American Psychological Association

Psychological Review

© 2018 American Psychological Association
0033-295X/18/$12.00

2018, Vol. 125, No. 5, 822–843
http://dx.doi.org/10.1037/rev0000118

Modulation of Attention and Action in the Medial Prefrontal Cortex of Rats

Melissa J. Sharpe
University of New South Wales; Princeton University; and
National Institute on Drug Abuse, Baltimore, Maryland

Simon Killcross
University of New South Wales

Theories of functioning in the medial prefrontal cortex are distinct across appetitively and aversively motivated procedures. In the appetitive domain, it is argued that the medial prefrontal cortex is important for producing adaptive behavior when circumstances change. This view advocates a role for this region in using higher-order information to bias performance appropriate to that circumstance. Conversely, literature born out of aversive studies has led to the theory that the prelimbic region of the medial prefrontal cortex is necessary for the expression of conditioned fear, whereas the infralimbic region is necessary for a decrease in responding following extinction. Here, the argument is that these regions are primed to increase or decrease fear responses and that this tendency is gated by subcortical inputs. However, we believe the data from aversive studies can be explained by a supraordinate role for the medial prefrontal cortex in behavioral flexibility, in line with the appetitive literature. Using a dichotomy between the voluntary control of behavior and the execution of well-trained responses, we attempt to reconcile these theories. We argue that the prelimbic region exerts voluntary control over behavior via top-down modulation of stimulus–response pathways according to task demands, contextual cues, and how well a stimulus predicts an outcome. Conversely, the infralimbic region promotes responding based on the strength of stimulus–response pathways determined by experience with reinforced contingencies. This system resolves the tension between executing voluntary actions sensitive to recent changes in contingencies, and responses that reflect the animal's experience across the long run.

Keywords: prelimbic cortex, infralimbic cortex, fear conditioning, appetitive conditioning, context

It is too often the case that different domains within psychological research do not integrate theories with one another. An illustrative example of this is lack of cross talk between research on learning in laboratory rodents based in the appetitive domain and that based on aversive procedures. This is particularly true for theories of functioning in prefrontal cortex (although work examining the function of other regions, such as the amygdala, frequently suffers in the same way; Balleine & Killcross, 2006). On the one hand, there is a long history of research in the appetitive literature looking at the role of the prefrontal cortex in tasks thought to require executive function. These studies have predominantly looked at the impact of damage or functional inactivation of the prefrontal cortex in appetitive learning tasks requiring the maintenance of information in working memory, the ability to shift responding to stimulus or response dimensions, the efficient allocation of attention toward stimuli, goal-sensitive behaviors, and the use of contextual cues to modulate behavior. This research has suggested that the prefrontal cortex is involved in allowing behavior to flexibly change when there is a shift in experimental circumstances. On the other hand, there is another literature that examines the role of the prefrontal cortex in aversive conditioning. These studies have resulted in a number of contradictory findings, where damage or inactivation of prefrontal cortex has been found to result in enhanced fear to a stimulus paired with an aversive outcome, greater levels of contextual conditioning, lower levels of fear responding, or have no effect. These two different fields are almost entirely independent of one another, and cross-references between them are uncommon.

The purpose of this review is to integrate theories of functioning in the medial regions of the prefrontal cortex across appetitive and aversive domains. Specifically, we aim to use the knowledge born out of appetitive literature to resolve the contradictions that currently reside in the aversive literature. In doing so, we will focus on two regions of the rodent medial prefrontal cortex (mPFC), the prelimbic (PL) and infralimbic (IL) cortices, often the focus of prefrontal research in appetitive and aversive studies. In the first

---

Melissa J. Sharpe, School of Psychology, University of New South Wales; Princeton Neuroscience Institute, Princeton University; and National Institute on Drug Abuse, Baltimore, Maryland. Simon Killcross, School of Psychology, University of New South Wales.

All authors contributed significantly to the development of this article and have approved the final version of this article. This work was supported by a National Health and Medical Research Council (Australia) CJ Martin awarded to Melissa J. Sharpe. The opinions expressed in this article are the authors' own and do not reflect the view of the National Institutes of Health/Department of Health and Human Services.

Correspondence concerning this article should be addressed to Melissa J. Sharpe, National Institute on Drug Abuse, 251 Bayview Boulevard, Baltimore, MD 21224. E-mail: mjsharpe@princeton.edu

section of this article, we will review the role of these regions in the resolution of response conflict and the contextual control of behavior. Second, we will describe new data from our laboratory suggesting a role of the PL and IL regions in the modulation of attention toward predictive cues where we interpret mPFC activity as exerting top-down modulation of an attentional response much as it would an explicit behavioral response. Then we will discuss studies from the aversive literature which have examined the effects of pretraining lesions and functional inactivation of mPFC on fear learning and expression, as well as the literature which implicates the PL and IL cortices in the contextual dependence of fear.

In laying out this research we hope to propose a novel model of functioning in the mPFC that generalizes across both appetitive and aversive paradigms. That is, we believe that the role of the PL and IL cortices in using different types of information across many divergent tasks can be unified by their involvement in a single cognitive framework. Specifically, we suggest that it is the PL cortex that exerts voluntary control over behavior by top-down modulation of stimulus–response pathways according to the current task demands, present contextual cues, and how well a stimulus currently predicts an outcome. This allows behavior to adapt rapidly to recent changes in contingencies by reference to the current information that predicts those changes (be that in working memory or physically observable in the environment). In contrast, the IL cortex appears to oppose this tendency in favor of acting on the basis of dominant stimulus–response pathways, directly proportional to the amount of experience with the stimulus–response association, allowing long-run experience to counteract any recent (and potentially unreliable) changes in experienced contingencies. In this manner, the type of information that is being recruited by these systems is not what unifies their role in learning and behavior. Rather, what unifies the role of the PL and IL cortices is how this information is used. We will illustrate our model by adapting a well-established framework (Cohen, Dunbar, & McClelland, 1990) from the human cognitive neuroscience literature that continues to reflect the dichotomy between the voluntary control of behavior and the execution of well-trained responses. In making this connection, we will attempt not only to illustrate the basic concepts and processes underlying our model but also to expand upon our existing understanding on how prefrontal cortex contributes to flexible behavior across species.

The Framework: Cohen et al.'s (1990) Connectionist Account of the Stroop Task

It has long been established that the possession of a prefrontal cortex (PFC) endows humans and experimental animals with behavior of greater flexibility and complexity (Uylings, Groenewegen, & Kolb, 2003). In line with this notion, there have been many theories which have attempted to encapsulate the idea of the PFC acting in a supervisory manner over other basic processes (Goldman-Rakic, O'Scalaidhe, & Chafee, 2000; Kesner, 2000). One early framework of this nature was developed by Norman and Shallice (1980) where frontal regions are argued to exert voluntary control over automatic lower-order processing. Although the specifics of the framework are not pertinent here, the main premise is that selection of well-trained responses in routine situations is automatic and decentralized, whereas selection of an appropriate response in times of novelty or conflict requires the allocation of control over lower-order processing to produce a response that is appropriate to those circumstances. Given that the classical view of frontal functioning involves the coordination of complex behavior, decision making, and responding to novelty, Shallice (1982) argued that the role of the frontal lobes was to exert this sort of voluntary control over behavior that facilitates the selection of appropriate action in nonroutine circumstance. Many theories have since expanded on this basic framework to explain the results of many tasks thought to reflect the ability to exert executive control over automatic processes, such as the Stroop task, the Tower of London from the Cambridge Neuropsychological Test Automated Battery, and the Wisconsin card sort task (WCST; Braver & Cohen, 2000; Cohen, Aston-Jones, & Gilzenrat, 2004; Cohen et al., 1990; Miller & Cohen, 2001).

The Stroop effect was one of the first such empirical demonstrations of a distinction between automatic and controlled processes (Stroop, 1935; Cohen et al., 1990). In this task, participants are shown words written in different colored inks. The written words and colors can be either congruous or incongruous with one another. When participants are shown incongruous pairs (e.g., GREEN in red ink), errors and reaction times are significantly increased when they are required to name the color of the ink (i.e., say "red") and disregard the word's written form (i.e., green). This effect is typically attributed to the different levels of experience of word reading over color naming (MacLeod & Dunbar, 1988). Hence, for example, processing of the word is thought to arrive at the response stage faster than the color name. When the task requires that the subject read the word, there is no conflict as word reading is accomplished more quickly and does not have to compete with naming of the color. When subjects are required to name the color, the influence of having processed the written word has to be overcome before the color can be accurately named. In this way, word reading is thought to be a relatively automatic process whereas the ability to overcome this tendency and name the color of the ink is thought to reflect voluntary control over response selection.

To account for this effect, Cohen et al. (1990) developed a parallel distributed processing (PDP) model. Here, experience with particular tasks develop networks of stimulus–response mappings, viewed as connections between different nodes, where the connection strength between these nodes is directly proportional to the experience with the task. Critically, this connection occurs through an intermediary node that allows task demands to influence activation of these mappings (see Figure 1). In terms of the Stroop task, the greater experience with word reading facilitates a stronger mapping between the sight of the written word (i.e., the stimulus) and the response to read the written word (i.e., the response). This means that the sight of the written word will robustly activate the response to read the written word. In contrast, the relative inexperience with color naming is thought to be reflected in the weak mapping between the sight of the written word and the response to name the color of the ink in which the word is written. Thus, when the task requires that people disregard the written word and name the color of the ink, the stronger activation of the response to read the written word conflicts with the relative weak activation of the response to name the color of the ink. It is argued that this conflict is overcome through an active rehearsal of task demands (i.e., name the color of the ink) which biases activation of the

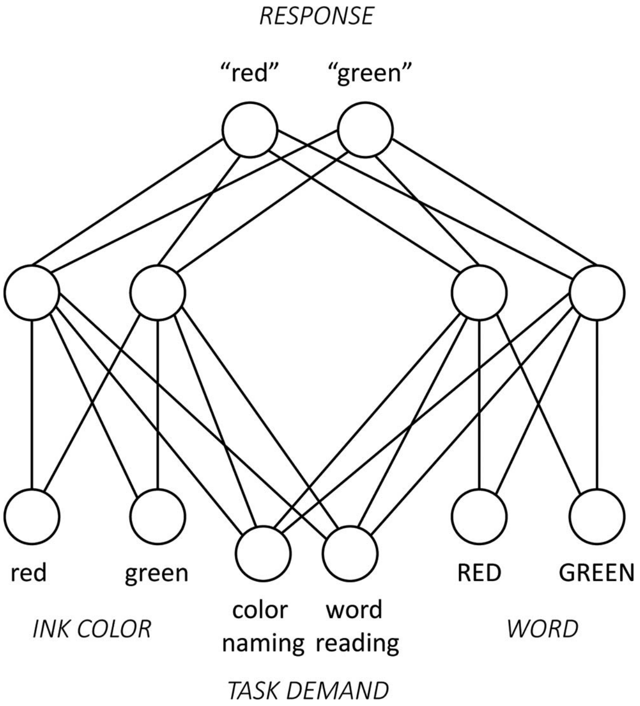

Figure 1. Cohen et al.'s (1990) connectionist model of the Stroop task. Units at the bottom are input (stimulus) units and units at the top are output (response) units. Experience with particular tasks promotes a network of representational nodes that links stimuli input units with response output units. The strength of the connection between nodes is directly proportional to experience with each stimulus–response combination. A response occurs when sufficient activation has accumulated that one of the output units exceeds a response threshold. Activity in these pathways can be influenced by prefrontal input, which allows an active representation of task demands to boost activity of the corresponding stimulus–response pathway via hidden units lying between stimulus and response units.

response to name the color of the ink, by impact at the level of the intermediary nodes. It is through this intermediary node that the PFC is argued to exert voluntary control over stimulus–response mappings (Braver & Cohen, 2000; Miller & Cohen, 2001). That is, PFC is argued to actively maintain a neural representation of task demands, which allows the specific demands of the task to name the color to influence response selection and overcome the prepotent tendency to read the written word. Later developments of this model have proposed that the PFC also contributes to the ability to maintain the patterns of activity that represent current task demands to bias activity of the relevant pathway to overcome response conflict across a range of different scenarios (Braver & Cohen, 2000; Cohen et al., 2004; Miller & Cohen, 2001).

Within this framework, task demands are conceptualized as higher-order information in the environment, which can be used to resolve conflict between multiple responses. That is, a task demand is any available information in working memory or the physical environment that can modulate stimulus–response mappings to influence production of a response that is appropriate to the current circumstance. On the other hand, a stimulus is information within the environment that frequently may enter into a direct association with a response. Presence of this stimulus will directly activate its associated response. In the Stroop example described above (see Figure 1), the task demand is the instruction given to the participants that tells them to either read the word or name the color and will influence the ability of the stimulus to activate a representation of its associated response. In the case of rats, the task demand may be contextual cues that signal the appropriate stimulus–response association in effect, whereas the stimulus may be of an auditory or visual nature that enters into a direct association with the response. However, it is also important to note here that although contextual cues can modulate activation of particular stimulus–response associations, such information can also function as a stimulus that enters into a direct association with the response. And the converse is also true, an auditory or visual stimulus can also function to act as a higher-order cue to modulate available stimulus–response associations. In summary, although different types of information may be processed differently in the brain, the actual content of any information available in the environment or the experimental circumstance does not determine whether it will function as a task demand or as a stimulus. Rather, the experimental design will dictate whether it functions as a task demand to modulate stimulus–response associations or as a stimulus to directly activate an associated response. For consistency, throughout this article we will refer to information functioning as a task demand as a "cue," whereas we will refer to information entering into a direct association with the response as a "stimulus".

