Neuroscience and Biobehavioral Reviews xxx (2013) xxx–xxx

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev

Review

From conflict management to reward-based decision making: Actors and critics in primate medial frontal cortex

Massimo Silvetti[a,**], William Alexander[a,b], Tom Verguts[a], Joshua W. Brown[b,*]

[a] Ghent University, Department of Experimental Psychology, Henri Dunantlaan 2, B-9000 Gent, Belgium
[b] Indiana University, Department of Psychological and Brain Sciences, 1101 E Tenth Street, Bloomington, IN 47405, USA

ARTICLE INFO

Article history:
Received 28 May 2013
Received in revised form 28 October 2013
Accepted 4 November 2013

Keywords:
Anterior cingulate
Reinforcement learning
Conflict monitoring
Cognitive control
Computational modeling

ABSTRACT

The role of the medial prefrontal cortex (mPFC) and especially the anterior cingulate cortex has been the subject of intense debate for the last decade. A number of theories have been proposed to account for its function. Broadly speaking, some emphasize cognitive control, whereas others emphasize value processing; specific theories concern reward processing, conflict detection, error monitoring, and volatility detection, among others. Here we survey and evaluate them relative to experimental results from neurophysiological, anatomical, and cognitive studies. We argue for a new conceptualization of mPFC, arising from recent computational modeling work. Based on reinforcement learning theory, these new models propose that mPFC is an Actor–Critic system. This system is aimed to predict future events including rewards, to evaluate errors in those predictions, and finally, to implement optimal skeletal-motor and visceromotor commands to obtain reward. This framework provides a comprehensive account of mPFC function, accounting for and predicting empirical results across different levels of analysis, including monkey neurophysiology, human ERP, human neuroimaging, and human behavior.

Contents

1. Introduction 00
2. ACC functional anatomy 00
   2.1. ACC as a limbic area 00
   2.2. ACC: motor properties 00
   2.3. ACC connectivity with associative cortical areas and subcortical nuclei 00
   2.4. ACC functional connectivity and cytoarchitecture 00
3. ACC in cognitive neuroscience 00
4. ACC and reward processing: neurophysiology 00
5. A unifying perspective: RL 00
   5.1. RL and the Actor–Critic framework 00
   5.2. ACC as an integrated Actor–Critic system 00
   5.3. A RL-based computational account of ACC functions 00
      5.3.1. The predicted response–outcome (PRO) model 00
      5.3.2. The reward value and prediction model (RVPM) 00
      5.3.3. ACC functions interpreted from a RL-critic perspective 00
6. A need for change: recent ACC findings interpreted from a RL-critic perspective 00
   6.1. ACC as a Critic: single-unit recordings in nonhuman primates 00
   6.2. Role of ACC in conflict monitoring, pain processing, social interactions and attention allocation 00
   6.3. RL and the stability/plasticity dilemma: the role of neuromodulators 00
   6.4. mPFC and RL: alternative theories 00
7. Prediction error beyond the ACC 00

* Corresponding author. Tel.: +1 812 855 9282.
** Corresponding author.
E-mail addresses: massimo.silvetti@ugent.be (M. Silvetti), jwmbrown@indiana.edu (J.W. Brown).

0149-7634/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.neubiorev.2013.11.003

8. Concluding remarks 00
Acknowledgments 00
References 00

1. Introduction

The medial prefrontal cortex (mPFC) has been intensively studied in recent years, both in humans and nonhuman primates. It has been cast as a system for adaptive control of behavior (Ridderinkhof et al., 2004). For example, it has been proposed that mPFC performs a key role in error processing (Critchley et al., 2005), in the estimation of the probability of committing an error (Brown and Braver, 2005), or in estimating the amount of conflict between two or more available options (Botvinick et al., 2001). Despite intense interest, consensus on its role in cognition has not been reached. Besides different theoretical views, there are also differences in experimental paradigms and different anatomical structures in mPFC.

In this paper, we discuss different theories and tasks relevant to mPFC functioning, paying attention to the historical evolution of these findings, and focusing on one specific structure of mPFC, whose function remained for years as much elusive as ubiquitous in behavioral neuroscience: the anterior cingulate cortex (ACC). Then, we discuss the recent application of a machine learning framework on mPFC investigation, namely Reinforcement Learning (RL). Finally, we discuss two very recent and similar neurocomputational models of mPFC. We show how these models, both belonging to the RL framework, are able to account for the varied and disparate data that have been described. Based on these models we develop a framework according to which this area signals a need for change. Finally, we point out novel directions for investigation.

2. ACC functional anatomy

2.1. ACC as a limbic area

A useful way to focus the main functions of the ACC is to place it inside a specific network of sensory–limbic–motor areas, which may integrate limbic and motivational factors with sensorimotor functions (Bush et al., 2000). ACC is classically identified as part of the limbic system (MacLean, 1955; Papez, 1937). This area, including most of the cingulate cortex, is widely connected with the hippocampus via mainly efferent fibers to the parahippocampal gyrus (Nieuwenhuys et al., 1981). The hippocampus connects to the mammillary bodies (part of the hypothalamus) through the fornix. Finally the mammillary bodies are connected to the anterior thalamic nuclei through the mammillothalamic tract, which projects back to ACC, closing the Papez circuit (1937). Electrical stimulation of the ACC evokes both emotional (Meyer et al., 1973) and autonomic responses, such as variation of blood pressure and heart rate, pupil dilation, pyloric contraction and penile erection (Devinsky et al., 1995). From the anatomical viewpoint, the rostral ACC (Brodmann area 25) sends extensive efferents toward both sympathetic (Hurley et al., 1991) and parasympathetic (Terreberry and Neafsey, 1983; Willett et al., 1986; Hurley et al., 1991) nuclei. Although there is a rostro-caudal gradient in ACC from visceral to motor connectivity (Fig. 1), the caudal part of ACC is to some extent able to evoke autonomic responses via connections with BA25 (Vogt and Pandya, 1987). Experimental conditioning paradigms showed that the ACC is necessary to associate appropriate anticipatory autonomic nervous system responses to future stressful stimuli (e.g. Frysztak and Neafsey, 1991; Critchley and Mathias, 2003), suggesting a role of the ACC in regulating autonomic response as a function of effortful cognitive or motor tasks.

Finally, the ACC receives and processes nociceptive information from thalamic nuclei (Vogt et al., 1979). Nociceptive information in ACC does not adhere to somatotopic organization, and it seems linked to avoidance learning and emotional and autonomic response to pain (Gabriel et al., 1991).

2.2. ACC: motor properties

The ACC contains wide populations of motor and premotor neurons, with a caudal-to-rostral gradient (i.e. reversed with respect to the visceromotor gradient), showing maximal density of motor units in its caudal part, located ventrally to the (pre-) supplementary motor area (SMA and pre-SMA) (Dum and Strick, 2002; Fig. 1). Cingulate motor unit targets are the same as those of the dorsolateral premotor cortex: the ACC establishes reciprocal connections with the primary motor cortex and the SMA, and also direct efferents toward the spinal cord (Dum and Strick, 1991; Bates and Goldman-Rakic, 1993). Electrical stimulation of the ACC typically evokes complex motor patterns (Talairach et al., 1973), somatotopically organized and involving the mouth and forelimbs primarily (Wang et al., 2008; Luppino et al., 1991). ACC motor properties are not limited to limbs. Indeed stimulation of this area can evoke also vocalizations, often coordinated with corresponding autonomic reactions in emotional motor responses (Paus, 2001). Microstimulation of bat ACC evoked vocalizations organized in a tonotopic fashion, similarly to the somatotopic organization in primates, providing a highly specialized map of frequencies that are used for echolocation (Gooler and O'Neill, 1987). Interestingly, while autonomic and motor functions of ACC are similar across mammals (including humans), electrical stimulation in humans (in which vocalization evolved into language) seldom evokes vocal behavior (Devinsky et al., 1995). However, ACC lesions can lead to akinetic mutism (Paus, 2001), but only if the lesions are sufficiently widespread (Fellows and Farah, 2005). The examples of humans

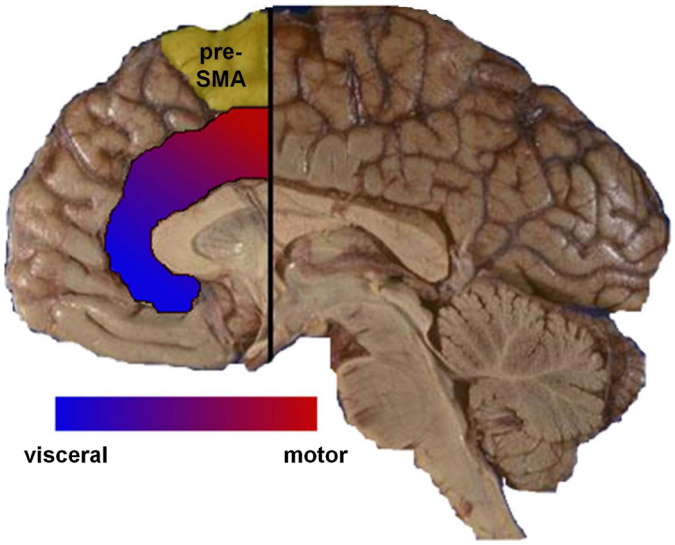

Fig. 1. Anatomical–functional gradients in the ACC. Color nuances indicate the density of neurons specialized in different functions. Pre-SMA area is evidenced as reference point. Vertical line: vertical line on anterior commissure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Please cite this article in press as: Silvetti, M., et al., From conflict management to reward-based decision making: Actors and critics in primate medial frontal cortex. Neurosci. Biobehav. Rev. (2013), http://dx.doi.org/10.1016/j.neubiorev.2013.11.003

and bats indicate how ACC motor functions are finely tuned by evolution in different species.

2.3. ACC connectivity with associative cortical areas and subcortical nuclei

Besides the connections with visceromotor and skeletomotor systems, the ACC is also widely connected with associative cortical areas, such as the anterior insula, the dorsolateral prefrontal cortex (DLPFC), the extrastriate visual areas and the parietal areas (Devinsky et al., 1995; Margulies et al., 2007). Functional connectivity studies suggested that cortico-cortical ACC connections are also organized according to a rostro-caudal gradient (Margulies et al., 2007; Beckmann et al., 2009) (Fig. 2). Caudal ACC is connected mainly with sensorimotor areas; more rostrally, the connectivity shifts toward the DLPFC and inferior parietal lobule (IPL); and finally, the most rostral ACC part is connected with orbitofrontal (OFC) and temporopolar cortex (Margulies et al., 2007). Wide reciprocal connections were shown also between ACC and basal ganglia (Haber et al., 1995; Kunishio and Haber, 1994; Selemon and Goldman-Rakic, 1985). Again, a rostro-caudal organization is observed, with a wider connectivity between rostral ACC and ventral striatum (nucleus accumbens) and between caudal ACC and dorsal striatum (Kunishio and Haber, 1994). Finally, the ACC is reciprocally connected with the catecholaminergic mesencephalic nuclei. This area originates efferents to and receives dopaminergic afferents from the ventral tegmental area (VTA) (Geisler et al., 2007; Oades and Halliday, 1987), and also has reciprocal connections with the noradrenergic locus coeruleus (LC) (Aston-Jones and Cohen, 2005; Jones and Moore, 1977; Samuels and Szabadi, 2008). The ACC is also one of the few regions to influence the midbrain dopamergic cells via the striatum (especially the striosomes, subregions within the striatum characterized by differences in cortical input and by response to neurochemical markers; Eblen and Graybiel, 1995). Although the number of direct connections from cortex to LC is low, both medial and lateral prefrontal cortices strongly modulate LC activity, by projecting efferents to LC surrounding areas that are rich in LC dendritic terminations (Jodo et al., 1998; Samuels and Szabadi, 2008).

