Published in final edited form as:
Annu Rev Neurosci. 2010 ; 33: 173–202. doi:10.1146/annurev.neuro.051508.135256.

Emotion, Cognition, and Mental State Representation in Amygdala and Prefrontal Cortex

C. Daniel Salzman[1,2,3,4,5,6] and **Stefano Fusi**[1]

C. Daniel Salzman: cds2005@columbia.edu; Stefano Fusi: sf2237@columbia.edu

[1]Department of Neuroscience, Columbia University, New York, NY 10032

[2]Department of Psychiatry, Columbia University, New York, NY 10032

[3]W.M. Keck Center on Brain Plasticity and Cognition, Columbia University, New York, NY 10032

[4]Kavli Institute for Brain Sciences, Columbia University, New York, NY 10032

[5]Mahoney Center for Brain and Behavior, Columbia University, New York, NY 10032

[6]New York State Psychiatric Institute, New York, NY 10032

Abstract

Neuroscientists have often described cognition and emotion as separable processes implemented by different regions of the brain, such as the amygdala for emotion and the prefrontal cortex for cognition. In this framework, functional interactions between the amygdala and prefrontal cortex mediate emotional influences on cognitive processes such as decision-making, as well as the cognitive regulation of emotion. However, neurons in these structures often have entangled representations, whereby single neurons encode multiple cognitive and emotional variables. Here we review studies using anatomical, lesion, and neurophysiological approaches to investigate the representation and utilization of cognitive and emotional parameters. We propose that these mental state parameters are inextricably linked and represented in dynamic neural networks composed of interconnected prefrontal and limbic brain structures. Future theoretical and experimental work is required to understand how these mental state representations form and how shifts between mental states occur, a critical feature of adaptive cognitive and emotional behavior.

Keywords

neurophysiology; orbitofrontal cortex; value; reward; aversive; reinforcement learning

Introduction

The past century has witnessed a debate concerning the nature of emotion. When the brain is confronted with a stimulus that evokes emotion, does it first respond by activating a range of visceral and behavioral responses, which are only then followed by the conscious experience of emotion? For example, when we encounter a threatening snake, does autonomic reactivity, as well as behaviors such as freezing or fleeing, emerge prior to the feeling of fear? This view, championed by the psychologists William James and Carle Lange around

Errata: An online log of corrections to *Annual Review of Neuroscience* articles may be found at http://neuro.annualreviews.org/

Disclosure Statement: The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

the turn of the twentieth century (James 1884, 1894; Lange 1922), has attracted renewed interest because of the influential work of Damasio and colleagues (Damasio 1994). Alternatively, do visceral and behavioral responses occur as a result of central processing in the brain—processing that gives rise to emotional feelings—which then regulates or controls a variety of bodily responses [a possibility raised decades ago by Walter Cannon (1927) and Philip Bard (1928)]?

Neuroscientists have often sidestepped this debate by operationally defining a particular aspect of emotion—e.g., learning about fear—and using a specific behavioral or physiological assay—e.g., freezing—to investigate the neural basis of the process (Salzman et al. 2005). This approach is agnostic about which response comes first: the visceral and behavioral expression of emotion or the feeling of emotion. But it has proven powerful in helping to identify and characterize the neural circuitry responsible for specific aspects of emotional expression and regulation. These investigations have shown that one brain area, the amygdala, plays a vital role in many emotional processes (Baxter & Murray 2002, Lang & Davis 2006, LeDoux 2000, Phelps & LeDoux 2005) and that the amygdala and its interconnections with the prefrontal cortex (PFC) likely underlie many aspects of the interactions between emotion and cognition (Barbas & Zikopoulos 2007, Murray & Izquierdo 2007, Pessoa 2008, Price 2007).

Today, we still lack a resolution to the original debate concerning the relationship between emotional feelings and the bodily expression of emotions, in large part because both viewpoints appear to be supported in some circumstances. Emotional feelings do not necessarily involve visceral and behavioral components and vice versa (Lang 1994). But neurobiological advances—in particular, emerging data on the intimate relationship between the PFC and limbic areas such as the amygdala—begin to suggest a solution. As discussed below, the amygdala is essential for many of the visceral and behavioral expressions of emotion; meanwhile, the PFC—especially its medial and orbital regions—appears to be responsible for many of the cognitive aspects of emotional responses. However, recent studies suggest that both the functional and the electrophysiological characteristics of the amygdala and the PFC overlap and intimately depend on each other. Thus, the neural circuits mediating cognitive, emotional, physiological, and behavioral responses may not truly be separable and instead are inextricably linked. Moreover, we lack a unifying conceptual framework for understanding how the brain links these processes and how these processes change in unison.

Mental States: Synthesizing Cognition and Emotion

Here, we propose a theoretical foundation for understanding emotion in the context of its intimate relation to the cognitive, physiological, and behavioral responses that constitute emotional expression. We review recent neurobiological data concerning the amygdala and the PFC and discuss how these data fit into a proposed framework for understanding interactions between emotion and cognition.

The concept of a mental state plays a central role in our theoretical framework. We define a mental state as a disposition to action—i.e., every aspect of an organism's inner state that could contribute to its behavior or other responses—which may comprise all the thoughts, feelings, beliefs, intentions, active memories, and perceptions, etc., that are present at a given moment. Thus mental states can be described by a large number of variables, and the set of all mental state variables could provide a quantitative description of one's disposition to behavior. Of note, the identification of mental state variables is constrained by the language we use to describe them. Consequently, mental state variables are not necessarily unique, and they are not necessarily independent from each other. Mental state variables

need not be conscious or unconscious because both types of variables can predispose one to action. Overall, an organism's mental state incorporates internal variables, such as hunger or fear, as well as the representation of a set of environmental stimuli present at a given moment, and the temporal context of stimuli and events. Any given mental state predisposes an organism to respond in certain ways; these actions may be cognitive (e.g., making a decision), behavioral (e.g., freezing or fleeing), or physiological (e.g., increasing heart rate). Mental state variables are useful theoretical constructs because they provide quantitative metrics for analyzing and understanding behavioral and brain processes.

The concept of a mental state is intimately related to, but distinct from, what we call a brain state. Each mental state corresponds to one or more states of the dynamic variables—firing rates, synaptic weights, etc.—that describe the neural circuits of the brain; the full set of values of these variables constitutes a brain state. How are the variables characterizing a mental state represented at the neural circuit level—i.e., the current brain state? This is one way to phrase a fundamental and long-standing question for neuroscientists. At one end of the spectrum is the possibility that each neuron encodes only one variable. For example, a neuron may respond only to the pleasantness of a sensory stimulus, and not to its identity, to its meaning, or to the context in which the stimulus appears. When neurons encode only one variable, other neurons may easily read out the information represented, and the representation can, in principle, be modified without affecting other mental state variables.

One of the disadvantages of the type of representation described immediately above is well illustrated by what is known as the “binding problem” (Malsburg 1999). If each neuron represents only one mental state variable, then it is difficult to construct representations of complex situations. For example, consider a scene with two visual stimuli, one associated with reward and the other with punishment. The brain state should contain the information that pleasantness is associated with the first stimulus and not with the other. If neurons represent only one mental state variable at a time, like stimulus identity, or stimulus valence, then “binding” the information about different variables becomes a substantial challenge. In this case, there must be an additional mechanism that links the activation of the neuron representing pleasantness to the activation of the neuron representing the first stimulus. One simple and efficient way to solve this problem is to introduce neurons with mixed selectivity to conjunctions of events, such as a neuron that responds only when the first stimulus is pleasant. In this scheme, the representations of pleasantness and stimulus identity would be entangled and more difficult to decode, but the number of situations that could be represented would be significantly larger. As discussed below, different brain areas may contain representations with different degrees of entanglement.

How do emotions fit into the conceptual framework of mental states arising from brain states? One influential schema for characterizing emotion posits that emotions can vary along two axes: valence (pleasant versus unpleasant or positive versus negative) and intensity (or arousal) (Lang et al. 1990, Russell 1980). These two variables can simply be conceived as components of the current mental state. Two mental states correspond to different emotions when at least one of the two mental state variables—valence or intensity—is significantly different. Thus, variables describing emotions have the same ontological status as do variables that describe cognitive processes such as memory, attention, decision-making, language, and rule-based problem-solving. Below, we describe neurophysiological data documenting that variables such as valence and arousal are strongly encoded in the amygdala–prefrontal circuit, along with variables related to other cognitive processes. We suggest that neural representations in the amygdala may be more biased toward encoding mental state variables characterizing emotions (valence and intensity). PFC neurons may encode a broader range of variables in an entangled fashion, reflecting the complexity of the behavior and cognition that are its putative outputs.

The concept of a mental state unites cognition and emotion as part of a common framework. How does this framework contribute to the debate about the relationship between emotions and bodily responses? We argue that the issues raised in the debate essentially dissolve when one conceptualizes emotions as part of mental states: Neither emotional feelings nor bodily responses necessarily come first or second. Rather, both of these aspects of emotion are outputs of the neural networks that represent mental states. Furthermore, all the thoughts, physiological responses, and behaviors that constitute emotion are part of an ongoing feedback loop that alters the dynamic, ever-fluctuating brain state and generates new mental states from moment to moment.