The Role of the Medial Prefrontal Cortex in Appetitive Procedures

The Resolution of Response Conflict

In an attempt to assess further the neural processes inherent in response conflict such as that seen in the Stroop task, Haddon, George, and Killcross (2008) designed a novel paradigm to evoke similar response conflict in rats (see Figure 2). In this task, rats are trained on two conditional discriminations, one auditory and one visual, in two distinct contexts across multiple sessions. In one context, discrete auditory cues (i.e., tone or clicks) predict whether a left or right lever response will be rewarded with one of two outcomes (i.e., sucrose solution or grain pellets). In the alternate context, discrete visual cues (i.e., steady or flashing light) dictate the correct lever response. During test sessions in both contexts, rats are presented with novel audio-visual cue compounds that dictated either the same or different responses during training (termed congruent and incongruent compounds, respectively). When rats are presented with congruent compounds, no conflict is present as both the auditory and the visual stimuli dictated the same correct lever press in training. However, when rats are presented with incongruent compounds, response conflict is evoked as the cues that form the compound dictated different correct lever responses during training. Haddon et al. (2008) found that rats are capable of resolving this conflict by using the task-setting contextual cues to disambiguate the response conflict and perform the correct lever-press response appropriate to the context in which they are being tested. That is, rats press more on the lever appropriate to the auditory cue components in the auditory training context, and

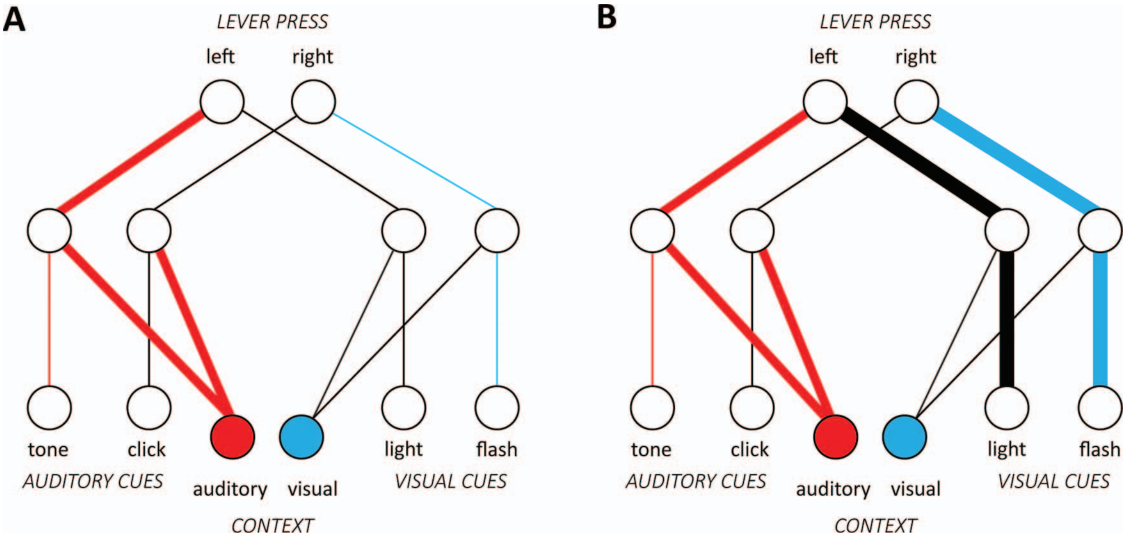

Figure 2. A schematic representation of the hierarchal associations thought to be at play in the rodent response conflict task. Circles represent processing units, conceptualized as populations of neurons that code task-relevant information. The lines represent connections between units, with colored lines indicating those connections currently activated by stimulus input; heavier lines indicate stronger underpinning associations or stronger activation of a pathway. During training two auditory stimuli (tone and click) predict whether a left or right lever response will be rewarded in one context (the auditory context). In another context, two visual stimuli (light and flash) predict which response will be rewarded (left or right). During testing animals are given novel audio-visual compounds that predicted either the same response (i.e., congruent; e.g., tone-light) or different responses (i.e., incongruent, e.g., tone-flash) during training. (A) Rats given an incongruent compound (here, tone-flash) during a test in the auditory context need to use contextual cues to resolve response conflict. These contextual cues are argued to bias activity in their corresponding stimulus–response pathway (here, tone-left), producing the response trained in that context. Inactivation of the prelimbic (PL) cortex prior to test disrupts the ability of animals to use the contextual cues to resolve response conflict. (B) Following overtraining of the visual discrimination, normal animals are unable to resolve response conflict when presented with the incongruent tone-flash in the undertrained auditory context. The stronger connection between the overtrained stimulus and response pairing (flash-right) overrides the ability of the contextual cues to influence activity in the contextually modulated tone-left stimulus–response pathway. Inactivation of the infralimbic (IL) cortex prior to test in the auditory context, however, allows animals to resolve the response conflict and perform the response associated with the tone-left pathway associated with the auditory context.

more on the lever appropriate to the visual cue components in the visual training context. In the same way that people have to use task instructions (i.e., name the color of the ink) to overcome the prepotent response of word-reading in the Stroop task, rats use contextual cues to resolve the conflict between the two elements of the incongruent auditory-visual compound stimuli.

The PL cortex facilitates the use of contextual cues to resolve response conflict. Lesion studies indicate that the medial prefrontal cortex is involved in the ability of animals to resolve response competition in the rodent Stroop task (Haddon & Killcross, 2006). Haddon and Killcross (2006) examined the impact of large pretraining lesions encompassing the entire medial prefrontal cortex on the ability of animals to use the task-setting contextual cues to resolve response conflict. These lesions did not disrupt acquisition of the conditional discriminations during the acquisition phase of the task. Further, they did not disrupt the ability of animals to respond appropriately during presentation of congruent compounds in the test session. However, when incongruent compounds were presented, animals with lesions of the mPFC failed to use the contextual cues present at test to resolve the ambiguous compound cue and perform the response associated with the stimulus elements trained in that context. This suggests that the mPFC usually facilitates the ability of animals to use contextual cues to influence response selection in ambiguous situations.

Further studies focusing on the contribution of the specific regions within the medial prefrontal cortex in this task have suggested that the PL and IL cortices play distinct roles in the resolution of response conflict. For example, Marquis, Killcross, and Haddon (2007) assessed the impact of functional inactivation of the PL cortex on the ability to resolve response conflict using this procedure. Inactivating the PL cortex before the test session had no impact on the ability of animals to perform the correct response when presented with congruent compounds. However, akin to the impact of large lesions of the mPFC, these animals were specifically impaired on incongruent trials. That is, inactivation of the PL cortex specifically disrupted the ability of animals to use the task-setting contextual cues to resolve response conflict. This implicates the PL region as the site within the mPFC that facilitates the ability of animals to use task-setting cues to direct responding in the presence of ambiguous information.

The IL cortex influences the ability of the well-trained response to dominate behavior. In contrast to the role of the PL cortex in promoting the ability of an animal to use contextual cues to resolve response conflict, an opposite pattern of results was found following inactivation of the IL cortex (Haddon & Killcross, 2011). In this study, the amount of training on each of the conditional discriminations was manipulated such that one discrimination received three times the training of the other. As a consequence of this differential training, and in line with the prototypical Stroop task in humans where people have trouble overcoming the tendency to read the word rather than name the color, rats show substantial interference from the overtrained stimulus when presented with incongruent compounds in the undertrained context. That is, animals are unable to use the undertrained task-setting contextual cues to resolve the conflict when these contextual cues required the animals to select an under- rather than overtrained response. However, when the IL cortex was inactivated prior to test, animals were rendered more able to use the contextual cues relevant to the undertrained stimulus–response associations to overcome the overtrained response. Thus, in contrast to the role of the PL cortex in facilitating the use of task-setting cues during response conflict, the IL cortex seems usually to be involved in attenuating the influence of these contextual cues in favor of allowing the simple stimulus–response associations to dominate behavior, independently of task-setting contextual cues.

In light of the evidence demonstrating that the PL and IL cortices contribute to the ability to resolve response conflict, the PL and IL regions may act as the sites within the PFC that facilitate the allocation of control in the manner described by Cohen et al.'s (1990) model. More specifically, PL lesions or inactivation prevent animals from using task-setting contextual cues to resolve response conflict. In contrast, IL lesion or inactivation allows animals to overcome the influence of overtraining and maintain the use of contextual cues to resolve conflict long after control animals. Hence, the PL cortex normally contributes to the ability to use task demands to modulate activity in stimulus–response pathways which dictate the correct lever response in the appropriate context. In contrast, the IL cortex appears normally to be involved in the process whereby with extended training stronger stimulus–response associations become progressively able to dominate the influence of task-setting cues.

The mPFC and Contextual Modulation of Instrumental and Pavlovian Behaviors

The PL cortex regulates expression of context-specificity of instrumental behavior. Other researchers have also found that the prelimbic region is involved in the contextual modulation of behavior in instrumental procedures (Trask, Shipman, Green, & Bouton, 2017; Willcocks & McNally, 2013). For example, Trask et al. (2017) found that inactivation of the PL cortex reduces the expression of ABA renewal. Specifically, Trask et al. (2017) trained rats to press a lever to receive food reward in one context (context A), extinguished the lever-press response in an alternative context (context B), and tested animals for levels of lever-press responding when they were placed back in the conditioning (context A). Under normal circumstances, rats will renew their responding when they are placed in context A as the cues present in extinction are no longer available and do not exert control on the inhibitory stimulus–response association developed in extinction (Bouton, 1993, 2004; Delamater, 1996; Rescorla, 1997). However, inactivation of the PL cortex during the renewal test selectively reduced the increase in lever pressing in context A. Much like the research investigating the resolution of response conflict described above, these findings suggest that the PL cortex is necessary to allow contextual cues to influence behavior to modulate the context-specificity of behavior. Specifically, we would argue that these results reflect a role for the PL cortex in using contextual cues to modulate the stimulus–response associations acquired during training, in a manner akin to that illustrated in Figure 2.

The role of the mPFC in the contextual modulation of responding and Pavlovian conditioning. Of course, there are many procedures that evoke response conflict that do not necessitate a choice between instrumental actions. Pavlovian procedures often endow a stimulus with ambiguous meaning. In the case of negative occasion setting, a target stimulus can signal either reinforcement or nonreinforcement, where presentation of another cue immediately prior to stimulus presentation signals that the stimulus will not lead to reward delivery (Holland, 1991). Essentially, animals need to use the initial cue to allow them to withhold responding to the target stimulus which has accrued excitatory associative strength via its relationship with reward on the other trials. As would be expected from the literature discussed above, the PL region of the mPFC is integral to allowing animals to withhold their response on cue-signaled trials that predict the absence of reinforcement (MacLeod & Bucci, 2010). Specifically, animals with PL lesions fail to perform appropriately in this task, where they exhibit similar levels of responding to the stimulus regardless of whether it is preceded by another cue which signals that reward will not be present (MacLeod & Bucci, 2010). Importantly, the responding of animals with PL lesions was intermediate to that of the control group. This suggests that in the absence of modulation by context from the PL cortex, rats respond on the basis of the associative strength of the stimulus (which is partially reinforced in the case of negative occasion setting). Interestingly, using a similar task, Moorman and Aston-Jones (2015) have recently demonstrated that activity in PL neurons increases during presentation of discriminative cues which signal whether or not an instrumental lever press will be rewarded. Significantly, the degree of activity correlates with successful performance on this task and is present across both discriminative cues (signaling lever presses that will, or will not, be rewarded) suggesting the PL cortex is not only involved in the reduction of responding on the basis of a task-setting cue predicting reward absence but is more generally involved the ability of task-setting cues to modulate both *increases* and *decreases* in performance.

In contrast, IL lesions impact on Pavlovian appetitive behaviors when animals should reduce responding following extinction learning. For example, lesions of the IL cortex have been found to enhance a renewal of responding toward a stimulus with a change in the contextual cues from that of extinction (Rhodes & Killcross, 2007). Here, Rhodes and Killcross (2007) used a Pavlovian version of the ABA renewal design described above where animals are presented with stimulus-reward pairings in one context (A), experience extinction to the stimulus in another context (B), and are then tested for responding to the stimulus back in the original conditioning context (A). Using this procedure, they found that animals with lesions of the IL cortex maintained low levels of responding in the extinction context but selectively exhibited enhanced renewal, relative to sham-lesioned controls, when placed back in the conditioning context. These data are in accordance with research showing that IL lesions also produce an increase in spontaneous recovery following the passage of time from the extinction episode (Rhodes & Killcross, 2004), commonly interpreted in terms of animals detecting a change in internal context from extinction, based on changing temporal cues (Bouton, 1993; Bouton, 2004; Harris, Jones, Bailey, & Westbrook, 2000). These data suggest that damage to the IL cortex appears to produce a heightened sensitivity to changes in contextual cues. Thus, the normal function of the IL cortex appears to be to offset the tendency to change responding on the basis of contextual cues in favor of the association developed in extinction (which is strong in the case of ABA renewal and spontaneous recovery following many extinction sessions). This is consistent with the data from the rodent Stroop task, where the IL cortex appears to oppose the influence exerted by the PL cortex in using present contextual cues to modulate responding.

As a brief note here, we would still interpret the role of the IL and PL cortices in Pavlovian behaviors as impacting upon a framework which entails stimulus–response associations (see Figure 3). Although Pavlovian behaviors are typically considered to be reflective of stimulus–outcome associations, it has been

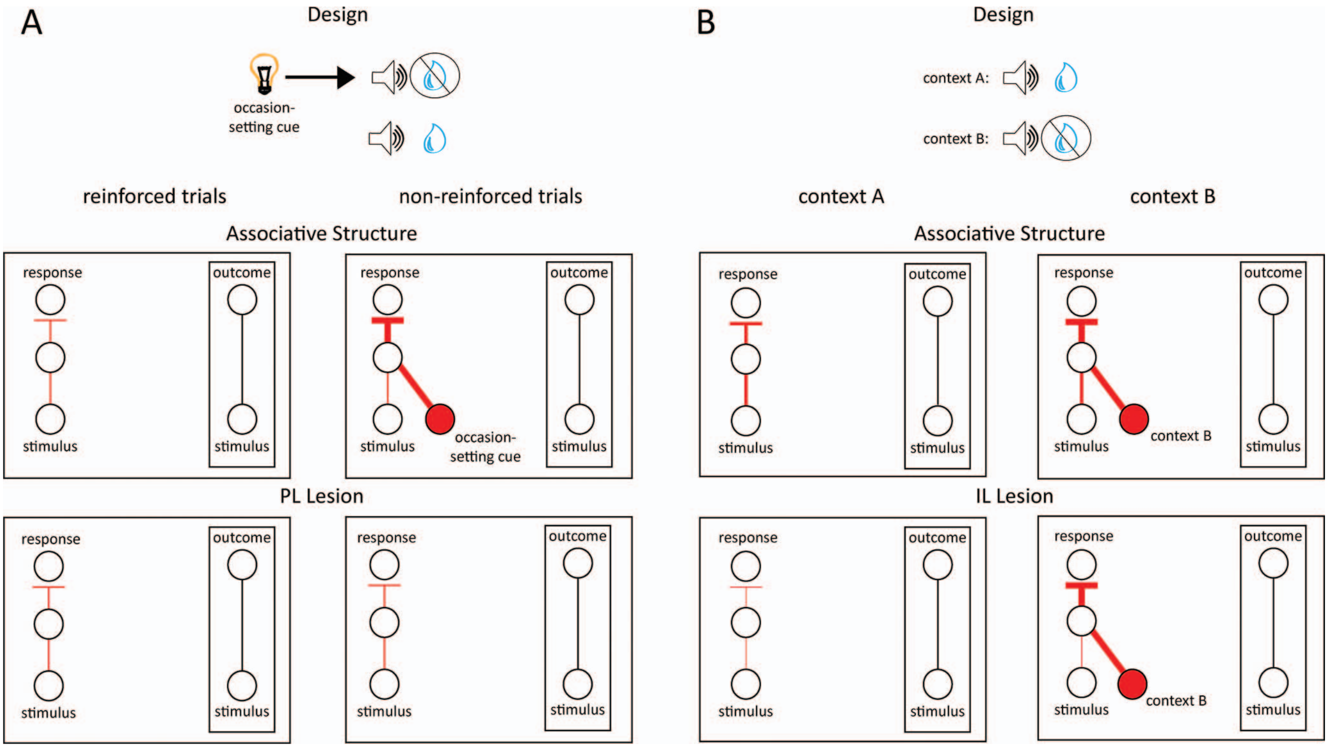

Figure 3. The role of the medial prefrontal cortex in the modulation of inhibitory stimulus–response associations during Pavlovian conditioning. We would view the role of the prelimbic (PL) and infralimbic (IL) cortices in Pavlovian conditioning as influencing the activation of inhibitory stimulus–response associations developed during extinction or partial reinforcement. Within this framework, stimulus-outcome associations formed during Pavlovian conditioning remain impervious to modulation and compete with the inhibitory stimulus–response association for behavioral expression. A: This model can account for the deficit in negative occasion setting found with PL lesions. Specifically, PL-lesioned rats fail to withhold responding on trials where a cue signals that an excitatory conditioned stimulus will not be rewarded. We would interpret this as a role for the PL cortex in allowing an occasion-setting cue to modulate activation of the inhibitory stimulus–response associations developed on nonrewarded trials. Specifically, under normal circumstances the PL cortex would function to increase activation of the inhibitory stimulus–response association in the presence of the occasion-setting cue and reduce responding. B: This model can also explain the finding that lesions of the IL cortex have been found to enhance the context-dependent renewal of responding following extinction. Here, IL-lesioned rats show normal low levels of responding in the extinction context but demonstrate enhanced renewal of responding when they are placed back in the conditioning context. We suggest that the IL cortex normally functions to promote the independent activity of stimulus–response associations (i.e., activity that is independent of top-down modulation). In intact animals, this would lead to some inhibitory stimulus–response activity acquired in context B generalizing to context A, leading to a modest reduction in responding compared to animals which did not receive extinction in context B. In the absence of IL cortex function, there is less generalization of the inhibitory stimulus–response to context A, leading to greater responding (i.e., enhanced renewal) due to the underpinning stimulus-outcome association. See the online article for the color version of this figure.