2.4. ACC functional connectivity and cytoarchitecture

From the literature we surveyed it emerges that the ACC is a hub in which multimodal information converges, providing multimodal output that includes emotional and autonomic regulations. Another striking feature of this area is its organization in anatomical–functional gradients, with marked limbic-autonomic functions in its rostral part and sensorimotor functions in its caudal part, with a central part that is mainly connected with supramodal associative cortical areas such as the DLPFC and the temporo-parietal junction (TPJ). Gradient-based organization has also been proposed in cognitive neuroscience studies, with the identification of cognitive and affective divisions (Bush et al., 2000; but see Shackman et al., 2011). Finally, additional insights on ACC function come from histological analysis. The ACC consists of agranular cortex (therefore made mainly of pyramidal cells), showing a type of neurons called spindle cells (Allman et al., 2011; Economo et al., 2008). The spindle cells are a type of bipolar neurons found only in the ACC of Hominidae, Elephants and Cetaceans, i.e. those species equipped with the largest brains. Like pyramidal neurons, by which spindle cells are typically surrounded, they are specialized for fast input–output integration, with projection targets outside the cingulate cortex. The presence of these cells only in species with very large brains, in conjunction with their computational features, suggests an evolutionary role for them in maintaining a constant speed of information exchange when the spatial distance between brain areas becomes very large (Allman et al., 2011). In summary, the cytoarchitectonics of ACC provide convergent evidence for the idea that this area is a hub where multimodal signals are quickly processed, and the result of such processing is sent to strongly influence a wide variety of brain areas.

3. ACC in cognitive neuroscience

Cognitive neuroscience research proceeded for a long time almost independently from the anatomo-functional findings described above. The ACC rose to the interest of cognitive science with early electroencephalography (EEG) work highlighting its involvement in error detection (Falkenstein et al., 1991; Gehring et al., 1993; Posner and Dehaene, 1994). The error related negativity (ERN) is an EEG event-related potential that develops during the response period of error trials. Subsequent studies showed that an analogous component, but with a smaller amplitude, can also be observed for correct trials (correct related negativity, CRN) (Falkenstein et al., 2000; Vidal et al., 2000). Source analysis demonstrated that both ERN and CRN have a common generator located in the ACC (Roger et al., 2010), a finding that has been confirmed with fMRI (Jessup et al., 2010). Some years after the first studies on ACC and error detection, further research identified additional experimental conditions able to evoke strong ACC responses. Increased ACC activity is observed when subjects are presented with incongruent stimuli cueing multiple, mutually incompatible responses (e.g. in a Stroop task, indicating the color of the word “RED” written in blue ink) (Bush et al., 1998; Scheffers and Coles, 2000; Van Veen et al., 2001). Further studies showed that ACC activity was greater in response incongruency rather than in perceptual or semantic incongruency (Van Veen and Carter, 2005). Moreover, analogous with the rostro-caudal gradient from the analysis of ACC connectivity, ACC responsiveness to incongruency appears to be modulated by the emotional content of the incongruent information, with highly emotional incongruent stimuli evoking activation in more rostral zone of the ACC (Bush et al., 2000; but see Shackman et al., 2011) and more “cognitive” incongruencies in caudal ACC. These results led to the formulation of the conflict monitoring theory of (caudal) ACC (Botvinick et al., 2001). According to this perspective, the ACC encodes the amount of conflict, defined as the simultaneous activation of several possible responses.

While the conflict model has gained a great deal of influence due to its computational elegance and ability to account for a number of effects observed within ACC, subsequent theories have been developed that challenge the conflict monitoring theory. In a revision of the classical error detection theory, Brown and Braver (2005) proposed that ACC activity reflects the estimated probability of committing an error. In contrast to the conflict model, the Error Likelihood model predicts that ACC activity increases with the likelihood that, within a given context, a behavioral error will be committed, regardless of whether or not an error actually occurs. This prediction was tested using a modified version of the stop signal task in which, rather than merely refraining from responding, subjects were instead asked to make a response different from that which was initially cued (by means of a “change” signal). Consistent with the error likelihood model, increased ACC activity was observed in correctly solved trials associated with a higher probability of error, even on those trials in which a subject did not receive the change signal. However, this pattern was reversed in error trials: ACC activity following response errors was greater for task conditions in which the likelihood of committing an error was low, as compared to high-error likelihood conditions (Brown and Braver, 2005). This latter finding, incompatible with both the error likelihood and conflict accounts of ACC, would not be explained computationally for several years, as described in later sections.

Please cite this article in press as: Silvetti, M., et al., From conflict management to reward-based decision making: Actors and critics in primate medial frontal cortex. Neurosci. Biobehav. Rev. (2013), http://dx.doi.org/10.1016/j.neubiorev.2013.11.003

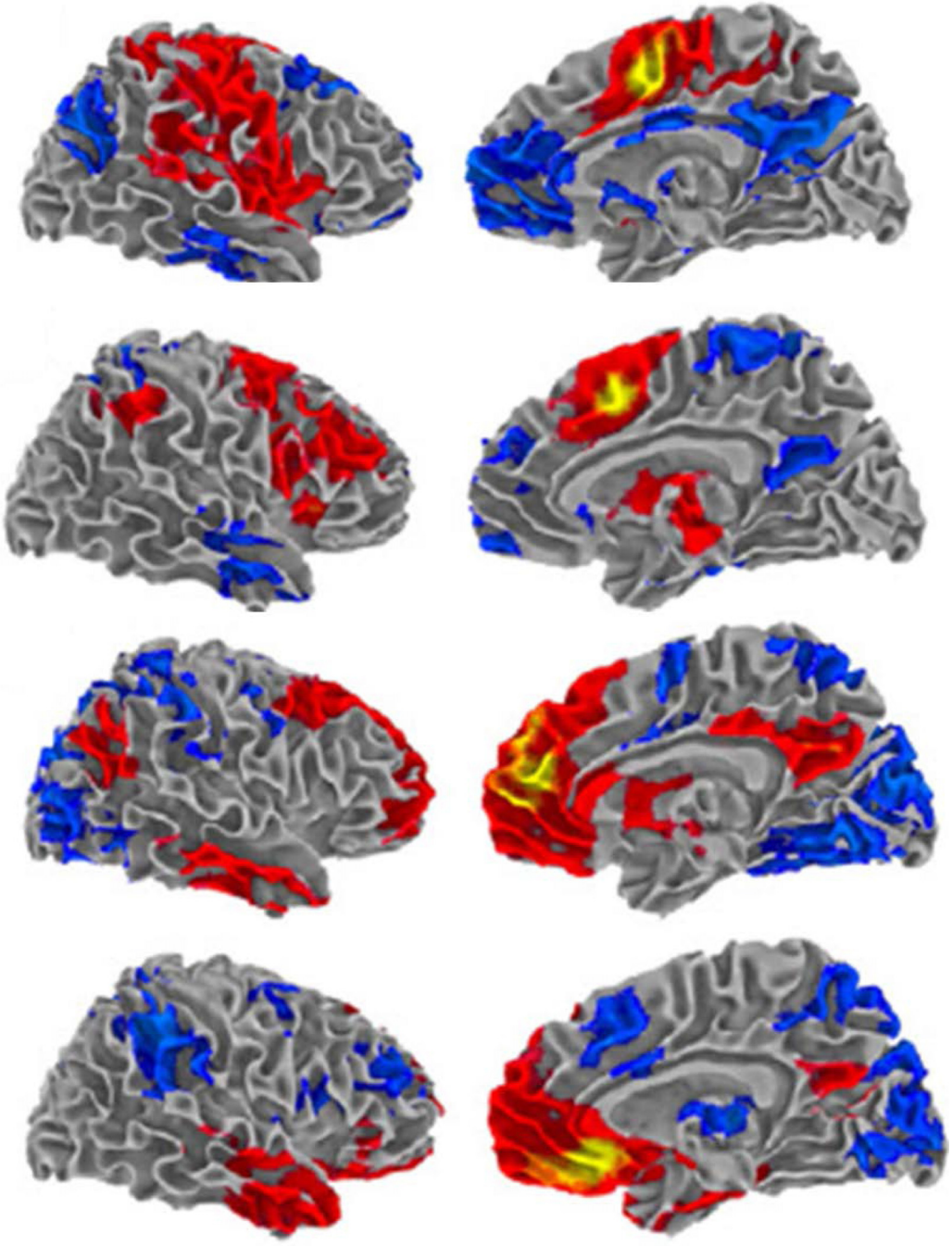

Fig. 2. ACC functional connectivity. Four different ACC regions of interest (ROIs; yellow areas right column) are differently connected with other cortical areas, following a caudal–rostral organization. Red: positive metabolic correlations; Blue: negative metabolic correlations. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)
Modified from Margulies et al. (2007).

Deriving from the longstanding clinical observation of akinetic mutism following ACC lesions (Németh et al., 1988), a further theory on ACC cognitive functions is that it energizes the cognitive system when effort needs to be exerted (e.g. Kouneiher et al., 2009). Despite the imprecision of these concepts, it is an empirical finding that situations requiring high effort robustly activate ACC (e.g. Sohn et al., 2007; Krebs et al., 2012). Finally, a recent cognitive theory on ACC functions suggests that this area could be involved in the response to environmental volatility as a signal to modulate learning rates (Behrens et al., 2007). Here, volatility refers to non-stationarity, i.e. the extent to which the probabilities linking situations and outcomes in the environment fluctuate across time. The latter perspective (Holroyd and Coles, 2002; Holroyd et al., 2005) contained a theoretical assumption that has been crucial for the most recent developments about ACC computational functions, namely its involvement in reinforcement learning (RL). Due to the growing experimental evidence on RL processing by ACC, and the theoretical maturation that this computational field recently reached (Sutton and Barto, 1998), RL reveals to be very promising in the investigation on ACC functions. For these reasons it will be the main theoretical stream guiding the rest of this review.

Please cite this article in press as: Silvetti, M., et al., From conflict management to reward-based decision making: Actors and critics in primate medial frontal cortex. Neurosci. Biobehav. Rev. (2013), http://dx.doi.org/10.1016/j.neubiorev.2013.11.003

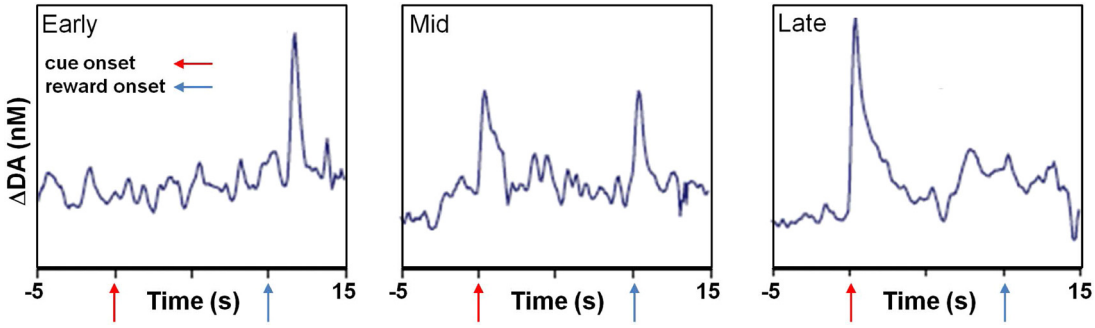

Fig. 3. Response temporal shifting of dopaminergic neurons in the VTA. During the early stage of a conditioning paradigm, dopamine release is locked on reward onset (left plot). After a period of transition (central plot), dopamine release is locked to cue onset (right plot).
Modified from Stuber et al. (2008).

