doi.org/10.1016/j.neuron.2012.05.016

López, J. C. et al. (2003) 'Spatial reversal learning deficit after medial cortex lesion in turtles', Neuroscience Letters. doi: 10.1016/S0304-3940(03)00186-1.

Lubow, R. E. and Moore, A. U. (1959) 'Latent inhibition: The effect of nonreinforced pre-exposure to the conditional stimulus', Journal of Comparative and Physiological Psychology. doi: 10.1037/h0046700.

Lucantonio, F., Gardner, M. P. H., Mirenzi, A., Newman, L. E., Takahashi, Y. K., & Schoenbaum, G. (2015). Neural Estimates of Imagined Outcomes in Basolateral Amygdala Depend on Orbitofrontal Cortex. *The Journal of Neuroscience : the Official Journal of the Society for Neuroscience*, *35*(50), 16521–16530. http://doi.org/10.1523/JNEUROSCI.3126-15.2015

Malvaez, M., Shieh, C., Murphy, M. D., Greenfield, V. Y., & Wassum, K. M. (2019). Distinct cortical-amygdala projections drive reward value encoding and retrieval. Nat Neurosci, 22(5), 762-769. doi:10.1038/s41593-019-0374-7

McDonald, R. J., Ko, C. H. and Hong, N. S. (2002) 'Attenuation of context-specific inhibition on reversal learning of a stimulus–response task in rats with neurotoxic hippocampal damage', Behavioural Brain Research, 136(1), pp. 113–126. doi: 10.1016/S0166-4328(02)00104-3.

Milad, M. R., & Quirk, G. J. (2002). Infralimbic cortex neurons signal memory for extinction of conditioned fear. Nature.

Miller, N. E. (1948). Studies of fear as an acquirable drive: I. Fear as motivation and fear-reduction as reinforcement in the learning of new responses. Journal of Experimental Psychology, 38(1), 89-101. doi:10.1037/h0058455

Mobbs, D., Headley, D. B., Ding, W., & Dayan, P. (2020). Space, Time, and Fear: Survival Computations along Defensive Circuits. Trends Cogn Sci. doi:https://doi.org/10.1016/j.tics.2019.12.016

Moscarello, J. M., & LeDoux, J. E. (2013). Active avoidance learning requires prefrontal suppression of amygdala-mediated defensive reactions. J Neurosci, 33(9), 3815-3823. doi:10.1523/JNEUROSCI.2596-12.2013

Mowrer, O. H., & Lamoreaux, R. R. (1946). Fear as an intervening variable in avoidance conditioning. Journal of Comparative Psychology, 39(1), 29-50. doi:10.1037/h0060150

O'Doherty, J. et al. (2001) 'Abstract reward and punishment representations in the human orbitofrontal cortex', Nature Neuroscience, 4(1), pp. 95–102. doi: 10.1038/82959.

Okada, K. et al. (2014) 'Enhanced flexibility of place discrimination learning by targeting striatal cholinergic interneurons', Nature Communications, 5(1), p. 3778. doi: 10.1038/ncomms4778.

Oleson, E. B., Gentry, R. N., Chioma, V. C., & Cheer, J. F. (2012). Subsecond dopamine release in the nucleus accumbens predicts conditioned punishment and its successful avoidance. *J Neurosci, 32*(42), 14804-14808. doi:10.1523/JNEUROSCI.3087-12.2012

Oualian, C. and Gisquet-Verrier, P. (2010) 'The differential involvement of the prelimbic and infralimbic cortices in response conflict affects behavioral flexibility in rats trained in a

new automated strategy-switching task', Learning and Memory. doi: 10.1101/lm.1858010.

Overmier, J. B., & Bull, J. A. (1969). On the independence of stimulus control of avoidance. Journal of Experimental Psychology, 79(3, Pt.1), 464-467. doi:10.1037/h0026985

Parkes, S. L. et al. (2018) 'Insular and Ventrolateral Orbitofrontal Cortices Differentially Contribute to Goal-Directed Behavior in Rodents', Cerebral Cortex, 28(7), pp. 2313–2325. doi: 10.1093/cercor/bhx132.