repeatedly shown that extinction facilitates the development of inhibitory stimulus–response associations (Delamater, 1996; Rescorla, 1997). That is, when a previously reinforced stimulus is presented in the absence of reward it results in the development of an inhibitory stimulus–response association which competes with the original excitatory stimulus-outcome association (Delamater, 1996; Rescorla, 1997). After prolonged extinction training, the inhibitory stimulus–response association dominates and presentation of the stimulus reduces the likelihood of response and produces extinction. Consistent with this view of extinction learning, we would interpret the finding that animals with PL lesions fail to inhibit responding in negative occasion setting as an inability to use the cue to modulate activation of the inhibitory stimulus–response association developed on nonreinforced trials (see Figure 3A). In much the same way, we would argue that the finding that IL lesions produce enhanced renewal following context change is reflective of an inability to promote responding of the strong inhibitory stimulus–response association regardless of the presence of contextual cues (see Figure 3B). In this model, the presence of PL top-down influence over the inhibitory stimulus–response association developed in extinction ensures responding remains low in context B in IL-lesioned animals (consistent with the experimental findings). With a functioning IL cortex to support the inhibitory stimulus–response association developed in extinction, some of the extinction seen in context B will generalize to context A, leading to limited renewal. However, when rats have no functioning IL cortex and they are placed back in context A, the IL cannot promote the inhibitory stimulus–response association, and the PL cortex cannot promote any top-down influence as this is conditional on being in context B. This allows the unmodulated stimulus–outcome association to dominate responding and produce an increase in renewal. In both cases, PL and IL cortices act to modulate the influence of the inhibitory stimulus–response association developed on nonreinforced trials where the excitatory stimulus–outcome association competes for behavioral expression and is unaffected by top-down modulation.

The Modulation of Attention

In addition to the role of the mPFC in the modulation of behavioral responding discussed above, the mPFC is also thought to play a central role in attentional processing when multiple stimuli compete for learning. Early rat studies assessed the role of the PL cortex in sustained attention using the five-choice serial reaction time (RT) task (Chudasama & Muir, 2001; Muir, Everitt, & Robbins, 1996), and subsequently moved toward suggesting that the PL cortex is involved in shifting attention when task demands change, as indexed by performance in the attentional-set shifting and response switching tasks (Birrell & Brown, 2000; Floresco, Block, & Tse, 2008; Gisquet-Verrier & Delatour, 2006; Ragozzino, Wilcox, Raso, & Kesner, 1999). For example, Birrell and Brown (2000) found that medial frontal lesions in rats selectively disrupted extradimensional set shifting, while leaving reversal learning and intradimensional set shifting intact. The ability to perform an extradimensional set shift involves disregarding a stimulus set (e.g., color) that previously predicted reward to shift attention to a new stimulus set (e.g., shape) which currently predicts reward. Furthermore, Floresco et al. (2008) used inactivation of the PL cortex to generalize this finding to switching response sets, which may also be interpreted in terms of attentional control (Sharpe & Killcross, 2014). These studies suggested that the PL cortex contributes to the attentional processing of stimuli that affords flexibility under changing contingencies.

Recently, results from our lab have attempted to clarify the specific role of the PL cortex in the modulation of attention toward stimuli during learning (Sharpe & Killcross, 2014). Given that previous studies had suggested the PL cortex is specifically involved in directing a preferential degree of attention toward certain stimuli over other present stimuli rather than changing attention toward single stimuli (e.g., shifting between attentional sets; Birrell & Brown, 2000), we chose to model the PL deficit on the basis of the competitive attentional process described in Mackintosh (1975) attentional model. This model was specifically developed to deal with scenarios in which multiple stimuli compete with one another and proposes that attention is a competitive process whereby an increment in attention toward one stimulus (or set of stimuli) necessitates a decrement in attention toward another stimulus (or set of stimuli). This is consistent the previous literature and also the current framework being developed in this review where higher-order information is used to arbitrate between different response choices. However, this does not discount the general role in learning for the attentional mechanism proposed by Pearce and Hall (1980) which advocates that attention will increase to a stimulus that is an uncertain predictor of an outcome and decrease to those stimuli which are reliable predictors of an outcome. In fact, it is now widely advocated that both these attentional mechanisms contribute in parallel during learning, where a Mackintosh (1975) mechanism may dominate in times of high attentional competition whereas the Pearce-Hall model may dominate in circumstances where few stimuli are present, ideas developed in so-called “hybrid” attentional theories which integrate these attentional mechanisms (Pearce & Mackintosh, 2010; Le Pelley, 2004). Given the role of the PL and IL cortices appears to be competition resolution, these areas may be specifically involved in a process described by Mackintosh’s (1975) attentional model, although future research is needed to assess this claim.

As the main premise underlying Mackintosh’s (1975) model is that attention is preferentially directed toward stimuli that are better predictors of an outcome than other stimuli present, we performed a number of experiments where we manipulated the predictive power of stimuli to assess the ability of animals with PL lesions to modulate the degree of attention directed toward such stimuli. In one experiment, we used an overshadowing procedure to examine the ability of animals with PL lesions to downregulate attention during learning (Sharpe & Killcross, 2014). Here, we assessed overshadowing by pairing an audiovisual compound with an outcome. Significantly, we arranged that the visual element of the compound was much less salient than the auditory component (Mackintosh, 1975, 1976; Pearce & Mackintosh, 2010). We did this because this treatment has been shown to bias humans and other animals toward use of an attentional mechanism to exhibit an overshadowing effect which favors the downregulation of attention toward the less salient element of the compound, without impacting on learning about the more salient stimulus (Le Pelley, Mitchell, Beesley, George, & Wills, 2016; Mackintosh, 1973, 1975, 1976). In our design then, attention to the visual element should be quickly downregulated as it becomes less predictive of the outcome than the more salient auditory stimulus which rapidly accrues associative strength (Mackintosh, 1975; Pearce & Mackintosh, 2010). Consequently, less learning accrues toward the visual element of the compound when compared to a matched visual stimulus paired on its own with the outcome, and the compound visual stimulus is said to be overshadowed. This form of overshadowing is to be contrasted with the mutual overshadowing predicted by models which argue that overshadowing occurs because the finite amount of associative strength is distributed across the stimuli presented in compound (Rescorla & Wagner, 1972; for more details, see Sharpe & Killcross, 2015a).

Using this overshadowing procedure, which produces a reliance on an attentional mechanism, we found that lesions of the PL cortex disrupted exhibition of the overshadowing effect (Sharpe & Killcross, 2014). Specifically, lesions of the PL cortex led to rats exhibiting greater learning about the overshadowed cues. This suggested that PL-lesioned animals failed to downregulate attention toward the visual element of the compound stimulus. We then confirmed the attentional nature of the deficit by demonstrating that animals with PL lesions were capable of exhibiting the blocking effect (an effect whereby prior learning about one cue blocks learning about a second cue when they are presented in compound), demonstrating that the rats could still distribute the finite amount of learning supported by an outcome across the two elements of a compound stimulus (Rescorla & Wagner, 1972) and that this could not be the reason we did not see overshadowing. However, after demonstration of successful blocking, PL-lesioned rats subsequently learnt about the blocked stimulus more quickly when it was paired individually with the outcome. This showed that rats with PL lesions failed to downregulate attention toward the redundant, blocked cue. That is, our dissociation between the overshadowing and blocking procedures demonstrates that the modulation of attention was required for overshadowing using stimuli of different salience, but not for the blocking procedure (as shown by the fact that PL-lesioned rats did not down-regulate attention to the blocked cue but still exhibited blocking). This suggested that the PL cortex is necessary for attentional modulation but not the distribution of learning across stimuli presented in compound with a motivationally significant outcome.

The argument above makes the assumption that PL-lesioned animals can resort to a Rescorla–Wagner (1972) to learn about relationships between stimuli on occasions when normal subjects will instead modulate attention toward stimuli to “solve” the task. To explicitly test this hypothesis, we designed an experiment that would pit these two mechanisms against one another. Specifically, we used a blocking of unblocking procedure where a Rescorla–Wagner (1972) and Mackintosh (1975) mechanism would predict an opposing pattern of results (Sharpe & Killcross, 2014). Here, rats were first presented with two stimuli, which were paired individually with an outcome (e.g., A+, C+). Then, rats received a blocking phase where one of these conditioned stimuli was paired in compound with another novel stimulus (i.e., AB+). Finally, rats were presented with the blocking compound and another novel compound comprising the other conditioned stimulus in an unblocking phase, where both compounds were followed by an upshift in the magnitude of reinforcement, which will drive additional learning (i.e., AB++, CD++).

Under these circumstances, a Rescorla–Wagner (1972) and Mackintosh (1975) process make distinct predictions about the effect that the blocking phase will have on subsequent learning about critical stimuli B and D in the unblocking phase. Specifically, according to a Rescorla–Wagner (1972) mechanism, the blocking phase will not have any effect on learning about stimuli in the unblocking phase. This is because attention toward stimuli does not change across the course of an experiment in the Rescorla–Wagner (1972) model and so the increment in learning about stimuli B and D in the unblocking phase is unaffected by the blocking phase. On the other hand, Mackintosh’s (1975) model argues that the redundancy of stimulus B in the blocking phase results in a downregulation of attention toward stimulus B. Consequently, learning about stimulus B is slower than stimulus D in the unblocking phase. Therefore, during a test session where B and D are presented individually under extinction, responding to B will be lower than that of D. In fact, this is what subjects in our control group showed; a reduction in responding to B relative to D demonstrating they downregulated attention toward stimulus B in the blocking phase, in line with the Mackintosh (1975) attentional model. However, rats with PL lesions did not demonstrate this difference (despite the ability to demonstrate blocking under normal circumstances; Sharpe & Killcross, 2014). Rather, they showed elevated levels of learning about B relative to D, a pattern of results predicted by Rescorla and Wagner (1972), where enhanced learning about B is explained by residual learning not allocated to stimulus A during the first phase accrues to B in the blocking phase, and this transfers to the unblocking phase. Taken together, these studies confirm that the PL cortex contributes to the ability of animals to downregulate attention toward redundant cues, and rats will resort to a Rescorla–Wagner (1972) mechanism to learn about stimuli in the absence of an ability to change attention toward stimuli.

Further emphasizing the coordinated roles of the PL and IL regions, George, Duffaud, and Killcross (2010) reported data from an optional set-shifting procedure which suggested that the IL region of the mPFC is involved in the maintenance of attention toward stimuli that have reliably signaled reward in the past (as opposed to the role of the PL cortex in changing attention toward stimuli). In this procedure, instrumental contingencies are set up so that stimuli of one stimulus set dimension (e.g., auditory stimuli; see Table 1) is informative in predicting which instrumental response will be reinforced (Table 1: Stage I). During a brief shift stage of the experiment, stimuli from another dimensional set also become relevant in predicting which instrumental response will be

Table 1
Design Used in George et al. (2010)

Phase and exemplar	Relevant	Irrelevant
Stage 1: Initial discrimination		
A1/V1: R1+		
A1/V2: R1+	Auditory	Visual
A2/V1: R2+		
A2/V2: R2+		
Stage 2: Shift discrimination	Auditory and visual	
A3/V3: R1+		
A4/V4: R2+		
Optional shift		
A3/V4 A4/V3		

Note. A = auditory; V = visual; R = response. Bold typeface references relevant stimuli.

reinforced (e.g., visual stimuli; Table 1: Stage II). In a test session (Table 1: Optional shift), George et al. (2010) found that sham-lesioned and PL-lesioned animals performed the instrumental responses associated with the stimulus dimension established as relevant in the first phase of the experiment. This suggested they had developed and maintained an attentional set despite the brief shift stage of the experiment. However, lesions of the IL cortex resulted in equivalent test responding on the basis of both stimulus dimensions, suggesting they did not maintain the attentional set from the first stage. This demonstrates that the IL region of the mPFC normally facilitates the formation or maintenance of an attentional set toward stimuli that have reliably predicted outcomes in the past, contrasting with the role of the PL region in modulating attentional changes in response to changing contingencies and promoting set-shifting. To illustrate this comparison using the attentional blocking example above, we would argue that the IL cortex would attempt to maintain the amount of attention directed toward the blocked cue (which starts out high as the novel stimulus is first presented; Mackintosh, 1975, 1976), opposing the influence of the PL cortex which attempts to use the relatively better predictive power of the predictive stimulus to increase attention directed toward it which results in a subsequent decrease in attention toward the novel, and consequently blocked, stimulus.

So how might we reconcile the role of the mPFC in attention with data implicating this region in the resolution of response conflict and contextual modulation? Interestingly, Cohen et al.'s (1990) model was initially described in terms of attentional processing. That is, the influence of task demands on the promotion of the correct response was considered to be the result of a subject's voluntary increase in attention toward the task-setting cues which subsequently biases activity of the relevant stimulus–response pathway. Thus, one way to reconcile these different findings would be integrate the attentional data within a hierarchical framework. Although we would maintain that attention will change as a consequence of changes in contingency or task demands as posited by Cohen et al. (1990), we conceptualize the nature of this attentional change in a slightly different way. Rather than the PFC facilitating attentional change through a voluntary increase in attention to, or activity of, the task-setting cues, which then bias activation of relevant stimulus–response pathways, we argue that a stimulus' predictive power (i.e., the strength of the relationship with the outcome) can also influence the ability of a target stimulus to elicit an attentional response. More specifically, in this situation some stimulus–response pathways consist of covert or overt orienting responses which will then change how behavior is controlled by that stimulus. Effectively, we propose that the associative strength of a stimulus will be used in a top-down manner to influence the ability of that stimulus to elicit an attentional response. Thus, attention will be high to good predictors of an outcome and low to poor predictors of an outcome, consistent with a Mackintosh (1975) attentional mechanism. In essence, we conceptualize an attentional response in this model as one of a host of responses that can be elicited by a stimulus and modulated by context, task demands, or the value of a goal (Le Pelley, Mitchell, & Johnson, 2013; see Figure 4). It is worth noting here not only the departure from Cohen and colleagues' original attentional model (Cohen et al., 1990, 2004; Miller & Cohen, 2001), but also the proposal of an explicit mechanism which explains how an attentional response may be influenced by top-down control alluded to in other frameworks (Le Pelley et al., 2013; Verbruggen, McLaren, & Chambers, 2014). While conceptualizing attention in this manner is novel, it is essentially a way in which we may reconcile the human cognitive neuroscience (which has traditionally been viewed in a top-down framework), with that of the associative learning literature.

