frontiers in Neuroscience

MINI REVIEW
published: 27 March 2018
doi: 10.3389/fnins.2018.00182

Solving the Credit Assignment Problem With the Prefrontal Cortex

*Alexandra Stolyarova**

Department of Psychology, University of California, Los Angeles, Los Angeles, CA, United States

In naturalistic multi-cue and multi-step learning tasks, where outcomes of behavior are delayed in time, discovering which choices are responsible for rewards can present a challenge, known as the *credit assignment problem*. In this review, I summarize recent work that highlighted a critical role for the prefrontal cortex (PFC) in assigning credit where it is due in tasks where only a few of the multitude of cues or choices are relevant to the final outcome of behavior. Collectively, these investigations have provided compelling support for specialized roles of the orbitofrontal (OFC), anterior cingulate (ACC), and dorsolateral prefrontal (dlPFC) cortices in contingent learning. However, recent work has similarly revealed shared contributions and emphasized rich and heterogeneous response properties of neurons in these brain regions. Such functional overlap is not surprising given the complexity of reciprocal projections spanning the PFC. In the concluding section, I overview the evidence suggesting that the OFC, ACC and dlPFC communicate extensively, sharing the information about presented options, executed decisions and received rewards, which enables them to assign credit for outcomes to choices on which they are contingent. This account suggests that lesion or inactivation/inhibition experiments targeting a localized PFC subregion will be insufficient to gain a fine-grained understanding of credit assignment during learning and instead poses refined questions for future research, shifting the focus from focal manipulations to experimental techniques targeting cortico-cortical projections.

Keywords: orbitofrontal, dorsolateral prefrontal, anterior cingulate, learning, reward, reinforcement, plasticity, behavioral flexibility

OPEN ACCESS

Edited by:
Aldo Genovesio,
Sapienza Università di Roma, Italy

Reviewed by:
Mehdi Khamassi,
UMR7222 Institut des Systèmes Intelligents et Robotiques (ISIR), France
Sabrina Ravel,
Centre National de la Recherche Scientifique (CNRS), France

****Correspondence:***
Alexandra Stolyarova
astolyarova@psych.ucla.edu

Specialty section:
This article was submitted to Decision Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *27 September 2017*
Accepted: *06 March 2018*
Published: *27 March 2018*

Citation:
Stolyarova A (2018) Solving the Credit Assignment Problem With the Prefrontal Cortex. Front. Neurosci. 12:182. doi: 10.3389/fnins.2018.00182

INTRODUCTION

When an animal is introduced to an unfamiliar environment, it will explore the surroundings randomly until an unexpected reward is encountered. Reinforced by this experience, the animal will gradually learn to repeat those actions that produced the desired outcome. The work conducted in the past several decades has contributed a detailed understanding of the psychological and neural mechanisms that support such reinforcement-driven learning (Schultz and Dickinson, 2000; Schultz, 2004; Niv, 2009). It is now broadly accepted that dopamine (DA) signaling conveys prediction errors, or the degree of surprise brought about by unexpected rewards, and interacts with cortical and basal ganglia circuits to selectively reinforce the advantageous choices (Schultz, 1998a,b; Schultz and Dickinson, 2000; Niv, 2009). Yet, in naturalistic settings, where rewards are delayed in time, and where multiple cues are encountered, or where several decisions are made before the outcomes of behavior are revealed, discovering which choices are responsible for rewards can present a challenge, known as the *credit assignment problem* (Mackintosh, 1975; Rothkopf and Ballard, 2010).

In most everyday situations, the rewards are not immediate consequences of behavior, but instead appear after substantial delays. To influence future choices, the teaching signal conveyed by DA release needs to reinforce synaptic events occurring on a millisecond timescale, frequently seconds before the outcomes of decisions are revealed (Izhikevich, 2007; Fisher et al., 2017). This apparent difficulty in linking preceding behaviors caused by transient neuronal activity to a delayed feedback has been termed the *distal reward* or *temporal* credit assignment problem (Hull, 1943; Barto et al., 1983; Sutton and Barto, 1998; Dayan and Abbott, 2001; Wörgötter and Porr, 2005). Credit for the reward delayed by several seconds can frequently be assigned by establishing an eligibility trace, a molecular memory of the recent neuronal activity, allowing modification of synaptic connections that participated in the behavior (Pan et al., 2005; Fisher et al., 2017). On longer timescales, or when multiple actions need to be performed sequentially to reach a final goal, intermediate steps themselves can acquire motivational significance and subsequently reinforce preceding decisions, such as in temporal-difference (TD) learning models (Sutton and Barto, 1998).

Several excellent reviews have summarized the accumulated knowledge on mechanisms that link choices and their outcomes through time, highlighting the advantages of eligibility traces and TD models (Wörgötter and Porr, 2005; Barto, 2007; Niv, 2009; Walsh and Anderson, 2014). Yet these solutions to the distal reward problem can impede learning in multi-choice tasks, or when an animal is presented with many irrelevant stimuli prior to or during the delay. Here, I only briefly overview the work on the distal reward problem to highlight potential complications that can arise in credit assignment based on eligibility traces when learning in multi-cue environments. Instead, I focus on the *structural* (or *spatial*) credit assignment problem, requiring animals to select and learn about the most meaningful features in the environment and ignore irrelevant distractors. Collectively, the reviewed evidence highlights a critical role for the prefrontal cortex (PFC) in such contingent learning.

Recent studies have provided compelling support for specialized functions of the orbitofrontal (OFC) and dorsolateral prefrontal (dlPFC) cortices in credit assignment in multi-cue tasks, with fewer experiments targeting the anterior cingulate cortex (ACC). For example, it has seen suggested that the dlPFC aids reinforcement-driven learning by directing attention to task-relevant cues (Niv et al., 2015), the OFC assigns credit for rewards based on the causal relationship between trial outcomes and choices (Jocham et al., 2016; Noonan et al., 2017), whereas the ACC contributes to unlearning of action-outcome associations when the rewards are available for free (Jackson et al., 2016). However, this work has similarly revealed shared contributions and emphasized rich and heterogeneous response properties of neurons in the PFC, with different subregions monitoring and integrating the information about the task (i.e., current context, available options, anticipated rewards, as well as delay and effort costs) at variable times within a trial (upon stimulus presentation, action selection, outcome anticipation, and feedback monitoring; ex., Hunt et al., 2015; Khamassi et al., 2015). In the concluding section, I overview the evidence suggesting that contingent learning in multi-cue environments relies on dynamic cortico-cortical interactions during decision making and outcome valuation.