How do mental states that integrate emotion and cognition arise from the activity of neural circuits? Below, we describe a potential anatomical substrate—the amygdala–prefrontal circuit—for emotional-cognitive interactions in the brain and how neurons in these areas could dynamically contribute to a subject's mental state. First, we review the bidirectional connections between the amygdala and the PFC that could form the basis of many interactions between cognition and emotion. Second, we review neurobiological studies that used lesions and pharmacological inactivation to investigate the function of the amygdala–PFC circuitry. Third, we review neurophysiological data from the amygdala and the PFC that reveal encoding of variables critical for the representation of mental states and for the learning algorithms—specifically, reinforcement learning (RL)—that emphasize the importance of encoding these parameters for adaptive behavior. For all these topics, we focus on data collected from nonhuman primates. Compared with their rodent counterparts, non-human primates are much more similar to humans in terms of both behavioral repertoire and anatomical development.

Finally, we describe a theoretical proposal that explains how mental states might emerge in the brain and how the amygdala and the PFC could play an integral role in this process. In particular, we propose that the interactions between emotion and cognition may be understood in the context of mental states that can be switched or gated by internal or external events.

An Anatomical Substrate for Interactions Between Emotion and Cognition

This review focuses on interactions between the amygdala and the PFC because of their long-established roles in mediating emotional and cognitive processes (Holland & Gallagher 2004, Lang & Davis 2006, LeDoux 2000, Miller & Cohen 2001, Ochsner & Gross 2005, Wallis 2007). The amygdala is most often discussed in the context of emotional processes; yet it is extensively interconnected with the PFC, especially the posterior orbitofrontal cortex (OFC), and the anterior cingulate cortex (ACC). Here we provide a brief overview of amygdala and PFC anatomy, with an emphasis on the potential anatomical basis of interactions between cognitive and emotional processes.

Amygdala

The amygdala is a structurally and functionally heterogeneous collection of nuclei lying in the anterior medial portion of each temporal lobe. Sensory information enters the amygdala from advanced levels of visual, auditory, and somatosensory cortices, from the olfactory system, and from polysensory brain areas such as the perirhinal cortex and the parahippocampal gyrus (Amaral et al. 1992, McDonald 1998, Stefanacci & Amaral 2002). Within the lateral nucleus, the primary target of projections from unimodal sensory cortices, different sensory modalities are segregated anatomically. But, owing in part to intrinsic connections, multimodal encoding subsequently emerges in the lateral, basal, accessory basal, and other nuclei of the amygdala (Pitkanen & Amaral 1998, Stefanacci & Amaral 2000). Output from the amygdala is directed to a wide range of target structures, including

the PFC, the striatum, sensory cortices (including primary sensory cortices, connections which are probably unique to primates), the hippocampus, the perirhinal cortex, the entorhinal cortex, and the basal forebrain, and to subcortical structures responsible for aspects of physiological responses related to emotion, such as autonomic responses, hormonal responses, and startle (Davis 2000). In general, subcortical projections originate from the central nucleus, and projections to cortex and the striatum originate from the basal, accessory basal, and in some cases the lateral nuclei (Amaral et al. 1992, 2003; Amaral & Dent 1981; Carmichael & Price 1995a; Freese & Amaral 2005; Ghashghaei et al. 2007; Stefanacci et al. 1996; Stefanacci & Amaral 2002; Suzuki & Amaral 1994).

Prefrontal Cortex

The PFC, located in the anterior portion of the cerebral cortex and defined by projections from the mediodorsal nucleus of the thalamus (Fuster 2008), is composed of a group of interconnected brain areas. The distinctive feature of primate PFC is the emergence of dysgranular and granular cortices, which are completely absent in the rodent. In rodents, prefrontal cortex is entirely agranular (Murray 2008, Preuss 1995, Price 2007, Wise 2008). Therefore, much of the primate PFC does not have a clear-cut homolog in rodents. The PFC is often grouped into different subregions; Petrides & Pandya (1994) have described these as dorsal and lateral areas (Walker areas 9, 46, and 9/46), ventrolateral areas (47/12 and 45), medial areas (32 and 24), and orbitofrontal areas (10, 11, 13, 14, and 47/12). Of note, there are extensive interconnections between different PFC areas, allowing information to be shared within local networks (Barbas & Pandya 1989, Carmichael & Price 1996, Cavada et al. 2000), and information also converges from sensory cortices in multiple modalities (Barbas et al. 2002). In general, dorsolateral areas receive input from earlier sensory areas (Barbas et al. 2002). Orbitofrontal areas receive inputs from advanced stages of sensory processing from every modality, including gustatory and olfactory (Carmichael & Price 1995b, Cavada et al. 2000, Romanski et al. 1999). Thus, extrinsic and intrinsic connections make the PFC a site of multimodal convergence of information about the external environment.

In addition, the PFC receives inputs that could inform it about internal mental state variables, such as motivation and emotions. Orbital and medial PFC are closely connected with limbic structures such as the amygdala (see below) and also have direct and indirect connections with the hippocampus and rhinal cortices (Barbas & Blatt 1995, Carmichael & Price 1995a, Cavada et al. 2000, Kondo et al. 2005, Morecraft et al. 1992). Medial and part of orbital PFC has connections to the hypothalamus and other subcortical targets that could mediate autonomic responses (Ongur et al. 1998). Neuromodulatory input to the PFC from dopaminergic, serotonergic, noradrenergic, and cholinergic systems could also convey information about internal state (Robbins & Arnsten 2009). Finally, outputs from the PFC, especially from dorsolateral PFC, are directed to motor systems, consistent with the notion that the PFC may form, represent, and/or transmit motor plans (Bates & Goldman-Rakic 1993, Lu et al. 1994). Altogether, the PFC receives inputs that provide information about many external and internal variables, including those related to emotions and to cognitive plans, providing a potential anatomical substrate for the representation of mental states.

Anatomical Interactions Between the PFC and Amygdala

Although there are diffuse bidirectional projections between amygdala and much of the PFC [see, e.g., figure 4 of Ghashghaei et al. (2007)], the densest interconnections are between the amygdala and orbital areas (e.g., caudal area 13) and medial areas (e.g., areas 24 and 25). The extensive anatomical connections among the amygdala, the PFC, and related structures are summarized in Figure 1. Amygdala input to the PFC often terminates in both superficial and deep layers. OFC output to the amygdala originates in deep layers, and in some cases

also in superficial layers, suggesting both feedforward and feedback modes of information transmission (Ghashghaei et al. 2007).

Previous work has established that the OFC output to the amygdala is complex and segregated, targeting multiple systems in the amygdala (Ghashghaei & Barbas 2002). Some OFC output is directed to the intercalated masses, a ribbon of inhibitory neurons in the amygdala that inhibits activity in the central nucleus (Ghashghaei et al. 2007, Pare et al. 2003). In addition, the OFC projects directly to the central nucleus, providing a means by which the OFC can activate this output structure in addition to inhibiting it (Ghashghaei & Barbas 2002, Stefanacci & Amaral 2000, 2002). Finally, the OFC projects to the basal, accessory basal, and lateral nuclei, where it may influence computations occurring within the amygdala (Ghashghaei & Barbas 2002, Stefanacci & Amaral 2000, 2002). Overall, the bidirectional communication between the amygdala and the OFC, as well as the connections with the rest of the PFC, provides a potential basis for the integration of cognitive, emotional, and physiological processes into a unified representation of mental states.

The Role of the Amygdala and the PFC in Representing Mental States: Lesion Studies

Recent studies using lesions or pharmacological inactivation combined with behavioral studies in monkeys have begun to reveal the specific roles of the primate amygdala and various regions of the PFC in cognitive and emotional processes. We focus here on studies that have helped demonstrate the roles of these brain structures in processes such as valuation, rule-based actions, emotional processes, attention, goal-directed behavior, and working memory—processes that are likely to set some of the variables that constitute a subject's mental state.