We would also briefly note here that viewing attention as the result of top-down modulation allows illustration of a framework whereby factors other than a stimulus' predictive power can modulate attention paid toward a cue. For example, many studies have shown that contextual factors can change how much attention is paid toward a stimulus (Chun & Jiang, 1998; Hall & Channell, 1985; Sharpe & Killcross, 2015c). Take the classic example of latent inhibition; an effect whereby prior exposure to a stimulus reduces the degree of attention paid toward a stimulus, as indicated by both a reduction in orienting responses and slower rates of learning when that stimulus is paired with reward. If the context is changed after exposure of the stimulus, attention is again reinstated toward the stimulus and learning proceeds normally (Hall & Channell, 1985). This demonstrates that contextual cues can modulate the attention that is directed toward a stimulus. In the current framework, we would view the context as modulating the ability of a stimulus to elicit an attentional response in addition to the behavioral response in a top-down manner. Importantly, the PL cortex also abolishes the ability of contexts to differentially modulate attention which is directed toward a stimulus (Sharpe & Killcross, 2015c), consistent with a role for the PL cortex in top-down modulate of attention and responding as advocated in this framework.

From the perspective of an attentional response that changes as a result of how well a cue predicts an outcome, we could then interpret the role of the PL cortex in attention as changing the ability of a stimulus to elicit an attentional response as a consequence of a stimulus' predictive power. That is, when a stimulus is a redundant or poor predictor of an event, the PL cortex is important for reducing the degree to which that stimulus will elicit an attentional response. Further, we would anticipate that the PL cortex also contributes to increases in attention toward good predictors of an outcome and, therefore, is generally involved in the modulation of attention rather than explicitly a downregulation of attention (though future research is necessary to test whether this is the case). This would fulfill the requirements of a system that allows a preferential degree of attention to be directed toward better predictors in the environment. In contrast, the IL cortex seems to be important for maintaining attention despite recent changes in that stimulus' predictive power. This allows the mPFC to function as an attentional system that shifts the degree of attention on the basis of how well a stimulus predicts an outcome but remains impervious to brief changes in contingency that may be insignificant. As mentioned, an attentional system that biases attention toward stimuli that are better predictors of an outcome is consistent with the theory of selective attention encapsulated by Mackintosh's (1975) theory of attention in associative learning (Sharpe & Killcross, 2014). Thus, conceptualizing the PL and IL cortices in biasing attention according to how well a stimulus predicts an outcome fits within the hierarchal framework presented here while integrating this top-down framework with other accepted theories of selective attention and learning in humans and other animals (Le Pelley, 2004).

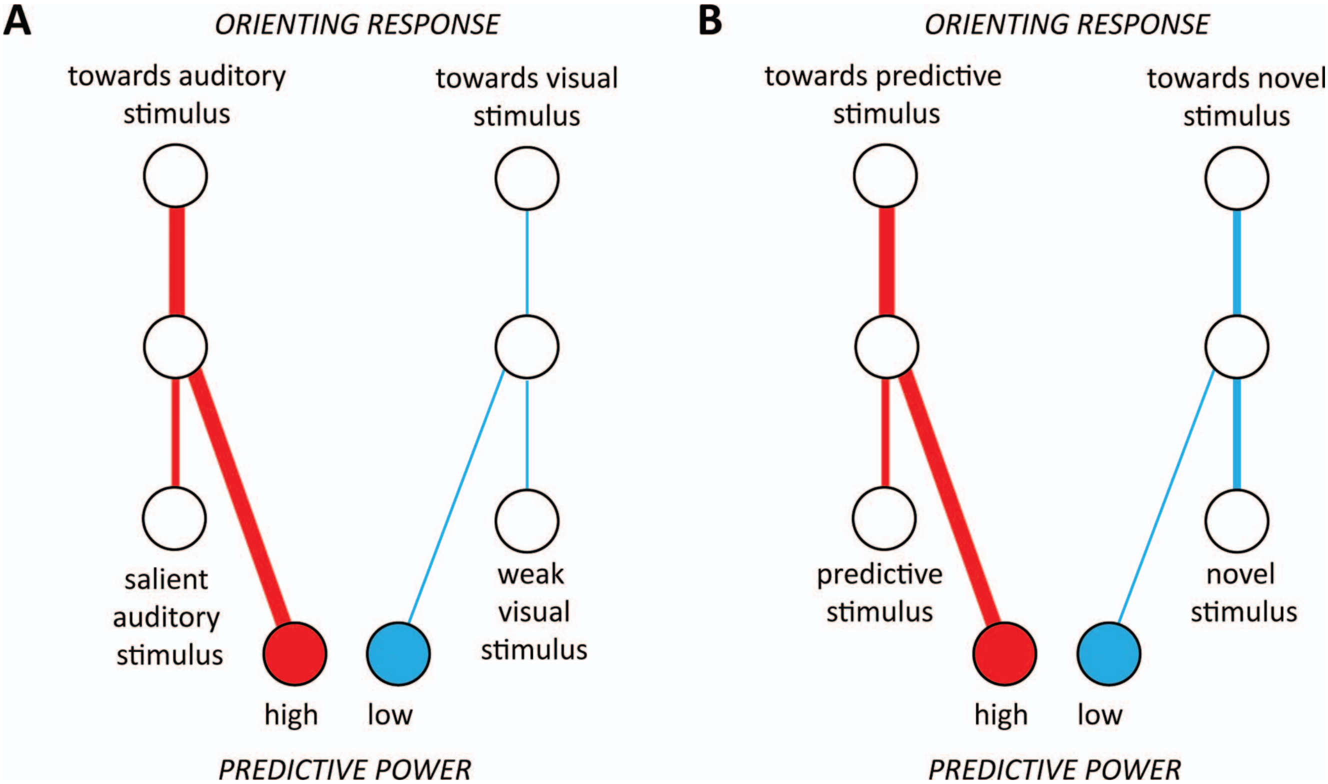

Figure 4. Top-down modulation of an attentional response by the predictive power of a stimulus. Mackintosh's (1975) model can be incorporated into a parallel distributed processing framework. Specifically, we would propose that the associative history of a stimulus can modulate the ability of that stimulus to elicit a covert or overt orienting response which will subsequently influence the rate of learning about a stimulus and the outcome it predicts. This is a departure from Cohen et al.'s (1990) attentional mechanism which argues that a voluntary increase in attention toward task demands, goal value, or contextual cues subsequently biases activation of the related stimulus–response pathway. Rather, we view the attentional response as simply another form of response which can be elicited by a stimulus, just as it might elicit an overt response. A: The model applied to an attentional interpretation of the overshadowing effect. Here, it is proposed that the inherent high salience of the auditory stimulus allows the stimulus to elicit a greater attentional response than the less salient visual stimulus. This increases the rate of learning and allows this stimulus to become a better predictor of the outcome, resulting in the cue acquiring strong predictive power. In turn this enhanced predictive power boosts the attentional response farther beyond that elicited by inherent salience alone. Consequently, the salient auditory stimulus commands a very high attentional response relative to the weak visual stimulus producing specific overshadowing of the visual stimulus. As a competitive system, the loss of attention toward the visual stimulus is a consequence of high attention paid toward the auditory stimulus, in line with Mackintosh's (1975) original framework. The prelimbic (PL) cortex is necessary for demonstration of this effect, where we would argue is the effect of PL inactivation is due to the consequent inability of the high predictive power of the auditory stimulus to modulate the ability of the stimulus to elicit a high attentional response, in turn failing to produce a decrease in attention toward the visual stimulus. B: The model applied to attentional downregulation in the blocking effect. After a stimulus has been established as predictive, that high predictive power promotes the maintenance of attention toward the predictive stimulus. Simultaneously, attention to the novel stimulus decreases. Although the PL cortex is not necessary for demonstration of the blocking effect per se, lesions of the PL cortex do prevent this downregulation of attention toward the novel stimulus. We would interpret this as a consequence of the lesion-induced inability of predictive power to modulate the ability of a stimulus to elicit an attentional response, where attention would decrease toward the novel stimulus as a direct consequence of an increases toward the predictive stimulus. See the online article for the color version of this figure.

Interim Summary

In the sections above we have reviewed the appetitive literature that implicates the involvement of the PL and IL cortices in a wide range of tasks. Specifically, these regions have been found to be involved in using contextual cues to influence a choice between responses in a rodent version of the Stroop task, occasion-setting procedures, and extinction protocols. In these tasks, the PL cortex promotes performance of an up-to-the-moment context-specific response, whereas the IL cortex promotes performance of well-trained responses that are context-independent and reflect the animal's long-term experience with the contingencies present. Further, the PL and IL cortices have also been implicated in the modulation of attention toward predictive cues. Here, the PL cortex mediates a change in attention directed toward a stimulus that is directly influenced by how well that stimulus predicts a motivationally significant outcome, whereas the IL cortex appears to contribute to the maintenance of attention toward stimuli regardless of how well they predict the outcome.

Although these tasks all necessitate the use of different information, we would argue that the manner in which the PL and IL cortices interact to influence behavior is the same. Specifically, in each case the PL cortex appears to use information available in the environment to produce a response that reflects the animal's most recent experience. That is, this region will exert voluntary control over responding through the use of current contextual cues, task demands, and how well a stimulus currently predicts an outcome, to elicit a behavioral response that responds acutely to the animal's current circumstance. In essence, it produces a rule that allows the animal to adapt rapidly to changing circumstances. In contrast, the IL cortex opposes this influence, favoring execution of responses that reflect the animal's long-term experience with ongoing contingencies and that is independent of current context, task demands, or most recent associative changes. So in contrast to the PL cortex, which is promoting adaptation to potentially important changes, the IL cortex is promoting behaviors that reflect long-term averages or probabilities. Thus, although the tasks in which these regions have been implicated are diverse, the way that they use information and interact with one another suggests a more integrated function that balances the need to respond flexibly in a changing world with the requirement to exploit what is already known about that world. We have incorporated these ideas into a well-established dichotomy in the cognitive neuroscience literature between automatic and controlled processing. Specifically, we would argue that the PL cortex exerts voluntary control over behavior through the top-down modulation of stimulus–response associations by context, task demands, and predictive power of a stimulus, whereas the IL cortex directly opposes this control to promote execution of strongest stimulus–response associations.

The Role of the mPFC in Aversive Procedures

The literature investigating work on the mPFC is essentially segregated into two different fields. As discussed above, on the one hand there is a history of research in the appetitive domain looking at the role of the mPFC in tasks requiring modulation of behavior to promote flexibility, supporting our theory that the mPFC may be involved in exerting top-down control over lower-order processing in a framework modeled on Cohen and colleagues (1990) PDP model. On the other hand, there is another literature that examines the role of the mPFC in aversive conditioning. Here, manipulating activity in the PL cortex has produced a number of contradictory results. Specifically, damage or inactivation of the PL cortex has been found to result in enhanced fear to a stimulus paired with an aversive outcome, or to produce greater levels of contextual conditioning to background cues, and lower levels of fear responding (Corcoran & Quirk, 2007; Holson, 1986; Lacroix, Spinelli, Heidbreder, & Feldon, 2000; Morgan & LeDoux, 1995). On the other hand, experiments manipulating activity in the IL cortex have produced consistent effects on the maintenance of extinction (Milad & Quirk, 2002; Quirk, Russo, Barron, & Lebron, 2000). However, implicating this region in extinction per se is at odds with the literature in the appetitive domain. That is, the absence of IL activity does not influence the maintenance of extinction when explicit contextual cues signaling extinction are present (Rhodes & Killcross, 2004, 2007). Rather, IL lesions appear to make animals more sensitive to changes in contextual cues following extinction, in line with the framework being proposed in this review. In this next section, we will reconcile the data from the appetitive and aversive domains by integrating these diverse findings from the fear literature into our framework.

The Role for the PL Cortex in Modulating Attention During Fear Conditioning

To test whether the PL cortex may also be necessary for modulating attention toward stimuli during fear procedures, we assessed the impact of PL inactivation during an aversively motivated version of the overshadowing task described above (Sharpe & Killcross, 2015a). Thus, as in the appetitively motivated procedures described above, we presented an audiovisual stimulus followed by delivery of a mild footshock and compared this learning with that attributed to two elemental stimuli (one auditory, one visual) each paired with shock individually. Here, we formally tested whether our procedure promoted the use of an attentional mechanisms to demonstrate the overshadowing effect. To do this, we demonstrated that our control group of rats showed overshadowing of the visual stimulus that formed part of the audio-visual compound. That is, learning about the visual stimulus when in compound was reduced relative to learning about the visual stimulus when paired individually with shock. However, learning about the auditory stimuli did not differ whether they were trained in compound or individually. As described in The Modulation of Attention section above, this finding is only consistent with an attentional account of overshadowing (Mackintosh, 1975, 1976). That is, Mackintosh's (1975) model predicts that attention will specifically reduce toward the weaker stimulus in a compound (in this case the visual stimulus) and he explicitly demonstrated this to be the case (Mackintosh, 1976). Further, as Mackintosh's (1975) model uses a separable error term, the finite amount of learning that can be supported by the unconditioned stimulus (US; in this case, the shock) does not need to be shared across the stimuli conditioned in compound. Thus, any small amount of learning attributed to the weak visual stimulus does not detract from learning about the salient auditory stimulus. This is in contrast to a nonattentional account of overshadowing (e.g., Rescorla & Wagner, 1972), which necessitates that any learning accruing toward the weaker stimulus detracts from learning about the auditory stimulus and cannot predict any learning about the visual stimulus at all without predicting a matching reduction of learning about the more salient auditory stimulus.

Using this overshadowing procedure, which produced a demonstrable reliance on an attentional mechanism, we found that functional inactivation of the PL cortex during learning produced a deficit in overshadowing. That is, when we inactivated activity in the PL cortex during conditioning where the audio-visual compound and elemental stimuli were paired with shock, we found a selective deficit in the ability of rats to demonstrate overshadowing of the visual stimulus. This was despite normal high levels of learning about both auditory stimuli. Interestingly, using the same procedure we found that inactivation of the PL cortex at test after learning had no impact on the expression of these associations. Specifically, all rats showed overshadowing of the visual stimulus and high levels of responding toward all other stimuli. These data demonstrated that the PL cortex is necessary for animals to modulate attention during fear learning but not for responding at test when attentional modulation is not necessary as stimuli are presented individually (rather than in compound) and are no longer in competition. This dovetails perfectly with the role of this region in modulating attention in appetitive procedures (see The Modulation of Attention section).