Parnaudeau, S. et al. (2015) 'Mediodorsal thalamus hypofunction impairs flexible goal-directed behavior', Biological Psychiatry. doi: 10.1016/j.biopsych.2014.03.020.

Patterson, T. K., Craske, M. G., & Knowlton, B. J. (2019). Enhanced Avoidance Habits in Relation to History of Early-Life Stress. Front Psychol, 10, 1876. doi:10.3389/fpsyg.2019.01876

Pavlov, I. P. (1927) 'Conditioned reflexes. An investigation of the physiological activity of the cerebral cortex. Translated and edited by E. V. ANREP', Conditioned reflexes. An investigation of the physiological activity of the cerebral cortex. Translated and edited by E. V. ANREP.

Pearce, J. M. and Hall, G. (1980) 'A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli', Psychological Review. doi: 10.1037/0033-295X.87.6.532.

Poremba, A., & Gabriel, M. (1999). Amygdala neurons mediate acquisition but not maintenance of instrumental avoidance behavior in rabbits. J Neurosci, 19(21), 9635-9641.

Ragozzino, M. E. et al. (2009) 'Acetylcholine activity in selective striatal regions supports behavioral flexibility', Neurobiology of Learning and Memory. doi: 10.1016/j.nlm.2008.09.008.

Rayburn-Reeves, R. M., Molet, M. and Zentall, T. R. (2011) 'Simultaneous discrimination reversal learning in pigeons and humans: anticipatory and perseverative errors', Learning & Behavior, 39(2), pp. 125–137. doi: 10.3758/s13420-010-0011-5.

Rescorla, R. A. (1968). Pavlovian conditioned fear in Sidman avoidance learning. J Comp Physiol Psychol, 65(1), 55-60. doi:10.1037/h0025412

Rescorla, R. A. (1996). Preservation of Pavlovian associations through extinction. *The Quarterly Journal of Experimental Psychology: ….*

Rescorla, R. A. (2001). Experimental extinction.

Rescorla, R. A. (2002) 'Comparison of the rates of associative change during acquisition and extinction', Journal of Experimental Psychology: Animal Behavior Processes. doi: 10.1037/0097-7403.28.4.406.

Rescorla, R. A. (2003). Protection from extinction. Learn Behav, 31(2), 124-132. doi:10.3758/bf03195975

Rescorla, R. A. (2005) 'Spontaneous recovery of excitation but not inhibition', Journal of Experimental Psychology: Animal Behavior Processes. doi: 10.1037/0097-7403.31.3.277.

Rescorla, R. A. (2006). Spontaneous recovery from overexpectation. *Learning & Behavior*, *34*(1), 13–20.

Rescorla, R. A. (2007) 'Spontaneous recovery after reversal and partial reinforcement', Learning and Behavior. doi: 10.3758/BF03206425.

Rescorla, R. A. (2007). Renewal after overexpectation. *Learning & Behavior*, *35*(1), 19–26.

Rescorla, R. A., & Solomon, R. L. (1967). Two-process learning theory: Relationships between Pavlovian conditioning and instrumental learning. Psychol Rev, 74(3), 151-182. doi:10.1037/h0024475

Rodriguez-Romaguera, J., Greenberg, B. D., Rasmussen, S. A., & Quirk, G. J. (2016). An Avoidance-Based Rodent Model of Exposure With Response Prevention Therapy for Obsessive-Compulsive Disorder. Biol Psychiatry, 80(7), 534-540. doi:10.1016/j.biopsych.2016.02.012

Roesch, M. R., Calu, D. J., Esber, G. R., & Schoenbaum, G. (2010). All that glitters ... dissociating attention and outcome expectancy from prediction errors signals. *Journal of Neurophysiology*, *104*(2), 587–595. http://doi.org/10.1152/jn.00173.2010

Rosas, J. M., García-Gutiérrez, A., Psicológica, J. C.-A.(2007). AAB and ABA renewal as a function of the number of extinction trials in conditioned taste aversion. *Psicológica, 28, 129-150.*

Rudebeck, P. H. and Murray, E. A. (2011) 'Balkanizing the primate orbitofrontal cortex: Distinct subregions for comparing and contrasting values', Annals of the New York Academy of Sciences. doi: 10.1111/j.1749-6632.2011.06267.x.