SOLVING THE TEMPORAL CREDIT ASSIGNMENT PROBLEM

When outcomes follow choices after short delays (**Figure 1A**), the credit for distal rewards can frequently be assigned by establishing an eligibility trace, a sustained memory of the recent activity that renders synaptic connections malleable to modification over several seconds. Eligibility traces can persist as elevated levels of calcium in dendritic spines of post-synaptic neurons (Kötter and Wickens, 1995) or as a sustained neuronal activity throughout the delay period (Curtis and Lee, 2010) to allow for synaptic changes in response to reward signals. Furthermore, spike-timing dependent plasticity can be influenced by neuromodulator input (Izhikevich, 2007; Abraham, 2008; Fisher et al., 2017). For example, the magnitude of short-term plasticity can be modulated by DA, acetylcholine and noradrenaline, which may even revert the sign of the synaptic change (Matsuda et al., 2006; Izhikevich, 2007; Seol et al., 2007; Abraham, 2008; Zhang et al., 2009). Sustained neural activity has been observed in the PFC and striatum (Jog et al., 1999; Pasupathy and Miller, 2005; Histed et al., 2009; Kim et al., 2009, 2013; Seo et al., 2012; Her et al., 2016), as well as the sensory cortices after experience with consistent pairings between the stimuli and outcomes separated by predictable delays (Shuler and Bear, 2006).

On extended timescales, when multiple actions need to be performed sequentially to reach a final goal, the distal reward problem can be solved by assigning motivational significance to intermediate choices that can subsequently reinforce preceding decisions, such as in TD learning models (Montague et al., 1996; Sutton and Barto, 1998; Barto, 2007). Assigning values to these intervening steps according to expected future rewards allows to break complex temporal credit assignment problems into smaller and easier tasks. There is ample evidence for TD learning in humans and other animals that on the neural level is supported by transfer of DA responses from the time of reward delivery to preceding cues and actions (Montague et al., 1996; Schultz, 1998a,b; Walsh and Anderson, 2014).

Both TD learning and eligibility traces offer elegant solutions to the distal reward problem, and models based on cooperation between these two mechanisms can predict animal behavior as well as neuronal responses to rewards and predictive stimuli (Pan et al., 2005; Bogacz et al., 2007). Yet assigning credit based on eligibility traces can be suboptimal when an animal interacts with many irrelevant stimuli prior to or during the delay (**Figure 1B**). Under such conditions sensory areas remain responsive to distracting stimuli and the arrival of non-specific reward signals can reinforce intervening cues that did not meaningfully contribute, but occurred close, to the outcome of behavior (FitzGerald et al., 2013; Xu, 2017).

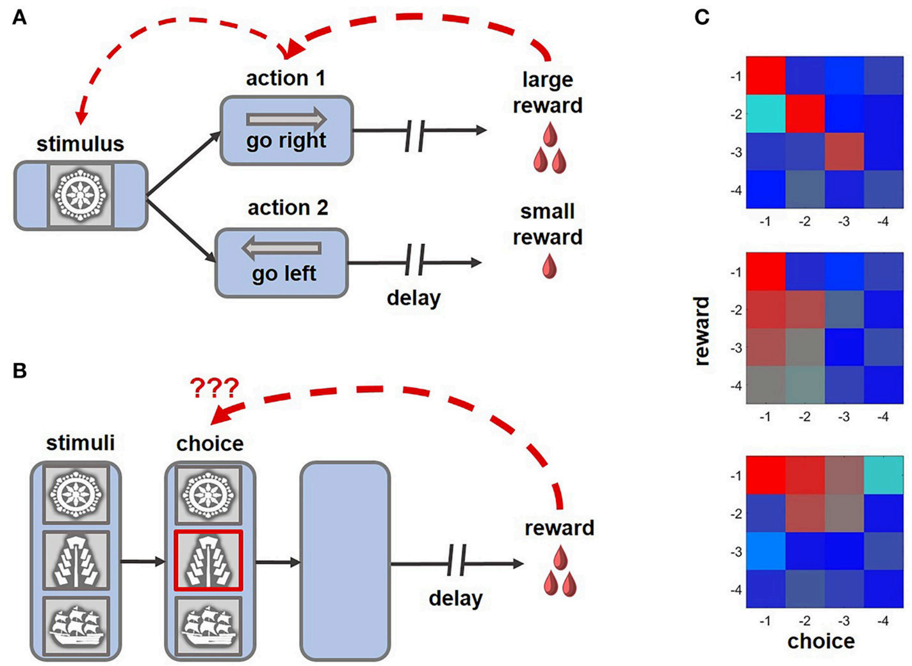

FIGURE 1 | Example tasks highlighting the challenge of credit assignment and learning strategies enabling animals to solve this problem. **(A)** An example of a distal reward task that can be successfully learned with eligibility traces and TD rules, where intermediate choices can acquire motivational significance and subsequently reinforce preceding decisions (ex., Pasupathy and Miller, 2005; Histed et al., 2009). **(B)** In this version of the task, multiple cues are present at the time of choice, only one of which is meaningful for obtaining rewards. After a brief presentation, the stimuli disappear, requiring an animal to solve a complex structural and temporal credit assignment problem (ex., Noonan et al., 2010, 2017; Niv et al., 2015; Asaad et al., 2017; while the schematic of the task captures the challenge of credit assignment, note that in some experimental variants of the behavioral paradigm stimuli disappeared before an animal revealed its choice, whereas in others the cues remained on the screen until the trial outcome was revealed). Under such conditions, learning based on eligibility traces is suboptimal, as non-specific reward signals can reinforce visual cues that did not meaningfully contribute, but occurred close, to beneficial outcomes of behavior. **(C)** On reward tasks, similar to the one shown in **(B)**, the impact of previous decisions and associated rewards on current behavior can be assessed by performing regression analyses (Jocham et al., 2016; Noonan et al., 2017). Here, the color of each cell in a matrix represents the magnitude of the effect of short-term choice and outcome histories, up to 4 trials into the past (red-strong influence; blue-weak influence on the current decision). Top: an animal learning based on the causal relationship between outcomes and choices (i.e., contingent learning). Middle: each choice is reinforced by a combined history of rewards (i.e., decisions are repeated if beneficial outcomes occur frequently). Bottom: the influence of recent rewards spreads to unrelated choices.

THE ROLE OF THE PFC IN STRUCTURAL CREDIT ASSIGNMENT

Several recent studies have investigated the neural mechanisms of appropriate credit assignment in challenging tasks where only a few of the multitude of cues predict rewards reliably. Collectively, this work has provided compelling support for causal contributions of the PFC to structural credit assignment. For example, Asaad et al. (2017) examined the activity of neurons in monkey dlPFC while subjects were performing a delayed learning task. The arrangement of the stimuli varied randomly between trials and within each block either the spatial location or stimulus identity was relevant for solving the task. The monkeys' goal was to learn by trial-and-error to select one of the four options that led to rewards according to current rules. When stimulus identity was relevant for solving the task, neural activity in the dlPFC at the time of feedback reflected both the relevant cue (regardless of its spatial location) and the trial outcome, thus integrating the information necessary for credit assignment. Such responses were strategy-selective: these neurons did not encode cue identity at the time of feedback when it was not necessary for learning in the spatial location task, in which making a saccade to the same position on the screen was reinforced within a block of trials. Previous research has similarly indicated that neurons in the dlPFC respond selectively to behaviorally-relevant and attended stimuli (Lebedev et al., 2004; Markowitz et al., 2015) and integrate information about prediction errors, choice values as well as outcome uncertainty prior to trial feedback (Khamassi et al., 2015).