4. ACC and reward processing: neurophysiology

The mosaic composing the wide variety of ACC functions became even more complex with the publication of several single-cell recording studies on reward processing in this area. Although the involvement of ACC in conditional learning has been recognized for at least two decades (Frysztak and Neafsey, 1991), only recently have data from single-cell recording revealed neural populations providing specific RL information coding in the ACC. Initial single-unit neurophysiological studies found ACC neural populations whose activity was a function of the prediction of future rewards given a specific state of the environment (e.g. an external cue, such as a light) or a specific action planned by the animal (such as pulling a joystick) (Amiez et al., 2006; Matsumoto et al., 2007). These neurons coded for global value of expected reward, i.e. their discharge rate was modulated by both reward probability and expected reward magnitude. Furthermore, ACC cells incorporate timing information, such that their activity increased as the time of anticipated reward drew closer (Shidara and Richmond, 2002). Other studies revealed neurons coding for the difference between such predictions and actual environmental outcomes (Amiez et al., 2005; Matsumoto et al., 2007). This variable is called prediction error (PE), and it can be positive, if the outcome is better than expected, or negative, when the outcome is worse than expected. Some cells also code unsigned prediction errors, i.e. the absolute value of the prediction error (Hayden et al., 2011a). Continued neurophysiological research on ACC reward processing in the following years revealed neural populations exhibiting firing patterns representing many combinations of reward prediction and PE coding, intermixed with units presenting specific coding for one type of information (Kennerley et al., 2011). Both single-unit recording and lesion studies have also shown that the ACC contains neural populations estimating expected costs to be paid in order to obtain a reward (Kennerley et al., 2011; Rudebeck et al., 2006), and that costs are estimated in terms of both expected effort to engage and delay to wait in order to obtain a reward. One of the main features of ACC is that its neural populations estimate the value of a specific action or stimulus after cost discounting, i.e. integrating the information about both reward magnitude/probability and estimated costs (Kennerley et al., 2011). Single cell recordings in ACC also revealed neurons whose activity was modulated by reward prediction (Matsumoto et al., 2003).

5. A unifying perspective: RL

In the following paragraphs we will summarize the core aspects of the RL framework. We will show how this theoretical approach is able in principle to provide a unified explanation to several experimental findings about the ACC, from the microscopic-single cells level to the macroscopic-behavioral findings. Finally we will describe two recent neurocomputational models of ACC that implemented the RL based approach to the ACC functions and created new perspectives and predictions.

5.1. RL and the Actor–Critic framework

RL is a theoretical framework from the field of machine learning (Sutton and Barto, 1998). Here, its main aim is to find strategies for optimizing a given goal (e.g. maximizing reward). A secondary aim is to account for (human and nonhuman) behavioral data and formalize classical theories of learning. During the 1990s, the identification of the mesolimbic dopaminergic pathway as the main macrocircuit involved in reward coding, promoted the application of the RL formalism to model reward neurophysiology. Schultz and colleagues (Schultz, 1998; Schultz et al., 1997) used the RL framework to explain the temporal shift from primary reward to the onset of reward predicting cues, which was found for the activity of brainstem dopaminergic neurons (in ventral tegmental area, VTA and substantia nigra, SN; see Fig. 3). The simplest RL approach consists of the formulation of reward expectation (V) given a specific state of the environment (S), which is updated by comparison with the actual environmental outcome (i.e. the actual achievement or not of a reward R). The result of this comparison is defined to be the prediction error (δ). This rule can be expressed as:

$$V(S)_t = V(S)_{t-1} + \alpha\delta_t \quad (1)$$

where $V(S)_t$ is the value (reward expectation) of the environmental state S at time t, α is a parameter defining the learning rate, and δ_t represents the prediction error at time t, which is formally defined as:

$$\delta_t = R_t - V(S)_{t-1} \quad (2)$$

where R_t is the reward achieved at time t.

Eqs. (1) and (2) describe an implementation of what is defined as the Critic system in RL, i.e. a system deputized to evaluate environmental states. Together these equations comprise the Rescorla–Wagner model of classical conditioning. As originally formulated, the Rescorla–Wagner model describes associative learning at the trial level; environmental states and their associated outcomes are, from the model's perspective, presented simultaneously. However, it is well-established that the temporal structure of individual trials, including duration of stimulus and feedback, temporal contiguity, and event ordering can influence associative learning in a manner that is not captured by the Rescorla–Wagner model. This led to a generalization of the Rescorla–Wagner rule: the TD learning algorithm (Sutton, 1988). This algorithm allowed learning not only by comparing a prediction with external feedback (which may or may not be immediately available), but additionally

Please cite this article in press as: Silvetti, M., et al., From conflict management to reward-based decision making: Actors and critics in primate medial frontal cortex. Neurosci. Biobehav. Rev. (2013), http://dx.doi.org/10.1016/j.neubiorev.2013.11.003

by comparing the current prediction with an earlier prediction. In this case the learning signal is the TD prediction error (denoted as $\delta^{TD)}$. The TD error signal can be written as follows:

$$\delta_t^{TD} = R_t + \gamma V(S)_t - V(S)_{t-1} \quad (3)$$

where γ is a discount factor, reflecting the loss in value of rewards as a function of delay. During simulated conditioning experiments, the TD error is initially observed only at the time when an unpredicted reward R is presented. In the course of a simulated experiment, as the TD model learns to associate task stimuli with future reward, the TD error at the time the reward occurs attenuates, and instead is observed at the presentation of the first stimulus that reliably predicts future reward delivery. Although initially conceived as a temporal generalization of models of associative learning, the TD algorithm effectively became a neural model when it was observed that the activity of midbrain dopamine neurons appeared to reflect TD prediction errors (Schultz et al., 1993). TD models are a generalization of Rescorla–Wagner models in the sense that the latter can only estimate actions/states value that just precede the final outcome (the delivery or not of reward). Conversely, TD models can estimate values of actions or environmental states far away from the final outcome, as they can update values by comparing them with other value estimations (δ_t^{TD}), without the need of reaching the final outcome (e.g. estimating how good is a specific position inside a maze, although such a position is not directly near the maze exit).

The models described above pertain to a critic system, which evaluates cues in the environment in terms of discounted future rewards that are predicted by those cues. Note that "environment" is interpretable very broadly as anything outside the critic system itself, and thus may also contain context representations, actions, etc. (cf. Sutton and Barto, 1998). This critic system must be coupled with an Actor, a system deputized to make decisions as a function of the reward predictions formulated by the Critic. In machine learning the Actor is often modeled by simple algorithms (e.g. Softmax) that introduce a certain amount of stochasticity, dependent on a "temperature" parameter, in selecting the action associated with the highest reward expectation. For higher temperature values, behavior tends to be more exploratory in nature, allowing the system to sample broadly from a state space. Low temperature values encourage exploitative behavior in which the system more frequently selects the action it predicts to be most valuable.

5.2. ACC as an integrated Actor–Critic system

Comparing the descriptions of ACC functional anatomy (Section 2) and RL basics, we hypothesize that the ACC is an Actor–Critic system for action selection based on outcome expectations in general and reward expectations in particular. The functional organization of ACC according to a rostro-caudal gradient suggests that the most rostral and caudal portions of the ACC (BA 24 and 32) perform the role of Actor, providing reward-based motor commands. These areas are respectively specialized in visceromotor and skeletal-motor control. The mid-third of the ACC (anterior rostral cingulate zone RCZa, also in BA 24 and 32) (Picard and Strick, 1996), on the other hand, appears to perform mainly evaluative functions, suggesting the role of Critic (Fig. 4). The RCZ (Critic) provides the caudal ACC and the rostral ACC (Actors) with value information to make adaptive choices and to fine tune the Actor parameters (e.g. change its learning rate or temperature parameter). As described below, the wide connectivity between ACC and brainstem also suggests a regulation of Actor parameters (e.g. learning rate or temperature) by the Critic system via catecholaminergic modulation, for example by long term potentiation (TLP) facilitation due to norepinephrinic modulatory effect (Izumi and Zorumski, 1999; Katsuki et al., 1997). A recent fMRI study (Nee et al., 2011) also suggested such an ACC

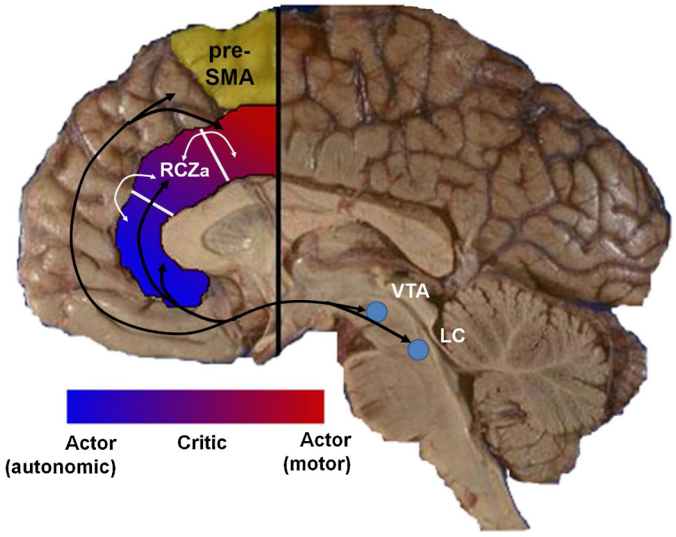

Fig. 4. ACC subdivision according to the Actor–Critic framework. The Critic module (RCZa) provides outcome predictions to the Actor modules (white bidirectional arrows), which select motor output according to them. The Critic also modulates the activity of catecholaminergic brainstem nuclei, which in turn regulate cortical activity (including Critic and Actor themselves). An alternative hypothesis is that the Critic directly modulates the Actors' functions.

organization. It systematically investigated the rostro-caudal organization of mPFC by studying the activation subsequent to errors, incongruency and task switching. The RCZa was activated by expectation violations across tasks, confirming the Critic role of this area. Caudal regions of mPFC were involved specifically in response selection (Actor function), and finally, the most rostral region was specifically activated by errors. As we will describe in the next sections, the Critic system is multi-componential and deployed across the brain, and the ACC Critic functions are only a part of this general evaluative mechanism. The same must be said about the Actor system, which cannot be confined to ACC motor portions; instead, it extends to dorsolateral prefrontal and subcortical areas, with which the ACC is widely connected.

5.3. A RL-based computational account of ACC functions

Recently, two neuro-computational models attempted to create a bridge connecting the wide variety of experimental findings on the ACC: from the microscopic level of single cell recording to the macroscopic level of cognitive neurosciences (Alexander and Brown, 2011; Silvetti et al., 2011). These models are both based on RL related findings from single cell recordings data in monkey mPFC. The core feature of their framework is that they focused on the Critic function of mPFC. In the next two sections we will provide a short description of both models.