Amygdala

Historically, lesions of primate amygdala produced a wide range of behavioral and emotional effects (Aggleton & Passingham 1981, Jones & Mishkin 1972, Kluver & Bucy 1939, Spiegler & Mishkin 1981, Weiskrantz 1956); but in recent years, scientists have increasingly recognized the importance of using anatomically precise lesions that spare fibers of passage. Many older studies had employed aspiration or radiofrequency lesions, which destroy both gray and white matter. By contrast, recent studies using excitotoxic chemical injections, which specifically kill cell bodies, have revised our understanding of cognitive and emotional functions that require the amygdala (Baxter & Murray 2000, Izquierdo & Murray 2007). Some conclusions, however, have been confirmed over many studies using both old and new techniques. In particular, scientists have most prominently used two types of behavioral tasks to establish the amygdala's role in forming or updating associations between sensory stimuli and reinforcement. First, consistent with findings from rodents, the primate amygdala is required for fear learning induced by Pavlovian conditioning (Antoniadis et al. 2009). Second, the amygdala is required for updating the value of a rewarding reinforcer during a devaluation procedure (Machado & Bachevalier 2007, Malkova et al. 1997, Murray & Izquierdo 2007). In this type of task, experimenters satiate an animal on a particular type of reward and test whether satiation changes subsequent choice behavior such that the animal chooses the satiated food type less often; amygdala lesions eliminate the effect of satiation. Pharmacological inactivation of the amygdala has confirmed the amygdala's role in updating a representation of a reinforcer's value; however, once this updating process finishes, the amygdala does not appear to be required (Wellman et al. 2005). In addition, the amygdala is important for other aspects of appetitive conditioned reinforcement (Parkinson et al. 2001) and for behavioral and physiological responding to emotional stimuli such as snakes and intruders in a manner

consistent with its playing a role in processing both emotional valence and intensity (Izquierdo et al. 2005, Kalin et al. 2004, Machado et al. 2009). Finally, experiments using ibotenic acid instead of aspiration lesions in the amygdala have led to revisions in our understanding of the amygdala for reversal-learning task performance (during which stimulus-reinforcement contingencies are reversed). Recent evidence indicates that the amygdala is not required for reversal learning on tasks involving only rewards, unlike previous accounts (Izquierdo & Murray 2007). Overall, these data link the amygdala to functions that rely on neural processing related to both emotional valence and intensity.

Prefrontal Cortex

A long history of studies have used lesions to establish the importance of the PFC in goal-directed behavior, rule-guided behavior, and executive functioning more generally (Fuster 2008, Miller & Cohen 2001, Wallis 2007). These complex cognitive processes form an integral part of our mental state. In addition, lesions of orbitofrontal cortex (OFC) cause many emotional and cognitive deficits reminiscent of amygdala lesions, including deficits in reinforcer devaluation and in behavioral and hormonal responses to emotional stimuli (Izquierdo et al. 2004, 2005; Kalin et al. 2007; Machado et al. 2009; Murray & Izquierdo 2007). Recently, investigators have employed detailed trial-by-trial data analysis to enhance the understanding of the effects of lesions; this work led investigators to propose that ACC and OFC are more involved in the valuation of actions and stimuli, respectively (Kennerley et al. 2006, Rudebeck et al. 2008, Rushworth & Behrens 2008).

In addition, a recent study separately examined lesions of the dorsolateral PFC, the ventrolateral PFC, the principal sulcus (PS), the ACC, and the OFC on a task analogous to the Wisconsin Card Sorting Test (Buckley et al. 2009) used to assay PFC function in humans (Stuss et al. 2000). In the authors' version of the task, monkeys must discover by trial and error the current rule that is in effect; subjects needed to employ working memory for the rule, as well as to utilize information about recent reward history to guide behavior. Lesions in different PFC regions caused distinct deficit profiles: Deficits in working memory, reward-based updating of value representations, and active utilization of recent choice-outcome values were ascribed primarily to PS, OFC, and ACC lesions, respectively (Buckley et al. 2009). Of note, this study used aspiration lesions of the targeted brain regions, which almost certainly damaged fibers of passage located nearby.

A classic finding following aspiration lesions of the OFC is a deficit in learning about reversals of stimulus-reward contingencies (Jones & Mishkin 1972). However, a recent study used ibotenic acid to place a discrete lesion in OFC areas 11 and 13 (Kazama & Bachevalier 2009) and failed to find a deficit in reversal learning. We therefore may need to revise our understanding of how OFC contributes to reversal learning (similar to revisions made with reference to amygdala function, see above); however, the lack of an effect in the recent study may have been due to the anatomically restricted nature of the lesion. This issue will require further investigation.

Prefrontal-Amygdala Interactions

The amygdala is reciprocally connected with the PFC, primarily OFC and ACC, but also diffusely to other parts of the PFC (Figure 1). Studies have begun to examine possible functional interactions between the amygdala and the OFC in mediating different aspects of reinforcement-based and emotional behavior. In one powerful set of experiments, Baxter and colleagues (2000) performed a crossed surgical disconnection of the amygdala and the OFC by lesioning amygdala on one side of the brain and the OFC in the other hemisphere [connections between the amygdala and the OFC are ipsilateral (Ghashghaei & Barbas 2002)]. As noted above, bilateral lesions of monkey amygdala or the OFC impair reinforcer

devaluation; consistent with this finding, the authors found that surgical disconnection also impaired reinforcer devaluation, indicating that the amygdala and the OFC must interact to update the value of a reinforcer. Notably, in humans, neuroimaging studies on rare patients with focal amygdala lesions have revealed that the BOLD signal related to reward expectation in the ventromedial PFC is dependent on a functioning amygdala (Hampton et al. 2007). Investigators have also described functional interactions between the amygdala and the OFC in rodents (Saddoris et al. 2005, Schoenbaum et al. 2003); however, as noted above, rodent OFC may not necessarily correspond to any part of the primate granular/dysgranular PFC (Murray 2008, Preuss 1995, Wise 2008).

The lesion studies described above support the notion that the PFC and the amygdala, often in concert with each other, participate in executive functions such as attention, rule representation, working memory, planning, and valuation of stimuli and actions. In addition, these structures mediate aspects of emotional processing, including processing related to emotional valence and intensity. Together these variables form an integral part of what we have termed a mental state. However, one must exercise some caution when interpreting the results of lesion studies: Owing to potential redundancy in neural coding among brain circuits, a negative result does not necessarily imply that the lesioned area is not normally involved in the function in question. As discussed in the next section, neurons in many parts of the PFC have complex, entangled physiological properties. Given redundancy in encoding, it is therefore not surprising that lesions in these parts of the PFC often do not impair functioning related to the full range of response properties.

Neurophysiological Components of Mental States

We have defined mental states as action dispositions, where actions are broadly defined to include cognitive, physiological, or behavioral responses. Here, we focus on neural signals in the PFC and the amygdala that may encode key cognitive and emotional features of a mental state: the valuation of stimuli, the valence and intensity of emotional reactions to stimuli, our knowledge of the context of sensory stimuli and the requisite rules in that context, and our plans for interacting with stimuli in the environment. We review recent neurophysiological recordings from behaving nonhuman primates that demonstrate coding of all these variables, and they often feature entangled encoding of multiple variables.

Neural Representations of Emotional Valence and Arousal in the Amygdala and the OFC

In recent years, a number of physiological experiments have been directed at understanding the coding properties of neurons in the amygdala and the OFC. The amygdala has long been investigated with respect to aversive processing and its prominent role in fear conditioning, primarily in rodents (Davis 2000, LeDoux 2000, Maren 2005). However, a number of scientists have recognized that the amygdala also plays a role in appetitive processing (Baxter & Murray 2002). Early neurophysiological experiments in monkeys established the amygdala as a potential locus for encoding the affective properties of stimuli (Fuster & Uyeda 1971; Nishijo et al. 1988a, b; Sanghera et al. 1979; Sugase-Miyamoto & Richmond 2005).

To determine whether neurons in the primate amygdala preferentially encoded rewarding or aversive associations, Paton and colleagues (2006) recorded single neuron activity while monkeys learned that visual stimuli—novel abstract fractal images—predicted liquid rewards or aversive air puffs directed at the face, respectively. The experiments employed a Pavlovian procedure called trace conditioning, in which there is a brief temporal gap (the trace interval) between the presentation of a conditioned stimulus (CS) and an unconditioned stimulus (US) (Figure 2a). Monkeys exhibited two behaviors that demonstrated their learning of the stimulus-outcome contingencies: anticipatory licking (an approach behavior)

and anticipatory blinking (a defensive behavior). After monkeys learned the initial CS-US associations, reinforcement contingencies were reversed. Neurophysiological recordings revealed that the amygdala contained some neurons that respond more strongly when a CS is paired with a reward (positive value-coding neurons), and other neurons respond more strongly when the same CS is paired with an aversive stimulus (negative value-coding neurons). Although individual neurons exhibited this differential response during different time intervals (e.g., during the CS interval or parts of the trace interval), across the population of neurons, the value-related signal was temporally extended across the entire trial (Figure 2b,c). Positive and negative value-coding neurons appeared to be intermingled in the amygdala; both types of neurons dispersed within (and perhaps beyond) the basolateral complex (Belova et al. 2008, Paton et al. 2006).

Theoretical accounts of reinforcement learning often posit a neural representation of the value of the current situation as a whole (state value). Data from Belova et al. (2008) suggest that the amygdala could encode the value of the state instantiated by the CS presentation. Neural responses to the fixation point, which appeared at the beginning of trials, were consistent with a role of the amygdala in encoding state value. One can argue that the fixation point is a mildly positive stimulus because monkeys choose to look at it to initiate trials; and indeed, positive value-coding neurons tend to increase their firing in response to fixation point presentation, and negative value-coding neurons tend to decrease firing (Figure 3). Neural signaling after reward or air-puff presentation also indicates that amygdala neurons track state value, as differential levels of activity, on a population level, extending well beyond the termination of USs (Figure 2b,c). All these signals related to reinforcement contingencies could be used to coordinate physiological and behavioral responses specific to appetitive and aversive systems; therefore, they form a potential neural substrate for positive and negative emotional variables.