It is worth noting here that other researchers have found that inactivation of the PL cortex at test after conditioning reduces the expression of fear toward a single stimulus. One of the first studies that looked at the impact of posttraining inactivation of the PL cortex was conducted by Corcoran and Quirk (2007). Following conditioning, Corcoran and Quirk (2007) inactivated the PL cortex immediately prior to an extinction test session. Rats without PL function were found to express less fear toward a stimulus across this session. Since then, data from the same lab have demonstrated that microstimulation of the PL cortex increases the degree of fear elicited by a stimulus during an extinction test (Vidal-Gonzalez, Vidal-Gonzalez, Rauch, & Quirk, 2006). Similarly, Burgos-Robles, Vidal-Gonzalez, and Quirk (2009) have also demonstrated that PL neurons exhibit sustained activity across presentation of a stimulus and that the degree of activity in these neurons correlated with measures of conditioned freezing. Taken together, these data have led the authors to describe the PL cortex as being involved in promoting a sustained conditioned response across a lengthy presentation of a stimulus (see Figure 5; Sotres-Bayon, Sierra-Mercado, Pardilla-Delgado, & Quirk, 2012). However, these data are inconsistent with our findings in the overshadowing procedure described above, as well as several published demonstrations showing that lesions or inactivation of the PL cortex do not disrupt the basic acquisition or expression of cue-outcome associations (Balleine & Dickinson, 1998; Chudasama & Muir, 2001; Killcross & Coutureau, 2003; Marquis et al., 2007; Ragozzino et al., 1999; Sharpe & Kilcross, 2015c; Sharpe & Kilcross, 2014).

It may be that the discrepancy between these findings rests in the difference between the parameters used across studies. Specifically, in the Corcoran and Quirk (2007) study rats were given all conditioning trials in a single session with short intertrial intervals (ITIs) between stimulus-shock pairings. This is in contrast to the parameters used in our overshadowing procedure where we preexposed rats to the conditioning context prior to conditioning, which was then conducted across multiple sessions using long ITIs. The significance of the difference in these parameters lies in the degree of competition that may come from the context. That is, just as discrete stimuli compete to become associated with an outcome in the overshadowing procedure, the background context and discrete stimulus can compete to become associated with the outcome in any conditioning procedure (Maes & Vossen, 1993; Rescorla, 1984). The degree of competition from the context will depend on the amount of experience the animals have with the context in the absence of presentation of the outcome. Thus, conducting conditioning across a single session and using short ITIs between stimulus-shock pairings will increase competition from the context, as animals have less experience with the context outside of cue-outcome pairings (Maes & Vossen, 1993; Rescorla, 1984). In an environment of high competition, to attribute more learning toward the discrete stimulus as the more temporally accurate predictor, animals must devote more attention toward this uniquely predictive stimulus and actively downregulate attention toward the contextual stimuli as a set of less predictive elements (Mackintosh, 1975; Pearce & Mackintosh, 2010; see Figure 6). This allows greater levels of learning to accrue to the discrete stimulus as the better predictor of the US, and relatively little learning to accrue to contextual stimuli. Given the importance of the PL cortex in modulating attentional competition, it is likely that increasing competition from the context will produce an increasing reliance of PL-dependent attentional modulation and, therefore, greater effects of PL inactivation.

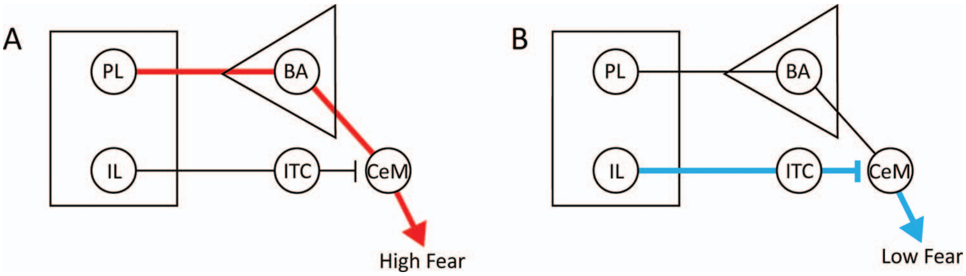

Figure 5. Quirk and colleagues argue that the prelimbic (PL) cortex is necessary for the expression of conditioned fear, whereas the intralimbic (IL) cortex is necessary for the maintenance of extinction. Findings that inactivation of the PL cortex reduces fear expression, while inactivation of the IL cortex increases fear expression have lent to the development of a model whereby the PL cortex is primed to increase fear responding whereas the IL cortex is important for the maintenance of extinction. A: Following conditioning, it is argued that the basal amygdala (BA) signals the presence of a fearful stimulus to the PL cortex, which in turn potentiates inputs to BA to produce a sustained freezing response to the stimulus via increased activation of neurons in the central nucleus (CeM) of the amygdala. B: Following extinction, stimulus presentation also recruits IL activity and freezing is suppressed through the amygdala intercalated cells (ITC) which inhibit freezing elicited by the CeM. See the online article for the color version of this figure.

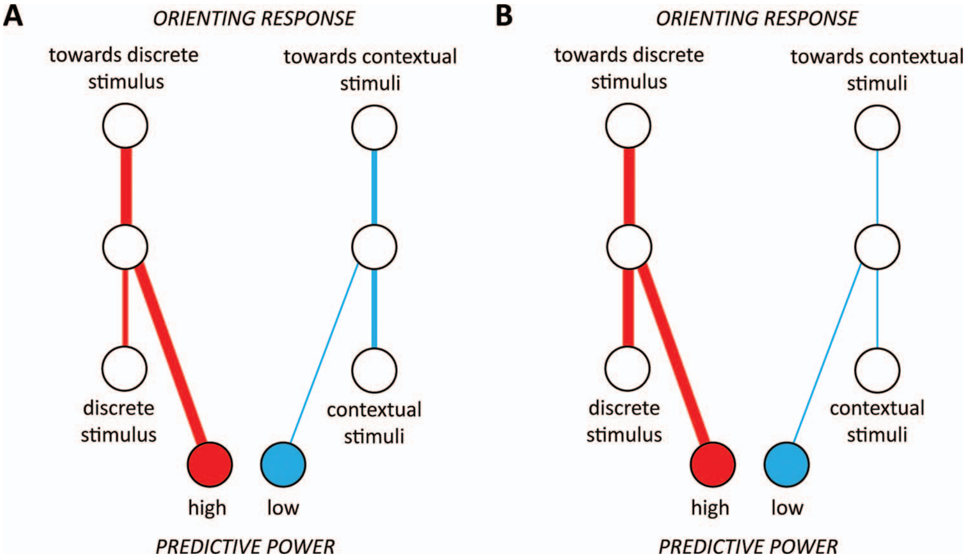

Figure 6. The prelimbic (PL) cortex modulates attention during fear learning. Lesions or inactivation of the PL cortex produce deficits in learning about fearful stimuli only when there is high competition between discrete and contextual stimuli. Consistent with results from the appetitive literature, we would argue that this is due to a role of the prelimbic cortex to use the predictive power of a stimulus to increase the ability of the stimulus to elicit an attentional response. In this model, the degree of attention directed toward the discrete or contextual stimuli is directly proportional to the competition between discrete and contextual stimuli, determined by the parameters of the experiment (e.g., current intertrial intervals and history of context exposure in the absence of shock). A: When competition between the contextual and discrete stimuli is high, the PL cortex can use the strength of the unmodulated stimulus-outcome associations to influence the attentional response elicited by a stimulus. This allows a greater degree of attention to be devoted to the discrete stimulus as a better predictor of an outcome and competition between the stimuli to be resolved which influences future learning. B: When competition between discrete and contextual stimuli is low, the prelimbic cortex is not needed to resolve conflict between stimuli as more attention is devoted toward the discrete stimulus and the majority of learning is directed toward it. See the online article for the color version of this figure.

Thus, it may be that findings resulting from posttraining inactivation of the PL cortex can also be reconciled by appealing to a role for the PL cortex in directing attention toward predictive stimuli (Sharpe & Killcross, 2014, 2015a). That is, just as attention can influence how much learning accrues toward a stimulus, the degree of attention that is directed toward a stimulus can also influence the level and duration of conditioned responding it elicits (Le Pelley, 2004; Mackintosh, 1975; Pearce & Mackintosh, 2010). In the Corcoran and Quirk (2007) study, during conditioning rats received conditioning in a single session using 2-min ITIs. These parameters would likely have promoted a high degree of competition between contextual and discrete cues. Despite this competition, as the PL cortex was functioning normally during conditioning, all animals would have been capable of resolving the competition and would have learnt that the discrete stimulus was the better predictor of shock relative to the context, and the discrete stimulus would therefore have gained the majority of associative strength. During the test session, the parameters remained the same (i.e., short ITIs). Here, as we have shown, the PL cortex is necessary to use the predictive value of a stimulus to modulate the degree of attentional response elicited by that predictive stimulus (see Figure 4). Hence, animals without PL function would fail to devote preferential attention to the predictive discrete stimulus relative to the context (which comprises many elemental stimuli and, therefore, receives more unmodulated attention and associative strength). Thus, animals without PL function would direct attention indiscriminately to both contextual and discrete stimuli in the chamber at test. As the stimuli that comprise the context had acquired little associative strength across training these animals are in effect basing their responding on a lower average associative strength, leading to a deficit in conditioned fear expression.

Interestingly, this interpretation can also reconcile other discrepancies in the fear literature regarding PL function. Specifically, lesions of the PL cortex have been found to produce no effect on conditioned fear, produce an enhancement of conditioned fear to both a context and a stimulus, or produce greater learning about a context at the expense of a discrete stimulus (Holson, 1986; Lacroix et al., 2000; Morgan & LeDoux, 1995). In each case the findings can be accounted for by differential levels of competition from the context produced by varying parameters. Studies that have shown no effect of PL lesions on fear learning have used parameters that would tend to produce low competition from the context (Holson, 1986), whereas those that have found increased learning about both a stimulus and a context have used parameters that tend to promote moderate competition from the context (Morgan & LeDoux, 1995), and those which have found an increase in learning about the context at the expense of the stimulus have used parameters that tend to promote high competition from the context

(Lacroix et al., 2000). Thus, studies that have found effects of PL damage on fear learning have used parameters which promote high competition from the context. This would likely necessitate attentional modulation to resolve. In much the same way as two discrete stimuli competing for attention, the contextual stimulus can compete with a discrete auditory stimulus. As in the case with two discrete stimuli, attention will increase to the better predictor of the outcome to preferentially devote learning toward the better predictor of shock. In the case of fear conditioning where the discrete stimulus is often temporally adjacent to delivery of the outcome, attention (and subsequently, learning) will increase to the discrete stimulus and decrease to the redundant contextual stimuli.

To test explicitly test the hypothesis that the role of the PL cortex in fear learning depends on competition from the context, we recently assessed the impact of pretraining lesions of the PL cortex on the expression of fear toward a stimulus in circumstances providing differential degrees of competition from the context (Sharpe & Killcross, 2015a). Specifically, we examined the impact of pretraining lesions of the PL cortex on responding to a stimulus when animals had little experience with the context outside of cue-shock pairings and when they had more experience with the context outside of these pairings. According to the argument above, if the PL cortex is specifically involved in modulating attention to resolve competition between cues during fear conditioning, PL lesions should only impact on fear learning when competition between the context and the discrete stimulus is high. In line with this, we found that when animals received short ITIs between stimulus-shock pairings and low context exposure, lesions of the PL cortex enhanced learning about the context at the expense of learning about the conditional stimulus (CS). This would be expected if a Rescorla–Wagner (1972) mechanism was used by PL-lesioned rats in the absence of attention modulation when competition between stimuli was high (as in the case in blocking, see The Modulation of Attention section). Specifically, as associative strength is distributed across all stimuli that comprise the context and the single discrete auditory stimulus, more associative strength becomes devoted to the many stimuli comprising the context relative to the discrete stimulus. However, when a separate group of animals received high levels of context exposure and were trained with long ITIs between stimulus-shock pairings, PL lesions had no effect on the ability of animals to exhibit high levels of fear toward the discrete stimulus and low levels of learning about the context. These data demonstrate that the effect of pretraining lesions of the PL cortex on fear conditioning procedures is dependent on whether the parameters used promote competition between the discrete stimulus and context, which will then require attentional modulation to resolve. That is, the PL cortex is involved in fear learning and expression only when attentional modulation is required, in line with our findings examining overshadowing in a fear conditioning procedure, as well as appetitive studies which demonstrate that the PL region is necessary to modulate attention toward predictive stimuli.

The Role of the PL Cortex in the Context Specificity of Extinction

Recent studies have also suggested that the PL cortex plays a role in the contextual regulation of fear expression following extinction (Orsini, Kim, Knapska, & Maren, 2011; Zelikowsky et al., 2013). For example, Orsini et al. (2011) found that disconnection of the hippocampus (HPC) and the PL cortex disrupted the ability of animals to exhibit renewal of conditioned fear following extinction. Orsini et al. (2011) paired a discrete stimulus with shock in one context (context A), extinguished the stimulus in an alternative context (context B), and tested animals for levels of fear expressed toward the stimulus in a third familiar context (context C; i.e., an ABC renewal paradigm). Similarly to the ABA renewal procedure described above (see The mPFC and Contextual Modulation of Instrumental and Pavlovian Behaviors section), rats will usually renew their expression of fear when they are placed in context C as the cues present in extinction are no longer available and do not exert control on the inhibitory stimulus–response association developed in extinction (Bouton, 1993, 2004; Delamater, 1996; Rescorla, 1997; see Figure 3). Prior to the extinction test session in context C, Orsini et al. (2011) gave rats unilateral lesions (contralateral or ipsilateral) of the HPC and PL cortex. Rats with contralateral lesions, where the HPC and PL cortex are functionally disconnected, failed to exhibit the renewal of fear when the context was different from that experienced in extinction. This suggests that these animals were not capable of using the contextual cues to exhibit a renewed fear response to the extinguished stimulus. These data may suggest that the PL cortex plays a role in allowing contexts to modulate responding in an aversive setting as is the case in appetitive procedures. Here, a role for the PL cortex in modulation behavior through use of contextual cues is to be contrasted with the role described in the section above where the PL cortex modulates attention toward a discrete stimulus over other present contextual stimuli (see Figure 3). In the latter case, the context acts as a stimulus which is capable of entering into a direct association with the outcome rather than as a modulatory cue (see Figure 6).

Data from our lab have recently extended results implicating PL–HPC interactions in contextual modulation to investigate the impact of specific manipulation of PL activity on the ability of animals to use contextual cues to regulate fear responding (Sharpe & Killcross, 2015c; Sharpe & Killcross, 2015b). In one study, we examined the impact of lesions and functional inactivation of the PL cortex on the ability of animals to exhibit ABA renewal. Here, animals received tone-shock pairings in context A, extinction in context B, and were then placed back into context A. Here, we found that pretraining lesions of the PL cortex did not impact on the development of extinction of conditioned fear, again supporting the idea that the PL cortex is not directly involved in the expression of fear per se. However, when these rats were again placed in context A following extinction, they failed to renew the expression of conditioned fear. These data demonstrate that the PL cortex is necessary to use contextual cues to modulate responding when the stimulus is presented after extinction, similarly to findings that have implicated the PL cortex in using contextual cues to resolve response conflict in the rodent version of the Stroop task (Marquis et al., 2007). This suggests that the role of the PL cortex in using contextual cues to modulate performance in fear learning is similar to that seen in appetitive studies, where we would envision its role as using contextual cues to exert top-down control over an inhibitory stimulus–response pathway acquired in extinction to demonstrate context-specific renewal (see Figure 3).

