Integrative Modeling of Prefrontal Cortex

William H. Alexander[1], Eliana Vassena[1,2], James Deraeve[1], and Zachary D. Langford[1]

Abstract

■ pFC is generally regarded as a region critical for abstract reasoning and high-level cognitive behaviors. As such, it has become the focus of intense research involving a wide variety of subdisciplines of neuroscience and employing a diverse range of methods. However, even as the amount of data on pFC has increased exponentially, it appears that progress toward understanding the general function of the region across a broad array of contexts has not kept pace. Effects observed in pFC are legion, and their interpretations are generally informed by a particular perspective or methodology with little regard with how those effects may apply more broadly. Consequently, the number of specific roles and functions that have been identified makes the region a very crowded place indeed and one that appears unlikely to be explained by a single general principle. In this theoretical article, we describe how the function of large portions of pFC can be accommodated by a single explanatory framework based on the computation and manipulation of error signals and how this framework may be extended to account for additional parts of pFC. ■

INTRODUCTION

In his studies of color phenomena (Newton, 1730), Isaac Newton investigated the composition of white light. Before Newton's work, color was generally believed to derive from combinations of light and dark. In his experiments, he demonstrated that white light, rather than indicating the absence of color, is in fact composed of all colors. In a famous experiment, white light was refracted through a prism to produce the color spectrum, after which the entire spectrum was refracted through a second prism, resulting in a white light produced by reintegration of the color spectrum. This experiment provides a concise and clear example of the processes of analysis and synthesis (Ritchey, 1991). One prism decomposes white light into a spectrum (analysis), whereas the second prism reconstitutes the color bands into white light (synthesis). More generally, the process of analysis attempts to understand a phenomenon through decomposition into its constituent parts—literally breaking it up into simpler, more tractable entities. In turn, synthesis attempts to take individual components and combine them into a unified whole (Figure 1).

The metaphorical prisms used by neuroscientists for analysis are the methods by which observations regarding a particular brain region are recorded (Grinvald & Hildesheim, 2004). Such tools decompose the world into a wide spectrum of data. For example, microelectrode arrays produce data with high temporal resolution recorded from a limited number of neurons. EEG and electrocorticography yield data with similarly high temporal resolution but reflecting the activity of ensembles of neurons over broad neural regions. Data from fMRI, conversely, have a relatively coarse temporal resolution but can provide much greater spatial detail over a larger area than other methods. The manner in which data are recorded can have profound implications for its interpretations; in extreme cases, data from the same region, recorded at different temporal and spatial resolutions, can yield interpretations that are almost diametrically opposed (Ford, Gati, Menon, & Everling, 2009).

Although it is generally assumed that the data obtained by these diverse methods reflect some aspect of the underlying neural mechanisms, the meaning ascribed to them is informed by a second, metaphorical prism, the process of synthesis: the particular theory that is brought to bear on interpreting brain function. For a particular region of the brain, as for example, the ACC, a social neuroscientist may find that it is primarily involved in processing social cues (Rotge et al., 2015), a neuroeconomist may discover deep connections with quantities important for decision-making such as value and uncertainty (Rangel, Camerer, & Montague, 2008), and all the while, an affective neuroscientist might insist that the same region is a vital hub of emotional information such as happiness, pain, regret, and so on (Lieberman & Eisenberger, 2015). This is not to say that any of these interpretations are necessarily wrong; however, the fractionation of interpretation induced by specialized subfields may result in a disjointed and incomplete understanding of the neural mechanisms underlying human behavior. At worst, this trend might produce an overly complex "integrative"

[1]Ghent University, [2]Radboud University

doi:10.1162/jocn_a_01138

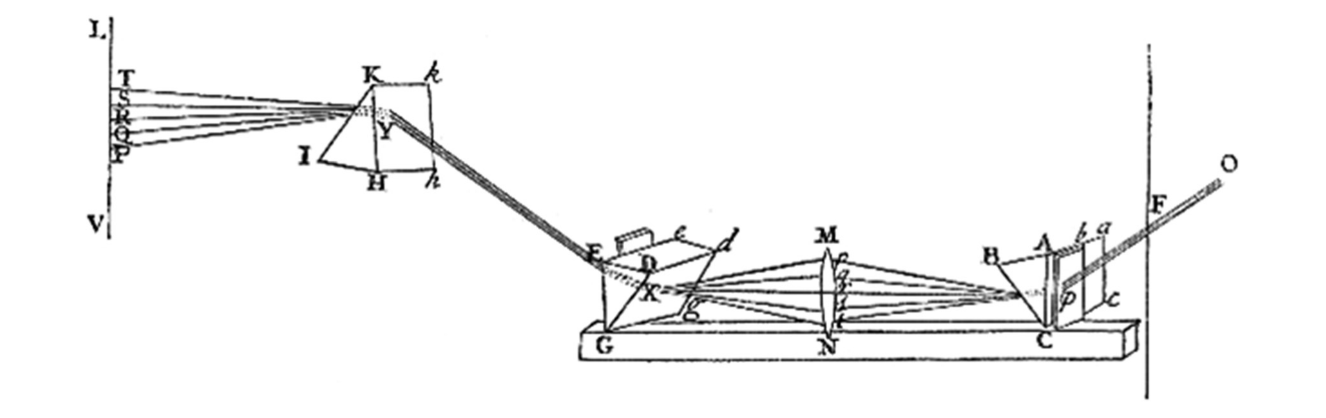

Figure 1. Figure from *Opticks* (Newton, 1730) depicting the apparatus used to decompose and reintegrate white light.

account that attempts to explain different functions as the product of multiple, spatially overlapping modules subserving specific and dissociable roles (Alexander & Brown, 2015b).

If the range of methods and perspectives deployed in recording and interpreting brain activity reflects the process of decomposing a signal into more easily understood constituents—analysis—then, what are the tools by which the constituent elements are reintegrated? Generally, this is the work of the theorist who proposes models and frameworks by which sets of data might be understood as emerging from some common underlying mechanism. Models can be specified in a variety of fashions, from simple graphical or written descriptions explaining how a system may function to more formal computational or mathematical descriptions. Synthesis through modeling tends to be a more catholic pursuit than analysis—to be worthwhile, a model should explain a range of analytic results rather than only one. However, even with this broader scope, synthesis can still be constrained by the perspective of the single theorist. A neuroscientist who is interested in intracellular signaling cascades will not pursue a theory of behavior, whereas a psychologist will generally not be interested in protein phosphorylation. Likewise, a cognitive neuroscientist may be able to explain the role of a region across a variety of tasks, such as ACC, but might be at a loss as to why the same region also responds to pain (Jahn, Nee, Alexander, & Brown, 2016).