As discussed earlier, however, valence is only one dimension of emotion; a second dimension is emotional intensity, or arousal. Recent data also link the amygdala to this second dimension. Belova and colleagues (2007) measured responses to rewards and aversive air puffs when they were either expected or unexpected. Surprising reinforcement is generally experienced as more arousing than when the same reinforcements occur predictably; consistent with this notion, expectation often modulated responses to reinforcement in the amygdala—in general, neural responses were enhanced when reinforcement was surprising. For some neurons, this modulation occurred only for rewards or for air puffs, but not for both (Figure 4a–d). These neurons therefore could participate in valence-specific emotional and cognitive processes. However, many neurons modulated their responses to both rewards and air puffs (Figure 4e, f). These neurons could underlie processes such as arousal or enhanced attention, which occur in response to intense emotional stimuli of both valences. Consistent with this role, neural correlates of skin conductance responses, which are mediated by the sympathetic nervous system, have been reported in the amygdala (Laine et al. 2009). Moreover, this type of valence-insensitive modulation of reinforcement responses by expectation could be appropriate for driving reinforcement learning through attention-based learning algorithms (Pearce & Hall 1980).

Of course, the amygdala does not operate in isolation; in particular, its close anatomical connectivity and functional overlap with the OFC raises the question of how OFC processing compares with and interacts with amygdala processing. Using a paradigm similar to that described above, Morrison and Salzman discovered that the OFC contains neurons that prefer rewarding or aversive associations, as in the amygdala, and that, across the population, the signals extend from shortly after CS onset until well after US offset (Figure 2d,e; data largely collected from area 13) (Morrison & Salzman 2009, Salzman et al. 2007). OFC responses to the fixation point are also modulated according to whether a cell has a

positive and negative preference, in a manner similar to the amygdala (S. Morrison & C.D. Salzman, unpublished data). Together, these data suggest that the OFC could also participate in a representation of state value.

Both positive and negative valences are represented in the amygdala and the OFC, but how might OFC and amygdala interact with each other? Unpublished data indicate that the appetitive system—composed of cells that prefer positive associations—updates more quickly in the OFC, adapting to changes in reinforcement contingencies faster than the appetitive system in the amygdala (S. Morrison & C.D. Salzman, personal communication, 2009). However, the opposite is true for the aversive system: Negative-preferring amygdala neurons adapt to changes in reinforcement contingencies more rapidly than do their counterparts in the OFC. Thus, the computational steps that update representations in appetitive and aversive systems are not the same in the amygdala and the OFC, even though the neurons appear to be anatomically interspersed in both structures. In contrast, after reinforcement contingencies are well learned in this task, the OFC signals upcoming reinforcement more rapidly than does the amygdala in both appetitive and aversive cells. This finding is consistent with a role for the OFC in rapidly signaling stimulus values and/or expected outcomes once learning is complete—a signal that could be used to exert prefrontal control over limbic structures such as the amygdala or to direct behavioral responding more generally.

The studies described above used Pavlovian conditioning—a procedure in which no action is required of the subject to receive reinforcement—to characterize neural response properties in relation to appetitive and aversive processing. However, many other studies have used decision-making tasks to quantify the extent to which neural response properties are related to reward values (Dorris & Glimcher 2004; Kennerley et al. 2008; Kim et al. 2008; Lau & Glimcher 2008; McCoy & Platt 2005; Padoa-Schioppa & Assad 2006, 2008; Platt & Glimcher 1999; Roesch & Olson 2004; Samejima et al. 2005; Sugrue et al. 2004; Wallis 2007; Wallis & Miller 2003); moreover, similar tasks are often used to examine human valuation processes using fMRI (Breiter et al. 2001, Gottfried et al. 2003, Kable & Glimcher 2007, Knutson et al. 2001, Knutson & Cooper 2005, McClure et al. 2004, Montague et al. 2006, O'Doherty et al. 2001, Rangel et al. 2008, Seymour et al. 2004). The strength of decision-making tasks is that the investigator can directly compare the subjects' preferences, on a fine scale, with neuron signaling. For example, Padoa-Schioppa & Assad (2006) trained monkeys to indicate which of two possible juice rewards they wanted; they offered the juices in different amounts by presenting visual tokens that indicated both juice type and juice amount (Figure 5a). Using this task, they discovered that multiple signals were present in different populations of neurons in the OFC. Some OFC neurons encoded what the authors termed “chosen value”: Firing was correlated with the value of the chosen reward; some neurons preferred higher and lower values, respectively (Figure 5b,c). These cell populations are reminiscent of the positive and negative value-coding neurons uncovered using the Pavlovian procedure described above. However, because negative valences were not explored in these experiments, these neurons may represent motivation, arousal, or attention, which are correlated with reward value (Maunsell 2004, Roesch & Olson 2004). Other OFC neurons encoded the value of one of the rewards offered (offer value cells; Figure 5d) and others still simply encoded the type of juice offered (taste neurons; Figure 5e), consistent with previous identification of taste-selective neurons in the OFC (Pritchard et al. 2007, Wilson & Rolls 2005). Further data suggested that the OFC responses were menu-invariant—i.e., if a cell prefers A to B, and B to C, it will also prefer A to C (Padoa-Schioppa & Assad 2008). This characteristic is called transitivity; it implies the ability to use the representation of value as a context-independent economic currency that could support decision-making. However, this finding may depend on the exact design of the task because other studies, focusing on partially overlapping regions of the OFC, have

reported neural responses that reflect relative reward preferences, i.e., responses that vary with context and do not meet the standard of transitivity (Tremblay & Schultz 1999).

This rich variety of response properties in the OFC and the amygdala still represents only a subset of the types of encoding that have been observed in these brain areas. For example, amygdala neurons recorded during trace conditioning often exhibited image selectivity (Paton et al. 2006), and similar signals have been observed in the OFC (S. Morrison & C.D. Salzman, personal communication). Moreover, investigators have also described amygdala neural responses to faces, vocal calls, and combinations of faces and vocal calls (Gothard et al. 2007, Kuraoka & Nakamura 2007, Leonard et al. 1985). Meanwhile, the OFC neurons also encode gustatory working memory and modulate their responses depending on reward magnitude, reward probability, and the time and effort required to obtain a reward (Kennerley et al. 2008). Overall, in addition to encoding variables related to valence and arousal/intensity—two variables central to the representation of emotion—amygdala and OFC neurons encode a variety of other variables in an entangled fashion (Paton et al. 2006, Rigotti et al. 2010a).

Neural Representations of Cognitive Processes in the PFC

We have reviewed briefly the encoding of valence and arousal in the amygdala and the OFC, and now we turn our attention to the encoding of other mental state variables in the PFC. Our goal is not to discuss systematically every aspect of PFC neurophysiology, but instead to highlight response properties that may play an especially vital role in setting the variables that constitute a mental state: encoding of rules, which are essential for appropriately contextualizing environmental stimuli and other variables; flexible encoding of stimulus-stimulus associations across time and sensory modality; and encoding of complex motor plans.

Encoding of rules in the PFC—Understanding rules for behavior forms the basis for much of our social interaction; therefore, rules must routinely be represented in our brains. A critical feature of our cognitive ability is the ability to apply abstract, as opposed to concrete, rules, i.e., rules that can be generalized and flexibly applied to new situations. In a striking demonstration of this type of rule encoding in the PFC, Wallis and Miller recorded from three parts of the PFC (dorsolateral, ventrolateral, and the OFC) while monkeys performed a task requiring them to switch flexibly between two abstract rules (Figure 6) (Wallis et al. 2001). In this task, monkeys viewed two sequentially presented visual cues that could be either matching or nonmatching. In different blocks of trials, monkeys had to apply either a match rule or a nonmatch rule—indicated by the presentation of another cue at the start of the trial—to guide their responding. The visual stimuli utilized in the blocks were identical; thus, the only difference between the blocks was the rule in effect, and this information must be a part of the monkey's mental state. Many neurons in all three parts of the PFC exhibited selective activity depending on the rule in effect; some neurons preferred match and others nonmatch (Figure 6). Of note, rule-selective activity was only one type of selectivity that was present: Neurons often responded selectively to the stimuli themselves, as well as to interactions between the stimuli and the rules. Therefore, it appears that these neurons represent abstract rules along with other variables in an entangled manner.