Following the finding that the PL cortex is necessary to modulate performance during the renewal test in the conditioning context, we were interested to assess whether the PL cortex would also be necessary for the development of modulatory associations that would allow the context to influence behavior (Sharpe & Killcross, 2015b). To investigate this question, we silenced PL activity during extinction in ABA renewal. That is, after rats experienced the tone paired with shock in context A, we placed rats in context B and presented the tone stimulus in the absence of shock. Interestingly, we found that inactivation of the PL cortex during extinction accelerated extinction learning. That is, without PL function rats exhibited a faster suppression of fear to the tone. This is as one would expect if rats without PL function could not use contexts to exert top-down control over responding and resorted to a simpler mechanism to express extinction. Specifically, we speculated that rats would resort to a Rescorla–Wagner (1972) mechanism (known to be intact in PL-lesioned animals; Sharpe & Killcross, 2014) which would predict that the cues comprising context B gain negative associative strength as they are "extinguished" with the conditioned stimulus (Delamater & Westbrook, 2014). This would result in faster extinction as the loss of associative strength to a compound is faster than that of a discrete stimulus alone (Delamater & Westbrook, 2014; Rescorla & Wagner, 1972). Further, this effectively protects the discrete stimulus from extinction as the summed associative strength of both the context and discrete stimulus must reach zero, causing the context to accrue negative associative strength and, therefore, the discrete conditioned stimulus to maintain a significant proportion of its associative strength. Counterintuitively, this would predict that rats without PL function during extinction would still exhibit renewal of fear when placed back in the conditioning context A. That is, when the rats are no longer in context B, they would exhibit responding that is based on the discrete stimulus alone, revealing the maintenance of associative strength. In fact, this is exactly what we found. Specifically, although rats without PL function exhibited faster extinction to the discrete stimulus, when placed back in context A they were still capable of showing robust renewal of fear to the tone. These data are consistent with a role for the PL cortex in the modulation of associations by context, where in the absence of PL activity, rats resort to an elemental mechanism to exhibit context-specific renewal of conditioned fear.

Of course, the account described above dictates that if rats could not use a simpler elemental mechanism rats without PL function would fail to demonstrate context-specific behavior. To test this hypothesis, we used a biconditional contextual discrimination. Here, we presented rats with two stimuli across two different contexts. In one context, one of the cues was presented with mild shock whereas the other was presented without reinforcement. In the other context, these associations were reversed such that the stimulus that previously predicted shock was now presented without reinforcement and the other stimulus now predicted shock. Under normal circumstances, rats will respond differentially to the stimuli depending on which contextual cues are present. That is, they will exhibit fear to one of the stimuli when it is presented in the context where it predicts shock, and the other when placed in the other context. Note that in this experimental design, as opposed to the ABA or ABC designs, rats cannot use the stimuli that comprise the context in an elemental fashion as both contexts have stimuli that are either reinforced or nonreinforced. As expected, in this circumstance we found inactivation of the PL cortex either during either learning or during an extinction test produced indiscriminate levels of fear toward both stimuli, regardless of context. That is, animals without PL function at either the learning or test phase of this task failed to use the contexts to modulate learning or responding toward the stimuli. These data extended current understanding of PL function, showing that the PL cortex is not only necessary to use contextual cues to modulate performance online but also to influence the development of context-specific modulation of associations during learning. In thinking about the framework we have proposed, we would argue that the PL cortex is necessary for the development of the higher-order association whereby the context comes to modulate the ability of a cue to elicit a behavioral response in addition to using contexts online to produce context-specific effects at test after normal learning has taken place.

The Role of the IL Cortex in Generalizing Fear Learning Across Contexts

Consistent with the appetitive literature, lesions or inactivation of the IL cortex have been found to increase the spontaneous recovery of fear after extinction. In fact, one of the first studies to assess the role of the IL cortex in fear investigated the impact of pretraining IL lesions on the recovery of fear after extinction (Quirk et al., 2000). Here, lesions did not impact acquisition or initial extinction of conditioned fear, in much the same manner as PL lesions. However, when IL-lesioned rats were placed back in the conditioning context following extinction they exhibited higher levels of recovery of fear. Since then, recording studies have shown that IL neurons are active the day after extinction, where increased activity is associated with lower levels of freezing during recovery tests (Milad & Quirk, 2002). Further, stimulation of IL neurons during recovery also reduces levels of freezing and extinction has been found to modify the intrinsic excitability of IL neurons (Milad & Quirk, 2002; Santini, Quirk, & Porter, 2008). These results have led to the theory that the IL cortex is necessary to reduce fear expression after extinction (in direct contrast to the role of PL cortex in increasing fear expression).

However, an increase in spontaneous recovery after extinction in IL-lesioned rats has also been reported in the appetitive literature (Rhodes & Killcross, 2004), and is entirely consistent with a role for the IL cortex in generalizing extinction across contexts (discussed in The IL Cortex Influences the Ability of the Well-Trained Response to Dominate Behavior section). That is, these data implicated a role for the IL cortex in promoting the inhibitory stimulus–response association developed in extinction that is independent of context. This is strongly supported by the nature of extinction deficit seen in IL-lesioned animals. Specifically, a critical component of the enhanced recovery seen in IL-lesioned animals is a transient increase in the recovery of fear at the beginning of the second extinction session which then decreases in line with that of controls across successive trials (Milad & Quirk, 2002; Quirk et al., 2000). This transient increase at the start of successive extinction sessions in IL-lesioned rats is also seen in appetitive conditioning (Rhodes & Killcross, 2004). Here, analyses showed that although IL-lesioned animals show increased recovery at the start of session relative to sham-lesioned controls, both groups show an equivalent reduction in responding across the remainder of the extinction session (Rhodes & Killcross, 2004). That is, after an initial increase in responding, IL-lesioned rats quickly adjust to normal levels of extinction performance, showing equivalent savings as sham-lesioned controls from one extinction session to the next. This is consistent with a role for the IL cortex in generalizing extinction across contexts. Specifically, in the absence of a physical change in context, rats will interpret the passage of time as a potential shift in context. Accordingly, at the beginning of successive extinction sessions rats will recover responding to an extinguished stimulus as they judge the temporal context to have shifted from that of initial extinction. As the test session proceeds, presentations of the stimulus in the absence of reinforcement reestablishes the extinction context and performance comes to reflect the previous and ongoing extinction learning. Thus, a selective increase in responding at the beginning of a session in IL-lesioned animals is consistent with an enhanced sensitivity to perceived changes in contextual cues resulting in an increase in initial responding which subsides after presentations of the stimulus without reinforcement reestablishes the extinction context.

Interpreting the enhanced recovery seen in animals without IL function as enhanced sensitivity to changes in contextual cues is supported by studies investigating the impact of IL lesions on ABA renewal. As has been demonstrated in appetitive studies (Rhodes & Killcross, 2007), it has been shown that IL lesions produce enhanced renewal to an extinguished stimulus following context change in an aversive setting (Zelikowsky et al., 2013). Here, IL lesions did not affect recovery after extinction when rats were tested in context B. That is, when contextual cues were available these animals were capable of withholding the response, demonstrating that IL cortex is not necessary for extinction when contextual cues signal that the extinction context is in effect. However, these rats showed enhanced recovery when placed back in the acquisition context A. These data confirm that lesions of the IL cortex render an animal particularly sensitive to changes in contextual cues that come to modulate extinction. To put this another way, the normal function of the IL cortex may be to allow extinction memories to generalize across contexts (i.e., to oppose the PL-mediated modulation of responding by contextual cues) so as to protect fear memories from recovery, rather than maintaining an extinction memory per se. This is in line with the role of the IL cortex in promoting the strongest stimulus–response pathway described in the appetitive section above. Here, we would argue that the IL cortex promotes activation of the inhibitory stimulus–response association acquired in extinction to offset the influence of a perceived change in context cues in reinstating performance to an extinguished stimulus. This again places the function of the IL in opposition to that of the PL, maintaining control of behavior by simple training history, and resisting top-down modulation by contextual cues.

Interestingly, inactivation of the IL cortex during or after the first extinction session has also been found to produce an increase in recovery when rats are tested the following day. That is, IL inactivation during the initial extinction session does not impair a reduction in responding during that session, however these rats show heightened recovery at the beginning of the extinction session the following day (Laurent & Westbrook, 2009). Similarly, IL inactivation in the consolidation window after the initial extinction session also produces enhanced recovery across the first few trials of the extinction session the following day (Laurent & Westbrook, 2009). These findings beg the question of how IL inactivation the day before test can result in a heightened sensitivity to a change in the contextual cues in the subsequent extinction session. One possibility is that the level of influence contextual cues have in modulating responding develops across extinction learning and is predetermined prior to next extinction session. Thus, during the initial extinction session activity in the PL and IL cortices interact to establish how much control contextual cues will have in biasing their associated responses. In the absence of IL activity, PL activity may promote a heavier influence of contextual cues over the association developed in extinction and produce enhances sensitivity to changes in contextual cues in subsequent extinction sessions. Thus, associations allowing contextual cues to regulate responding may be continually updated to change how these cues may influence responding in future learning episodes. In this way, the interaction between IL and PL activity may balance the benefit of adapting responding to a perceived change in context with that of increasing knowledge that the cue will no longer produce reinforcement regardless of a change in context.

Interim Summary

In the section described above we have outlined the current discrepancies residing within the aversive literature. First, we discussed findings from our laboratory which have shown that the PL cortex modulates attention toward predictive cues to resolve competition and devote learning toward the best predictor of an outcome during aversive learning, as is the case in appetitive conditioning procedures. Here, we discussed how a role for the PL cortex in modulating attention could explain the finding that PL inactivation reduces the expression of fear. Specifically, we would argue that findings that have shown that lesions or inactivation of PL cortex reduces fear expression have used parameters which promote high competition which produce a reliance of PL activity to resolve and devote learning and responding to the best predictors of shock. We experimentally tested this hypothesis by varying parameters to promote either high or low competition from contextual cues and assessed the impact of PL lesions in each case. We found that PL lesions only impacted on the expression of conditioned fear when there was high competition between contextual and discrete cues. When competition from the context is low, we saw that PL lesions or inactivation had no effect on the ability of animals to express fear to a discrete stimulus. Second, we and others have shown that the PL cortex is involved in utilizing contextual information to regulate learning and responding in aversive paradigms in much the same way as that seen in appetitive procedures. Finally, we have also discussed the literature that implicate the IL cortex in the maintenance of extinction. We believe that the specific pattern of the deficit seen in extinction, where responding is increases selectively across the first few trials of the session, is due to an enhanced sensitivity to contextual cues. In each case, we would argue that the role of the PL and IL cortices can be explained using an expansion of the PDP model and framework put forth by Cohen and colleagues (Cohen et al., 1990; Miller & Cohen, 2001). Specifically, that the PL cortex in involved in the process whereby contextual cues and predictive power of a stimulus can modulate the ability of the associated stimulus to elicit a response. Again, the role of the IL cortex in generalizing memories across contexts can be explained by a role for this region in promoting the strongest stimulus–response pathway independent of context, potentially through direct opposition of PL-mediated modulation of responding by context via intrinsic connections.

Can We Integrate the Role of the mPFC in Goal-Directed Actions and Habits Into This Framework?

One of the first functions attributed to the mPFC cortex is the coordination of actions and habits (Balleine & Dickinson, 1998; Coutureau & Killcross, 2003; Killcross & Coutureau, 2003). Specifically, the PL cortex is necessary for instrumental responding which is sensitive to the current value of a goal and the contingency between action and that goal (Balleine & Dickinson, 1998). That is, animals without PL lesions do not adjust their behavior following goal devaluation or when the goal is presented noncontingently (where the probability of reward is independent of responding). This demonstrates that the PL cortex is necessary for the voluntary, goal-directed actions. On the other hand, it is now well established that the IL cortex contributes to the formation of stimulus–response habits (Coutureau & Killcross, 2003; Killcross & Coutureau, 2003; Smith, Virkud, Deisseroth, & Graybiel, 2012). In accordance with this, lesions of the IL cortex have been found to result in the persistence of sensitivity to goal value after extended training, indicating a lack of transition toward a habitual system (Killcross & Coutureau, 2003), and neuronal inactivation of the IL has been demonstrated to reinstate goal sensitivity in well-trained rats that would otherwise demonstrate habitual responding (Coutureau & Killcross, 2003). More recently, Smith et al. (2012) have also demonstrated, using optogenetic perturbation, that the IL cortex appears to be acting online to promote performance of habits at the time of behavioral expression. As suggested by Killcross and Coutureau (2003), this dissociation between the contribution of the IL and PL to instrumental performance indicates that the mPFC acts as a system which facilitates the trade-off between flexible but taxing goal-directive performance (supported by the activity of the PL cortex) and the control of behavior by autonomous and involuntary habits (supported by the activity of the IL cortex).

The balance between goal-directed behavior and habitual behavior is similar to the framework advocated here. That is, the current article advocates that activity in the PL and IL cortex exerts top-down control over behavior to achieve a trade-off between behavior which is sensitive to current changes in the environment and that which reflects the long-run history of experience. As such, it is tempting to integrate within this framework the role of the PL and IL cortices in the balance between responding which is sensitive to changes in the value of an outcome being worked toward (i.e., goal-directed behavior) and making a response which has reliably resulted in procurement of a rewarding outcome in the past (habitual behavior). For instance, one might argue that the goal could act as a cue to exert top-down control over stimulus–response pathways in much the same way as contextual cues. Accordingly, the value of a goal could influence performance of a response which is associated with that goal, so reducing the value of the goal would reduce the likelihood of making a response associated with the devalued goal.

However, there are two reasons to think this is not possible. First, it has been demonstrated that the PL cortex is only involved in the acquisition of goal-directed behaviors. That is, inhibition of PL activity only impacts on goal-directed behavior during instrumental training and not during a test session after subjects have learnt normally (Ostlund & Balleine, 2005). This is problematic because the role of the PL cortex in other behaviors described in this review (such as attention and contextual modulation) extends to performance of these behaviors even when these behaviors are well learnt. The second issue is that goal-directed behavior is generally conceptualized as governed by an association between the response and the outcome (i.e., an R–O association; Balleine & Dickinson, 1998; Colwill & Rescorla, 1990). As such, thinking about the goal as exerting top-down control over stimulus–response pathways is at odds with current accounts of the hierarchical structure of goal-directed learning, for which there is considerable evidence. Taking these concerns into account, it may be that the role of the PL and IL cortices in goal-directed behavior cannot be integrated into the current framework and is subserved by processing of a qualitatively different nature in these regions that may be governed by distinct connections between the medial prefrontal cortex and the rest of the brain. This would be supported by the literature in humans, showing that goal-directed behavior is dependent on different neural circuits than both contextual and attentional modulation (see the next section for discussion). However, a view for top-down modulation of behavior by goals may have relevance for more complex settings where goals are used as discriminanda and also require the medial prefrontal cortex (de Wit, Kosaki, Balleine, & Dickinson, 2006; Dwyer, Dunn, Rhodes, & Killcross, 2010).