5.3.1. The predicted response–outcome (PRO) model

The architecture of the PRO model (Alexander and Brown, 2011) consists of three modules: a Critic, an Actor, and a response–outcome associator that provides an interface between the first two. In the PRO model, the Critic (Fig. 5) implements a temporal-difference algorithm (Montague et al., 1996; Sutton and Barto, 1998) that learns predictions (V) for several possible action/outcome conjunctions contingent on task stimuli. A crucial aspect of the PRO model is that predictions are learned for outcomes regardless of affective valence (contrasted with more common formulations in which RL models learn to predict the value of future rewards). As in other models based on RL, learning is driven by positive and negative prediction errors which are used to update

Please cite this article in press as: Silvetti, M., et al., From conflict management to reward-based decision making: Actors and critics in primate medial frontal cortex. Neurosci. Biobehav. Rev. (2013), http://dx.doi.org/10.1016/j.neubiorev.2013.11.003

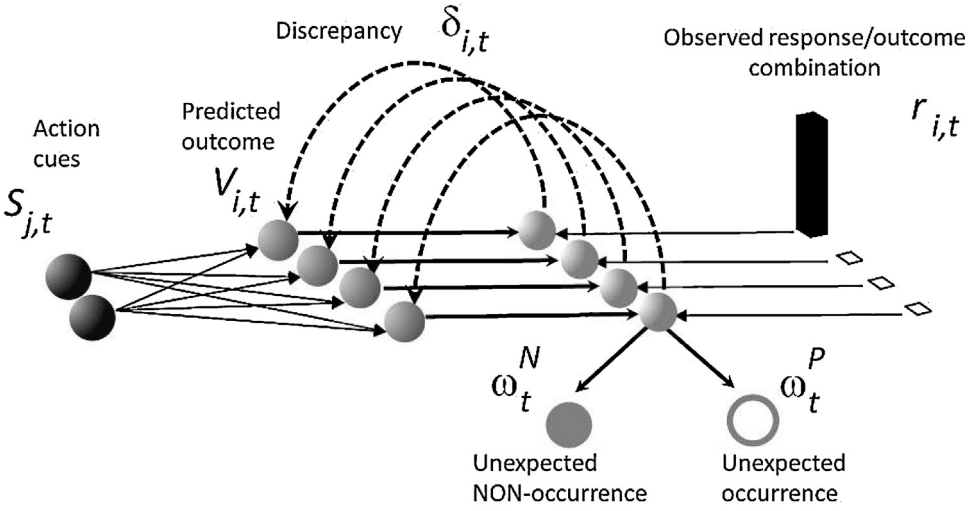

Fig. 5. Schema summarizing the critic component of the PRO model structure. Action cues *S* drive a set of outcome predictions *V*, which are compared against actual outcomes *r*. The resulting TD error signal δ provides a basis for positive (ω^P) and negative (ω^N) rectified prediction error signals.
Adapted with permission from Alexander and Brown (2011).

adjustable weights between task-related stimuli (*S*) and outcome predictions (*V*); however, since the PRO model learns predictions of both affectively positive and negative outcomes, the interpretation of prediction errors differs from standard RL. In typical RL models, positive and negative prediction errors are associated with outcomes that are better than or worse than expected, respectively. In the PRO model, however, positive prediction errors (ω^P) indicate that an event occurred that was not predicted, while negative prediction errors (ω^N) indicate that a predicted outcome failed to occur as expected. In the context of cognitive control, outcomes are conceived of as feedback related to performance on a task (e.g. error or correct). Alexander and Brown propose that specifically this second component (ω^N) is responsible for most of the classical cognitive neuroscience results on ACC functions. They proposed that the two ω components together modulate the learning of a response–outcome associator which also predicts action/outcome combinations, but without a timing component. Finally, this outcome representation system can then provide proactive control through inhibition of an actor system if the outcome representation system predicts that the actor will choose a low-valence action. This approach allows the PRO model to simulate a variety of effects such as incongruency and error effects, as described in earlier work (Alexander and Brown, 2011). For example, incongruency effects as in the Stroop or Flanker tasks result from a greater aggregate level of prediction signals (*V*) about the likely outcomes of the task (both correct and error outcomes are likely). Likewise, error effects reflect a surprising discrepancy (aggregate ω^N) between an expected outcome (usually correct performance) versus an actual outcome (e.g. an error).

5.3.2. The reward value and prediction model (RVPM)

The RVPM (Silvetti et al., 2011, Fig. 6) similarly proposes that ACC plays the role of a Critic system. The core of the model consists of three modules, one composed of neural units coding for external cues, one representing the ACC itself and the third modeling the brainstem dopaminergic nuclei (VTA). Based on neurophysiological work (Amiez et al., 2006, 2005; Matsumoto et al., 2007), the ACC module consists of three types of cells, namely value cells (*V*), coding for reward expectations, but also positive (δ^+) and negative (δ^+) prediction error cells (the analog of ω units in PRO model), which are used for updating the value estimates (Eq. (2)). The ACC module computes reward expectations given a specific state of the environment (cue) and a subsequent occurrence of a reward (VTA signal). Here environment can be intended as both what is external to the organism (e.g. a stimulus), and what is external to the ACC (e.g. an action representation). In contrast with the PRO model, the RVPM estimates exclusively reward expectations and not outcome expectations. Therefore the activity of *V* unit has always a positive valence, and this also slightly changes the meaning of δ units, given that δ^+will encode for unexpected reward, while $\delta^-$for unexpected lack of reward. The authors further proposed that *V* and δ signals are transmitted down to dopaminergic brainstem structures (e.g. VTA, SN), driving the temporal difference dynamics typical of these structures (Schultz, 1998). Through widespread cortical and subcortical efferents, these brainstem structures can thus update learning elsewhere in the brain. As described more in detail in Silvetti et al. (2011), the RVPM model assumes that reward signals can be generated also after the correct execution of a task where there is no external reinforcer (e.g. Stroop task). This assumption is corroborated by neuroimaging data, showing the activation of the mesolimbic dopaminergic system after correct responses in tasks where no explicit reward was provided (Koepp et al., 1998; Satterthwaite et al., 2012).

5.3.3. ACC functions interpreted from a RL-critic perspective

Both models ascribe a critic function to ACC. Most importantly, both explain "cognitive" data from an RL perspective (Fig. 7), thus naturally connecting data across species (human and monkey) and tasks (cognitive or reward-related). To account for error effects (Fig. 7a), both models propose that an error tends to be less expected, as unexpectedness (of any kind) activates ACC. In general, this finding accounts also for the fact that the size of the ERN depends on the probability of accuracy (Holroyd and Coles, 2002; Nunez Castellar et al., 2010), and also for the fact that the feedback-related negativity (FRN) reverses when errors are more frequent than correct responses (hence correct responses are more surprising; Ferdinand et al., 2012; Oliveira et al., 2007) and similar fMRI findings (Jessup et al., 2010).

To account for error likelihood effects (Brown and Braver, 2005; Fig. 7b), the models replicated the results that correct but high error likelihood trials activate ACC more strongly than correct but low error likelihood trials. At the same time both models showed a reversed ACC response for error trials, with higher activation for erroneous low error likelihood trials. As for the error-related ACC response, the error likelihood effect on error signals is explained by the two models by means of δ activity, which codes the discrepancy

Please cite this article in press as: Silvetti, M., et al., From conflict management to reward-based decision making: Actors and critics in primate medial frontal cortex. Neurosci. Biobehav. Rev. (2013), http://dx.doi.org/10.1016/j.neubiorev.2013.11.003

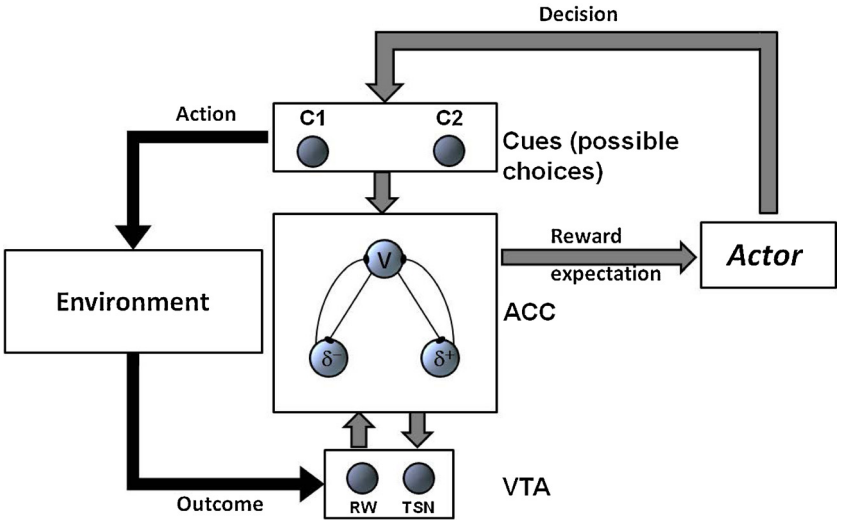

Fig. 6. Schema summarizing the RVPM – Actor structure (gray arrows) and its interaction with the external environment (black arrows). The ACC module receives input from both the Cue module (coding for actions or external cues) and the VTA module (providing the reward signal, RW unit). The action-reward expectation is then sent to an Actor that selects the decision according to ACC evaluation and an internal policy. A second ACC output consists in a recurrent signal to the VTA module. The ACC signal is integrated by the Temporal Shifting Unit (TSN), generating the shifting of dopaminergic activity form reward period to cue period. Once the action is performed, the environment provides the outcome, closing the loop agent-environment.

between expectations (e.g. low likelihood of committing an error) and the real outcomes (e.g. error response).

The effect of ACC activation in volatile environments (Behrens et al., 2007; Fig. 7c) is explained by more frequent errors in prediction in volatile environments, when the systems are forced to continuously update the mapping between environmental states and their likely outcomes (Silvetti et al., 2013).

While both the RVPM and PRO models account for the preceding results through the identical computational mechanism of discrepancy between expected and actual outcomes, the models diverge in their explanations of the effects of conflict, in which increased ACC activity is observed on trials in which multiple, mutually incompatible responses are cued (incongruent) vs. trials on which a single response is cued (congruent). According to the RVPM (Fig. 7d), incongruent trials are, generally, slower and more error-prone, resulting in increased δ units activation for incongruent trials that are performed successfully. In contrast, the PRO model suggests that when incongruent response cues are present as in the case of conflict, ACC signals that both the correct and the incorrect responses are likely. The net result is greater overall activity in the prediction layer, relative to the case of congruent trials, in which only a single (correct) outcome is likely. This implies that when multiple responses and outcomes are possible even in the absence of conflict, then ACC activity should be elevated, and recent fMRI results are consistent with this prediction (Brown, 2009). This is consistent with an outcome prediction signal that may influence decision-making by the Actor (Jahn et al., 2011).

The divergent accounts of conflict effects in ACC are one example in which the RVPM and PRO models might drive future research in the area of cognitive control. Although both models agree on the basic computational mechanism underlying ACC activity, that of discrepancy between predictions and outcomes, differences between the two models suggest additional work that needs to be done in order to more adequately characterize the function of ACC. One such difference is how information regarding the timing of outcomes relative to a predictive stimulus is incorporated within ACC. In the RVPM model, temporal information is learned in parallel with value predictions through a spectral timing mechanism (Brown et al., 1999), suggesting that the activity of at least some brain regions projecting to ACC (and possibly ACC itself) should be observed only around the time an outcome is expected, and that activity in such regions should be independent of how surprising (or not) that outcome is. In contrast, the PRO model represents timing information as a discounted function of the time remaining before an outcome is expected. In this case, regions encoding temporal information should reflect a gradual increase in activity beginning at stimulus onset and peaking around the time an outcome is predicted.

Besides this difference, the two models consider also the incongruency-related ACC activation as deriving from two different mechanisms. Rather than being mutually exclusive, these mechanisms likely represent two related features of the ACC. The RVPM attributes the incongruency-related activation to a combined negative and positive prediction error signal occurring during the outcome period. In contrast, the PRO model suggests that such activation is linked to the outcome anticipation period, due to the simultaneous activation of different units coding for the outcome expectation of competing responses.

Finally, an additional difference between the two models is the role of value in ACC activity. As reviewed above, the concept of value has played a central role in research on ACC, and the RVPM model reflects this in characterizing the region as being primarily concerned with predicting the value of future outcomes and signaling discrepancies between expected and actual value. The PRO model, on the other hand, suggests that ACC predicts multiple future outcomes without regard to the valence associated with those outcomes, implying a role for ACC in more general predictive vs specifically value-based learning and decision-making. Future research on this issue may provide definitive evidence regarding which model is more consistent with the evidence, or, as seems more likely, suggest how value predictions from the RVPM model and outcome predictions from the PRO model might be reconciled to provide a more robust and comprehensive perspective on the role of ACC.