The activity within the dlPFC has been linked to structural credit assignment through selective attention and representational learning (Niv et al., 2015). Under conditions of reward uncertainty and unknown relevant task features, human participants opt for computational efficiency and engage in a serial-hypothesis-testing strategy (Wilson and Niv, 2011), selecting one cue and its anticipated outcome as the main focus of their behavior, and updating the expectations associated exclusively with that choice upon feedback receipt (Akaishi et al., 2016). Niv and colleagues tested participant on a three-armed bandit task, where relevant stimulus dimensions (i.e., shape, color or texture) predicting outcome probabilities changed between block of trials (Niv et al., 2015). In such multidimensional environment, reinforcement-driven learning was aided by attentional control mechanisms that engaged the dlPFC, intraparietal cortex, and precuneus.

In many tasks, the credit for outcomes can be assigned according to different rules: based on the causal relationship between rewards and choices (i.e., contingent learning), their temporal proximity (i.e., when the reward is received shortly after a response), or their statistical relationship (when an action has been executed frequently before beneficial outcomes; Jocham et al., 2016; **Figure 1C**). The analyses presented in papers discussed above did not allow for the dissociation between these alternative strategies of credit assignment. By testing human participants on a task with continuous stimulus presentation, instead of a typical trial-by-trial structure, Jocham et al. (2016) demonstrated that the tendency to repeat choices that were immediately followed by rewards and causal learning operate in parallel. In this experiment, activity within another subregion of the PFC, the OFC, was associated with contingent learning. Complementary work in monkeys revealed that the OFC contributes causally to credit assignment (Noonan et al., 2010): animals with OFC lesions were unable to associate a reward with the choice on which it was contingent and instead relied on temporal and statistical learning rules. In another recent paper, Noonan and colleagues (2017) extended these observations to humans, demonstrating causal contributions of the OFC to credit assignment across species. The participants were tested on a three-choice probabilistic learning task. The three options were presented simultaneously and maintained on the screen until the outcome of a decision was revealed, thus requiring participants to ignore irrelevant distractors. Notably, only patients with lateral OFC lesions displayed any difficulty in learning the task, whereas damage to the medial OFC or dorsomedial PFC preserved contingent learning mechanisms. However, it is presently unknown whether lesions to the dlPFC or ACC affect such causal learning.

In another test of credit assignment in learning, contingency degradation, the subjects are required to track causal relationships between the stimuli or actions and rewards. During contingency degradation sessions, the animals are still reinforced for responses, but rewards are also available for free. After experiencing non-contingent rewards, control subjects reliably decrease their choices of the stimuli. However, lesions to both the ACC and OFC inhibit contingency degradation (Jackson et al., 2016). Taken together, these observations demonstrate causal contributions of the PFC to appropriate credit assignment in multi-cue environments.

COOPERATION BETWEEN PFC SUBREGIONS SUPPORTS CONTINGENT LEARNING IN MULTI-CUE TASKS

Despite the segregation of temporal and structural aspects of credit assignment in earlier sections of this review, in naturalistic settings the brains frequently need to tackle both problems simultaneously. Here, I overview the evidence favoring a network perspective, suggesting that dynamic cortico-cortical interactions during decision making and outcome valuation enable adaptive solutions to complex spatio-temporal credit assignment problems. It has been previously suggested that feedback projections from cortical areas occupying higher levels of processing hierarchy, including the PFC, can aid in attribution of outcomes to individual decisions by implementing attention-gated reinforcement learning (Roelfsema and van Ooyen, 2005). Similarly, recent theoretical work has shown that even complex multi-cue and multi-step problems can be solved by an extended cascade model of synaptic memory traces, in which the plasticity is modulated not only by the activity within a population of neurons, but also by feedback about executed decisions and resulting rewards (Urbanczik and Senn, 2009; Friedrich et al., 2010, 2011). Contingent learning, according to these models, can be supported by the communication between neurons encoding available options, committed choices and outcomes of behavior during decision making and feedback monitoring. For example, at the time of outcome valuation, information about recent choices can be maintained as a memory trace in the neuronal population involved in action selection or conveyed by an efference copy from an interconnected brain region (Curtis and Lee, 2010; Khamassi et al., 2011, 2015). Similarly, reinforcement feedback is likely communicated as a global reward signal (ex., DA release) as well as projections from neural populations engaged in performance monitoring, such as those within the ACC (Friedrich et al., 2010; Khamassi et al., 2011). The complexity of reciprocal and recurrent projections spanning the PFC (Barbas and Pandya, 1989; Felleman and Van Essen, 1991; Elston, 2000) may enable this network to implement such learning rules, integrating the information about the task, executed decisions and performance feedback.

In many everyday decisions, the options are compared across multiple features simultaneously (ex., by considering current context, needs, available reward types, as well as delay and effort costs). Neurons in different subregions of the PFC exhibit rich response properties, signaling these features of the task at various time epochs within a trial. For example, reward selectivity in response to predictive stimuli emerges earlier in the OFC and may then be passed to the dlPFC that encodes both the expected outcome and the upcoming choice (Wallis and Miller, 2003). Similarly, on trials where options are compared based on delays to rewards, choices are dependent on interactions between the OFC and dlPFC (Hunt et al., 2015). Conversely, when effort costs are more meaningful for decisions, it is the ACC that influences choice-related activity in the dlPFC (Hunt et al., 2015). The OFC is required not only for the evaluation of stimuli, but also more complex abstract rules, based on rewards they predict (Buckley et al., 2009). While both the OFC and dlPFC encode abstract strategies (ex., persisting with recent choices or shifting to a new response), such signals appear earlier in the OFC and may be subsequently conveyed to the dlPFC where they are combined with upcoming response (i.e., left vs. right saccade) encoding (Tsujimoto et al., 2011). Therefore, the OFC may be the first PFC subregion to encode task rules and/or potential rewards predicted by sensory cues; via cortico-cortical projections, this information may be subsequently communicated to the dlPFC or ACC (Kennerley et al., 2009; Hayden and Platt, 2010) to drive strategy-sensitive response planning.

The behavioral strategy that the animal follows is influenced by recent reward history (Cohen et al., 2007; Pearson et al., 2009). If its choices are reinforced frequently, the animal will make similar decisions in the future (i.e., exploit its current knowledge). Conversely, unexpected omission of expected rewards can signal a need for novel behaviors (i.e., exploration). Neurons in the dlPFC carry representations of planned as well as previous choices, anticipate outcomes, and jointly encode the current decisions and their consequences following feedback (Seo and Lee, 2007; Seo et al., 2007; Tsujimoto et al., 2009; Asaad et al., 2017). Similarly, the ACC tracks trial-by-trial outcomes of decisions (Procyk et al., 2000; Shidara and Richmond, 2002; Amiez et al., 2006; Quilodran et al., 2008) as well as reward and choice history (Seo and Lee, 2007; Kennerley et al., 2009, 2011; Sul et al., 2010; Kawai et al., 2015) and signals errors in outcome prediction (Kennerley et al., 2009, 2011; Hayden et al., 2011; Monosov, 2017). At the time of feedback, neurons in the OFC encode committed choices, their values and contingent rewards (Tsujimoto et al., 2009; Sul et al., 2010). Notably, while the OFC encodes the identity of expected outcomes and the value of the chosen option after the alternatives are presented to an animal, it does not appear to encode upcoming decisions (Tremblay and Schultz, 1999; Wallis and Miller, 2003; Padoa-Schioppa and Assad, 2006; Sul et al., 2010; McDannald et al., 2014), therefore it might be that feedback projections from the dlPFC or ACC are required for such activity to emerge at the time of reward feedback.