In the work by Wallis and Miller, the rule in effect was cued on every trial, and the monkeys switched from one rule to the other on a trial-by-trial basis. In contrast, Mansouri and colleagues (2006) used a task in which the rule switched in an uncued manner on a block-by-block basis, and monkeys had to discover the rule in effect in a given block (an analog of the Wisconsin Card Sorting Task). In one block of trials, monkeys had to apply a color-match rule to match two stimuli, and in the other block, monkeys had to apply a shape-

match rule. The authors discovered that neural activity in the dorsolateral PFC encoded the rule in effect; different neurons encoded color and shape rules (Figure 7a,b). Rule encoding occurred during the trial itself but also during the fixation interval, and even during the intertrial interval (ITI) (Figure 7c). This observation implies that a neural signature of the rule in effect was maintained throughout a block of trials—even when the monkey was not performing a trial—as if the monkey had to keep the rule in mind. We suggest that this representation of rules therefore represents a distinctive component of a mental state.

Temporal integration of sensory stimuli and actions—One's current situation is defined not only in terms of the stimuli currently present, but also by the temporal context in which those stimuli appear, as well as by the associations those stimuli have with other stimuli. Fuster (2008) proposed that a cardinal function of the PFC is to provide a representation that reflects the temporal integration of relevant sensory information. Indeed, Fuster and colleagues (2000) have demonstrated this type of encoding in areas 6, 8, and 9/46 of the dorsolateral PFC. In this study, monkeys performed a task in which they had to associate an auditory tone (high or low) with a subsequently presented colored target (red or green). The authors discovered cells that responded selectively to associated tones and colors, e.g., cells that fired strongly only for the high tone and its associated target. Meanwhile, failure to represent the correct association accurately was correlated with behavioral errors; thus, PFC neurons' ability to form and represent cross-temporal and cross-modality representations was linked to subsequent actions.

The integration of sensory stimuli in the environment, as described above, is key for setting mental state variables; moreover, if we recall that a mental state can be defined as a disposition to action, any representation of planned actions must clearly be an important element of our mental state. Neural signals related to planned actions have been reported in numerous parts of the PFC in several tasks (Fuster 2008, Miller & Cohen 2001). In recent years, scientists have used more complex motor tasks to explore encoding of sequential movement plans. In the dorsolateral PFC, Tanji and colleagues have described activity related to cursor movements that will result from a series of planned arm movements (Mushiake et al. 2006). This activity therefore reflects future events that occur as a result of planned movements. Other studies of PFC neurons have discovered neural ensembles that predict a sequence of planned movements (Averbeck et al. 2006); when the required sequence of movements changes from block to block, the neural ensemble coding changes, too. In a manner reminiscent of rule encoding, this coding of planned movements was also present during the ITI, as if these cells were keeping note of the planned movement sequence throughout the block of trials (Averbeck & Lee 2007). Thus, the PFC not only tracks stimuli across time, but also represents the temporal integration of planned actions and the events that hinge on them. Dorsolateral PFC may well interact with the OFC and the ACC, and, via these areas, the amygdala, to make decisions based on the values of both environmental stimuli and internal variables and then to execute these decisions via planned action sequences.

Neural Networks and Mental States: A Conceptual and Theoretical Framework for Understanding Interactions between Cognition and Emotion

We have so far reviewed how neurons in the amygdala and the PFC may encode neural signals representing variables—some more closely tied to emotional processes, and others to cognitive processes—that are components of mental states and how these representations are often entangled (i.e., more than one variable is encoded by a single neuron). But how do these neurons interact within a network to represent mental states in their entirety? Moreover, how can cognitive processes regulate emotional processes?

A central element of emotional regulation involves developing the ability to alter one's emotional response to a stimulus. In general, one can consider at least two basic ways in which this can occur. First, learning mechanisms may operate to change the representation of the emotional meaning of a stimulus. Indeed, one could simply forget or overwrite a previously stored association. Moreover given a stimulus previously associated with a particular reinforcement, such that the stimulus elicits an emotional response, re-experiencing the stimulus in the absence of the associated reinforcement can induce extinction. Extinction is thought to be a learning process whereby previously acquired responses are inhibited. In the case of fear extinction, scientists currently believe that original CS-US associations continue to be stored in the brain (so that they are not forgotten or overwritten), and inhibitory mechanisms develop that suppress the fear response (Quirk & Mueller 2008). Second, mechanisms must exist that can change or switch one's emotional responses depending on one's knowledge of his/her context or situation. A simple example of this phenomenology occurs when playing the game of blackjack. Here, the same card, such as a jack of clubs, can be rewarding, if it makes a total of 21 in your hand, or upsetting, if it makes a player go bust. Emotional responses to the jack of clubs can thereby vary on a moment-to-moment basis depending on the player's knowledge of the situation (e.g., his/her understanding of the rules of the game and of the cards already dealt). Emotional variables here depend critically on the cognitive variables representing one's understanding of the game and one's current hand of cards. Although mechanisms for this type of emotional regulation remain poorly understood, it presumably involves PFC-amygdala neural circuitry.

What type of theoretical framework could describe these different types of emotional regulation? Are there qualitative differences between the neural mechanisms that underlie them? Here we briefly describe one possible approach for explaining this phenomenology. Our proposal is built on the assumption that each mental state corresponds to a large number of states of dynamic variables that describe neurons, synapses, and other constituents of neural circuits. These components must interact such that neural circuit dynamics can actively maintain a representation of the current disposition of behavior, i.e., the current mental state. Complex interactions between these components must therefore correspond to the interactions between mental state variables such as emotional and cognitive parameters. Indeed, when brain states change, these changes typically and inherently involve correlated modifications of multiple mental state variables. In this section, we discuss how a class of neural mechanisms could underlie the representation of mental states and the potential interaction between cognition and emotion. We construct a conceptual framework whereby cognition-emotion interactions can occur via two sorts of mechanisms: associative learning and switching between mental states representing different contexts or situations.

A natural candidate mechanism for representing mental states is the reverberating activity that has been observed at the single neuron level in the form of selective persistent firing rates, such as that which has been described in the PFC (e.g., Figure 7) and other structures (Miyashita & Chang 1988, Yakovlev et al. 1998). Each mental state could be represented by a self-sustained, stable pattern of reverberating activity. Small perturbations of these activity patterns are damped by the interactions between neurons so that the state of the network is attracted toward the closest pattern of persistent activity representing a particular mental state. For this reason, these patterns are called attractors of the neural dynamics. Attractor networks have been proposed as models for associative and working memory (Amit 1989, Hopfield 1982), for decision-making (Wang 2002), and for rule-based behavior (O'Reilly & Munakata 2000, Rolls & Deco 2002). Here, we suggest a scenario in which attractors represent stable mental states and every external or internal event encountered by an organism may steer the activity from one attractor to a different one. This type of mechanism could provide stable yet modifiable representations for the mental states, just

like the on and off states of a switch. Thus mental states could be maintained over relatively long timescales but could also rapidly change in response to brief events.

Attractor networks can be utilized to model associative learning. Consider again the experiment performed by Paton and colleagues (2006), described in the section on neural representation of emotional variables. In a simple model, one can assume that learning involves modifying connections from neurons representing the CS (for simplicity, called external neurons) to some of the neurons representing the mental state, in particular those that represent the value of the CS in relation to reinforcement (called internal valence neurons). When the CSs are novel, a monkey does not know what to expect (reward or air puff). The monkey may know that it will be one of the two outcomes. Therefore, the CS in that particular context could induce a transition into one of the preexistent attractors representing the possible states. Some of these states correspond to the expectation of positive or negative reinforcement, and other states could correspond to neutral valence states. The external input starts a biased competition between all these different states. If the reinforcement received differs from the expected one, then the synapses connecting external and internal neurons will be modified such that the competition between mental states will generate a bias toward the correct association (see e.g., Fusi et al. 2007). This learning process is typical of situations in which there are one-to-one associations and, for example, the same CS always has the same value. The monkey can simply learn the stimulus-value associations by trial and error; with appropriate synaptic learning rules, the external connections are modified as needed.

In the situation described above, one CS always predicts reward, and the other punishment. But such conditions do not always exist. For example, Paton and colleagues reversed reinforcement contingencies after learning had occurred. In principle, learning these reversed contingencies could involve modifying the external connections to the neural circuit, thereby having new associations overwrite or override the previous associations. However, reversal tasks may not simply erase or unlearn associations; instead, reversal tasks may rely on processes similar to those invoked during extinction (Bouton 2002, Myers & Davis 2007). Increasing evidence implicates the amygdala-PFC circuit as playing a fundamental role in extinction (Gottfried & Dolan 2004, Izquierdo & Murray 2005, Likhtik et al. 2005, Milad & Quirk 2002, Olsson & Phelps 2004, Pare et al. 2004, Quirk et al. 2000).