Recent years have seen an increased emphasis on the synthetic process in neuroscience, frequently in the search for unifying principles underlying brain function, by which the diversity of data might be reintegrated. Proposed unifying frameworks include predictive coding, free energy, and the Bayesian brain hypothesis. Although the success of reinforcement learning and deep learning architectures at approximating human level performance at a variety of tasks provides existence proofs that relatively simple mechanisms can be used to understand human cognition, it remains an open question as to whether the variety of effects observed in brain and behavior can be reduced to a simple underlying principle. Indeed, the range of effects observed across different neuroscientific methodologies seems to provide evidence to the contrary.

Nevertheless, to the extent that the goal of neuroscience is to understand the function of the brain, it is insufficient to develop comprehensive models of highly circumscribed data. However, the sheer proliferation of data, in type and in quantity, tends to resist easy integration, sometimes resulting in the attempt of imposing some degree of order on an otherwise chaotic landscape through sufficiently sophisticated analyses and machine learning methods (although it remains unclear how effective these approaches are at recovering function; Jonas & Kording, 2017). In attempting to uncover the principles underlying brain function, then, it is necessary to negotiate between competing demands: The range of data incorporated should be sufficiently broad to specify a general underlying mechanism, yet not so broad as to render integration unlikely.

An Integrative Account of Medial pFC

One region that typifies this tradeoff is medial pFC (mPFC), especially ACC. mPFC/ACC activity is routinely observed across a range of experimental paradigms and is frequently associated with processing behavioral error (Gehring, Goss, Coles, Meyer, & Donchin, 1993). However, a host of other interpretations have been ascribed to the region, often deriving from the particular subdiscipline from which a study hails. From the affective neuroscience literature, the region has been assigned roles in processing somatic pain, joy, regret, perseverance, and other functions primarily associated with emotionally relevant information (Lieberman & Eisenberger, 2015; Parvizi, Rangarajan, Shirer, Desai, & Greicius, 2013; Chandrasekhar, Capra, Moore, Noussair, & Berns, 2008; Coricelli et al., 2005). In the domain of social neuroscience, ACC has been observed to be involved in processing social exclusion, monitoring the outcomes of another's choices, or learning from observing others (Hill, Boorman, & Fried, 2016; Rotge et al., 2015; Apps, Balsters, & Ramnani, 2012). Meanwhile, in the cognitive domain, ACC has been implicated in processing behavioral conflict, predicting the likelihood of an error, determining the value of exerting effort, or selecting optimal control signals (Holroyd & McClure, 2015; Verguts, Vassena, & Silvetti, 2015; Holroyd & Yeung, 2012; Brown & Braver, 2005; Botvinick, Braver, Barch, Carter, & Cohen, 2001).

Given the diverse array of effects observed in the region, it is an open question as to whether a single explanatory framework could be brought to bear to interpret signals generated by ACC. By and large, theorizing regarding ACC function (and pFC function in general) has tended to avoid overarching accounts and instead focused on the role of ACC under particular contexts (Holroyd & McClure, 2015; Shenhav, Straccia, Cohen, & Botvinick, 2014; Shenhav, Botvinick, & Cohen, 2013; Holroyd & Yeung, 2012; Kolling, Behrens, Mars, & Rushworth, 2012; Grinband et al., 2011; Silvetti, Seurinck, & Verguts, 2011; Brown & Braver, 2005; Yeung, Cohen, & Botvinick, 2004; Holroyd & Coles, 2002; Botvinick et al., 2001). Computational and mathematical models of the region have typically concerned themselves with the function of ACC within constrained empirical perspectives such as cognitive control or value-based decision-making. Indeed, the impetus behind the development of the predicted response–outcome (PRO) model (Alexander & Brown, 2010, 2011) was to provide an account of the function of ACC under relatively simple cognitive control tasks. The PRO model states that ACC learns to predict the likely outcomes of actions and signals deviations between observed and expected outcomes. Although the PRO model successfully captured effects related primarily to cognitive control, the formulation of the model as signaling surprising deviations from expectations suggested that it could be applied in a more general manner. In follow-up modeling work (Brown & Alexander, 2017; Alexander, Fukunaga, Finn, & Brown, 2015; Alexander & Brown, 2014) based on the PRO model, as well as tests of model predictions performed by other researchers (Jahn, Nee, Alexander, & Brown, 2014; Chang, Gariépy, & Platt, 2013; Talmi, Atkinson, & El-Deredy, 2013; Ferdinand, Mecklinger, Kray, & Gehring, 2012; Bryden, Johnson, Tobia, Kashtelyan, & Roesch, 2011), the twin functions of the PRO model—prediction and error signaling—have been applied to a broad range of perspectives, ranging from decision-making; social, affective, and clinical neuroscience; and perception and attention. A non-comprehensive list of effects captured by the PRO model is included elsewhere in this issue (Brown & Alexander, 2017). Given the breadth of effects encompassed by the PRO model, the range of perspectives to which the PRO account of ACC can be applied, and its ability to address observations from the level of single units to behavior, the PRO model remains the most comprehensive account of ACC function to date.

Building Out the Brain

Beyond merely addressing the function of ACC, however, the formulation of the PRO model carries implications regarding the function of regions of the brain with which ACC interacts. The PRO model generates two main signals, one related to predicting future events and another related to signaling surprising deviations. If, as the PRO model suggests, these two signals constitute the main outputs of ACC, regions connected to ACC should interact with at least one, and possibly both, of those signals (Figure 2A). Furthermore, the error and prediction signals generated by the PRO model are vector valued, carrying information regarding all possible outcomes that may be observed after a stimulus and reporting the amount by which an observed event deviates from all predictions.