Functional Homologues of the mPFC in Humans

Functional Homologue for the PL Cortex

A number of studies have suggested that the dorsolateral prefrontal cortex (DLPFC) shares functional similarity with the rodent PL cortex (Egner & Hirsch, 2005; MacDonald, Cohen, Stenger, & Carter, 2000; Rogers, Andrews, Grasby, Brooks, & Robbins, 2000). For example, MacDonald et al. (2000) conducted a functional MRI (fMRI) study using a task-switching version of the Stroop task. They found that the DLPFC was specifically involved on trials requiring the implementation of cognitive control to overcome response conflict. That is, when the task required that participants to name the color of an incongruous pair, the DLPFC was preferentially activated. Further, the degree of activation of the DLPFC predicted the interference effect. More specifically, greater activation in this region was significantly correlated with a smaller interference effect. This suggests that the DLPFC may be a good candidate for a functional analogue of the PL cortex in terms of conflict resolution, enabling current task demands to influence selection of the appropriate response.

The DLPFC is also involved in attentional set-shifting which may also implicate the DLPFC in aspects of attentional processing attributed to the PL cortex (Birrell & Brown, 2000; Rogers et al., 2000; Sharpe & Killcross, 2014). Rogers et al. (2000) used positron emission topography (PET) to measure change in regional cerebral blood flow while subjects performed the WCST. The WCST is a task commonly used to measure executive function and requires subjects to solve a series of discriminations requiring shifting attention within and between different stimulus dimensions (i.e., intra- and extradimensional set shifts). The authors found that the pattern of activation in the DLPFC differed depending on the type of attentional shift that was required to complete the discrimination. The DLPFC was preferentially activated when participants were required to make an extradimensional set shift in which they had to shift attention away from a previously relevant stimulus dimension and toward a different stimulus set. This pattern was not seen with intradimensional set shifts or when participants were required to reverse their discriminations. Further, DLPFC function has also been causally shown to be involved in attentional set shifting in marmosets, where lesions encompassing the DLPFC produce a selective deficit in extradimensional set shifts (Dias, Robbins, & Roberts, 1996, 1997). This mirrors the pattern of deficits seen with animals with PL lesions (Birrell & Brown, 2000), which has subsequently been attributed to a specific deficit in downregulating attention toward the previously relevant stimulus dimension (Sharpe & Killcross, 2014). These similarities suggest that that the DLPFC may also play a similar role to the PL cortex in attention as well as implementing control over behavior in the Stroop task. Thinking about the DLPFC as a potential homologue for the PL cortex is also consistent with studies which have demonstrated that the PL cortex receives the same anatomical projections inherent in the primate DLPFC, particularly as defined by their thalamic afferents (Uylings et al., 2003). However, it is of course important to note that other researchers have shown that the prefrontal cortex of the rodent lacks critical features of that seen in humans and nonhuman primates (Öngür & Price, 2000).

On the other hand, the ventromedial prefrontal cortex (vmPFC) appears to be a neural locus for the execution of goal-directed actions in humans (Balleine & O'Doherty, 2010; de Wit, Corlett, Aitken, Dickinson, & Fletcher, 2009; Gläscher, Hampton, & O'Doherty, 2009). Using fMRI, Gläscher et al. (2009) demonstrated that the vmPFC was selectively involved when participants were required to choose between two instrumental actions associated with two different monetary rewards. Further, Tanaka, Balleine, and O'Doherty (2008) also showed that activity in this region was sensitive to the contingency between different actions and outcomes. Using a free-operant design, activity in the vmPFC was found to be selectively elevated when participants were performing on a schedule with a high contingency between the action and outcome relative to a low-contingency schedule. In contrast, research has suggested that habit learning is predominantly subserved by the lateral striatum (Tricomi, Balleine, & O'Doherty, 2009). Tricomi et al. (2009) trained participants on a task using a variable interval schedule for food reward and used fMRI to examine how neural activation may change as behavior becomes habitual. In the third session, participants were given an outcome devaluation test and shown to be performing habitually. During this session, activity in the lateral striatum was elevated compared to the first session where subjects were found to be responding in a goal-directed manner. These studies suggest that the vmPFC may be involved in allowing the value of rewards to influence action selection, whereas processing may become confined to subcortical structures as training progresses and dominated by habitual responding. Implicating both the DLPFC and vmPFC in functions attributed to only the PL cortex in rodents suggests, unsurprisingly, that the human PFC is more differentiated than that of the rodent.

Functional Homologue for the IL Cortex

There is a relative paucity of data regarding a homologue for the IL cortex in humans. This may be in part due to the traditional view of cognitive control that conceptualizes top-down influence over responding as allowing behavior to be governed by rules and task sets. However, as this review suggests, top-down control may also involve a competing process that opposes the influence of rules and task set to promote execution of more automatic responses. Nevertheless, research stemming from the aversive literature may suggest a potential neural locus for some of the processes which have been attributed to the IL cortex. As discussed in this review, the rodent IL cortex has been implicated in the consolidation of fear extinction and an inhibition of fear. Research from the cognitive neuroscience domain has suggested that the human subgenual vmPFC may be the region involved in the consolidation of extinction memory and inhibition of fear in humans (Hartley, Fischl, & Phelps, 2011; Nieuwenhuis & Takashima, 2011), prompting the authors to suggest that the human vmPFC may be a likely candidate for an analogue of the rodent IL cortex. Given the vmPFC has also been implicated in allowing the value of a goal to influence behavior (Balleine & O'Doherty, 2010; De Wit et al., 2009; Gläscher et al., 2009), it may be surprising that the same region may be involved in processes attributed to both the PL and IL cortex in rodents. Although there is some evidence that it may be different regions of the vmPFC that are involved in fear inhibition and goal-directed responding (Hartley et al., 2011), it may be that the functional distinctions observed in the rodent brain are not respected in the human brain. Rather, it appears as though the human brain is organized according to functional domains, where regions such as the vmPFC are involved in goal-directed behavior and the inhibition of fear responding, whereas the DLPFC appears to be involved in using rules and task sets to influence behavior. Thus, certain functions from the medial wall of the rodent brain may have migrated to the more lateral regions of the human PFC where they retain the functional divisions present in the rodent PL and IL cortices in a domain specific manner. In line with this, an analogue for the IL cortex in opposing the influence of rules and task-set over stimulus–response pathways may more likely be found in the DLPFC. Further research is needed to investigate the neural loci for specific functions inherent in the rodent medial PFC within distinct regions of the vmPFC and DLPFC of the human brain.

What About the Cingulate?

In rodents, the anterior cingulate cortex (ACC) is considered to be a critical region of the medial prefrontal cortex and shares extensive projections with both the PL and IL cortices (Öngür & Price, 2000; Uylings et al., 2003). In our hands, we have seen that selective lesions of the ACC in rodents produces a deficit in the detection of response conflict (Haddon & Killcross, 2006; Marquis et al., 2007). For example, lesions of the ACC produce a transient within-trial deficit in the rodent version of the Stroop task (Haddon & Killcross, 2006). Specifically, ACC-lesioned animals show impaired responding in the initial 10s period of incongruent test compounds. That is, when a conflict arises between two well-trained responses as incongruent stimuli are presented, animals without ACC function show a selective deficit at the beginning of cue presentation. This deficit subsides across the later 50s of the stimulus where ACC-lesioned animals show appropriate conflict resolution through use of available contextual cues. These data may suggest that lesions of the ACC produce a deficit in the initial detection of response conflict while not being necessary for the resolution of conflict itself. This is supported by research examining activity in the ACC in humans during the Stroop task (Carter et al., 1998; Cohen et al., 2004). Such studies have shown that activity in the human ACC is increased during presentation of incongruent compounds, when conflict is high. Activity in the ACC does not correlate with performance itself but rather the degree of conflict present, in contrast to activity in the DLPFC (MacDonald et al., 2000). This research suggests that the ACC may be involved in detecting response conflict to engage DLPFC/PL systems to resolve response present conflict and perform the contextually appropriate response.

Conclusions

Here we have proposed a framework that attempts to describe a role for the medial prefrontal cortex in rats that generalizes across both the appetitive and aversive domains. We have adopted a model that describes voluntary control of behavior in terms of hierarchal control (Cohen et al., 1990; Miller & Cohen, 2001; Cohen et al., 2004). Within this framework, the PFC maintains patterns of activity that represent current task demands which bias activity in basic stimulus–response pathways. We argue that the rodent mPFC contributes to the exertion of control over automatic behaviors in the manner described within this framework. Moreover, we believe that this region is capable of directing control according to a range of factors including contextual cues (extinction experiments), predictive history of a stimulus (i.e., attentional modulation; overshadowing and blocking experiments), and task-setting cues (Stroop-analogue experiments). In addition, rather than conceptualizing attention as a change in focus directed toward task-setting cues which subsequently influences activation in stimulus–response pathways (Cohen et al., 1990; Miller & Cohen, 2001), we conceptualize an attentional response as one of a host of responses that can be elicited by a stimulus and modulated by task. Thus, the predictive history of a stimulus can modulate activity in the corresponding stimulus–response pathway that produces an increase in an attentional response to that stimulus. Furthermore, the dichotomous nature of the interaction between the PL and IL cortices suggests behavior is controlled by two competing processes, one that prioritizes responding according to current circumstances and competing task demands, and one that opposes this influence by promoting performance of well-established behavioral and attentional responses. This framework has implications for loss of voluntary control of behavior relevant to models of drug abuse and neuropsychiatric disorders such as schizophrenia, long thought to reflect deficits in prefrontal control over appropriate behavior.

References

Balleine, B. W., & Dickinson, A. (1998). Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates. *Neuropharmacology, 37,* 407–419. http://dx.doi.org/10.1016/S0028-3908(98)00033-1

Balleine, B. W., & Killcross, S. (2006). Parallel incentive processing: An integrated view of amygdala function. *Trends in Neurosciences, 29,* 272–279. http://dx.doi.org/10.1016/j.tins.2006.03.002

Balleine, B. W., & O'Doherty, J. P. (2010). Human and rodent homologies in action control: Corticostriatal determinants of goal-directed and habitual action. *Neuropsychopharmacology, 35,* 48–69. http://dx.doi.org/10.1038/npp.2009.131

Birrell, J. M., & Brown, V. J. (2000). Medial frontal cortex mediates perceptual attentional set shifting in the rat. *The Journal of Neuroscience, 20,* 4320–4324. http://dx.doi.org/10.1523/JNEUROSCI.20-11-04320.2000

Bouton, M. E. (1993). Context, time, and memory retrieval in the interference paradigms of Pavlovian learning. *Psychological Bulletin, 114,* 80–99. http://dx.doi.org/10.1037/0033-2909.114.1.80

Bouton, M. E. (2004). Context and behavioral processes in extinction. *Learning & Memory, 11,* 485–494. http://dx.doi.org/10.1101/lm.78804

Braver, T. S., & Cohen, J. D. (2000). On the control of control: The role of dopamine in regulating prefrontal function and working memory. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 713–737). Cambridge, MA: MIT Press.

Burgos-Robles, A., Vidal-Gonzalez, I., & Quirk, G. J. (2009). Sustained conditioned responses in prelimbic prefrontal neurons are correlated with fear expression and extinction failure. *The Journal of Neuroscience, 29,* 8474–8482. http://dx.doi.org/10.1523/JNEUROSCI.0378-09.2009

Carter, C. S., Braver, T. S., Barch, D. M., Botvinick, M. M., Noll, D., & Cohen, J. D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science, 280,* 747–749. http://dx.doi.org/10.1126/science.280.5364.747

Chudasama, Y., & Muir, J. L. (2001). Visual attention in the rat: A role for the prelimbic cortex and thalamic nuclei? *Behavioral Neuroscience, 115,* 417–428. http://dx.doi.org/10.1037/0735-7044.115.2.417

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36,* 28–71. http://dx.doi.org/10.1006/cogp.1998.0681

Cohen, J. D., Aston-Jones, G., & Gilzenrat, M. S. (2004). A systems-level perspective on attention and cognitive control: guided activation, adaptive gating, conflict monitoring, and exploitation versus exploration. In M. I. Posner (Ed.), *Cognitive neuroscience of attention* (pp. 71–90). New York, NY: Guilford Press.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review, 97,* 332–361. http://dx.doi.org/10.1037/0033-295X.97.3.332

Colwill, R. M., & Rescorla, R. A. (1990). Evidence for the hierarchical structure of instrumental learning. *Animal Learning & Behavior, 18,* 71–82. http://dx.doi.org/10.3758/BF03205241

Corcoran, K. A., & Quirk, G. J. (2007). Activity in prelimbic cortex is necessary for the expression of learned, but not innate, fears. *The Journal of Neuroscience, 27,* 840–844. http://dx.doi.org/10.1523/JNEUROSCI.5327-06.2007

Coutureau, E., & Killcross, S. (2003). Inactivation of the infralimbic prefrontal cortex reinstates goal-directed responding in overtrained rats. *Behavioural Brain Research, 146,* 167–174. http://dx.doi.org/10.1016/j.bbr.2003.09.025

Delamater, A. R. (1996). Effects of several extinction treatments upon the integrity of Pavlovian stimulus-outcome associations. *Animal Learning & Behavior, 24,* 437–449. http://dx.doi.org/10.3758/BF03199015

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://dx.doi.org/10.1016/j.nlm.2013.09.016

de Wit, S., Corlett, P. R., Aitken, M. R., Dickinson, A., & Fletcher, P. C. (2009). Differential engagement of the ventromedial prefrontal cortex by goal-directed and habitual behavior toward food pictures in humans. *The Journal of Neuroscience, 29,* 11330–11338. http://dx.doi.org/10.1523/JNEUROSCI.1639-09.2009

de Wit, S., Kosaki, Y., Balleine, B. W., & Dickinson, A. (2006). Dorsomedial prefrontal cortex resolves response conflict in rats. *The Journal of Neuroscience, 26,* 5224–5229. http://dx.doi.org/10.1523/JNEUROSCI.5175-05.2006

Dias, R., Robbins, T. W., & Roberts, A. C. (1996). Primate analogue of the Wisconsin Card Sorting Test: Effects of excitotoxic lesions of the prefrontal cortex in the marmoset. *Behavioral Neuroscience, 110,* 872–886. http://dx.doi.org/10.1037/0735-7044.110.5.872

Dias, R., Robbins, T. W., & Roberts, A. C. (1997). Dissociable forms of inhibitory control within prefrontal cortex with an analog of the Wisconsin Card Sort Test: Restriction to novel situations and independence from "on-line" processing. *The Journal of Neuroscience, 17,* 9285–9297. http://dx.doi.org/10.1523/JNEUROSCI.17-23-09285.1997

Dwyer, D. M., Dunn, M. J., Rhodes, S. E., & Killcross, A. S. (2010). Lesions of the prelimbic prefrontal cortex prevent response conflict produced by action-outcome associations. *The Quarterly Journal of Experimental Psychology, 63,* 417–424. http://dx.doi.org/10.1080/17470210903411049

Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience, 8,* 1784–1790. http://dx.doi.org/10.1038/nn1594

Floresco, S. B., Block, A. E., & 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, 190,* 85–96. http://dx.doi.org/10.1016/j.bbr.2008.02.008

George, D. N., Duffaud, A. M., & Killcross, S. (2010). Neural correlates of attentional set. In C. J. Mitchell & M. E. Le Pelley (Eds.), *Attention and associative learning: From brain to behaviour* (pp. 351–383). New York, NY: Oxford University Press.