For the second type of emotional regulation, during which emotional responses to stimuli depend on knowledge of one's situation or context, we need a qualitatively different learning mechanism. Consider a hypothetical variant of the experiment by Paton et al. (2006), in which the associations are reversed and changed multiple times. For example, stimulus A may initially be associated with a small reward and B with a small punishment. Then, in a second block of trials, A becomes associated with a large reward, and B with a large punishment. Assume that as the experiment proceeds, subjects go back and forth between these two types of blocks of trials so that the two contexts are alternated many times. In this case, if we can store a representation of both the two alternating contexts, we can adopt a significantly more efficient computational strategy. Instead of learning and forgetting associations, we can simply switch from one context to the other. For example, on the first trial of a block, if a large punishment follows B, the monkey can predict that seeing A on subsequent trials will result in its receiving a large reward. Overall, in the first context, A and B can lead to only small rewards and punishments, respectively. In the second context, A and B always predict large rewards and punishments. To implement this switching type of computational strategy, internal synaptic connections within the neural network must be modified to create the neural representations of the mental states corresponding to the two contexts (Rigotti et al. 2010a).

To illustrate how a model employing an attractor neural network can describe the case of mental state switching described above, we employ an energy landscape metaphor [see e.g., Amit (1989) and Figure 8]. Each network state can be described as a vector containing the activation states of all neurons, and it can be characterized by its energy value. If we know the energy for each state, then we can predict the network behavior because the network state will evolve toward the state corresponding to the closest minimum of the energy. In Figure 8, we represent each network state as a point on a plane and the corresponding energy as a surface that resembles a hilly landscape. To describe the hypothetical experiment under discussion, which involves switching between contexts, we assume that two variables, context and valence (with two contexts and five different valences represented), characterize each mental state and that only one brain state corresponds to each mental state. As a consequence, for each point on the context-valence plane there is only one energy value (Figure 8, red surface). The network naturally relaxes toward the bottom of the valleys (minima of the energy), which represent different mental states. At the neural level, each of these points corresponds to a particular pattern of persistent activity. The six valleys in Figure 8 correspond to six potential mental states created after modifying internal connections to represent the two different contexts and the related CS-US associations. As a result of the interactions due to recurrent connections, the cognitive variable corresponding to context constrains the set of accessible emotional states. In both contexts, interactions with external neurons representing CS A or B can tilt temporarily the energy surface and bring the neural network into a different valley, corresponding to a different mental state. The final destination depends on the initial mental state representing the context. The valences of the states differ in the two contexts because valences associated with large rewards and punishments exist only in the second context.

This example illustrates the cognitive regulation of emotion because changes in a cognitive variable (context) cause a change in the possible associated emotional parameters (valence). Analogous mechanisms could underlie how other cognitive variables can influence emotional responses. For example, different social situations can demand different emotional responses to similar sensory stimuli, and knowledge of the social situation (essentially a context variable) can thereby constrain the emotional responses possible.

We based the forgoing discussion on the assumption that the mental states are represented by attractors of the neural dynamics. Alternative and complementary solutions are based on neural representations of mental states that change in time (Buonomano & Maass 2009, Jaeger & Hass 2004). For these neural systems, every trajectory or set of trajectories in the space of all possible brain states represents a particular mental state. These dynamic systems can generate complex temporal sequences that are important for motor planning (Susillo & Abbott 2009). However, they cannot instantaneously generalize to situations in which events are timed differently and they can be difficult to decode. Generally speaking, all known models of mental states provide a useful conceptual framework for understanding the principles of the dynamics of neural circuits, but they fall short of capturing the richness and complexity of real biological neural networks. For example, brain states are not encoded solely in the neuronal spiking activity. Investigators only now have begun to study interactions among dynamic variables operating on diverse timescales, all contributing to a particular brain state [see e.g., Mongillo et al. (2008) for a working memory model based on short-term synaptic facilitation].

The theoretical framework we propose provides a means for representing mental states in the distributed activity of networks of neurons encoding entangled representations. Because we have defined mental states as action dispositions, it is natural to wonder how mental states are linked to the selection and execution of actions. In recent years, reinforcement learning (RL) algorithms have provided an elegant framework for understanding both how

subjects choose their actions to maximize reward and minimize punishment and how the brain may represent modeled parameters during this process (Daw et al. 2006, Dayan & Abbott 2001, Sutton & Barto 1998). In particular, scientists have attempted to link two types of RL algorithms onto specific neural structures: a model-based algorithm and a model-free algorithm (Daw et al. 2005). Model-based algorithms are suited for goal-directed actions and likely involve the PFC, whereas model-free algorithms could mediate the generation of habitual behavior (Graybiel 2008) and may involve the striatum. Of note, both types of RL algorithms actually require an already formed representation of states (i.e., of the relevant variables of one's current situation) to enable one to assign values to them to guide action selection. If one drives in an unfamiliar neighborhood packed with restaurants and needs to choose a restaurant, one must first build a mental map of the environment as it is experienced before assigning values to possible destinations. This process of creating mental state representations is not provided for by RL algorithms, which involve only the assignment and updating of values to already created states.

Creating a representation of mental states involves forging links between the many mental state variables that neurons represent. Recent work on the neural basis of object perception represents an initial step toward understanding how variables may be combined to support the formation of a mental state representation. The perception of objects requires one to develop a representation that is invariant for many viewing conditions, such as the precise retinal position of the object, the size, scale, or pose of the object, or the amount of clutter the object appears within the visual field. Di Carlo and colleagues have now provided evidence that unsupervised learning arising from the temporal contiguity of stimuli experienced during natural viewing leads to the formation of an invariant representation of a visual object (Li & DiCarlo 2008). Object perception corresponds to only one component of a mental state, but the scientific approach pioneered by Di Carlo's group may provide a path for understanding how other mental state variables also become linked to create a unified representation of a particular state. Indeed, some theoreticians have proposed that simple mechanisms such as temporal contiguity might underlie new mental state formation (O'Reilly & Munakata 2000, Rigotti et al. 2010a).

Conclusions

The conceptual framework we have put forth posits that mental states are composed of many variables that together correspond to an action disposition. These variables include parameters, such as valence and arousal, which are often ascribed to emotional processes, as well as parameters ascribed to cognitive processes, such as perceptions, memories, and plans. Of course, mental state parameters also include variables encoding our visceral state. Much in the way that Wittgenstein argued that philosophical controversies dissolve once one carefully disentangles the different ways in which language is being used (Wittgenstein 1958), we argue that the debate between scientists about the origin of emotional feelings—whether visceral processes precede or follow emotional feeling—dissolves. Instead, all these parameters may be linked and together form the representation of our mental state.

This conceptual framework has broad implications for understanding interactions between cognition and emotion in the brain. On the one hand, emotional processes can influence cognitive processes; on the other hand, cognitive processes can regulate or modify our emotions. Both of these interactions can be implemented by changing mental state variables (either emotional or cognitive ones); emotions and thoughts shift together, corresponding to the new mental state. Of course, different mechanisms may exist for implementing these interactions between cognition and emotion, such as mechanisms involving learning and extinction, as well as mechanisms that support the creation of new mental state representations, such as when one learns a new rule. The process of understanding the

complex encoding properties of the amygdala, the PFC, and related brain structures, as well as understanding their functional interactions, is in its infancy. Somehow the intricate connectivity of these brain structures gives rise to mental states and accounts for interactions between cognition and emotion that are fundamental to our well-being and our existence.

Acknowledgments

We thank B. Lau and S. Morrison for helpful discussion, as well as comments on the manuscript, and H. Cline for invaluable support. C.D.S. gratefully acknowledges funding support from NIH (R01 MH082017, R01 DA020656, and RC1 MH088458) and the James S. McDonnell Foundation. S.F. receives support from DARPA SyNAPSE, the Gatsby Foundation, and the Sloan-Swartz Foundation.

Literature Cited

Aggleton, J., editor. The Amygdala—A Functional Analysis. Oxford: Oxford Univ. Press; 2000.

Aggleton JP, Passingham RE. Syndrome produced by lesions of the amygdala in monkeys (*Macaca mulatta*). J Comp Physiol Psychol. 1981; 95:961–77. [PubMed: 7320283]

Amaral, D.; Price, J.; Pitkanen, A.; Carmichael, S. Anatomical organization of the primate amygdaloid complex. In: Aggleton, J., editor. The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction. New York: Wiley-Liss; 1992. p. 1-66.

Amaral DG, Behniea H, Kelly JL. Topographic organization of projections from the amygdala to the visual cortex in the macaque monkey. Neuroscience. 2003; 118:1099–120. [PubMed: 12732254]

Amaral DG, Dent JA. Development of the mossy fibers of the dentate gyrus: I. A light and electron microscopic study of the mossy fibers and their expansions. J Comp Neurol. 1981; 195:51–86. [PubMed: 7204652]

Amit, D. Modeling Brain Function—The World of Attractor Neural Networks. New York: Cambridge Univ. Press; 1989.