Rygula, R. et al. (2010) 'Differential contributions of the primate ventrolateral prefrontal and orbitofrontal cortex to serial reversal learning', Journal of Neuroscience. doi: 10.1523/JNEUROSCI.2631-10.2010.

Salzman, C. D. et al. (2007) 'Flexible neural representations of value in the primate brain', in Annals of the New York Academy of Sciences. doi: 10.1196/annals.1401.034.

Schoenbaum, G. (2003) 'Lesions of Orbitofrontal Cortex and Basolateral Amygdala Complex Disrupt Acquisition of Odor-Guided Discriminations and Reversals', Learning & Memory, 10(2), pp. 129–140. doi: 10.1101/lm.55203.

Schoenbaum, G., Chiba, A. A. and Gallagher, M. (1999) 'Neural Encoding in Orbitofrontal Cortex and Basolateral Amygdala during Olfactory Discrimination Learning', The Journal of Neuroscience, 19(5), pp. 1876–1884. doi: 10.1523/JNEUROSCI.19-05-01876.1999.

Schoenbaum, G., Nugent, S. L., et al. (2002) 'Orbitofrontal lesions in rats impair reversal but not acquisition of go, no-go odor discriminations', Neuroreport, 13(6), pp. 885–890. doi: 10.1097/00001756-200205070-00030.

Schoenbaum, G., Nugent, S., et al. (2002) 'Teaching old rats new tricks: age-related impairments in olfactory reversal learning', Neurobiology of Aging, 23(4), pp. 555–564. doi: 10.1016/S0197-4580(01)00343-8.

Sengupta, A., Winters, B., Bagley, E. E., & McNally, G. P. (2016). Disrupted Prediction Error Links Excessive Amygdala Activation to Excessive Fear. *The Journal of Neuroscience : the Official Journal of the Society for Neuroscience*, *36*(2), 385–395. http://doi.org/10.1523/JNEUROSCI.3670-15.2016

Sierra-Mercado, D., Padilla-Coreano, N., & Quirk, G. J. (2011). Dissociable roles of prelimbic and infralimbic cortices, ventral hippocampus, and basolateral amygdala in the expression and extinction of conditioned fear. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology*, *36*(2), 529–538. http://doi.org/10.1038/npp.2010.184

Sissons, H. T. and Miller, R. R. (2009) 'Spontaneous recovery of excitation and inhibition.', Journal of Experimental Psychology: Animal Behavior Processes, 35(3), pp. 419–426. doi: 10.1037/a0014815.

Solomon, R. L., Kamin, L. J., & Wynne, L. C. (1953). Traumatic avoidance learning: the outcomes of several extinction procedures with dogs. The Journal of Abnormal and Social Psychology, 48(2), 291-302. doi:10.1037/h0058943

Stalnaker, T. A. et al. (2016) 'Cholinergic interneurons use orbitofrontal input to track beliefs about current state', Journal of Neuroscience. doi: 10.1523/JNEUROSCI.0157-16.2016.

Takahashi, Y. K., Chang, C. Y., Lucantonio, F., Haney, R. Z., Berg, B. A., Yau, H.-J., et al. (2013). Neural estimates of imagined outcomes in the orbitofrontal cortex drive behavior and learning. *Neuron*, *80*(2), 507–518. http://doi.org/10.1016/j.neuron.2013.08.008

Takahashi, Y. K., Roesch, M. R., Stalnaker, T. A., Haney, R. Z., Calu, D. J., Taylor, A. R., et al. (2009). The orbitofrontal cortex and ventral tegmental area are necessary for learning from unexpected outcomes. *Neuron*, *62*(2), 269–280. http://doi.org/10.1016/j.neuron.2009.03.005

Thompson, R., Kao, L. and Yang, S. (1981) 'Rapid forgetting of individual spatial reversal problems in rats with parafascicular lesions', Behavioral and Neural Biology. doi: 10.1016/S0163-1047(81)92189-0.

Thorpe, S. J., Rolls, E. T. and Maddison, S. (1983) 'The orbitofrontal cortex: Neuronal activity in the behaving monkey', Experimental Brain Research, 49(1). doi: 10.1007/BF00235545.