To capture the interactions between PFC subregions in reinforcement-driven learning, Khamassi and colleagues have formulated a computation model in which action values are stored and updated in the ACC and then communicated to the dlPFC that decides which action to trigger (Khamassi et al., 2011, 2013). This model relies on meta-learning principles (Doya, 2002), flexibly adjusting the exploration-exploitation parameter based on performance history and variability in the environment that are monitored by the ACC. The explore-exploit parameter then influences action-selection mechanisms in the dlPFC, prioritizing choice repetition once the rewarded actions are discovered and encouraging switching between different options when environmental conditions change. In addition to highlighting the dynamic interactions between the dlPFC and ACC in learning, the model similarly offers an elegant solution to the credit assignment problem by restricting value updating only to those actions that were selected on a given trial. This is implemented by requiring the prediction error signals in the ACC to coincide with a motor efference copy sent by the premotor cortex. The model is endorsed with an ability to learn meta-values of novel objects in the environment based on the changes in the average reward that follow the presentation of such stimuli. While the authors proposed that such meta-value learning is implemented by the ACC, it is plausible that the OFC also plays a role in this process based on its contributions to stimulus-outcome and state learning (Wilson et al., 2014; Zsuga et al., 2016). Intriguingly, this model could reproduce monkey behavior and neural responses on two tasks: four-choice deterministic and two-choice probabilistic paradigms, entailing a complex spatio-temporal credit assignment problem as the stimuli disappeared from the screen prior to action execution and outcome presentation (Khamassi et al., 2011, 2013, 2015). Model-based analyses of neuronal responses further revealed that information about prediction errors, action values and outcome uncertainty is integrated both in the dlPFC and ACC, but at different timepoints: before trial feedback in the dlPFC and after feedback in the ACC (Khamassi et al., 2015).

Collectively, these findings highlight the heterogeneity of responses in each PFC subregion that differ in temporal dynamics within a single trial and suggest that the cooperation between the OFC, ACC and dlPFC may support flexible, strategy- and context-dependent choices. This network perspective further suggests that individual PFC subregions may be less specialized in their functions than previously thought. For example, in primates both the ACC and dlPFC participate in decisions based on action values (Hunt et al., 2015; Khamassi et al., 2015). And more recently, it has been demonstrated that the OFC is involved in updating action-outcome values as well (Fiuzat et al., 2017). Analogously, while it has been proposed that the OFC is specialized for stimulus-outcome and ACC for action-outcome learning (Rudebeck et al., 2008), lesions to the ACC have been similarly reported to impair stimulus-based reversal learning (Chudasama et al., 2013), supporting shared contributions of the PFC subregions to adaptive behavior. Indeed, these brain regions communicate extensively, sharing the information about presented options, executed decisions and received rewards (**Figure 2**), which can enable them to assign credit for outcomes to choices on which they are contingent (Urbanczik and Senn, 2009; Friedrich et al., 2010, 2011). Attention-gated learning likely relies on the cooperation between PFC subregions as well: for example, coordinated and synchronized activity between the ACC and dlPFC aids in goal-directed attentional shifting and prioritization of task-relevant information (Womelsdorf et al., 2014; Oemisch et al., 2015; Voloh et al., 2015).

Functional connectivity within the PFC can support contingent learning on shorter timescales (ex., across trials within the same task), when complex rules or stimulus-action-outcome mappings are switching frequently (Duff et al., 2011; Johnson et al., 2016). Under such conditions, the same stimuli can carry different meaning depending on task context or due to changes in the environment (ex., serial discrimination-reversal problems) and the PFC neurons with heterogeneous response properties may be better targets for modification, allowing the brain to exert flexible, rapid and context-sensitive control over behavior (Asaad et al., 1998; Mansouri et al., 2006). Indeed, it has been shown that rule and reversal learning induce plasticity in OFC synapses onto the dorsomedial PFC (encompassing the ACC) in rats (Johnson et al., 2016). When motivational significance of reward-predicting cues fluctuates frequently, neuronal responses and synaptic connections within the PFC tend to update more rapidly (i.e., across block of trials) compared to subcortical structures and other cortical regions (Padoa-Schioppa and Assad, 2008; Morrison et al., 2011; Xie and Padoa-Schioppa, 2016; Fernández-Lamo et al., 2017; Saez et al., 2017). Similarly, neurons in the PFC promptly adapt their responses to incoming information based on the recent history

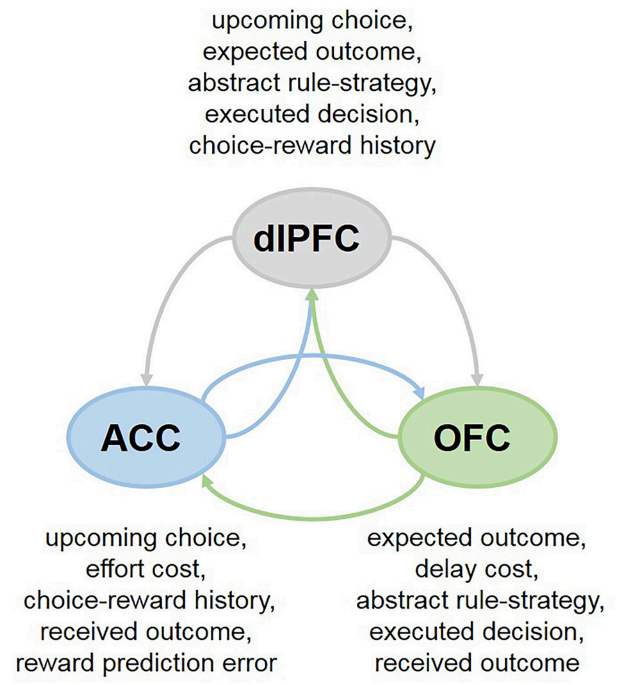

FIGURE 2 | Cooperation between PFC subregions in multi-cue tasks**.** In many everyday decisions, the options are compared across multiple features simultaneously (ex., by considering current context, needs, available reward types, as well as delay and effort costs). Neurons in different subregions of the PFC exhibit rich response properties, integrating many aspects of the task at hand. The OFC, ACC and dlPFC communicate extensively, sharing the information about presented options, executed decisions and received rewards, which can enable them to assign credit for outcomes to choices on which they are contingent.

of inputs (Freedman et al., 2001; Meyers et al., 2012; Stokes et al., 2013). Critically, changes in the PFC activity closely track behavioral performance (Mulder et al., 2003; Durstewitz et al., 2010), and interfering with neural plasticity within this brain area prevents normal responses to contingency degradation (Swanson et al., 2015).