6. A need for change: recent ACC findings interpreted from a RL-critic perspective

In the previous section, we described classical data that were crucial for the models to capture. In recent years, a stream of novel findings on ACC has emerged, some of them appearing after publication of the two models. Some of these data naturally fit in

Please cite this article in press as: Silvetti, M., et al., From conflict management to reward-based decision making: Actors and critics in primate medial frontal cortex. Neurosci. Biobehav. Rev. (2013), http://dx.doi.org/10.1016/j.neubiorev.2013.11.003

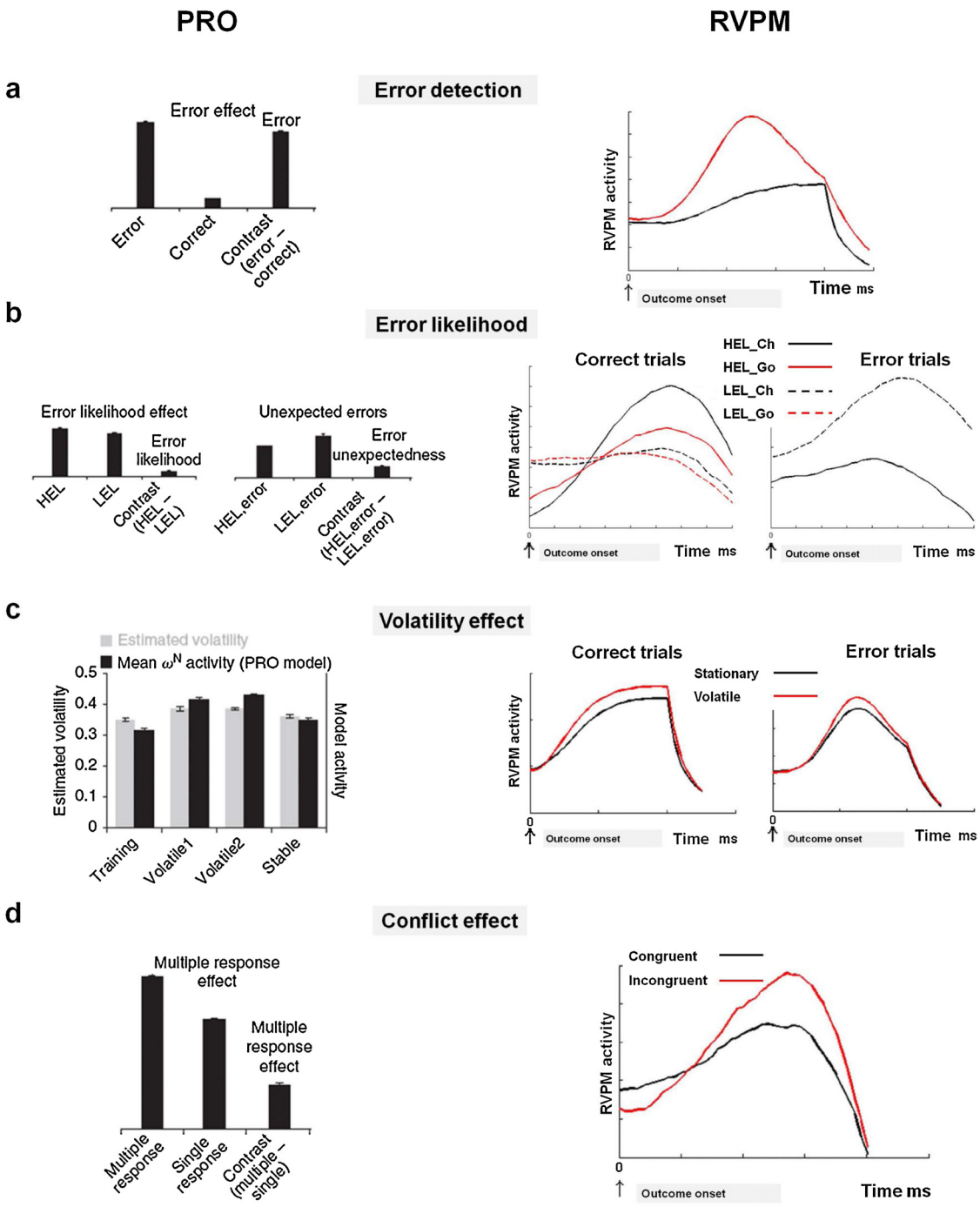

Fig. 7. Simulation results of both PRO (left column) and RVPM (right column) in four different experimental paradigms. PRO model activity is plotted as mean activation over multiple model iterations, while RVPM activity is plotted as the activation of the whole ACC module per iteration (average time course for each trial type). (a) Model activity during correct and error trials, when error trials are less likely that correct. The error related activity is due to negative prediction error signal. (b) Effects of error likelihood and unexpected errors. For correct trials high error likelihood (HEL) trials evoked higher activation than low error likelihood (LEL) trials, while errors on LEL trials elicit higher activation than errors on HEL trials. (c) Effects of environmental volatility. Both the RVPM and PRO models explain increased ACC activity in environments with high volatility as the increased frequency of prediction errors relative to stationary environments. (d) Conflict effects. Although both models reproduce conflict-type effects commonly observed in ACC, the mechanism differs. In the PRO model, conflict effects are due to multiple concurrent predictions regarding possible outcomes. In RVPM, conflict effects arise from errors in prediction.

Adapted from Alexander and Brown (2011) and Silvetti et al. (2011).

the RL framework, others less easily so. A significant challenge in attempting to integrate these findings into a unified framework lies in the diversity of experimental methodologies used to probe ACC, as well as the variety of theoretical views brought to bear on interpreting observed activity. This challenge is perhaps best exemplified in a recent review (Cole et al., 2009) in which it was suggested that the absence of clear conflict-related activity in monkey ACC (commonly measured with single-unit electrophysiology) may result from interspecies differences between monkeys and humans (in which ACC activity is commonly measured using EEG or fMRI, i.e. methodologies that reflect that activity of ensembles of neurons). As described in the previous section, both the

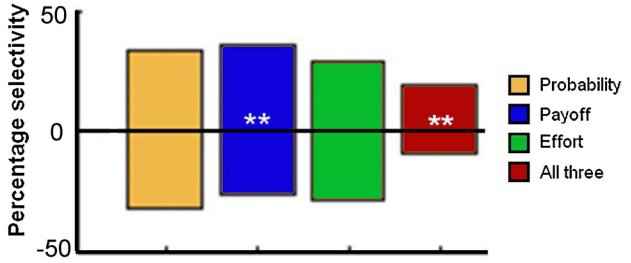

Fig. 8. Percentage of ACC neurons coding positively (positive y axis) or negatively (negative y axis) for reward probability, reward magnitude (payoff), effort and the combination of all the three (general reward value). Only for reward magnitude and general reward value, the ACC showed a preference for positive coding (e.g. the higher the discharge rate the higher the reward value; asterisks). For the other two variables, neural populations with positive and negative coding were equally represented.

Modified from Kennerley et al. (2011).

RVPM and PRO models are able to provide an account by which this discrepancy might be resolved without resorting to putative differences in the functional role of ACC between species. Similarly, theoretical viewpoints associated with different subfields such as social or affective neuroscience may lead to interpretations of ACC as being involved in e.g. processing another person's beliefs (Apps et al., 2013) vs. detecting painful or aversive stimuli (Chandrasekhar et al., 2008), interpretations that do not bear an immediate resemblance to one another. Drawing on the models and these recent data, we here propose a critic framework of ACC consistent with, yet significantly extending, the earlier proposals, and discuss how it may reconcile a wide range of findings on ACC.

6.1. ACC as a Critic: single-unit recordings in nonhuman primates

Kennerley et al. (2011) performed single-cell recordings while monkeys performed a probabilistic selection task. This work is important for understanding ACC functions, because the authors performed a census of the different types of ACC neurons. The main datum emerging from this study is that in the ACC there are neural populations specialized in processing various aspects of both expected costs and rewards. More precisely, possible outcomes are coded by both positive and negative valence (like the units of PRO model, Fig. 8 first three bars), and neural populations coding for costs (effort) consequent to specific actions were found (Fig. 8, third bar). It is worth noting that effort coding by ACC was found also in humans (Krebs et al., 2012; Sohn et al., 2007), as was risk avoidance coding (Fukunaga et al., 2012). Finally, the authors showed the presence of neural populations multiplexing all the other signals to obtain a general cost-discounted and positive coding of reward expectations and prediction error (like the units of RVPM, Fig. 8, fourth bar). These findings are consistent with both PRO and RVPM architectures, showing how the differences between the two models are due to different focusing on specific aspects of ACC role of Critic system.

6.2. Role of ACC in conflict monitoring, pain processing, social interactions and attention allocation

Other findings, consistent with the overall RL framework, are by Grinband et al. (2011). This study demonstrated that conflict effects in human ACC reverse when congruency (traditionally, the manipulation of choice for conflict) was pitted against response time. In particular, the contrast of slow but congruent versus fast but incongruent trials robustly activated the RCZ-ACC. Grinband et al. interpreted their data as consistent with a "time on task" model, according to which the total amount of time spent on the task determines the amount of ACC activation. This finding is in agreement with the RVPM, according to which even when accuracy is constant, a difference in reaction times (RTs) is still able to evoke prediction error activity. Similarly, the PRO model interprets the correlation between time on task and ACC activity as reflecting the failure of a response to occur within the time frame it is usually observed. Similarly, Carp et al. (2010) demonstrated that when RT was included in the regression model, the congruency effect in mPFC disappeared. However, when they did a similar analysis controlling error versus correct trials by RT differences, the effect did not disappear. Consistent with the RL models, the accuracy effect (ACC more active during error than during correct trials) does not seem to result from RT differences (see also Jessup et al., 2010). It must be noted that Yeung et al. (2011) demonstrated that when an actor system spends more time on a task, the conflict level rises proportionally, demonstrating that it is difficult to separate conflict from time on task. Overall, however, the PRO and RVPM models can account for all of the effects accounted for by the conflict model as well as other effects that are not accounted for by the conflict model. In addition to its involvement in monitoring future outcomes of decisions and actions, ACC has also been implicated in processing prediction error linked to sensory input. EEG studies have identified a negative ERP component within ACC present when a stimulus is unexpectedly absent from a sequence (Crottaz-Herbette and Menon, 2006; Waberski et al., 2001), similar to the mismatch negativity (MMN) observed in sensory cortices. A recent paper finds that the network underlying error processing significantly overlaps with brain regions that signal the occurrence of novel events, including ACC, anterior insula, and inferior frontal gyrus (Wessel et al., 2012). These findings seem to indicate the involvement of the mPFC in computation of prediction errors regarding environmental states (state-prediction error) rather than reward-prediction error. Nonetheless, occurrence of unexpected environmental states typically activates dorsolateral brain regions (Corbetta et al., 2000; Doricchi et al., 2009; Glascher et al., 2010); mPFC involvement in state prediction error can be considered still an open issue. Although these findings are not specifically addressed by the RVPM and PRO models as initially conceived, the central concept of both models, prediction and surprise, is easily extended to include predicting sensory events and signaling discrepancies between expected and actual stimuli.

The ACC is also typically active in pain prediction and actual pain processing, and is regarded as the core of the so-called 'pain matrix', the distributed network of regions in the brain commonly engaged by painful or aversive stimuli. In addition to ACC, regions contributing to the processing of pain include somatosensory cortex, insula, and prefrontal cortex (Bantick et al., 2002; Iannetti and Mouraux, 2010), areas which overlap with the 'task-positive' network, suggesting that pain processing and cognitive control may have certain computational mechanisms in common. In the framework of RL, ACC activity related to anticipated and actual pain may reflect a general prediction and evaluation mechanism suggested by the RVPM and PRO model, while activity in other regions may indicate a subject's affective response to painful stimuli.