Antoniadis EA, Winslow JT, Davis M, Amaral DG. The nonhuman primate amygdala is necessary for the acquisition but not the retention of fear-potentiated startle. Biol Psychiatry. 2009; 65:241–48. [PubMed: 18823878]

Averbeck BB, Lee D. Prefrontal neural correlates of memory for sequences. J Neurosci. 2007; 27:2204–11. [PubMed: 17329417]

Averbeck BB, Sohn JW, Lee D. Activity in prefrontal cortex during dynamic selection of action sequences. Nat Neurosci. 2006; 9:276–82. [PubMed: 16429134]

Barbas H, Blatt GJ. Topographically specific hippocampal projections target functionally distinct prefrontal areas in the rhesus monkey. Hippocampus. 1995; 5:511–33. [PubMed: 8646279]

Barbas, H.; Ghashghaei, H.; Rempel-Clower, N.; Xiao, D. Anatomic basis of functional specialization in prefrontal cortices in primates. In: Grafman, J., editor. Handbook of Neuropsychology. Amersterdam: Elsevier Science B.V.; 2002. p. 1-27.

Barbas H, Pandya DN. Architecture and intrinsic connections of the prefrontal cortex in the rhesus monkey. J Comp Neurol. 1989; 286:353–75. [PubMed: 2768563]

Barbas H, Zikopoulos B. The prefrontal cortex and flexible behavior. Neuroscientist. 2007; 13:532–45. [PubMed: 17901261]

Bard P. A diencephalic mechanism for the expression of rage with special reference to the sympathetic nervous system. Am J Physiol. 1928; 84:490–515.

Bates JF, Goldman-Rakic PS. Prefrontal connections of medial motor areas in the rhesus monkey. J Comp Neurol. 1993; 336:211–28. [PubMed: 7503997]

Baxter M, Murray E. Reinterpreting the behavioural effects of amygdala lesions in nonhuman primates. See Aggleton. 2000:545–68.

Baxter M, Murray EA. The amygdala and reward. Nat Rev Neurosci. 2002; 3:563–73. [PubMed: 12094212]

Baxter M, Parker A, Lindner CC, Izquierdo AD, Murray EA. Control of response selection by reinforcer value requires interaction of amygdala and orbital prefrontal cortex. J Neurosci. 2000; 20:4311–19. [PubMed: 10818166]

Belova MA, Paton JJ, Morrison SE, Salzman CD. Expectation modulates neural responses to pleasant and aversive stimuli in primate amygdala. Neuron. 2007; 55:970–84. [PubMed: 17880899]

Belova MA, Paton JJ, Salzman CD. Moment-to-moment tracking of state value in the amygdala. J Neurosci. 2008; 28:10023–30. [PubMed: 18829960]

Bouton ME. Context, ambiguity, and unlearning: sources of relapse after behavioral extinction. Biol Psychiatry. 2002; 52:976–86. [PubMed: 12437938]

Breiter H, Aharon I, Kahneman D, Dale A, Shizgal P. Functional imaging of neural responses to expectancy and experience of monetary gains and losses. Neuron. 2001; 30:619–39. [PubMed: 11395019]

Buckley MJ, Mansouri FA, Hoda H, Mahboubi M, Browning PG, et al. Dissociable components of rule-guided behavior depend on distinct medial and prefrontal regions. Science. 2009; 325:52–58. [PubMed: 19574382]

Buonomano DV, Maass W. State-dependent computations: spatiotemporal processing in cortical networks. Nat Rev Neurosci. 2009; 10:113–25. [PubMed: 19145235]

Cannon W. The James-Lange theory of emotions: a critical examination and an alternative theory. Am J Psychol. 1927; 39:106–24.

Carmichael ST, Price JL. Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. J Comp Neurol. 1995a; 363:615–41. [PubMed: 8847421]

Carmichael ST, Price JL. Sensory and premotor connections of the orbital and medial prefrontal cortex of macaque monkeys. J Comp Neurol. 1995b; 363:642–64. [PubMed: 8847422]

Carmichael ST, Price JL. Connectional networks within the orbital and medial prefrontal cortex of macaque monkeys. J Comp Neurol. 1996; 371:179–207. [PubMed: 8835726]

Cavada C, Company T, Tejedor J, Cruz-Rizzolo RJ, Reinoso-Suarez F. The anatomical connections of the macaque monkey orbitofrontal cortex. A review Cerebral Cortex. 2000; 10:220–42.

Damasio, A. Descarte's Error: Emotion, Reason, and the Human Brain. New York: Harcourt Brace; 1994.

Davis M. The role of the amygdala in conditioned and unconditioned fear and anxiety. See Aggleton. 2000:213–87.

Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci. 2005; 8:1704–11. [PubMed: 16286932]

Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ. Cortical substrates for exploratory decisions in humans. Nature. 2006; 441:876–79. [PubMed: 16778890]

Dayan, P.; Abbott, LF. Theoretical Neuroscience. Cambridge, MA: MIT Press; 2001.

Dorris MC, Glimcher PW. Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. Neuron. 2004; 44:365–78. [PubMed: 15473973]

Freese JL, Amaral DG. The organization of projections from the amygdala to visual cortical areas TE and V1 in the macque monkey. J Comp Neurol. 2005; 486:295–317. [PubMed: 15846786]

Fusi S, Asaad WF, Miller EK, Wang XJ. A neural circuit model of flexible sensorimotor mapping: learning and forgetting on multiple timescales. Neuron. 2007; 54:319–33. [PubMed: 17442251]

Fuster, J. The Prefrontal Cortex. London: Elsevier; 2008.

Fuster JM, Bodner M, Kroger JK. Cross-modal and cross-temporal association in neurons of frontal cortex. Nature. 2000; 405:347–51. [PubMed: 10830963]

Fuster JM, Uyeda AA. Reactivity of limbic neurons of the monkey to appetitive and aversive signals. Electroencephalogr Clin Neurophysiol. 1971; 30:281–93. [PubMed: 4103500]

Ghashghaei H, Barbas H. Pathways for emotion: interactions of prefrontal and anterior temporal pathways in the amygdala of the rhesus monkey. Neuroscience. 2002; 115:1261–79. [PubMed: 12453496]

Ghashghaei HT, Hilgetag CC, Barbas H. Sequence of information processing for emotions based on the anatomic dialogue between prefrontal cortex and amygdala. Neuroimage. 2007; 34:905–23. [PubMed: 17126037]

Gothard KM, Battaglia FP, Erickson CA, Spitler KM, Amaral DG. Neural responses to facial expression and face identity in the monkey amygdala. J Neurophysiol. 2007; 97:1671–83. [PubMed: 17093126]

Gottfried J, O'Doherty J, Dolan RJ. Encoding predictive reward value in human amygdala and orbitofrontal cortex. Science. 2003; 301:1104–7. [PubMed: 12934011]

Gottfried JA, Dolan RJ. Human orbitofrontal cortex mediates extinction learning while accessing conditioned representations of value. Nat Neurosci. 2004; 7:1144–52. [PubMed: 15361879]

Graybiel AM. Habits, rituals, and the evaluative brain. Annu Rev Neurosci. 2008; 31:359–87. [PubMed: 18558860]

Hampton AN, Adolphs R, Tyszka MJ, O'Doherty JP. Contributions of the amygdala to reward expectancy and choice signals in human prefrontal cortex. Neuron. 2007; 55:545–55. [PubMed: 17698008]

Holland PC, Gallagher M. Amygdala-frontal interactions and reward expectancy. Curr Opin Neurobiol. 2004; 14:148–55. [PubMed: 15082318]

Hopfield JJ. Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci USA. 1982; 79:2554–58. [PubMed: 6953413]

Izquierdo A, Murray EA. Opposing effects of amygdala and orbital prefrontal cortex lesions on the extinction of instrumental responding in macaque monkeys. Eur J Neurosci. 2005; 22:2341–46. [PubMed: 16262672]

Izquierdo A, Murray EA. Selective bilateral amygdala lesions in rhesus monkeys fail to disrupt object reversal learning. J Neurosci. 2007; 27:1054–62. [PubMed: 17267559]

Izquierdo A, Suda RK, Murray EA. Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. J Neurosci. 2004; 24:7540–48. [PubMed: 15329401]

Izquierdo A, Suda RK, Murray EA. Comparison of the effects of bilateral orbital prefrontal cortex lesions and amygdala lesions on emotional responses in rhesus monkeys. J Neurosci. 2005; 25:8534–42. [PubMed: 16162935]

Jaeger H, Haas H. Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. Science. 2004; 304:78–80. [PubMed: 15064413]

James W. What is an emotion? Mind. 1884; 9:188–205.

James W. The physical basis of emotion. Psychol Rev. 1894; 1:516–29.