Concurrently, the possible function of interactive brain regions is further implied by the class of tasks that the PRO model is unable to address. As noted above, the

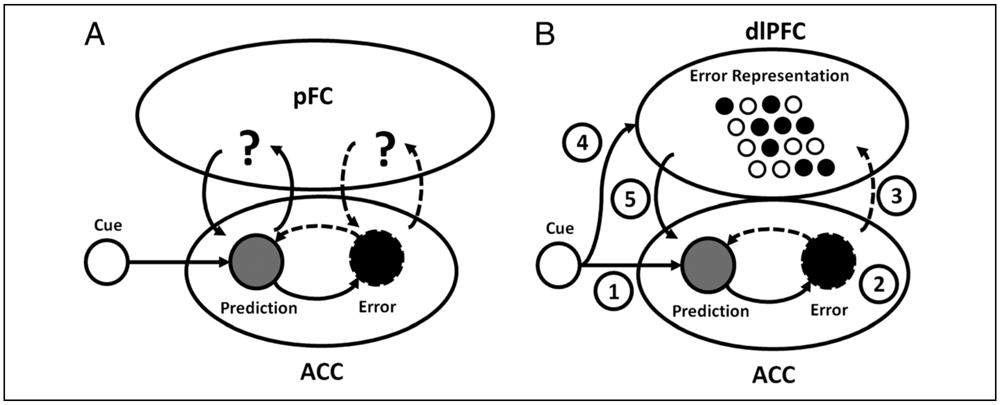

Figure 2. (A) According to the PRO model, ACC generates two principal signals, prediction and prediction error. If this account is correct, regions in pFC with which ACC interacts must do so through one or both of these signals, constraining the range of possible functions those regions may have. (B) The HER model specifies how dlPFC may interact with ACC by learning representations of the error signal generated by ACC and deploying active error representations to modulate predictive activity. (1) Task stimuli lead to predictions regarding likely outcomes. (2) Deviations between predicted and observed outcomes produce error signals, which are used to train distributed error representations in dlPFC. (4) Subsequent encounters with task stimuli leading to prediction errors reactivate error representations in dlPFC, (5) which are then used to modulate predictive activity in ACC.

PRO model was developed with the intent of capturing effects related to cognitive control. In typical cognitive control experiments, participants observe a stimulus indicating that a response is required, and after the generation of a response, the participant receives feedback regarding their performance, after which the next trial begins. Beyond a limited range of intertrial effects (Alexander & Brown, 2014), however, the PRO model is unable to address observations regarding ACC involvement in more sophisticated working memory tasks that involve the maintenance of information over protracted delays, often in the face of distracting, irrelevant information and potentially involving complex interrelationships among stimulus features that must be learned to inform correct behavior (Nee & Brown, 2013). An example of such a task is the AX Continuous Performance Task (CPT; Rosvold, Mirsky, Sarason, Bransome, & Beck, 1956), in which participants observe a sequence of stimuli (A, B, X, and Y) and are required to make a target response when an X appears, but only if the stimulus immediately preceding it was an A. To successfully perform this task, information related to the stimulus preceding an X must be maintained to correctly determine the response to the X.

Considering these two points, then, that (1) regions with which ACC interacts either receive or alter processing of prediction and/or error signals generated by ACC and (2) these regions are important for learning and performing complex cognitive tasks that require representing information regarding the relationships of task components, along with the assumption that prediction and error signaling constitute a general role for ACC across a range of experimental paradigms, we can begin to develop a clearer idea of the functions of additional regions of pFC. In this regard, dorsolateral pFC (dlPFC) is a likely candidate: dlPFC is densely and reciprocally interconnected with ACC (Medalla & Barbas, 2009, 2010; Barbas & Pandya, 1989) and is generally implicated in representing rules and complex task structure as well as in maintaining information over protracted delays, that is, working memory (Badre, Kayser, & D'Esposito, 2010; Chadderdon & Sporns, 2006; Koechlin, Ody, & Kouneiher, 2003). dlPFC is believed to be organized along a rostrocaudal abstraction gradient, with caudal regions representing concrete rules and rostral areas representing abstract context information, and it is frequently coactivated, with ACC, in tasks that involve complex interrelationships and learning models of the world (Nee & Brown, 2013; Badre & Frank, 2012; Gläscher, Daw, Dayan, & O'Doherty, 2010; Badre & D'Esposito, 2007, 2009).

How might dlPFC interact with prediction and error signals generated by ACC? In Alexander and Brown (2011), we noted that the vector-valued error signal used by the PRO model is appropriate for model-based reinforcement learning (Barto, Bradtke, & Singh, 1995; Sutton, 1990) as distinct from model-free reinforcement learning approaches that employ a scalar value signal to drive learning (Sutton & Barto, 1990). Previous work (Gläscher et al., 2010) has observed effects in dlPFC consistent with such a learning signal, suggesting that error signals generated by ACC may be used in dlPFC to learn representations of a task. Furthermore, working memory is important for informing and contextualizing behavioral responses; to respond correctly to an X in the AX CPT, information carried by the immediately preceding stimulus is required to modify predictions regarding the likely outcomes of the various responses one could make. Together, these observations led to the development of the hierarchical error representation (HER) model (Alexander & Brown, 2015a, 2016). The HER model proposes that error signals generated in ACC/mPFC are used to train representations in dlPFC, which are associated with task-relevant stimuli that reliably precede a prediction error (Figure 2B). When these representations are elicited by future presentations of the task stimuli with which they are associated, they are used to modulate prediction-related activity in ACC/mPFC. In typical reinforcement learning applications, the error signal specifies the direction and magnitude by which associations between a stimulus and its associated outcomes should be modified. In contrast, representations learned by the dlPFC in the HER model are predictions of prediction errors reported by ACC/mPFC; error signals constitute a kind of "proxy" outcome upon which representations in dlPFC converge during the course of learning. By using error signals themselves as outcomes that are the target of predictive processes, additional error signals reflecting the discrepancy between a predicted error and an actual error can be calculated, and these higher-order error signals may themselves be subject to further prediction and error calculations, and so on. Although this process of calculating increasingly abstract prediction errors could, in principle, continue arbitrarily, it is computationally limited by the capacity of computer systems on which the HER model is simulated, and biologically, it appears that the human brain is organized into three to five hierarchical processing levels in pFC, from premotor cortex at the base layer to rostral dlPFC (Reynolds, O'Reilly, Cohen, & Braver, 2012; Badre, 2008; Koechlin et al., 2003).