Studies from the perspective of social neuroscience have likewise observed activity in ACC, where the region has been implicated in processing negative affect related to social exclusion (Eisenberger et al., 2003), predicting the actions of others (Lamm et al., 2007), and theory of mind (Fletcher et al., 1995). At least one fMRI study has identified spatially segregated areas within ACC showing differential effects for evaluating discrepancies between one's own predictions regarding likely and actual outcomes vs. a confederate's predictions (Apps et al., 2013), suggesting that the role of ACC in prediction and evaluation may constitute a general function that operates across multiple modalities and contexts.

Please cite this article in press as: Silvetti, M., et al., From conflict management to reward-based decision making: Actors and critics in primate medial frontal cortex. Neurosci. Biobehav. Rev. (2013), http://dx.doi.org/10.1016/j.neubiorev.2013.11.003

While a number of findings regarding ACC activity from different sub-fields of neuroscience can be interpreted within the RL framework, the functional import of the ACC signals remains very much an open question. Both the RVPM and PRO model suggest specific functional roles for the Critic component to influence its adjacent (rostral and caudal) ACC actors. The critic itself consists of a prediction or value signal (V) and a prediction error signal (δ or ω). These may each have different roles; for example the prediction signals may themselves bias decision-making at a cognitive level (i.e. proactive control) (Fukunaga et al., 2012). Likewise, the prediction error may drive error correction actions or learning signals, i.e. reactive control (Braver et al., 2007). With regard to the prediction or value signals, one possibility is that the Critic delivers value information for the Actor regions to choose an optimal (i.e. high-value) response (as implemented in the RVPM). In this case, value signals may originate in the Critic and also be found in the Actor units they influence. By the same logic, prediction error signals may also be found in the Actor units they influence. To the extent that these signals are found in both the Actor and Critic, it may be difficult to distinguish the actor vs. critic empirically in different brain regions.

Finally, a large body of literature (Bush et al., 2000) has established an extensive link between ACC activity and attentional processes. In particular Crottaz-Herbette and Menon (2006) provided evidence for ACC involvement in attentional resource allocation, by modulation of modal areas specifically involved in task execution. The RL framework proposed in this manuscript may provide a link unifying these areas of study. Indeed, recent single-unit recordings studying attention in rats (Bryden et al., 2011) have found results consistent with a role for surprise signals generated by ACC in updating how attention is allocated. Allocation of attentive resources can be considered as a form of decision by the Actor made on the basis of the Critic evaluation.

6.3. RL and the stability/plasticity dilemma: the role of neuromodulators

Another (not mutually exclusive) possibility is that the Critic adjusts Actor parameters governing how the system integrates, evaluates, and acts upon new information obtained from an environment (Doya, 2002). These parameters may include learning rate, temperature, and time scale. In the case of learning rate, one method to optimize an agent's behavior is to modulate learning rate depending on the perceived volatility of an environment (Behrens et al., 2007). In relatively static environments, low learning rates protect valid information regarding the statistical contingencies of the environment from the influence of chance events. On the other hand, for environments in which contingencies are frequently shifting, high learning rates allow an agent to integrate new information rapidly. One possible role for the Critic, then, is to detect the need for change in learning rate. To detect such a need for change, it must compute values and evaluate how computed values deviate from expected values. In this regard, ACC may play a vital role in resolving the stability-plasticity dilemma (Carpenter and Grossberg, 1988). Briefly, the dilemma states that a system should be sometimes stable (resistant to irrelevant information overwriting valuable knowledge) but plastic at other times (when environmental contingencies change and new information has to be learned). Due to its role of Critic and its connections to both cortical and subcortical areas, we propose that ACC is excellently placed for this purpose, and indeed the PRO model can simulate effects of learning rate changes with increasing non-stationarity in the environment (Alexander and Brown, 2011; Behrens et al., 2007). The exact details of this regulation remain to be discovered and both a direct role of the ACC (as modeled in the PRO) or an action mediated by brainstem nuclei (as modeled in the RVPM) are plausible proposals open to further investigation.

In a similar manner, ACC may mediate the tradeoff between exploration and exploitation of an environment (Cohen et al., 2007). The exploration/exploitation tradeoff is typified by foraging tasks in which a foraging animal must choose between continuing to forage in a patch with a known rate of reward, or to explore other patches which may have higher or lower reward rates. If the animal chooses to exploit without sufficiently exploring its environment, it may choose a suboptimal patch. On the other hand, an animal which constantly chooses exploration may fail to remain in a high-yield patch. In theory, the "correct" balance between exploration and exploitation is that which optimizes some value function, e.g. the average reward (Charnov, 1976). In monkey, the activity of neurons in ACC above a particular threshold appears to signal when the animal will choose to forego exploiting a current food source in order to travel to a potentially richer source (Hayden et al., 2011b; Kolling et al., 2012). One mechanism by which ACC may instigate exploration is through descending projections to LC. In RL theory, the temperature parameter controls the tendency of an agent to engage in exploration, and has been associated with the neuromodulator norepinephrine (Doya, 2002). Although it is known that ACC and LC are reciprocally connected, additional research is needed to establish how interactions between the two regions may contribute to exploratory behavior.

6.4. mPFC and RL: alternative theories

The idea that the ACC embodies an RL architecture is not entirely new, although the conceptualizations of how ACC implements a RL framework here differ from previous proposals. Holroyd and Coles (2002) suggested that ACC is trained by mesencephalic dopamine signals to recognize which other brain regions are best suited to generate outputs in a given situation. In this hypothesized role in action selection, ACC functions as part of an Actor component in an Actor–Critic architecture. This is conceptually similar to the mixture of experts model of Jacobs et al. (1991) and overlaps with the proposed role of the basal ganglia in action selection (Redgrave et al., 1999). In another RL-based model, Holroyd et al. (2005) suggested that ACC might use RL signals to learn conjunctions of events that indicate an error has occurred. The PRO and RVPM models differ from previous RL accounts of ACC by positing that the TD/delta signals are computed within the ACC, whereas other models simulate the TD signals as a scalar signal that originates from the mesencephalic dopamine cells.

More recent theoretical work by Holroyd and Yeung (2012) proposes that ACC plays a critical role in selecting and maintaining high-level goals and behaviors in a hierarchical RL (HRL) framework. As in the RVPM and PRO models, the HRL framework outlined by Holroyd and Yeung follows an Actor–Critic architecture. However, in their view the Critic is implemented by a hierarchically organized network comprised of orbitofrontal cortex (OFC) and ventral striatum rather than ACC, while the Actor is implemented by a parallel hierarchy of DLPFC and dorsal striatum. ACC interacts with both Actor and Critic components by providing information about goals to the Critic as context by which to judge actions taken by the Actor, and providing high-level policies consistent with current goals to the Actor in order to instantiate appropriate task sets for action selection. Although some aspects of the HRL framework may be consistent with both the RVPM and PRO models – in particular, the inclusion of DLPFC and dorsal striatum within the Actor component was also suggested by Silvetti et al. (2011) – the question of whether ACC acts as a Critic or as a super-ordinate region coordinating the activities of both Actor and Critic is a key difference. Here we must stress that, according to our models, the Critic function is not necessarily related to the

Please cite this article in press as: Silvetti, M., et al., From conflict management to reward-based decision making: Actors and critics in primate medial frontal cortex. Neurosci. Biobehav. Rev. (2013), http://dx.doi.org/10.1016/j.neubiorev.2013.11.003

selection of specific motor actions, but it could involve also the evaluation of more abstract-higher order data structures, such as general options or even entire representations of the environment. This functional aspect makes our models able to explain also the mPFC-related findings about option selection, without the necessity of attributing to the mPFC the function of "options controller". Future theoretical and empirical work focusing on the participation of ACC in hierarchically organized behavior will be required to successfully distinguish between the two hypotheses.

Another relevant perspective on the ACC role in RL-based decision making comes from Boorman et al. (2013). The authors find a dissociation between ACC and ventromedial prefrontal cortex (vmPFC) in decision-making, with increased ACC activity appearing to reflect future deviations from a default response, while vmPFC encodes the propensity to repeat responses. Moreover the authors report ACC negative coding of long-term value of choices during the outcome period (highest activity for infrequently chosen options), while positive coding in vmPFC (highest activity for chosen option). These findings are consistent with the RL framework developed in this review inasmuch as the RVPM and PRO models learn long-term values and probabilities, respectively.

However, the models offer a different interpretation of increased ACC activity observed for infrequently chosen options during the outcome period. In the Boorman et al. task, default behavior is associated with repeatedly selecting the option with the highest probability of reward; other options, with lower reward probabilities, carry a greater chance of failure. According to the RVPM framework, negative coding of value during the outcome period may reflect average PE signal, that is low for highly expected outcomes and maximal for highly uncertain outcomes (see Silvetti et al., 2013). In the PRO model, this increase in activity with the infrequently chosen option is explained as the simultaneous prediction of the multiple possible outcomes associated with these options, namely reward and failure (note this entails that prediction-related activity is a concave-down function of probability, so that for example two likely events yield greater summed prediction activity than a single likely event). Moreover, the PRO model accounts for increased outcome-related ACC activity prior to a behavioral switch as due to the non-occurrence of an established default event.

Finally, a very recent article (Shenhav et al., 2013) also attempts to integrate RL and cognitive control by suggesting that dorsal ACC computes the expected value of control. This is similar to our own "need for change" proposal – both theories will require explicit computational formulation though to validate their implications.

7. Prediction error beyond the ACC

The two ACC models that we described here are both based on the basic RL concept of prediction error (δ). This signal is necessary to update the connections between the neural units coding for states (or actions) and the units (V) predicting the future outcomes. This kind of mechanism is a neurobiologically refined version of the classic feedback-based idea of learning proposed in cybernetics (Wiener, 1948), where the difference between desired (or predicted) value and the actual outcome was used as feedback signal to minimize error. Signals consistent with prediction-based learning are additionally observed in several brain regions not typically associated with reward processing during tasks that do not involve explicit reward administration (Glascher et al., 2010). It is possible that the mammalian brain contains several predicting systems, some used for the formulation of expectations about future environmental states, independently from reward (Doricchi et al., 2009; Glascher et al., 2010). As such, prediction error would be a component of any learning mechanism in the brain, as proposed by early cybernetics. Additionally, it may be that the reward prediction error computed by ACC is only one aspect of the region's function, and that ACC might also signal discrepancies between predicted and observed environmental states. In this case, ACC might serve as a general purpose critic for multiple brain regions, consistent with the high degree of connectivity of ACC with other cortical and subcortical sites.

Two other general theories on brain functioning could be considered as closely related to the concept of prediction error: the classical theory of Neural Darwinism (Edelman, 1978) and the recent theory of free energy (Friston, 2009). In Neural Darwinism, the selection of the most suitable synaptic links between single neurons and neural populations is analogous to the Darwinian process of natural selection of individual organisms. The fitness of single synapses is determined by a signal that Edelman called value. The general learning rule, based on a reward prediction error (i.e. value) signal, implements such a scheme. There is also overlap between the concept of prediction error and free energy (Friston, 2009). Here, neural ensembles act in order to reduce the discrepancy between a model of the environment encoded in their connectivity and the flow of data from the sensory channels. This is exactly what happens with the reduction of prediction error in both reward and non-reward related circuits. Moreover, such adaptation toward the minimum amount of error must work at several different timescales, in order to optimize the tradeoff between plasticity and stability.

8. Concluding remarks

We have proposed how the enormous variety of experimental findings on mPFC anatomy, physiology and cognitive processing may be reconciled within the conceptual framework of RL. We showed how two recent and similar RL neural models are able to provide a computational account to many of these findings. We also tried to generalize the importance of the computational concept founding the two models, the prediction error, beyond ACC functions and RL processing. Finally, we proposed that formulating predictions and updating them by prediction errors may be a neurobiologically sound and computationally explicit foundation for Neural Darwinism.