When the circumstances are stable overall and the same cues or actions remain reliable predictors of rewards, long-range connections between the PFC, association and sensory areas can support contingent learning on prolonged timescales. Neurons in the lateral intraparietal area demonstrate larger post-decisional responses and enhanced learning following choices that predict final outcomes of sequential behavior in a multi-step and -cue task (Gersch et al., 2014). Such changes in neuronal activity likely rely on information about task rules conveyed by the PFC directly or via interactions with neuromodulatory systems. These hypotheses could be tested in future work.

In summary, dynamic interactions between subregions of the PFC can support contingent learning in multi-cue environments. Furthermore, via feedback projections, the PFC can guide plasticity in other cortical areas associated with sensory and motor processing (Cohen et al., 2011). This account suggests that lesion experiments targeting a localized PFC subregion will be insufficient to gain fine-grained understanding of credit assignment during learning and instead poses refined questions for future research, shifting the focus from focal manipulations to experimental techniques targeting cortico-cortical projections. To gain novel insights into functional connectivity between PFC subregions, it will be critical to assess neural correlates of contingent learning in the OFC, ACC, and dlPFC simultaneously in the context of the same task. In humans, functional connectivity can be assessed by utilizing coherence, phase synchronization, Granger causality and Bayes network approaches (Bastos and Schoffelen, 2016; Mill et al., 2017). Indeed, previous studies have linked individual differences in cortico-striatal functional connectivity to reinforcement-driven learning (Horga et al., 2015; Kaiser et al., 2017) and future work could focus on examining cortico-cortical interactions in similar paradigms. To probe causal contributions of projections spanning the PFC, future research may benefit from designing multi-cue tasks for rodents and taking advantage of recently developed techniques (i.e., chemo- and opto-genetic targeting of projection neurons followed by silencing of axonal terminals to achieve pathway-specific inhibition; Deisseroth, 2010; Sternson and Roth, 2014) that afford increasingly precise manipulations of cortico-cortical connectivity. It should be noted, however, that most experiments to date have probed the contributions of the PFC to credit assignment in primates, and functional specialization across different subregions may be even less pronounced in mice and rats. Finally, as highlighted throughout this review, the recent progress in understanding the neural mechanisms of credit assignment has relied on introduction of more complex tasks, including multi-cue and probabilistic choice paradigms. While such tasks better mimic the naturalistic problems that the brains have evolved to solve, they also produce behavioral patterns that are more difficult to analyze and interpret (Scholl and Klein-Flügge, 2017). As such, computational modeling of the behavior and neuronal activity may prove especially useful in future work on credit assignment.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and approved it for publication.

FUNDING

This work was supported by UCLA's Division of Life Sciences Recruitment and Retention fund (Izquierdo), as well as the UCLA Distinguished University Fellowship (Stolyarova).

ACKNOWLEDGMENTS

The author thanks her mentor Dr. Alicia Izquierdo for helpful feedback and critiques on the manuscript, and Evan E. Hart, as well as the members of the Center for Brains, Minds and Machines and Lau lab for stimulating conversations on the topic.

REFERENCES

Abraham, W. C. (2008). Metaplasticity: tuning synapses and networks for plasticity. *Nat. Rev. Neurosci.* 9:387 doi: 10.1038/nrn2356

Akaishi, R., Kolling, N., Brown, J. W., and Rushworth, M. (2016). Neural mechanisms of credit assignment in a multicue environment. *J. Neurosci.* 36, 1096–1112. doi: 10.1523/JNEUROSCI.3159-15.2016

Amiez, C., Joseph, J. P., and Procyk, E. (2006). Reward encoding in the monkey anterior cingulate cortex. *Cereb. Cortex* 16, 1040–1055. doi: 10.1093/cercor/bhj046

Asaad, W. F., Lauro, P. M., Perge, J. A., and Eskandar, E. N. (2017). Prefrontal neurons encode a solution to the credit assignment problem. *J. Neurosci.* 37, 6995–7007. doi: 10.1523/JNEUROSCI.3311-16.2017.

Asaad, W. F., Rainer, G., and Miller, E. K. (1998). Neural activity in the primate prefrontal cortex during associative learning. *Neuron* 21, 1399–1407. doi: 10.1016/S0896-6273(00)80658-3

Barbas, H., and Pandya, D. N. (1989). Architecture and intrinsic connections of the prefrontal cortex in the rhesus monkey. *J. Comp. Neurol.* 286, 353–375 doi: 10.1002/cne.902860306

Barto, A. G. (2007). Temporal difference learning. *Scholarpedia J.* 2:1604. doi: 10.4249/scholarpedia.1604

Barto, A. G., Sutton, R. S., and Anderson, C. W. (1983). "Neuronlike adaptive elements that can solve difficult learning control problems," in *IEEE Transactions on Systems, Man, and Cybernetics, SMC-13*, 834–846

Bastos, A. M., and Schoffelen, J. M. (2016). A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Front. Syst. Neurosci.* 9:175. doi: 10.3389/fnsys.2015.00175

Bogacz, R., McClure, S. M., Li, J., Cohen, J. D., and Montague, P. R. (2007). Short-term memory traces for action bias in human reinforcement learning. *Brain Res.* 1153, 111–121. doi: 10.1016/j.brainres.2007.03.057

Buckley, M. J., Mansouri, F. A., Hoda, H., Mahboubi, M., Browning, P. G. F., Kwok, S. C., et al. (2009). Dissociable components of rule-guided behavior depend on distinct medial and prefrontal regions. *Science* 325, 52–58. doi: 10.1126/science.1172377

Chudasama, Y., Daniels, T. E., Gorrin, D. P., Rhodes, S. E., Rudebeck, P. H., and Murray, E. A. (2013). The role of the anterior cingulate cortex in choices based on reward value and reward contingency. *Cereb Cortex* 23, 2884–2898. doi: 10.1093/cercor/bhs266

Cohen, J. D., McClure, S. M., and Yu, A. J. (2007). Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 362, 933–942. doi: 10.1098/rstb.2007.2098

Cohen, M. X., Wilmes, K., and Vijver, I. v. (2011). Cortical electrophysiological network dynamics of feedback learning. *Trends Cogn. Sci.* 15, 558–566. doi: 10.1016/j.tics.2011.10.004

Curtis, C. E., and Lee, D. (2010). Beyond working memory: the role of persistent activity in decision making. *Trends Cogn. Sci.* 14, 216–222. doi: 10.1016/j.tics.2010.03.006

Dayan, P., and Abbott, L. F. (2001). *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems.* Cambridge, MA: MIT Press.