The purpose of learning to predict prediction errors themselves, as opposed to some other quantity, is to refine predictions regarding the likely outcomes of actions; being able to predict the kinds of prediction errors that are possible within a given context provides information sufficient to refine predictions of the likely outcomes given a current stimulus. The use of error signals in this fashion—as being the target of predictive processes in addition to governing prediction learning—is appealing for two reasons. First, and most pragmatically, learning representations of errors works: The HER model is able to learn to perform structured tasks from trial-and-error learning in a manner consistent with human behavior. Second, from an aesthetic point of view, the use of a common representation scheme used among regions in pFC is parsimonious and does not require intermediate transformations of information. The HER model is thus composed of a relatively simple computational motif that is hierarchically iterated. At each level, the model attempts to learn to predict the association between task stimuli and outcome signals arriving from lower hierarchical levels (or, at the base level, from the external environment), passing the results of error calculations upward along the hierarchy, while prediction information is passed downward to modulate the processing of lower hierarchical levels. However, although the calculation and maintenance of quantities related to error appear to be a useful scheme for interpreting the function of pFC, this aspect of the model remains speculative and in need of testing.

Although the computational motif on which it is based is relatively simple (in fact, it is functionally identical to the PRO model), the HER model is capable of learning complex cognitive tasks in a manner consistent with human behavior and evidence from neuroimaging studies (Alexander & Brown, 2015a). Rather than being limited to explaining results from a single task, the architecture of the HER model constitutes a general learning algorithm that can solve a range of tasks reported in the literature (Alexander & Brown, 2016), ranging from relatively simple examples such as the AX CPT or delayed-match-to-sample tasks to highly involved tasks using multiple stimulus categories with complex interrelationships (Koechlin et al., 2003) in a way that captures the function of ACC/mPFC and dlPFC as well as how the two regions interact during behavior (Kim, Johnson, Cilles, & Gold, 2011). The HER model additionally captures patterns of activity in single neurons observed in lateral pFC and mPFC during tasks involving maintenance of information and sequential decision-making (Procyk, Tanaka, & Joseph, 2000; Miller, Erickson, & Desimone, 1996). In brief, the HER model addresses itself to a broad range of tasks to account for data from multiple levels of description simultaneously.

Toward an Integrative Model of pFC

Together, the PRO and HER models provide one of the most comprehensive accounts of effects observed in pFC (cf. Brown & Alexander, 2017; Table 1). These effects range from single units in lateral pFC and mPFC, the activity of ensembles of neurons indexed by EEG and fMRI, the nature of representations deployed by pFC in the context of high-level cognitive tasks, and how the acquisition of these representations during learning contributes to behavioral markers of adaptive behavior. The ability of the models to capture these effects rests on the reconceptualization of activity in pFC as being fundamentally related to calculating, maintaining, and manipulating quantities related to prediction error: In the HER framework, mPFC calculates deviations between expected and observed outcomes, whereas dlPFC learns representations of the expected error reported by mPFC and associated with task-relevant stimuli.

Table 1. Effects Simulated by the HER Model So Far

	Region
fMRI	
Badre et al., 2010	LPFC
Kim et al., 2011	LPFC/mPFC
Koechlin et al., 2003	LPFC
Nee & Brown, 2012	LPFC
Nee & Brown, 2013	LPFC
Nee, Jahn, & Brown, 2013	LPFC
Nee & D'Esposito, 2016	LPFC
Reverberi, Görgen, & Haynes, 2011	LPFC
Reynolds et al., 2012	LPFC
Lesion	
Gehring & Knight, 2000	mPFC
Tsuchida & Fellows, 2008	LPFC/mPFC
Single unit	
Hayden, Pearson, & Platt, 2011	mPFC
Miller et al., 1996	LPFC
Procyk et al., 2000	mPFC
Shidara & Richmond, 2002	mPFC
Stoll et al., 2016	LPFC/mPFC
Behavioral	
Badre et al., 2010	NA
Krueger, 2011	NA
Krueger & Dayan, 2009	NA
Markant & Gureckis, 2012	NA
Stoll et al., 2016	NA

LPFC = lateral prefrontal cortex; NA = not applicable.

The integration suggested by the HER model then is twofold. First, the HER model bridges multiple levels of description, concurrently providing an account of the function of single neurons in pFC, the role those units play in neural ensembles, and ultimately, how their distributed activity conspires to produce observed patterns of behavior. Second, the architecture of the HER model suggests a relationship with theoretical frameworks that have been proposed as potentially unifying models of neocortex. Recent years have seen a renewed interest in the search for such a unifying framework that may be of use in interpreting the function and organization of the brain. Approaches such as hierarchical Bayesian inference, free energy, and predictive coding (Clark, 2013; Friston, 2010; Lee & Mumford, 2003; Rao & Ballard, 1999) have garnered significant interest in this respect and have achieved success in explaining effects observed in sensory and motor cortices. Generally, these approaches suggest a hierarchical organization of the brain in which information in the form of prediction errors is passed from inferior hierarchical levels to superior levels, whereas information required to "explain away" prediction errors generated at a lower level are passed downward from superior hierarchical levels. The HER model conforms to this overall framework, with prediction errors traveling through the hierarchy along bottom–up routes, while representations of prediction errors are passed in a top–down fashion to refine predictions of lower levels. Within each level of the hierarchy, mPFC and dlPFC serve complementary roles along the bottom–up and top–down processing pathways. In the bottom–up pathway, mPFC calculates error signals used to train error representations in dlPFC at superior hierarchical layers. In the top–down pathway, contextually relevant components of active error representations in dlPFC are selected by mPFC to modulate ongoing prediction-related activity at lower hierarchical layers (Alexander & Brown, 2015a). At the base layer of the hierarchy, the HER model interprets mPFC activity as being involved in predicting response–outcome conjunctions (as in the PRO model) and signaling discrepancies; top–down information thus serves to contextualize or "explain away" errors that would otherwise be reported without top–down modulation. Thus, the HER model provides a demonstration that predictive coding and related approaches may be extended into pFC.