Acknowledgments

JWB and WHA were supported in part by the Intelligence Advanced Research Projects Activity (IARPA) via Department of the Interior (DOI) contract number D10PC20023. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DOI or the U.S. Government. MS and TV were supported by Ghent University GOA grant BOF08/GOA/011. WHA was supported by FWO-Flanders Odysseus II Award #G.OC44.13N. We acknowledge the support of Ghent University Multidisciplinary Research Platform "The integrative neuroscience of behavioral control".

References

Alexander, W.H., Brown, J.W., 2011. Medial prefrontal cortex as an action-outcome predictor. Nat. Neurosci. 14, 1338–1344.
Allman, J.M., Tetreault, N.A., Hakeem, A.Y., Manaye, K.F., Semendeferi, K., Erwin, J.M., Park, S., Goubert, V., Hof, P.R., 2011. The von Economo neurons in the frontoinsular and anterior cingulate cortex. Ann. N. Y. Acad. Sci. 1225, 59–71.
Amiez, C., Joseph, J.P., Procyk, E., 2005. Anterior cingulate error-related activity is modulated by predicted reward. Eur. J. Neurosci. 21, 3447–3452.

Please cite this article in press as: Silvetti, M., et al., From conflict management to reward-based decision making: Actors and critics in primate medial frontal cortex. Neurosci. Biobehav. Rev. (2013), http://dx.doi.org/10.1016/j.neubiorev.2013.11.003

Amiez, C., Joseph, J.P., Procyk, E., 2006. Reward encoding in the monkey anterior cingulate cortex. Cereb. Cortex 16, 1040–1055.
Apps, M.A.J., Green, R., Ramnani, N., 2013. Reinforcement learning signals in the anterior cingulate cortex code for others' false beliefs. Neuroimage 64, 1–9.
Aston-Jones, G., Cohen, J.D., 2005. An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance. Annu. Rev. Neurosci. 28, 403–450.
Bantick, S.J., Wise, R.G., Ploghaus, A., Clare, S., Smith, S.M., Tracey, I., 2002. Imaging how attention modulates pain in humans using functional MRI. Brain 125, 310–319.
Bates, J.F., Goldman-Rakic, P.S., 1993. Prefrontal connections of medial motor areas in the rhesus monkey. J. Comp. Neurol. 336, 211–228.
Beckmann, M., Johansen-Berg, H., Rushworth, M.F.S., 2009. Connectivity-based parcellation of human cingulate cortex and its relation to functional specialization. J. Neurosci. 229 (4), 1175–1190.
Behrens, T.E., Woolrich, M.W., Walton, M.E., Rushworth, M.F., 2007. Learning the value of information in an uncertain world. Nat. Neurosci. 10, 1214–1221.
Boorman, E.D., Rushworth, M.F., Behrens, T.E., 2013. Ventromedial prefrontal and anterior cingulate cortex adopt choice and default reference frames during sequential multi-alternative choice. J. Neurosci. 33, 2242–2253.
Botvinick, M., Braver, T.S., Barch, D.M., Carter, C.S., Cohen, J.D., 2001. Conflict monitoring and cognitive control. Psychol. Rev. 108, 624–652.
Braver, T., Gray, J., Burgess, G., 2007. Explaining the many varieties of working memory variation: dual mechanisms of cognitive control. In: Variation in Working Memory.
Brown, J., Bullock, D., Grossberg, S., 1999. How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues. J. Neurosci. 19, 10502–10511.
Brown, J.W., 2009. Conflict effects without conflict in anterior cingulate cortex: multiple response effects and context specific representations. Neuroimage 47, 334–341.
Brown, J.W., Braver, T.S., 2005. Learned predictions of error likelihood in the anterior cingulate cortex. Science 80 (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. J. Neurosci. 31, 18266–18274.
Bush, G., Luu, P., Posner, M.I., 2000. Cognitive and emotional influences in anterior cingulate cortex. Trends Cogn. Sci. 4, 215–222.
Bush, G., Whalen, P.J., Rosen, B.R., Jenike, M.A., McInerney, S.C., Rauch, S.L., 1998. The counting Stroop: an interference task specialized for functional neuroimaging—validation study with functional MRI. Hum. Brain Mapp. 6, 270–282.
Carp, J., Kim, K., Taylor, S.F., Fitzgerald, K.D., Weissman, D.H., 2010. Conditional differences in mean reaction time explain effects of response congruency, but not accuracy, on posterior medial frontal cortex activity. Front Hum. Neurosci. 4, 231.
Carpenter, G., Grossberg, S., 1988. The ART of adaptive pattern recognition by a self-organizing neural network. Computer 21, 77–88.
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.
Charnov, E.L., 1976. Optimal foraging, the marginal value theorem. Theor. Popul. Biol. 9, 129–136.
Cohen, J.D., McClure, S.M., 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. B: Biol. Sci. 362, 933–942.
Cole, M.W., Yeung, N., Freiwald, W.A., Botvinick, M., 2009. Cingulate cortex: diverging data from humans and monkeys. Trends Neurosci. 32, 566–574.
Corbetta, M., Kincade, J.M., Ollinger, J.M., McAvoy, M.P., Shulman, G.L., 2000. Voluntary orienting is dissociated from target detection in human posterior parietal cortex. Nat. Neurosci. 3, 292–297.
Critchley, H.D., Mathias, C.J., 2003. Blood pressure, attention and cognition: drivers and air traffic controllers. Clin. Auton. Res. 13, 399–401.
Critchley, H.D., Tang, J., Glaser, D., Butterworth, B., Dolan, R.J., 2005. Anterior cingulate activity during error and autonomic response. Neuroimage 27, 885–895.
Crottaz-Herbette, S., Menon, V., 2006. Where and when the anterior cingulate cortex modulates attentional response: combined fMRI and ERP evidence. J. Cogn. Neurosci. 18, 766–780.
Devinsky, O., Morrell, M.J., Vogt, B.A., 1995. Contributions of anterior cingulate cortex to behaviour. Brain 118 (Pt. 1), 279–306.
Doricchi, F., Macci, E., Silvetti, M., Macaluso, E., 2009. Neural correlates of the spatial and expectancy components of endogenous and stimulus-driven orienting of attention in the posner task. Cereb. Cortex 20, 1574–1585.
Doya, K., 2002. Metalearning and neuromodulation. Neural Netw. 15, 495–506.
Dum, R.P., Strick, P.L., 1991. The origin of corticospinal projections from the premotor areas in the frontal lobe. J. Neurosci. 11, 667–689.
Dum, R.P., Strick, P.L., 2002. Motor areas in the frontal lobe of the primate. Physiol. Behav. 77, 677–682.
Eblen, F., Graybiel, A.M., 1995. Highly restricted origin of prefrontal cortical inputs to striosomes in the macaque monkey. J. Neurosci. 15, 5999–6013.
Economo, C., von Koskinas, G., Triarhou, L., 2008. Atlas of Cytoarchitectonics of the Adult Human Cerebral Cortex. Karger Publishers, Basel.
Edelman, G., 1978. The Mindful Brain. MIT Press, Cambridge, MA.
Eisenberger, N.I., Lieberman, M.D., Williams, K.D., 2003. Does rejection hurt? An FMRI study of social exclusion. Science 302, 290–292.
Falkenstein, M., Hohnsbein, J., Hoormann, J., Blanke, L., 1991. Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. Electroencephalogr. Clin. Neurophysiol. 78, 447–455.
Falkenstein, M., Hoormann, J., Christ, S., Hohnsbein, J., 2000. ERP components on reaction errors and their functional significance: a tutorial. Biol. Psychol. 51, 87–107.
Fellows, L.K., Farah, M.J., 2005. Dissociable elements of human foresight: a role for the ventromedial frontal lobes in framing the future, but not in discounting future rewards. Neuropsychologia 43, 1214–1221.
Ferdinand, N.K., Mecklinger, A., Kray, J., Gehring, W.J., 2012. The processing of unexpected positive response outcomes in the mediofrontal cortex. J. Neurosci. 32, 12087–12092.
Fletcher, P.C., Happé, F., Frith, U., Baker, S.C., Dolan, R.J., Frackowiak, R.S., Frith, C.D., 1995. Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. Cognition 57, 109–128.
Friston, K., 2009. The free-energy principle: a rough guide to the brain? Trends Cogn. Sci. 13, 293–301.
Frysztak, R.J., Neafsey, E.J., 1991. The effect of medial frontal cortex lesions on respiration, "freezing" and ultrasonic vocalizations during conditioned emotional responses in rats. Cereb. Cortex 1, 418–425.
Fukunaga, R., Brown, J.W., Bogg, T., 2012. Decision making in the Balloon Analogue Risk Task (BART): anterior cingulate cortex signals loss aversion but not the infrequency of risky choices. Cogn. Affect. Behav. Neurosci. 12, 479–490.
Gabriel, M., Vogt, B.A., Kubota, Y., Poremba, A., Kang, E., 1991. Training-stage related neuronal plasticity in limbic thalamus and cingulate cortex during learning: a possible key to mnemonic retrieval. Behav. Brain Res. 46, 175–185.
Gehring, W.J., Goss, B., Coles, M.G.H., Meyer, D.E., Donchin, E., 1993. A neural system for error detection and compensation. Psychol. Sci. 4, 385–390.
Geisler, S., Derst, C., Veh, R.W., Zahm, D.S., 2007. Glutamatergic afferents of the ventral tegmental area in the rat. J. Neurosci. 27, 5730–5743.
Glascher, 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.
Gooler, D.M., O'Neill, W.E., 1987. Topographic representation of vocal frequency demonstrated by microstimulation of anterior cingulate cortex in the echolocating bat, *Pteronotus parnelli* parnelli. J. Comp. Physiol. A. 161, 283–294.
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.
Haber, S.N., Kunishio, K., Mizobuchi, M., Lynd-Balta, E., 1995. The orbital and medial prefrontal circuit through the primate basal ganglia. J. Neurosci. 15, 4851–4867.
Hayden, B.Y., Heilbronner, S.R., Pearson, J.M., Platt, M.L., 2011a. Surprise signals in anterior cingulate cortex: neuronal encoding of unsigned reward prediction errors driving adjustment in behavior. J. Neurosci. 31, 4178–4187.
Hayden, B.Y., Pearson, J.M., Platt, M.L., 2011b. Neuronal basis of sequential foraging decisions in a patchy environment. Nat. Neurosci. 14, 933–939.
Holroyd, C.B., Coles, M.G., 2002. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. Psychol. Rev. 109, 679–709.
Holroyd, C.B., Yeung, N., 2012. Motivation of extended behaviors by anterior cingulate cortex. Trends Cogn. Sci. 16, 122–128.
Holroyd, C.B., Yeung, N., Coles, M.G.H., Cohen, J.D., 2005. A mechanism for error detection in speeded response time tasks. J. Exp. Psychol. Gen. 134, 163–191.
Hurley, K.M., Herbert, H., Moga, M.M., Saper, C.B., 1991. Efferent projections of the infralimbic cortex of the rat. J. Comp. Neurol. 308, 249–276.
Iannetti, G.D., Mouraux, A., 2010. From the neuromatrix to the pain matrix (and back). Exp. Brain Res. 205, 1–12.
Izumi, Y., Zorumski, C.F., 1999. Norepinephrine promotes long-term potentiation in the adult rat hippocampus in vitro. Synapse 31, 196–202.
Jacobs, R., Jordan, M., Nowlan, S., Hinton, G., 1991. Adaptive mixtures of local experts. Neural Comput. 3, 79–87.
Jahn, A., Nee, D.E., Brown, J.W., 2011. The neural basis of predicting the outcomes of imagined actions. Front. Neurosci. 5, 128.
Jessup, R.K., Busemeyer, J.R., Brown, J.W., 2010. Error effects in anterior cingulate cortex reverse when error likelihood is high. J. Neurosci. 30, 3467–3472.
Jodo, E., Chiang, C., Aston-Jones, G., 1998. Potent excitatory influence of prefrontal cortex activity on noradrenergic locus coeruleus neurons. Neuroscience 83, 63–79.
Jones, B.E., Moore, R.Y., 1977. Ascending projections of the locus coeruleus in the rat. II. Autoradiographic study. Brain Res. 127, 25–53.
Katsuki, H., Izumi, Y., Zorumski, C.F., 1997. Noradrenergic regulation of synaptic plasticity in the hippocampal CA1 region. J. Neurophysiol. 77, 3013–3020.
Kennerley, S.W., Behrens, T.E., Wallis, J.D., 2011. Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. Nat. Neurosci. 14, 1581–1589.
Koepp, M.J., Gunn, R.N., Lawrence, A.D., Cunningham, V.J., Dagher, A., Jones, T., Brooks, D.J., Bench, C.J., Grasby, P.M., 1998. Evidence for striatal dopamine release during a video game. Nature 393, 266–268.
Kolling, N., Behrens, T.E.J., Mars, R.B., Rushworth, M.F.S., 2012. Neural mechanisms of foraging. Science 336, 95–98.
Kouneiher, F., Charron, S., Koechlin, E., 2009. Motivation and cognitive control in the human prefrontal cortex. Nat. Neurosci. 12, 939–945.
Krebs, R.M., Boehler, C.N., Roberts, K.C., Song, A.W., Woldorff, M.G., 2012. The involvement of the dopaminergic midbrain and cortico–striatal–thalamic