Deisseroth, K. (2010). Optogenetics. *Nat. Methods* 8, 26–29. doi: 10.1038/nmeth.f.324

Doya, K. (2002). Metalearning and neuromodulation. *Neural. Netw.* 15, 495–506. doi: 10.1016/S0893-6080(02)00044-8

Duff, A., Sanchez Fibla, M., and Verschure, P. F. M. J. (2011). A biologically based model for the integration of sensory–motor contingencies in rules and plans: a prefrontal cortex based extension of the distributed adaptive control architecture. *Brain Res. Bull.* 85, 289–304. doi: 10.1016/j.brainresbull.2010.11.008

Durstewitz, D., Vittoz, N. M., Floresco, S. B., and Seamans, J. K. (2010). Abrupt transitions between prefrontal neural ensemble states accompany behavioral transitions during rule learning. *Neuron* 66, 438–448. doi: 10.1016/j.neuron.2010.03.029

Elston, G. N. (2000). Pyramidal cells of the frontal lobe: all the more spinous to think with. *J. Neurosci.* 20:RC95. Available online at: http://www.jneurosci.org/content/20/18/RC95.long

Felleman, D. J., and Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1, 1–47. doi: 10.1093/cercor/1.1.1

Fernández-Lamo, I., Delgado-García, J. M., and Gruart, A. (2017). When and where learning is taking place: multisynaptic changes in strength during different behaviors related to the acquisition of an operant conditioning task by behaving rats. *Cereb. Cortex* 14, 1–13. doi: 10.1093/cercor/bhx011

Fisher, S. D., Robertson, P. B., Black, M. J., Redgrave, P., Sagar, M. A., Abraham, W. C., et al. (2017). Reinforcement determines the timing dependence of corticostriatal synaptic plasticity *in vivo*. *Nat. Commun.* 8:334. doi: 10.1038/s41467-017-00394-x

FitzGerald, T. H. B., Friston, K. J., and Dolan, R. J. (2013). Characterising reward outcome signals in sensory cortex. *NeuroImage* 83, 329–334. doi: 10.1016/j.neuroimage.2013.06.061

Fiuzat, E. C., Rhodes, S. E., and Murray, E. A. (2017). The role of orbitofrontal-amygdala interactions in updating action-outcome valuations in macaques. *J. Neurosci.* 37, 2463–2470. doi: 10.1523/JNEUROSCI.1839-16.2017

Freedman, D. J., Riesenhuber, M., Poggio, T., and Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* 291, 312–316. doi: 10.1126/science.291.5502.312

Friedrich, J., Urbanczik, R., and Senn, W. (2010). Learning spike-based population codes by reward and population feedback. *Neural. Comput.* 22, 1698–1717. doi: 10.1162/neco.2010.05-09-1010

Friedrich, J., Urbanczik, R., and Senn, W. (2011). Spatio-temporal credit assignment in neuronal population learning. *PLoS Comput. Biol.* 7:e1002092. doi: 10.1371/journal.pcbi.1002092

Gersch, T. M., Foley, N. C., Eisenberg, I., and Gottlieb, J. (2014). Neural correlates of temporal credit assignment in the parietal lobe. *PloS One*, 9:e88725. doi: 10.1371/journal.pone.0088725

Hayden, B. Y., Heilbronner, S. R., Pearson, J. M., and Platt, M. L. (2011). Surprise signals in anterior cingulate cortex: neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. *J. Neurosci.* 31, 4178–4187. doi: 10.1523/JNEUROSCI.4652-10.2011

Hayden, B. Y., and Platt, M. L. (2010). Neurons in anterior cingulate cortex multiplex information about reward and action. *J. Neurosci.* 30, 3339–3346. doi: 10.1523/JNEUROSCI.4874-09.2010

Her, E. S., Huh, N., Kim, J., and Jung, M. W. (2016). Neuronal activity in dorsomedial and dorsolateral striatum under the requirement for temporal credit assignment. *Sci. Rep.* 6:27056. doi: 10.1038/srep27056

Histed, M. H., Pasupathy, A., and Miller, E. K. (2009). Learning substrates in the primate prefrontal cortex and striatum: sustained activity related to successful actions. *Neuron* 63, 244–253. doi: 10.1016/j.neuron.2009.06.019

Horga, G., Maia, T. V., Marsh, R., Hao, X., Xu, D., Duan, Y., et al. (2015). Changes in corticostriatal connectivity during reinforcement learning in humans. *Hum. Brain Mapp.* 36, 793–803. doi: 10.1002/hbm.22665

Hull, C. (1943). *Principles of Behavior*. New York, NY: Appleton-Century-Crofts.

Hunt, L. T., Behrens, T. E. J., Hosokawa, T., Wallis, J. D., and Kennerley, S. W. (2015). Capturing the temporal evolution of choice across prefrontal cortex. *eLife* 4:e11945. doi: 10.7554/eLife.11945

Izhikevich, E. M. (2007). Solving the distal reward problem through linkage of STDP and dopamine signaling. *Cereb. Cortex* 17, 2443–2452. doi: 10.1093/cercor/bhl152

Jackson, S. A. W., Horst, N. K., Pears, A., Robbins, T. W., and Roberts, A. C. (2016). Role of the perigenual anterior cingulate and orbitofrontal cortex in contingency learning in the marmoset. *Cereb. Cortex* 26, 3273–3284. doi: 10.1093/cercor/bhw067

Jocham, G., Brodersen, K. H., Constantinescu, A. O., Kahn, M. C., Ianni, A. M., Walton, M. E., et al. (2016). Reward-guided learning with and without causal attribution. *Neuron* 90, 177–190. doi: 10.1016/j.neuron.2016.02.018

Jog, M. S., Kubota, Y., Connolly, C. I., Hillegaart, V., and Graybiel, A. M. (1999). Building neural representations of habits. *Science* 286, 1745–1749. doi: 10.1126/science.286.5445.1745

Johnson, C. M., Peckler, H., Tai, L. H., and Wilbrecht, L. (2016). Rule learning enhances structural plasticity of long-range axons in frontal cortex. *Nat. Commun.* 7:10785. doi: 10.1038/ncomms10785

Kaiser, R. H., Treadway, M. T., Wooten, D. W., Kumar, P., Goer, F., Murray, L., et al. (2017). Frontostriatal and dopamine markers of individual differences in reinforcement learning: a multi-modal investigation. *Cereb. Cortex.* doi: 10.1093/cercor/bhx281. [Epub ahead of print].

Kawai, T., Yamada, H., Sato, N., Takada, M., and Matsumoto, M. (2015). Roles of the lateral habenula and anterior cingulate cortex in negative outcome monitoring and behavioral adjustment in nonhuman primates. *Neuron* 88, 792–804. doi: 10.1016/j.neuron.2015.09.030

Kennerley, S. W., Behrens, T. E. J., and Wallis, J. D. (2011). Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nat. Neurosci.* 14, 1581–1589. doi: 10.1038/nn.2961

Kennerley, S. W., Dahmubed, A. F., Lara, A. H., and Wallis, J. D. (2009). Neurons in the frontal lobe encode the value of multiple decision variables. *J. Cogn. Neurosci.* 21, 1162–1178. doi: 10.1162/jocn.2009.21100

Khamassi, M., Enel, P., Dominey, P. F., and Procyk, E. (2013). Medial prefrontal cortex and the adaptive regulation of reinforcement learning parameters. *Prog. Brain Res.* 202, 441–464. doi: 10.1016/B978-0-444-62604-2.00022-8

Khamassi, M., Lallée, S., Enel, P., Procyk, E., and Dominey, P. F. (2011). Robot cognitive control with a neurophysiologically inspired reinforcement learning model. *Front Neurorobot* 5:1. doi: 10.3389/fnbot.2011.00001

Khamassi, M., Quilodran, R., Enel, P., Dominey, P. F., and Procyk, E. (2015). Behavioral regulation and the modulation of information coding in the lateral prefrontal and cingulate cortex. *Cereb. Cortex* 25, 3197–3218. doi: 10.1093/cercor/bhu114