By recasting the function of large portions of pFC as relating to prediction errors, either through the explicit calculation of error or through maintaining predictions of potential future prediction errors, the HER model suggests that error calculation and representation may serve as a common code underlying neural activity and communication. This possibility stands in contrast to recent proposals (Shenhav et al., 2013; Levy & Glimcher, 2012) that quantities related to the prediction and calculation of value might constitute the common neural currency under which the function of brain regions should be interpreted. Although a large literature in neuroeconomics and judgment and decision-making has implicated aspects of the frontal lobes in value computations, especially, for example, ventromedial and orbitofrontal pFC (Grabenhorst & Rolls, 2011; Gläscher, Hampton, & O'Doherty, 2009; Rangel et al., 2008; Padoa-Schioppa & Assad, 2006; Kringelbach, 2005; Gottfried, O'Doherty, & Dolan, 2003), it is not automatic that value representation needs to be the only, or even primary, role of those regions (Stalnaker, Cooch, & Schoenbaum, 2015; Gläscher et al., 2010; Hampton, Bossaerts, & O'Doherty, 2006). One possibility is that effects that appear to relate to neuroeconomic quantities such as value may have an alternate interpretation under the framework of error and error representation. Alternately, it is possible that different processing streams in pFC utilize complementary but distinct forms of representation to support diverse cognitive behaviors. An open question therefore is whether predictive coding in general, and the HER model in particular, might be expanded to account for the function of additional regions of pFC without reference to explicit value signaling.

In this regard, one possible avenue by which the HER model might be extended relates to the status of internal representations used by the model. As detailed above, the PRO model was aimed initially at explaining effects observed within mPFC and with little regard as to how the signals postulated by the model might be deployed by regions with which mPFC interacts. Additional modeling work, building on the PRO model, specifies how prediction and error signals in the model may be used in supporting proactive and reactive control (Brown & Alexander, 2017) or in the acquisition and performance of cognitive tasks (Alexander & Brown, 2015a). In a similar fashion, the origin of internal representations used by the PRO and HER models as the bases for learning is left underspecified; the appearance of an external stimulus results in the activation of an internal representation corresponding to that stimulus. This mapping of external stimuli to internal representations in a one-to-one fashion is likely overly simplistic—besides the considerable processing needed to transform patterns of light hitting the retina into unitary internal representations (e.g., letters or numbers), additional processes are involved in governing whether the presence of an external stimulus is registered (e.g., attention) as well as contextual influences on how that stimulus, once registered, informs ongoing behavior. A significant challenge to be addressed then is whether unifying schemes such as predictive coding can be leveraged to explain the representation and contextualization of task stimuli in pFC.

It is possible that additional regions with which mPFC interacts may be involved in regulating access of internal stimulus representations to regions of pFC involved with outcome prediction and error calculations. One region that may potentially serve this role is the anterior insula cortex (AIC). AIC is reciprocally connected with mPFC/ACC (Augustine, 1996), and coactivation of the two regions is routinely observed, especially during the registration and processing of behavioral error (Ullsperger, Harsay, Wessel, & Ridderinkhof, 2010). It has been suggested, considering the dense innervation of AIC from amygdala (Augustine, 1996), that AIC is important for processing emotionally relevant information (Jones, Ward, & Critchley, 2010; Wiech et al., 2010; Singer, Critchley, & Preuschoff, 2009). However, considering that ACC has also been extensively implicated in processing affective information (Lieberman & Eisenberger, 2015; Rotge et al., 2015; Chandrasekhar et al., 2008; Bush, Luu, & Posner, 2000), it seems unlikely that the two regions are dissociated by their role in emotional processing. An alternative possibility is that AIC may be involved in the selection of information for further processing by ACC. AIC receives rich interoceptive signals related to bodily states (Barrett & Simmons, 2015; Critchley, Wiens, Rotshtein, Öhman, & Dolan, 2004) as well as information potentially related to the significance of sensory input (Han & Marois, 2014; Menon & Uddin, 2010; Nelson et al., 2010; Eckert et al., 2009; Corbetta & Shulman, 2002). Models of associative learning (Alexander, 2007; Kruschke, 2001; Pearce & Hall, 1980; Mackintosh, 1975) have suggested that error signals generated during learning might not only support the alteration of associations between a stimulus and its subsequent outcomes but also modulate the associability (or salience) of a stimulus. Error signals generated by AIC might therefore provide a means by which incoming information, interoceptive or exteroceptive, is triaged for further processing, whereas error signals in mPFC influence the associations learned regarding selected information. In support of this possibility, AIC is known to project to the nucleus basalis, the primary source of cholinergic input to cortex, although evidence for innervation of the nucleus basalis by cingulate is mixed (Russchen, Amaral, & Price, 1985; Mesulam & Mufson, 1984); acetycholine has been implicated as an important neuromodulator for estimating risk and selecting internally represented information (Smith, Saaj, & Allouis, 2012; Krichmar, 2008; Yu & Dayan, 2005).

The integration suggested by the HER model, although of potential interest, remains speculative for a number of reasons. First, although the HER model is able learn a number of tasks that have been deployed in the study of high-level cognitive behaviors (Alexander & Brown, 2016), these tasks represent only one "operating mode" of the brain. Specifically, in the kinds of tasks the HER model was developed to learn, participants are required to integrate a history of observations to determine the correct behavior given a currently observed stimulus. This type of task is exemplified by the 1-2 AX CPT (O'Reilly & Frank, 2006) in which the sequence of stimuli observed by a participant is externally controlled; when a potential target cue is displayed, participants can only refer to past observations to arrive at a decision as to whether to make a target or nontarget response. Contrast this with a situation in which participants may be asked to navigate from one point in a maze to another; in this case, participants must themselves determine the sequence of observations required to correctly solve the maze. Although the present version of the HER model is unable to address this kind of behavior, it is possible that the general hierarchical organization of pFC as instantiated in the model, as well as its interpretation of activity in pFC as relating to error representation and manipulation, may be suitable for this form of goal-oriented decision-making. Under this mode of operation, goals might be interpreted as discrepancies (or errors) between a desired and current state and behaviors selected on the basis of how efficiently this discrepancy is reduced.

A second potential limitation of the model in its current form, also related to the class of tasks the model was developed to perform, is its inability to account for behaviors related to the manipulation of internal representations. An example of this kind of behavior, pervasive in the working memory literature, is the *n*-back task (Kirchner, 1958), in which participants observe a sequence of stimuli and are required to report whether the current stimulus is a match for the stimulus observed *n* steps previously (typically *n* is a number from 1 to 3). Above and beyond passively integrating a history of observations, the *n*-back task requires participants, upon presentation of a new stimulus, not only to maintain the identity of previously observed stimuli but also to update those representations with information pertaining to the number of steps in the past they were observed. For example, if a picture of a dog was observed one step in the past, the presentation of a new stimulus requires participants to remember that the dog was now observed two steps in the past. Related to this kind of task are other cognitive behaviors, such as mental calculation, in which the results of simple calculations must be represented internally and used in further calculation to arrive at the correct solution. The active maintenance and manipulation of internal representations implied by tasks of this sort suggest the existence of a "visuospatial sketchpad" or "phonological loop" (Baddeley & Hitch, 1974) in which the results of internal representational manipulations can be stored for later use or reintegrated during further manipulation. Although the HER model in its current form does not incorporate a mechanism by which such manipulations of internal representations might be carried out, it is possible that future work might extend the model to include such operations.