Gisquet-Verrier, P., & Delatour, B. (2006). The role of the rat prelimbic/infralimbic cortex in working memory: Not involved in the short-term maintenance but in monitoring and processing functions. *Neuroscience, 141,* 585–596. http://dx.doi.org/10.1016/j.neuroscience.2006.04.009

Gläscher, J., Hampton, A. N., & O'Doherty, J. P. (2009). Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. *Cerebral Cortex, 19,* 483–495. http://dx.doi.org/10.1093/cercor/bhn098

Goldman-Rakic, P., O'Scalaidhe, S., & Chafee, M. (2000). Domain specificity in cognitive systems. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 733–742). Cambridge, MA: MIT Press.

Haddon, J. E., George, D. N., & Killcross, S. (2008). Contextual control of biconditional task performance: Evidence for cue and response competition in rats. *The Quarterly Journal of Experimental Psychology, 61,* 1307–1320. http://dx.doi.org/10.1080/17470210701515819

Haddon, J. E., & Killcross, S. (2006). Prefrontal cortex lesions disrupt the contextual control of response conflict. *The Journal of Neuroscience, 26,* 2933–2940. http://dx.doi.org/10.1523/JNEUROSCI.3243-05.2006

Haddon, J. E., & Killcross, S. (2011). Inactivation of the infralimbic prefrontal cortex in rats reduces the influence of inappropriate habitual responding in a response-conflict task. *Neuroscience, 199,* 205–212. http://dx.doi.org/10.1016/j.neuroscience.2011.09.065

Hall, G., & Channell, S. (1985). Differential effects of contextual change on latent inhibition and on the habituation of an orienting response. *Journal of Experimental Psychology: Animal Behavior Processes, 11,* 470–481. http://dx.doi.org/10.1037/0097-7403.11.3.470

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,* 174–185. http://dx.doi.org/10.1037/0097-7403.26.2.174

Hartley, C. A., Fischl, B., & Phelps, E. A. (2011). Brain structure correlates of individual differences in the acquisition and inhibition of conditioned fear. *Cerebral Cortex, 21,* 1954–1962. http://dx.doi.org/10.1093/cercor/bhq253

Holland, P. C. (1991). Acquisition and transfer of occasion setting in operant feature positive and feature negative discriminations. *Learning and Motivation, 22,* 366–387. http://dx.doi.org/10.1016/0023-9690(91)90002-P

Holson, R. R. (1986). Mesial prefrontal cortical lesions and timidity in rats: I. Reactivity to aversive stimuli. *Physiology & Behavior, 37,* 221–230. http://dx.doi.org/10.1016/0031-9384(86)90224-6

Kesner, R. P. (2000). Subregional analysis of mnemonic functions of the prefrontal cortex in the rat. *Psychobiology, 28,* 219–228.

Killcross, S., & Coutureau, E. (2003). Coordination of actions and habits in the medial prefrontal cortex of rats. *Cerebral Cortex, 13,* 400–408. http://dx.doi.org/10.1093/cercor/13.4.400

Lacroix, L., Spinelli, S., Heidbreder, C. A., & Feldon, J. (2000). Differential role of the medial and lateral prefrontal cortices in fear and anxiety. *Behavioral Neuroscience, 114,* 1119–1130. http://dx.doi.org/10.1037/0735-7044.114.6.1119

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

Le Pelley, M. E. (2004). The role of associative history in models of associative learning: A selective review and a hybrid model. *The Quarterly Journal of Experimental Psychology B: Comparative and Physiological Psychology, 57,* 193–243. http://dx.doi.org/10.1080/02724990344000141

Le Pelley, M. E., Mitchell, C. J., Beesley, T., George, D. N., & Wills, A. J. (2016). Attention and associative learning in humans: An integrative review. *Psychological Bulletin, 142,* 1111–1140. http://dx.doi.org/10.1037/bul0000064

Le Pelley, M. E., Mitchell, C. J., & Johnson, A. M. (2013). Outcome value influences attentional biases in human associative learning: Dissociable effects of training and instruction. *Journal of Experimental Psychology: Animal Behavior Processes, 39,* 39–55. http://dx.doi.org/10.1037/a0031230

MacDonald, A. W., III, Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science, 288,* 1835–1838. http://dx.doi.org/10.1126/science.288.5472.1835

Mackintosh, N. J. (1973). Stimulus selection: Learning to ignore stimuli that predict no change in reinforcement. In R. A. Hinde & J. Stevenson-Hinde (Eds.), *Constraints on learning: Limitations and predispositions*. Oxford, England: Academic Press.

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82,* 276–298. http://dx.doi.org/10.1037/h0076778

Mackintosh, N. J. (1976). Overshadowing and stimulus intensity. *Animal Learning & Behavior, 4,* 186–192. http://dx.doi.org/10.3758/BF03214033

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 126–135. http://dx.doi.org/10.1037/0278-7393.14.1.126

MacLeod, J. E., & Bucci, D. J. (2010). Contributions of the subregions of the medial prefrontal cortex to negative occasion setting. *Behavioral Neuroscience, 124,* 321–328. http://dx.doi.org/10.1037/a0019344

Maes, J. H., & Vossen, J. M. (1993). Competition for associative strength between a punctate signal and contextual stimuli: Effect of signal preexposure versus context preexposure. *Behavioural Processes, 30,* 29–46. http://dx.doi.org/10.1016/0376-6357(93)90010-O

Marquis, J.-P., Killcross, S., & Haddon, J. E. (2007). Inactivation of the prelimbic, but not infralimbic, prefrontal cortex impairs the contextual control of response conflict in rats. *European Journal of Neuroscience, 25,* 559–566. http://dx.doi.org/10.1111/j.1460-9568.2006.05295.x

Milad, M. R., & Quirk, G. J. (2002). Neurons in medial prefrontal cortex signal memory for fear extinction. *Nature, 420,* 70–74. http://dx.doi.org/10.1038/nature01138

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24,* 167–202. http://dx.doi.org/10.1146/annurev.neuro.24.1.167

Moorman, D. E., & Aston-Jones, G. (2015). Prefrontal neurons encode context-based response execution and inhibition in reward seeking and extinction. *Proceedings of the National Academy of Sciences of the United States of America, 112,* 9472–9477. http://dx.doi.org/10.1073/pnas.1507611112

Morgan, M. A., & LeDoux, J. E. (1995). Differential contribution of dorsal and ventral medial prefrontal cortex to the acquisition and extinction of conditioned fear in rats. *Behavioral Neuroscience, 109,* 681–688. http://dx.doi.org/10.1037/0735-7044.109.4.681

Muir, J. L., Everitt, B. J., & Robbins, T. W. (1996). The cerebral cortex of the rat and visual attentional function: Dissociable effects of mediofrontal, cingulate, anterior dorsolateral, and parietal cortex lesions on a five-choice serial reaction time task. *Cerebral Cortex, 6,* 470–481. http://dx.doi.org/10.1093/cercor/6.3.470

Nieuwenhuis, I. L., & Takashima, A. (2011). The role of the ventromedial prefrontal cortex in memory consolidation. *Behavioural Brain Research, 218,* 325–334. http://dx.doi.org/10.1016/j.bbr.2010.12.009

Norman, D., & Shallice, T. (1980). *Attention to action: Willed and automatic control of behaviour* (Report 99). San Diego, CA: Center for Human Information Processing, University of California.

Öngür, D., & Price, J. L. (2000). The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cerebral Cortex, 10,* 206–219. http://dx.doi.org/10.1093/cercor/10.3.206

Orsini, C. A., Kim, J. H., Knapska, E., & Maren, S. (2011). Hippocampal and prefrontal projections to the basal amygdala mediate contextual regulation of fear after extinction. *The Journal of Neuroscience, 31,* 17269–17277. http://dx.doi.org/10.1523/JNEUROSCI.4095-11.2011

Ostlund, S. B., & Balleine, B. W. (2005). Lesions of medial prefrontal cortex disrupt the acquisition but not the expression of goal-directed learning. *The Journal of Neuroscience, 25,* 7763–7770. http://dx.doi.org/10.1523/JNEUROSCI.1921-05.2005

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87,* 532–552. http://dx.doi.org/10.1037/0033-295X.87.6.532

Pearce, J. M., & Mackintosh, N. J. (2010). Two theories of attention: A review and a possible integration. In C. J. Mitchell & M. E. Le Pelley (Eds.), *Attention and associative learning: From brain to behaviour* (pp. 11–39). Oxford, UK: Oxford University Press.

Quirk, G. J., Russo, G. K., Barron, J. L., & Lebron, K. (2000). The role of ventromedial prefrontal cortex in the recovery of extinguished fear. *The Journal of Neuroscience, 20,* 6225–6231. http://dx.doi.org/10.1523/JNEUROSCI.20-16-06225.2000

Ragozzino, M. E., Wilcox, C., Raso, M., & Kesner, R. P. (1999). Involvement of rodent prefrontal cortex subregions in strategy switching. *Behavioral Neuroscience, 113,* 32–41. http://dx.doi.org/10.1037/0735-7044.113.1.32

Rescorla, R. A. (1984). Signaling intertrial shocks attenuates their negative effect on conditioned suppression. *Bulletin of the Psychonomic Society, 22,* 225–228. http://dx.doi.org/10.3758/BF03333812

Rescorla, R. A. (1997). Response inhibition in extinction. *The Quarterly Journal of Experimental Psychology: Section B, 50,* 238–252.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (Vol. 2, pp. 64–99). New York, NY: Appleton-Century-Crofts.

Rhodes, S. E. V., & Killcross, A. S. (2007). Lesions of rat infralimbic cortex enhance renewal of extinguished appetitive Pavlovian responding. *European Journal of Neuroscience, 25,* 2498–2503. http://dx.doi.org/10.1111/j.1460-9568.2007.05486.x

Rhodes, S. E. V., & Killcross, S. (2004). Lesions of rat infralimbic cortex enhance recovery and reinstatement of an appetitive Pavlovian response. *Learning & Memory, 11,* 611–616. http://dx.doi.org/10.1101/lm.79704

Rogers, R. D., Andrews, T. C., Grasby, P. M., Brooks, D. J., & Robbins, T. W. (2000). Contrasting cortical and subcortical activations produced by attentional-set shifting and reversal learning in humans. *Journal of Cognitive Neuroscience, 12,* 142–162. http://dx.doi.org/10.1162/089892900561931

Santini, E., Quirk, G. J., & Porter, J. T. (2008). Fear conditioning and extinction differentially modify the intrinsic excitability of infralimbic neurons. *Journal of Neuroscience, 28,* 4028–4036. http://dx.doi.org/10.1523/JNEUROSCI.2623-07.2008

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences, 298,* 199–209. http://dx.doi.org/10.1098/rstb.1982.0082

Sharpe, M. J., & Killcross, S. (2014). The prelimbic cortex contributes to the down-regulation of attention toward redundant cues. *Cerebral Cortex, 24,* 1066–1074. http://dx.doi.org/10.1093/cercor/bhs393

Sharpe, M. J., & Killcross, S. (2015a). The prelimbic cortex directs attention toward predictive cues during fear learning. *Learning & Memory, 22,* 289–293. http://dx.doi.org/10.1101/lm.038273.115

Sharpe, M. J., & Killcross, S. (2015b). The prelimbic cortex uses higherorder cues to modulate both the acquisition and expression of conditioned fear. *Frontiers in Systems Neuroscience, 8,* 235. http://dx.doi.org/10.3389/fnsys.2014.00235

Sharpe, M. J., & Killcross, S. (2015c). The prelimbic cortex uses contextual cues to modulate responding towards predictive stimuli during fear renewal. *Neurobiology of Learning and Memory, 118,* 20–29. http://dx.doi.org/10.1016/j.nlm.2014.11.005

Smith, K. S., Virkud, A., Deisseroth, K., & Graybiel, A. M. (2012). Reversible online control of habitual behavior by optogenetic perturbation of medial prefrontal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 109,* 18932–18937. http://dx.doi.org/10.1073/pnas.1216264109

Sotres-Bayon, F., Sierra-Mercado, D., Pardilla-Delgado, E., & Quirk, G. J. (2012). Gating of fear in prelimbic cortex by hippocampal and amygdala inputs. *Neuron, 76,* 804–812. http://dx.doi.org/10.1016/j.neuron.2012.09.028

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643–662. http://dx.doi.org/10.1037/h0054651

Tanaka, S. C., Balleine, B. W., & O'Doherty, J. P. (2008). Calculating consequences: Brain systems that encode the causal effects of actions. *The Journal of Neuroscience, 28,* 6750–6755. http://dx.doi.org/10.1523/JNEUROSCI.1808-08.2008

Trask, S., Shipman, M. L., Green, J. T., & Bouton, M. E. (2017). Inactivation of the prelimbic cortex attenuates context-dependent operant responding. *The Journal of Neuroscience, 37,* 2317–2324. http://dx.doi.org/10.1523/JNEUROSCI.3361-16.2017

Tricomi, E., Balleine, B. W., & O'Doherty, J. P. (2009). A specific role for posterior dorsolateral striatum in human habit learning. *European Journal of Neuroscience, 29,* 2225–2232. http://dx.doi.org/10.1111/j.1460-9568.2009.06796.x

Uylings, H. B. M., Groenewegen, H. J., & Kolb, B. (2003). Do rats have a prefrontal cortex? *Behavioural Brain Research, 146*(1–2), 3–17. http://dx.doi.org/10.1016/j.bbr.2003.09.028

Verbruggen, F., McLaren, I. P., & Chambers, C. D. (2014). Banishing the control homunculi in studies of action control and behavior change. *Perspectives on Psychological Science, 9,* 497–524.

Vidal-Gonzalez, I., Vidal-Gonzalez, B., Rauch, S. L., & Quirk, G. J. (2006). Microstimulation reveals opposing influences of prelimbic and infralimbic cortex on the expression of conditioned fear. *Learning & Memory, 13,* 728–733. http://dx.doi.org/10.1101/lm.306106

Willcocks, A. L., & McNally, G. P. (2013). The role of medial prefrontal cortex in extinction and reinstatement of alcohol-seeking in rats. *European Journal of Neuroscience, 37,* 259–268. http://dx.doi.org/10.1111/ejn.12031

Zelikowsky, M., Bissiere, S., Hast, T. A., Bennett, R. Z., Abdipranoto, A., Vissel, B., & Fanselow, M. S. (2013). Prefrontal microcircuit underlies contextual learning after hippocampal loss. *Proceedings of the National Academy of Sciences of the United States of America, 110,* 9938–9943. http://dx.doi.org/10.1073/pnas.1301691110

Received March 26, 2016
Revision received May 4, 2018
Accepted May 7, 2018 ■