Toyomitsu, Y., Nishijo, H., Uwano, T., Kuratsu, J., & Ono, T. (2002). Neuronal responses of the rat amygdala during extinction and reassociation learning in elementary and configural associative tasks. *The European Journal of Neuroscience*, *15*(4), 753–768.

Tye, K. M., Cone, J. J., Schairer, W. W., & Janak, P. H. (2010). Amygdala neural encoding of the absence of reward during extinction. *The Journal of Neuroscience : the Official Journal of the Society for Neuroscience*, *30*(1), 116–125. http://doi.org/10.1523/JNEUROSCI.4240-09.2010

Villaruel, F. R., Lacroix, F., Sanio, C., Sparks, D. W., Chapman, C. A., & Chaudhri, N. (2018). Optogenetic Activation of the Infralimbic Cortex Suppresses the Return of Appetitive Pavlovian-Conditioned Responding Following Extinction. *Cereb Cortex, 28*(12), 4210-4221. doi:10.1093/cercor/bhx275.

Wagner, A. R. (1961). Effects of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *Journal of Experimental Psychology: …*, *62*(3), 234–242. http://doi.org/10.1037/h0042251

Wenzel, J. M., Oleson, E. B., Gove, W. N., Cole, A. B., Gyawali, U., Dantrassy, H. M., . . . Cheer, J. F. (2018). Phasic Dopamine Signals in the Nucleus Accumbens that Cause Active Avoidance Require Endocannabinoid Mobilization in the Midbrain. Curr Biol, 28(9), 1392-1404 e1395. doi:10.1016/j.cub.2018.03.037

Wilson, R. C. et al. (2014) 'Orbitofrontal Cortex as a Cognitive Map of Task Space', Neuron, 81(2), pp. 267–279. doi: 10.1016/j.neuron.2013.11.005.

Yin, H. H., Knowlton, B. J., & Balleine, B. W. (2006). Inactivation of dorsolateral striatum enhances sensitivity to changes in the action-outcome contingency in instrumental conditioning. Behav Brain Res, 166(2), 189-196. doi:10.1016/j.bbr.2005.07.012

Zimmermann, K. S., Li, C.-C., Rainnie, D. G., Ressler, K. J., & Gourley, S. L. (2018). Memory Retention Involves the Ventrolateral Orbitofrontal Cortex: Comparison with the Basolateral Amygdala. *Neuropsychopharmacology : Official Publication of the American College of Neuropsychopharmacology*, *43*(3), 674–674. http://doi.org/10.1038/npp.2017.219

Zimmermann, K. S., Yamin, J. A., Rainnie, D. G., Ressler, K. J., & Gourley, S. L. (2017). Connections of the Mouse Orbitofrontal Cortex and Regulation of Goal-Directed Action Selection by Brain-Derived Neurotrophic Factor. Biol Psychiatry, 81(4), 366-377. doi:10.1016/j.biopsych.2015.10.026

Tables

Task/Paradigm	Stage 1	Stage 2	Test
Extinction	S1-O1	S1-Ø	S1
Overexpectation	S1-O1, S2-O1	S1+S2-O1	S1
Reward/No reward reversal	S1-O1, S2-Ø	S1-Ø, S2-O1	
Latent inhibition	S2-Ø	S2-O1	S2
Probabilistic reversal	S1-(O1 x 0.7) S2-(O1 x 0.3)	S1-(O1 x 0.3) S2-(O1 x 0.7)	

Table 1: Common associative learning phenomena and their designs

Learning phase	Association	Potential conflicting behaviours
Pavlovian conditioning	S- $O_{aversive}$	
Avoidance acquisition	S(R- $Ø_{aversive}$)	Stimulus-dependent reactions e.g. flight (S- $Ø_{aversive}$)
Avoidance expression (late)	S-R	Maintenance of outcome strength through failed trials [S(noR-$O_{aversive}$)]
Avoidance extinction	S(no R- $Ø_{aversive}$)	R blocking extinction of S through conditioned inhibition

Table 2: conflict during different phases of avoidance learning