More generally, although the HER model is incomplete, its reconceptualization of large portions of frontal cortex as engaging in error calculation and representation provides a lens through which additional regions might be viewed. A key challenge for future work is to investigate whether and how processes related to error computation might constitute a general functional principle of pFC. In much the same way that the PRO model provided critical constraints on how regions with which ACC interacts may function, the HER model may further inform our understanding of the organization and function of the rest of pFC. Although the success of the HER and PRO models in accounting for a wide array of effects, from single neurons to behavior, suggests that error-related processes may be a useful framework for interpreting brain function, it is possible that such a framework may in fact prove insufficient to explain the diversity of observations throughout pFC. In either eventuality, whether it serves as the basis for a broader understanding of pFC, or as a theory to be superseded by more comprehensive accounts, the HER model is a step toward the ultimate goal of understanding the function of pFC.

Acknowledgments

W. H. A. was supported by FWO-Flanders Odysseus II Award #G.OC44.13N. E. V. was supported by the Marie Sklodowska-Curie action with a standard IF-EF fellowship, within the H2020 framework (H2020-MSCA-IF2015, Grant number 705630).

Reprint requests should be sent to William H. Alexander, Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, Ghent, Belgium 9000, or via e-mail: william.alexander@ugent.be.

REFERENCES

Alexander, W. H. (2007). Shifting attention using a temporal difference prediction error and high-dimensional input. *Adaptive Behavior, 15,* 121–133.

Alexander, W. H., & Brown, J. W. (2010). Computational models of response–outcome predictions as a basis for cognitive control. *Topics in Cognitive Science, 2,* 658–677.

Alexander, W. H., & Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nature Neuroscience, 14,* 1338–1344.

Alexander, W. H., & Brown, J. W. (2014). A general role for medial prefrontal cortex in event prediction. *Frontiers in Computational Neuroscience, 8,* 69.

Alexander, W. H., & Brown, J. W. (2015a). Hierarchical error representation: A computational model of anterior cingulate and dorsolateral prefrontal cortex. *Neural Computation, 27,* 2354–2410.

Alexander, W. H., & Brown, J. W. (2015b). Reciprocal interactions of computational modeling and empirical investigation. In B. U. Forstmann & E.-J. Wagenmakers (Eds.), *An introduction to model-based cognitive neuroscience* (pp. 321–338). New York: Springer.

Alexander, W. H., & Brown, J. W. (2016). Frontal cortex function derives from hierarchical predictive coding. *bioRxiv,* 076505.

Alexander, W. H., Fukunaga, R., Finn, P., & Brown, J. W. (2015). Reward salience and risk aversion underlie differential ACC activity in substance dependence. *Neuroimage: Clinical, 8,* 59–71.

Apps, M. A. J., Balsters, J. H., & Ramnani, N. (2012). The anterior cingulate cortex: Monitoring the outcomes of others' decisions. *Social Neuroscience, 7,* 424–435.

Augustine, J. R. (1996). Circuitry and functional aspects of the insular lobe in primates including humans. *Brain Research. Brain Research Reviews, 22,* 229–244.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 8, pp. 47–89). New York: Academic Press.

Badre, D. (2008). Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends in Cognitive Sciences, 12,* 193–200.

Badre, D., & D'Esposito, M. (2007). Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *Journal of Cognitive Neuroscience, 19,* 2082–2099.

Badre, D., & D'Esposito, M. (2009). Is the rostro-caudal axis of the frontal lobe hierarchical? *Nature Reviews Neuroscience, 10,* 659–669.

Badre, D., & Frank, M. J. (2012). Mechanisms of hierarchical reinforcement learning in cortico-striatal circuits 2: Evidence from fMRI. *Cerebral Cortex, 22,* 527–536.

Badre, D., Kayser, A. S., & D'Esposito, M. (2010). Frontal cortex and the discovery of abstract action rules. *Neuron, 66,* 315–326.

Barbas, H., & Pandya, D. N. (1989). Architecture and intrinsic connections of the prefrontal cortex in the rhesus monkey. *The Journal of Comparative Neurology, 286,* 353–375.

Barrett, L. F., & Simmons, W. K. (2015). Interoceptive predictions in the brain. *Nature Reviews Neuroscience, 16,* 419–429.

Barto, A., Bradtke, S. J., & Singh, S. P. (1995). Learning to act using real-time dynamic programming. *Artificial Intelligence, 72,* 81–138.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. C. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108,* 624–652.

Brown, J. W., & Alexander, W. H. (2017). Foraging value, risk avoidance, and multiple control signals: How the anterior cingulate cortex controls value-based decision-making. *Journal of Cognitive Neuroscience, 29,* 1656–1673.

Brown, J. W., & Braver, T. S. (2005). Learned predictions of error likelihood in the anterior cingulate cortex. *Science, 307,* 1118–1121.

Bryden, D. W., Johnson, E. E., Tobia, S. C., Kashtelyan, V., & Roesch, M. R. (2011). Attention for learning signals in anterior cingulate cortex. *Journal of Neuroscience, 31,* 18266–18274.

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences, 4,* 215–222.

Chadderdon, G. L., & Sporns, O. (2006). A large-scale neurocomputational model of task-oriented behavior selection and working memory in prefrontal cortex. *Journal of Cognitive Neuroscience, 18,* 242–257.

Chandrasekhar, P. V. S., Capra, C. M., Moore, S., Noussair, C., & Berns, G. S. (2008). Neurobiological regret and rejoice functions for aversive outcomes. *Neuroimage, 39,* 1472–1484.

Chang, S. W. C., Gariépy, J.-F., & Platt, M. L. (2013). Neuronal reference frames for social decisions in primate frontal cortex. *Nature Neuroscience, 16,* 243–250.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36,* 181–204.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience, 3,* 201–215.