Please cite this article in press as: Silvetti, M., et al., From conflict management to reward-based decision making: Actors and critics in primate medial frontal cortex. Neurosci. Biobehav. Rev. (2013), http://dx.doi.org/10.1016/j.neubiorev.2013.11.003

circuits in the integration of reward prospect and attentional task demands. Cereb. Cortex 22, 607–615.
Kunishio, K., Haber, S.N., 1994. Primate cingulostriatal projection: limbic striatal versus sensorimotor striatal input. J. Comp. Neurol. 350, 337–356.
Lamm, C., Fischer, M.H., Decety, J., 2007. Predicting the actions of others taps into one's own somatosensory representations–a functional MRI study. Neuropsychologia 45, 2480–2491.
Luppino, G., Matelli, M., Camarda, R.M., Gallese, V., Rizzolatti, G., 1991. Multiple representations of body movements in mesial area 6 and the adjacent cingulate cortex: an intracortical microstimulation study in the macaque monkey. J. Comp. Neurol. 311, 463–482.
MacLean, P.D., 1955. The limbic system ("visceral brain") and emotional behavior. AMA. Arch. Neurol. Psychiatry 73, 130–134.
Margulies, D.S., Kelly, A.M.C., Uddin, L.Q., Biswal, B.B., Castellanos, F.X., Milham, M.P., 2007. Mapping the functional connectivity of anterior cingulate cortex. Neuroimage 37, 579–588.
Matsumoto, K., Suzuki, W., Tanaka, K., 2003. Neuronal correlates of goal-based motor selection in the prefrontal cortex. Science 80 (301), 229–232.
Matsumoto, M., Matsumoto, K., Abe, H., Tanaka, K., 2007. Medial prefrontal cell activity signaling prediction errors of action values. Nat. Neurosci. 10, 647–656.
Meyer, G., McElhaney, M., Martin, W., 1973. Stereotactic cingulotomy with results of acute stimulation and serial psychological testing. In: Laitinen, L., Livingsone, K. (Eds.), Surgical Approaches in Psychiatry. MTP, Baltimore, Lancaster, pp. 39–58.
Montague, P.R., Dayan, P., Sejnowski, T.J., 1996. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J. Neurosci. 16, 1936–1947.
Nee, D.E., Kastner, S., Brown, J.W., 2011. Functional heterogeneity of conflict, error, task-switching, and unexpectedness effects within medial prefrontal cortex. Neuroimage 54, 528–540.
Németh, G., Hegedüs, K., Molnár, L., 1988. Akinetic mutism associated with bicingular lesions: clinicopathological and functional anatomical correlates. Eur. Arch. Psychiatry Neurol. Sci. 237, 218–222.
Nieuwenhuys, R., Voogd, J., van Huijzen, C., 1981. The Human Central Nervous System. A Synopsis and Atlas. Springer-Verlag, Berlin.
Nunez Castellar, E., Kuhn, S., Fias, W., Notebaert, W., 2010. Outcome expectancy and not accuracy determines posterror slowing: ERP support. Cogn. Affect Behav. Neurosci. 10, 270–278.
Oades, R.D., Halliday, G.M., 1987. Ventral tegmental (A10) system: neurobiology. 1. Anatomy and connectivity. Brain Res. 434, 117–165.
Oliveira, F.T., McDonald, J.J., Goodman, D., 2007. Performance monitoring in the anterior cingulate is not all error related: expectancy deviation and the representation of action–outcome associations. J. Cogn. Neurosci. 19, 1994–2004.
Papez, J., 1937. A proposed mechanism of emotion. Arch. Neurol. Psychiatry 38 (4), 725–743.
Paus, T., 2001. Primate anterior cingulate cortex: where motor control, drive and cognition interface. Nat. Rev. Neurosci. 2, 417–424.
Picard, N., Strick, P.L., 1996. Motor areas of the medial wall: a review of their location and functional activation. Cereb. Cortex 6, 342–353.
Posner, M.I., Dehaene, S., 1994. Attentional networks. Trends Neurosci. 17, 75–79.
Redgrave, P., Prescott, T.J., Gurney, K., 1999. The basal ganglia: a vertebrate solution to the selection problem? Neuroscience 89, 1009–1023.
Ridderinkhof, K.R., Ullsperger, M., Crone, E.A., Nieuwenhuis, S., 2004. The role of the medial frontal cortex in cognitive control. Science 80 (306), 443–447.
Roger, C., Benar, C.G., Vidal, F., Hasbroucq, T., Burle, B., 2010. Rostral cingulate zone and correct response monitoring: ICA and source localization evidences for the unicity of correct- and error-negativities. Neuroimage 51, 391–403.
Rudebeck, P.H., Walton, M.E., Smyth, A.N., Bannerman, D.M., Rushworth, M.F., 2006. Separate neural pathways process different decision costs. Nat. Neurosci. 9, 1161–1168.
Samuels, E.R., Szabadi, E., 2008. Functional neuroanatomy of the noradrenergic locus coeruleus: its roles in the regulation of arousal and autonomic function part I: principles of functional organisation. Curr. Neuropharmacol. 6, 235–253.
Satterthwaite, T.D., Ruparel, K., Loughead, J., Elliott, M.A., Gerraty, R.T., Calkins, M.E., Hakonarson, H., Gur, R.C., Gur, R.E., Wolf, D.H., 2012. Being right is its own reward: load and performance related ventral striatum activation to correct responses during a working memory task in youth. Neuroimage 61, 723–729.
Scheffers, M.K., Coles, M.G., 2000. Performance monitoring in a confusing world: error-related brain activity, judgments of response accuracy, and types of errors. J. Exp. Psychol. Hum. Percept. Perform. 26, 141–151.
Schultz, W., 1998. Predictive reward signal of dopamine neurons. J. Neurophysiol. 80, 1–27.
Schultz, W., Apicella, P., Ljungberg, T., 1993. Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. J. Neurosci. 13, 900–913.
Schultz, W., Dayan, P., Montague, P.R., 1997. A neural substrate of prediction and reward. Science 80 (275), 1593–1599.
Selemon, L.D., Goldman-Rakic, P.S., 1985. Longitudinal topography and interdigitation of corticostriatal projections in the rhesus monkey. J. Neurosci. 5, 776–794.
Shackman, A.J., Salomons, T.V., Slagter, H.A., Fox, A.S., Winter, J.J., Davidson, R.J., 2011. The integration of negative affect, pain and cognitive control in the cingulate cortex. Nat. Rev. Neurosci. 12, 154–167.
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.
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. Front. Hum. Neurosci. 5, 75.
Silvetti, M., Seurinck, R., Verguts, T., 2013. Value and prediction error estimation account for volatility effects in ACC: a model-based fMRI study. Cortex 49, 1627–1635, http://dx.doi.org/10.1016/j.cortex.2012.05.008.
Sohn, M.-H., Albert, M.V., Jung, K., Carter, C.S., Anderson, J.R., 2007. Anticipation of conflict monitoring in the anterior cingulate cortex and the prefrontal cortex. Proc. Natl. Acad. Sci. U. S. A. 104, 10330–10334.
Stuber, G.D., Klanker, M., de Ridder, B., Bowers, M.S., Joosten, R.N., Feenstra, M.G., Bonci, A., 2008. Reward-predictive cues enhance excitatory synaptic strength onto midbrain dopamine neurons. Science 321 (5896), 1690–1692.
Sutton, R.S., 1988. Learning to predict by the method of temporal differences. Mach. Learn. 3, 9–44.
Sutton, R.S., Barto, A.G., 1998. Reinforcement Learning: An Introduction. MIT Press, Cambridge, MA.
Talairach, J., Bancaud, J., Geier, S., Bordas-Ferrer, M., Bonis, A., Szikla, G., Rusu, M., 1973. The cingulate gyrus and human behaviour. Electroencephalogr. Clin. Neurophysiol. 34, 45–52.
Terreberry, R.R., Neafsey, E.J., 1983. Rat medial frontal cortex: a visceral motor region with a direct projection to the solitary nucleus. Brain Res. 278, 245–249.
Van Veen, V., Carter, C.S., 2005. Separating semantic conflict and response conflict in the Stroop task: a functional MRI study. Neuroimage 27, 497–504.
Van Veen, V., Cohen, J.D., Botvinick, M.M., Stenger, V.A., Carter, C.S., 2001. Anterior cingulate cortex, conflict monitoring, and levels of processing. Neuroimage 14, 1302–1308.
Vidal, F., Hasbroucq, T., Grapperon, J., Bonnet, M., 2000. Is the "error negativity" specific to errors? Biol. Psychol. 51, 109–128.
Vogt, B.A., Pandya, D.N., 1987. Cingulate cortex of the rhesus monkey: II. Cortical afferents. J. Comp. Neurol. 262, 271–289.
Vogt, B.A., Rosene, D.L., Pandya, D.N., 1979. Thalamic and cortical afferents differentiate anterior from posterior cingulate cortex in the monkey. Science 204, 205–207.
Waberski, T.D., Kreitschmann-Andermahr, I., Kawohl, W., Darvas, F., Ryang, Y., Gobbelé, R., Buchner, H., 2001. Spatio-temporal source imaging reveals subcomponents of the human auditory mismatch negativity in the cingulum and right inferior temporal gyrus. Neurosci. Lett. 308, 107–110.
Wang, Y., Matsuzaka, Y., Mushiake, H., Shima, K., 2008. Spatial distribution of cingulate cortical cells projecting to the primary motor cortex in the rat. Neurosci. Res. 60, 406–411.
Wessel, J.R., Danielmeier, C., Morton, J.B., Ullsperger, M., 2012. Surprise and error: common neuronal architecture for the processing of errors and novelty. J. Neurosci. 32, 7528–7537.
Wiener, N., 1948. Cybernetics or Control and Communication in the Animal and the Machine. Hermann & Cie Editeurs, Paris.
Willett, C.J., Gwyn, D.G., Rutherford, J.G., Leslie, R.A., 1986. Cortical projections to the nucleus of the tractus solitarius: an HRP study in the cat. Brain Res. Bull. 16, 497–505.
Yeung, N., Cohen, J.D., Botvinick, M.M., 2011. Errors of interpretation and modeling: a reply to Grinband et al. Neuroimage 57, 316–319.

Please cite this article in press as: Silvetti, M., et al., From conflict management to reward-based decision making: Actors and critics in primate medial frontal cortex. Neurosci. Biobehav. Rev. (2013), http://dx.doi.org/10.1016/j.neubiorev.2013.11.003