Kim, H., Lee, D., and Jung, M. W. (2013). Signals for previous goal choice persist in the dorsomedial, but not dorsolateral striatum of rats. *J. Neurosci.* 33, 52–63. doi: 10.1523/JNEUROSCI.2422-12.2013

Kim, H., Sul, J. H., Huh, N., Lee, D., and Jung, M. W. (2009). Role of striatum in updating values of chosen actions. *J. Neurosci.* 29, 14701–14712. doi: 10.1523/JNEUROSCI.2728-09.2009

Kötter, R., and Wickens, J. (1995). Interactions of glutamate and dopamine in a computational model of the striatum. *J. Comput. Neurosci.* 2, 195–214. doi: 10.1007/BF00961434

Lebedev, M. A., Messinger, A., Kralik, J. D., and Wise, S. P. (2004). Representation of attended versus remembered locations in prefrontal cortex. *PLoS Biol.* 2:e365. doi: 10.1371/journal.pbio.0020365

Mackintosh, N. J. (1975). Blocking of conditioned suppression: role of the first compound trial. *J. Exp. Psychol.* 1, 335–345. doi: 10.1037/0097-7403.1.4.335

Mansouri, F. A., Matsumoto, K., and Tanaka, K. (2006). Prefrontal cell activities related to monkeys' success and failure in adapting to rule changes in a Wisconsin Card Sorting Test analog. *J. Neurosci.* 26, 2745–2756. doi: 10.1523/JNEUROSCI.5238-05.2006

Markowitz, D. A., Curtis, C. E., and Pesaran, B. (2015). Multiple component networks support working memory in prefrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 112, 11084–11089. doi: 10.1073/pnas.1504172112

Matsuda, Y., Marzo, A., and Otani, S. (2006). The presence of background dopamine signal converts long-term synaptic depression to potentiation in rat prefrontal cortex. *J. Neurosci.* 26, 4803–4810. doi: 10.1523/JNEUROSCI.5312-05.2006

McDannald, M. A., Esber, G. R., Wegener, M. A., Wied, H. M., Liu, T.-L., Stalnaker, T. A., et al. (2014). Orbitofrontal neurons acquire responses to "valueless" Pavlovian cues during unblocking. *eLife* 3:e02653. doi: 10.7554/eLife.02653

Meyers, E. M., Qi, X. L., and Constantinidis, C. (2012). Incorporation of new information into prefrontal cortical activity after learning working memory tasks. *Proc. Natl. Acad. Sci. U.S.A.* 109, 4651–4656. doi: 10.1073/pnas.1201022109

Mill, R. D., Bagic, A., Bostan, A., Schneider, W., and Cole, M. W. (2017). Empirical validation of directed functional connectivity. *Neuroimage* 146, 275–287. doi: 10.1016/j.neuroimage.2016.11.037

Monosov, I. E. (2017). Anterior cingulate is a source of valence-specific information about value and uncertainty. *Nat. Commun.* 8:134. doi: 10.1038/s41467-017-00072-y

Montague, P. R., Dayan, P., and Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* 16, 1936–1947.

Morrison, S. E., Saez, A., Lau, B., and Salzman, C. D. (2011). Different time courses for learning-related changes in amygdala and orbitofrontal cortex. *Neuron* 71, 1127–1140. doi: 10.1016/j.neuron.2011.07.016

Mulder, A. B., Nordquist, R. E., Orgüt, O., and Pennartz, C. M. A. (2003). Learning-related changes in response patterns of prefrontal neurons during instrumental conditioning. *Behav. Brain Res.* 146, 77–88. doi: 10.1016/j.bbr.2003.09.016

Niv, Y. (2009). Reinforcement learning in the brain. *J. Math. Psychol.* 53, 139–154. doi: 10.1016/j.jmp.2008.12.005

Niv, Y., Daniel, R., Geana, A., Gershman, S. J., Leong, Y. C., Radulescu, A., et al. (2015). Reinforcement learning in multidimensional environments relies on attention mechanisms. *J. Neurosci.* 35, 8145–8157. doi: 10.1523/JNEUROSCI.2978-14.2015

Noonan, M. P., Chau, B. K. H., Rushworth, M. F. S., and Fellows, L. K. (2017). Contrasting effects of medial and lateral orbitofrontal cortex lesions on credit assignment and decision-making in humans. *J. Neurosci.* 37, 7023–7035. doi: 10.1523/JNEUROSCI.0692-17.2017

Noonan, M. P., Walton, M. E., Behrens, T. E., Sallet, J., Buckley, M. J., and Rushworth, M. F. (2010). Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. *Proc. Natl. Acad. Sci. U.S.A.*107, 20547–20252. doi: 10.1073/pnas.1012246107

Oemisch, M., Westendorff, S., Everling, S., and Womelsdorf, T. (2015). Interareal spike-train correlations of anterior cingulate and dorsal prefrontal cortex during attention shifts. *J. Neurosci.* 35, 13076–13089. doi: 10.1523/JNEUROSCI.1262-15.2015

Padoa-Schioppa, C., and Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature* 441, 223–226 doi: 10.1038/nature04676

Padoa-Schioppa, C., and Assad, J. A. (2008). The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nat. Neurosci.* 11, 95–102. doi: 10.1038/nn2020

Pan, W. X., Schmidt, R., Wickens, J. R., and Hyland, B. I. (2005). Dopamine cells respond to predicted events during classical conditioning: evidence for eligibility traces in the reward-learning network. *J. Neurosci.* 25, 6235–6242. doi: 10.1523/JNEUROSCI.1478-05.2005

Pasupathy, A., and Miller, E. K. (2005). Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature* 433, 873–876. doi: 10.1038/nature03287

Pearson, J. M., Hayden, B. Y., Raghavachari, S., and Platt, M. L. (2009). Neurons in posterior cingulate cortex signal exploratory decisions in a dynamic multioption choice task. *Curr. Biol.* 19, 1532–1537. doi: 10.1016/j.cub.2009.07.048

Procyk, E., Tanaka, Y. L., and Joseph, J. P. (2000). Anterior cingulate activity during routine and non-routine sequential behaviors in macaques. *Nat. Neurosci.* 3, 502–508. doi: 10.1038/74880

Quilodran, R., Rothe, M., and Procyk, E. (2008). Behavioral shifts and action valuation in the anterior cingulate cortex. *Neuron* 57, 314–325. doi: 10.1016/j.neuron.2007.11.031

Roelfsema, P. R., and van Ooyen, A. (2005). Attention-gated reinforcement learning of internal representations for classification. *Neural. Comput.* 17, 2176–2214. doi: 10.1162/0899766054615699

Rothkopf, C. A., and Ballard, D. H. (2010). Credit assignment in multiple goal embodied visuomotor behavior. *Front. Psychol.* 1:173. doi: 10.3389/fpsyg.2010.00173

Rudebeck, P. H., Behrens, T. E., Kennerley, S. W., Baxter, M. G., Buckley, M. J., Walton, M. E., et al. (2008). Frontal cortex subregions play distinct roles in choices between actions and stimuli. *J. Neurosci.* 28, 13775–13785. doi: 10.1523/JNEUROSCI.3541-08.2008

Saez, R. A., Saez, A., Paton, J. J., Lau, B., and Salzman, C. D. (2017). Distinct roles for the amygdala and orbitofrontal cortex in representing the relative amount of expected reward. *Neuron* 95, 70.e3–77.e3. doi: 10.1016/j.neuron.2017.06.012

Scholl, J., and Klein-Flügge, M. (2017). Understanding psychiatric disorder by capturing ecologically relevant features of learning and decision-making. *Behav Brain Res*. doi: 10.1016/j.bbr.2017.09.050. [Epub ahead of print].