Coricelli, G., Critchley, H., Joffily, M., O'Doherty, J., Sirigu, A., & Dolan, R. J. (2005). Regret and its avoidance: A neuroimaging study of choice behavior. *Nature Neuroscience, 8,* 1255–1262.

Critchley, H. D., Wiens, S., Rotshtein, P., Öhman, A., & Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nature Neuroscience, 7,* 189–195.

Eckert, M. A., Menon, V., Walczak, A., Ahlstrom, J., Denslow, S., Horwitz, A., et al. (2009). At the heart of the ventral attention system: The right anterior insula. *Human Brain Mapping, 30,* 2530–2541.

Ferdinand, N. K., Mecklinger, A., Kray, J., & Gehring, W. J. (2012). The processing of unexpected positive response outcomes in the mediofrontal cortex. *Journal of Neuroscience, 32,* 12087–12092.

Ford, K. A., Gati, J. S., Menon, R. S., & Everling, S. (2009). BOLD fMRI activation for anti-saccades in nonhuman primates. *Neuroimage, 45,* 470–476.

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience, 11,* 127–138.

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science, 4,* 385–390.

Gehring, W. J., & Knight, R. T. (2000). Prefrontal-cingulate interactions in action monitoring. *Nature Neuroscience, 3,* 516–520.

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron, 66,* 585–595.

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.

Gottfried, J. A., O'Doherty, J., & Dolan, R. J. (2003). Encoding predictive reward value in human amygdala and orbitofrontal cortex. *Science, 301,* 1104–1107.

Grabenhorst, F., & Rolls, E. T. (2011). Value, pleasure and choice in the ventral prefrontal cortex. *Trends in Cognitive Sciences, 15,* 56–67.

Grinband, J., Savitskaya, J., Wager, T. D., Teichert, T., Ferrera, V. P., & Hirsch, J. (2011). The dorsal medial frontal cortex is sensitive to time on task, not response conflict or error likelihood. *Neuroimage, 57,* 303–311.

Grinvald, A., & Hildesheim, R. (2004). VSDI: A new era in functional imaging of cortical dynamics. *Nature Reviews Neuroscience, 5,* 874–885.

Hayden, B. Y., Pearson, J. M., & Platt, M. L. (2011). Neuronal basis of sequential foraging decisions in a patchy environment. *Nature Neuroscience, 14,* 933–999.

Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2006). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *Journal of Neuroscience, 26,* 8360–8367.

Han, S. W., & Marois, R. (2014). Functional fractionation of the stimulus-driven attention network. *Journal of Neuroscience, 34,* 6958–6969.

Hill, M. R., Boorman, E. D., & Fried, I. (2016). Observational learning computations in neurons of the human anterior cingulate cortex. *Nature Communications, 7,* 12722.

Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review, 109,* 679–709.

Holroyd, C. B., & McClure, S. M. (2015). Hierarchical control over effortful behavior by rodent medial frontal cortex: A computational model. *Psychological Review, 122,* 54–83.

Holroyd, C. B., & Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. *Trends in Cognitive Sciences, 16,* 122–128.

Jahn, A., Nee, D. E., Alexander, W. H., & Brown, J. W. (2014). Distinct regions of anterior cingulate cortex signal prediction and outcome evaluation. *Neuroimage, 95,* 80–89.

Jahn, A., Nee, D. E., Alexander, W. H., & Brown, J. W. (2016). Distinct regions within medial prefrontal cortex process pain and cognition. *Journal of Neuroscience, 36,* 12385–12392.

Jonas, E., & Kording, K. (2017). Could a neuroscientist understand a microprocessor? *PLoS Computational Biology, 13,* e1005268.

Jones, C. L., Ward, J., & Critchley, H. D. (2010). The neuropsychological impact of insular cortex lesions. *Journal of Neurology, Neurosurgery and Psychiatry, 81,* 611–618.

Kim, C., Johnson, N. F., Cilles, S. E., & Gold, B. T. (2011). Common and distinct mechanisms of cognitive flexibility in prefrontal cortex. *Journal of Neuroscience, 31,* 4771–4779.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology, 55,* 352–358.

Koechlin, E., Ody, C., & Kouneiher, F. (2003). The architecture of cognitive control in the human prefrontal cortex. *Science, 302,* 1181–1185.

Kolling, N., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2012). Neural mechanisms of foraging. *Science, 336,* 95–98.

Krichmar, J. L. (2008). The neuromodulatory system: A framework for survival and adaptive behavior in a challenging world. *Adaptive Behavior, 16,* 385–399.

Kringelbach, M. L. (2005). The human orbitofrontal cortex: Linking reward to hedonic experience. *Nature Reviews Neuroscience, 6,* 691–702.

Krueger, K. A. (2011). *Sequential learning in the form of shaping as a source of cognitive flexibility*. UCL (University College London).

Krueger, K. A., & Dayan, P. (2009). Flexbile shaping: How learning in small steps helps. *Cognition, 110,* 380–394.

Kruschke, J. K. (2001). Toward a unified model of attention in associative learning. *Journal of Mathematical Psychology, 45,* 812–863.

Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A, 20,* 1434.

Levy, D. J., & Glimcher, P. W. (2012). The root of all value: A neural common currency for choice. *Current Opinion in Neurobiology, 22,* 1027–1038.

Lieberman, M. D., & Eisenberger, N. I. (2015). The dorsal anterior cingulate cortex is selective for pain: Results from large-scale reverse inference. *Proceedings of the National Academy of Sciences, U.S.A., 112,* 15250–15255.

Mackintosh, N. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82,* 276–298.

Markant, D., & Gureckis, T. (2012). One piece at a time: Learning complex rules through self-directed sampling. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Medalla, M., & Barbas, H. (2009). Synapses with inhibitory neurons differentiate anterior cingulate from dorsolateral prefrontal pathways associated with cognitive control. *Neuron, 61,* 609–620.

Medalla, M., & Barbas, H. (2010). Anterior cingulate synapses in prefrontal areas 10 and 46 suggest differential influence in cognitive control. *Journal of Neuroscience, 30,* 16068–16081.

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: A network model of insula function. *Brain Structure and Function, 214,* 655–667.

Mesulam, M.-M., & Mufson, E. J. (1984). Neural inputs into the nucleus basalis of the substantia innominata (Ch4) in the rhesus monkey. *Brain, 107,* 253–274.

Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience, 16,* 5154–5167.

Nee, D. E., & Brown, J. W. (2012). Rostral-caudal gradient of abstraction revealed by multi-variate pattern analysis of working memory. *Neuroimage, 63,* 1285–1294.

Nee, D. E., & Brown, J. W. (2013). Dissociable frontal–striatal and frontal–parietal networks involved in updating hierarchical contexts in working memory. *Cerebral Cortex, 23,* 2146–2158.

Nee, D. E., & D'Esposito, M. (2016). The hierarchical organization of the lateral prefrontal cortex. *eLife, 5,* e12112.

Nee, D. E., Jahn, A., & Brown, J. W. (2013). Prefrontal cortex organization: Dissociating effects of temporal abstraction, relational abstraction, and integration with fMRI. *Cerebral Cortex,* bht091.

Nelson, S. M., Dosenbach, N. U. F., Cohen, A. L., Wheeler, M. E., Schlaggar, B. L., & Petersen, S. E. (2010). Role of the anterior insula in task-level control and focal attention. *Brain Structure and Function, 214,* 669–680.

Newton, I. (1730). *Opticks: Or a treatise of the reflections, refractions, inflections and colors of light*. London: William

Innys. Retrieved from http://www.gutenberg.org/files/33504/33504-h/33504-h.htm.

O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation, 18,* 283–328.

Padoa-Schioppa, C., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature, 441,* 223–226.

Parvizi, J., Rangarajan, V., Shirer, W. R., Desai, N., & Greicius, M. D. (2013). The will to persevere induced by electrical stimulation of the human cingulate gyrus. *Neuron, 80,* 1359–1367.

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.

Procyk, E., Tanaka, Y. L., & Joseph, J. P. (2000). Anterior cingulate activity during routine and non-routine sequential behaviors in macaques. *Nature Neuroscience, 3,* 502–508.

Rangel, A., Camerer, C., & Montague, P. R. (2008). Neuroeconomics: The neurobiology of value-based decision-making. *Nature Reviews Neuroscience, 9,* 545–556.

Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience, 2,* 79–87.

Reverberi, C., Görgen, K., & Haynes, J.-D. (2012). Compositionality of rule representations in human prefrontal cortex. *Cerebral Cortex, 22,* 1237–1246.

Reynolds, J. R., O'Reilly, R. C., Cohen, J. D., & Braver, T. S. (2012). The function and organization of lateral prefrontal cortex: A test of competing hypotheses. *PLoS One, 7,* e30284.

Ritchey, T. (1991). Analysis and synthesis: On scientific method—Based on a study by Bernhard Riemann. *Systems Research, 8,* 21–41.

Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome, E. D. J., & Beck, L. H. (1956). A continuous performance test of brain damage. *Journal of Consulting Psychology, 20,* 343–350.

Rotge, J.-Y., Lemogne, C., Hinfray, S., Huguet, P., Grynszpan, O., Tartour, E., et al. (2015). A meta-analysis of the anterior cingulate contribution to social pain. *Social Cognitive and Affective Neuroscience, 10,* 19–27.

Russchen, F. T., Amaral, D. G., & Price, J. L. (1985). The afferent connections of the substantia innominata in the monkey, Macaca fascicularis. *Journal of Comparative Neurology, 242,* 1–27.

Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron, 79,* 217–240.

Shenhav, A., Straccia, M. A., Cohen, J. D., & Botvinick, M. M. (2014). Anterior cingulate engagement in a foraging context reflects choice difficulty, not foraging value. *Nature Neuroscience, 17,* 1249–1254.

Shidara, M., & Richmond, B. J. (2002). Anterior cingulate: Single neuronal signals related to degree of reward expectancy. *Science, 296,* 1709–1711.

Silvetti, M., Seurinck, R., & Verguts, T. (2011). Value and prediction error in medial frontal cortex: Integrating the single-unit and systems levels of analysis. *Frontiers in Human Neuroscience, 5,* 75.

Singer, T., Critchley, H. D., & Preuschoff, K. (2009). A common role of insula in feelings, empathy and uncertainty. *Trends in Cognitive Sciences, 13,* 334–340.

Smith, B. J. H., Saaj, C. M., & Allouis, E. (2012). ANUBIS: Artificial neuromodulation using a Bayesian inference system. *Neural Computation, 25,* 221–258.

Stalnaker, T. A., Cooch, N. K., & Schoenbaum, G. (2015). What the orbitofrontal cortex does not do. *Nature Neuroscience, 18,* 620–627.

Stoll, F. M., Fontanier, V., & Procyk, E. (2016). Specific frontal neuronal dynamics contribute to decisions to check. *Nature Communications, 7,* 11990.

Sutton, R. S. (1990). Integrated architectures for learning, planning, and reacting based on approximating dynamic programming. In *Proceedings of the Seventh International Conference on Machine Learning* (pp. 216–224). Austin, TX: Morgan Kaufmann.

Sutton, R. S., & Barto, A. (1990). Time-derivative models of Pavlovian reinforcement. In M. Gabriel & J. Moore (Eds.), *Learning and computational neuroscience: Foundations of adaptive networks* (pp. 497–537). Cambridge, MA: The MIT Press.

Talmi, D., Atkinson, R., & El-Deredy, W. (2013). The feedback-related negativity signals salience prediction errors, not reward prediction errors. *Journal of Neuroscience, 33,* 8264–8269.

Tsuchida, A., & Fellows, L. K. (2008). Lesion evidence that two distinct regions within prefrontal cortex are critical for *n*-back performance in humans. *Journal of Cognitive Neuroscience, 21,* 2263–2275.

Ullsperger, M., Harsay, H. A., Wessel, J. R., & Ridderinkhof, K. R. (2010). Conscious perception of errors and its relation to the anterior insula. *Brain Structure and Function, 214,* 629–643.

Verguts, T., Vassena, E., & Silvetti, M. (2015). Adaptive effort investment in cognitive and physical tasks: A neurocomputational model. *Frontiers in Behavioral Neuroscience, 9,* 57.

Wiech, K., Lin, C., Brodersen, K. H., Bingel, U., Ploner, M., & Tracey, I. (2010). Anterior insula integrates information about salience into perceptual decisions about pain. *Journal of Neuroscience, 30,* 16324–16331.

Yeung, N., Cohen, J. D., & Botvinick, M. M. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review, 111,* 931–959.

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron, 46,* 681–692.