Schultz, W. (1998a). Predictive reward signal of dopamine neurons. *J. Neurophysiol.* 80, 1–27. doi: 10.1152/jn.1998.80.1.1

Schultz, W. (1998b). The phasic reward signal of primate dopamine neurons. *Adv. Pharmacol.* 42, 686–690. doi: 10.1016/S1054-3589(08)60841-8

Schultz, W. (2004). Neural coding of basic reward terms of animal learning theory, game theory, microeconomics and behavioural ecology. *Curr. Opin. Neurobiol.* 14, 139–147. doi: 10.1016/j.conb.2004.03.017

Schultz, W., and Dickinson, A. (2000). Neuronal coding of prediction errors. *Ann. Rev. Neurosci.* 23, 473–500. doi: 10.1146/annurev.neuro.23.1.473

Seo, H., Barraclough, D. J., and Lee, D. (2007). Dynamic signals related to choices and outcomes in the dorsolateral prefrontal cortex. *Cerebral Cortex* 17(Suppl. 1), i110–i117. doi: 10.1093/cercor/bhm064

Seo, H., and Lee, D. (2007). Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *J. Neurosci.* 27, 8366–8377. doi: 10.1523/JNEUROSCI.2369-07.2007

Seo, M., Lee, E., and Averbeck, B. B. (2012). Action selection and action value in frontal-striatal circuits. *Neuron* 74, 947–960. doi: 10.1016/j.neuron.2012.03.037

Seol, G. H., Ziburkus, J., Huang, S., Song, L., Kim, I. T., Takamiya, K., et al. (2007). Neuromodulators control the polarity of spike-timing-dependent synaptic plasticity. *Neuron* 55, 919–929. doi: 10.1016/j.neuron.2007.08.013

Shidara, M., and Richmond, B. J. (2002). Anterior cingulate: single neuronal signals related to degree of reward expectancy. *Science* 296, 1709–1711. doi: 10.1126/science.1069504

Shuler, M. G., and Bear, M. F. (2006). Reward timing in the primary visual cortex. *Science* 311, 1606–1609. doi: 10.1126/science.1123513

Sternson, S. M., and Roth, B. L. (2014). Chemogenetic tools to interrogate brain functions. *Ann. Rev. Neurosci.* 37, 387–407. doi: 10.1146/annurev-neuro-071013-014048

Stokes, M. G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., and Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron* 78, 364–375. doi: 10.1016/j.neuron.2013.01.039

Sul, J. H., Kim, H., Huh, N., Lee, D., and Jung, M. W. (2010). Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making. *Neuron* 66, 449–460. doi: 10.1016/j.neuron.2010.03.033

Sutton, R. S., and Barto, A. G. (1998). *Reinforcement Learning: An Introduction Vol. 1* Cambridge: MIT press

Swanson, A. M., Allen, A. G., Shapiro, L. P., and Gourley, S. L. (2015). GABAAα1-mediated plasticity in the orbitofrontal cortex regulates context-dependent action selection. *Neuropsychopharmacology* 40, 1027–1036. doi: 10.1038/npp.2014.292

Tremblay, L., and Schultz, W. (1999). Relative reward preference in primate orbitofrontal cortex. *Nature* 398, 704–708. doi: 10.1038/19525

Tsujimoto, S., Genovesio, A., and Wise, S. P. (2009). Monkey orbitofrontal cortex encodes response choices near feedback time. *J. Neurosci.* 29, 2569–2574. doi: 10.1523/JNEUROSCI.5777-08.2009

Tsujimoto, S., Genovesio, A., and Wise, S. P. (2011). Comparison of strategy signals in the dorsolateral and orbital prefrontal cortex. *J. Neurosci.* 31, 4583–4592. doi: 10.1523/JNEUROSCI.5816-10.2011

Urbanczik, R., and Senn, W. (2009). Reinforcement learning in populations of spiking neurons. *Nat. Neurosci.* 12, 250–252. doi: 10.1038/nn.2264

Voloh, B., Valiante, T. A., Everling, S., and Womelsdorf, T. (2015). Theta-gamma coordination between anterior cingulate and prefrontal cortex indexes correct attention shifts. *Proc. Natl. Acad. Sci. U.S.A.* 112, 8457–8462. doi: 10.1073/pnas.1500438112

Wallis, J. D., and Miller, E. K. (2003). Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *Eur. J. Neurosci.* 18, 2069–2081. doi: 10.1046/j.1460-9568.2003.02922.x

Walsh, M. M., and Anderson, J. R. (2014). Navigating complex decision spaces: problems and paradigms in sequential choice. *Psychol. Bull.* 140, 466–486. doi: 10.1037/a0033455

Wilson, R. C., and Niv, Y. (2011). Inferring relevance in a changing world. *Front Hum. Neurosci.* 5:189. doi: 10.3389/fnhum.2011.00189

Wilson, R. C., Takahashi, Y. K., Schoenbaum, G., and Niv, Y. (2014). Orbitofrontal cortex as a cognitive map of task space. *Neuron* 81, 267–279. doi: 10.1016/j.neuron.2013.11.005

Womelsdorf, T., Ardid, S., Everling, S., and Valiante, T. A. (2014). Burst firing synchronizes prefrontal and anterior cingulate cortex during attentional control. *Curr. Biol.* 24, 2613–2621. doi: 10.1016/j.cub.2014.09.046

Wörgötter, F., and Porr, B. (2005). Temporal sequence learning, prediction, and control: a review of different models and their relation to biological mechanisms. *Neural. Comput.* 17, 245–319. doi: 10.1162/0899766053011555

Xie, J., and Padoa-Schioppa, C. (2016). Neuronal remapping and circuit persistence in economic decisions. *Nat. Neurosci.* 19, 855–861. doi: 10.1038/nn.4300

Xu, Y. (2017). Reevaluating the sensory account of visual working memory storage. *Trends Cogn. Sci.* 21, 794–815 doi: 10.1016/j.tics.2017.06.013

Zhang, J. C., Lau, P.-M., and Bi, G.-Q. (2009). Gain in sensitivity and loss in temporal contrast of STDP by dopaminergic modulation at hippocampal synapses. *Proc. Natl. Acad. Sci. U.S.A.* 106, 13028–13033 doi: 10.1073/pnas.0900546106

Zsuga, J., Biro, K., Tajti, G., Szilasi, M. E., Papp, C., Juhasz, B., et al. (2016). 'Proactive' use of cue-context congruence for building reinforcement learning's reward function. *BMC Neurosci.* 17:70. doi: 10.1186/s12868-016-0302-7

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.