Jones B, Mishkin M. Limbic lesions and the problem of stimulus-reinforcement associations. Exp Neurol. 1972; 36:362–77. [PubMed: 4626489]

Kable JW, Glimcher PW. The neural correlates of subjective value during intertemporal choice. Nat Neurosci. 2007; 10:1625–33. [PubMed: 17982449]

Kalin NH, Shelton SE, Davidson RJ. The role of the central nucleus of the amygdala in mediating fear and anxiety in the primate. J Neurosci. 2004; 24:5506–15. [PubMed: 15201323]

Kalin NH, Shelton SE, Davidson RJ. Role of the primate orbitofrontal cortex in mediating anxious temperament. Biol Psychiatry. 2007; 62:1134–39. [PubMed: 17643397]

Kazama A, Bachevalier J. Selective aspiration or neurotoxic lesions of orbital frontal areas 11 and 13 spared monkeys' performance on the object discrimination reversal task. J Neurosci. 2009; 29:2794–804. [PubMed: 19261875]

Kennerley SW, Dahmubed AF, Lara AH, Wallis JD. Neurons in the frontal lobe encode the value of multiple decision variables. J Cogn Neurosci. 2008; 21:1162–78. [PubMed: 18752411]

Kennerley SW, Walton ME, Behrens TE, Buckley MJ, Rushworth MF. Optimal decision making and the anterior cingulate cortex. Nat Neurosci. 2006; 9:940–47. [PubMed: 16783368]

Kim S, Hwang J, Lee D. Prefrontal coding of temporally discounted values during intertemporal choice. Neuron. 2008; 59:161–72. [PubMed: 18614037] Erratum. Neuron. 59(3):522.

Kluver H, Bucy P. Preliminary analysis of functions of the temporal lobes in monkeys. Arch Neurol Psychiatry. 1939; 42:979–1000.

Knutson B, Adams CM, Fong GW, Hommer D. Anticipation of increasing monetary reward selectively recruits nucleus accumbens. J Neurosci. 2001; 21:RC159. [PubMed: 11459880]

Knutson B, Cooper JC. Functional magnetic resonance imaging of reward prediction. Curr Opin Neurol. 2005; 18:411–17. [PubMed: 16003117]

Kondo H, Saleem KS, Price JL. Differential connections of the perirhinal and parahippocampal cortex with the orbital and medial prefrontal networks in macaque monkeys. J Comp Neurol. 2005; 493:479–509. [PubMed: 16304624]

Kuraoka K, Nakamura K. Responses of single neurons in monkey amygdala to facial and vocal emotions. J Neurophysiol. 2007; 97:1379–87. [PubMed: 17182913]

Laine CM, Spitler KM, Mosher CP, Gothard KM. Behavioral triggers of skin conductance responses and their neural correlates in the primate amygdala. J Neurophysiol. 2009; 101:1749–54. [PubMed: 19144740]

Lang PJ. The varieties of emotional experience: a meditation on James-Lange theory. Psychol Rev. 1994; 101:211–21. [PubMed: 8022956]

Lang PJ, Bradley MM, Cuthbert BN. Emotion, attention, and the startle reflex. Psychol Rev. 1990; 97:377–95. [PubMed: 2200076]

Lang PJ, Davis M. Emotion, motivation, and the brain: reflex foundations in animal and human research. Prog Brain Res. 2006; 156:3–29. [PubMed: 17015072]

Lange, C. The Emotions. Baltimore, MD: Williams & Wilkins; 1922.

Lau B, Glimcher PW. Value representations in the primate striatum during matching behavior. Neuron. 2008; 58:451–63. [PubMed: 18466754]

LeDoux JE. Emotion circuits in the brain. Annu Rev Neurosci. 2000; 23:155–84. [PubMed: 10845062]

Leonard CM, Rolls ET, Wilson FA, Baylis GC. Neurons in the amygdala of the monkey with responses selective for faces. Behav Brain Res. 1985; 15:159–76. [PubMed: 3994832]

Li N, DiCarlo JJ. Unsupervised natural experience rapidly alters invariant object representation in visual cortex. Science. 2008; 321:1502–7. [PubMed: 18787171]

Likhtik E, Pelletier JG, Paz R, Pare D. Prefrontal control of the amygdala. J Neurosci. 2005; 25:7429–37. [PubMed: 16093394]

Lu MT, Preston JB, Strick PL. Interconnections between the prefrontal cortex and the premotor areas in the frontal lobe. J Comp Neurol. 1994; 341:375–92. [PubMed: 7515081]

Machado CJ, Bachevalier J. The effects of selective amygdala, orbital frontal cortex or hippocampal formation lesions on reward assessment in nonhuman primates. Eur J Neurosci. 2007; 25:2885–904. [PubMed: 17561849]

Machado CJ, Kazama AM, Bachevalier J. Impact of amygdala, orbital frontal, or hippocampal lesions on threat avoidance and emotional reactivity in nonhuman primates. Emotion. 2009; 9:147–63. [PubMed: 19348528]

Malkova L, Gaffan D, Murray EA. Excitotoxic lesions of the amygdala fail to produce impairment in visual learning for auditory secondary reinforcement but interfere with reinforcer devaluation effects in rhesus monkeys. J Neurosci. 1997; 17:6011–20. [PubMed: 9221797]

Mansouri FA, Matsumoto K, Tanaka K. Prefrontal cell activities related to monkeys' success and failure in adapting to rule changes in a Wisconsin Card Sorting Test analog. J Neurosci. 2006; 26:2745–56. [PubMed: 16525054]

Maren S. Synaptic mechanisms of associative memory in the amygdala. Neuron. 2005; 47:783–86. [PubMed: 16157273]

Maunsell JH. Neuronal representations of cognitive state: reward or attention? Trends Cogn Sci. 2004; 8:261–65. [PubMed: 15165551]

McClure SM, Laibson DI, Loewenstein G, Cohen JD. Separate neural systems value immediate and delayed monetary rewards. Science. 2004; 306:503–7. [PubMed: 15486304]

McCoy AN, Platt ML. Risk-sensitive neurons in macaque posterior cingulate cortex. Nat Neurosci. 2005; 8:1220–27. [PubMed: 16116449]

McDonald AJ. Cortical pathways to the mammalian amygdala. Prog Neurobiol. 1998; 55:257–332. [PubMed: 9643556]

Milad MR, Quirk GJ. Neurons in medial prefrontal cortex signal memory for fear extinction. Nature. 2002; 420:70–74. [PubMed: 12422216]

Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. Annu Rev Neurosci. 2001; 24:167–202. [PubMed: 11283309]

Miyashita Y, Chang H. Neuronal correlate of pictorial short-term memory in the primate temporal cortex. Nature. 1988; 331:68–70. [PubMed: 3340148]

Mongillo G, Barak O, Tsodyks M. Synaptic theory of working memory. Science. 2008; 319:1543–46. [PubMed: 18339943]

Montague PR, King-Casas B, Cohen JD. Imaging valuation models in human choice. Annu Rev Neurosci. 2006; 29:417–48. [PubMed: 16776592]

Morecraft RJ, Geula C, Mesulam MM. Cytoarchitecture and neural afferents of orbitofrontal cortex in the brain of the monkey. J Comp Neurol. 1992; 323:341–58. [PubMed: 1460107]

Morrison S, Salzman C. The convergence of information about rewarding and aversive stimuli in single neurons. J Neurosci. 2009; 29:11471–83. [PubMed: 19759296]

Murray, EA. Neuropsychology of primate reward processes. In: Squire, LR., editor. New Encyclopedia of Neuroscience. Vol. 6. Oxford; Academic Press; 2008. p. 993-99.

Murray EA, Izquierdo A. Orbitofrontal cortex and amygdala contributions to affect and action in primates. Ann NY Acad Sci. 2007; 1121:273–96. [PubMed: 17846154]

Mushiake H, Saito N, Sakamoto K, Itoyama Y, Tanji J. Activity in the lateral prefrontal cortex reflects multiple steps of future events in action plans. Neuron. 2006; 50:631–41. [PubMed: 16701212]

Myers KM, Davis M. Mechanisms of fear extinction. Mol Psychiatry. 2007; 12:120–50. [PubMed: 17160066]

Nishijo H, Ono T, Nishino H. Single neuron responses in amygdala of alert monkey during complex sensory stimulation with affective significance. J Neurosci. 1988a; 8:3570–83. [PubMed: 3193171]

Nishijo H, Ono T, Nishino H. Topographic distribution of modality-specific amygdalar neurons in alert monkey. J Neurosci. 1988b; 8:3556–69. [PubMed: 3193170]

Ochsner KN, Gross JJ. The cognitive control of emotion. Trends Cogn Sci. 2005; 9:242–49. [PubMed: 15866151]

O'Doherty J, Kringelbach ML, Rolls ET, Hornak J, Andrews C. Abstract reward and punishment representations in the human orbitofrontal cortex. Nat Neurosci. 2001; 4:95–102. [PubMed: 11135651]

Olsson A, Phelps EA. Learned fear of “unseen” faces after Pavlovian, observational, and instructed fear. Psychol Sci. 2004; 15:822–28. [PubMed: 15563327]

Ongur D, An X, Price JL. Prefrontal cortical projections to the hypothalamus in macaque monkeys. J Comp Neurol. 1998; 401:480–505. [PubMed: 9826274]

O'Reilly, R.; Munakata, Y. Computational Explorations in Cognitive Neuroscience. Cambridge, MA: MIT Press; 2000.

Padoa-Schioppa C, Assad JA. Neurons in the orbitofrontal cortex encode economic value. Nature. 2006; 441:223–26. [PubMed: 16633341]

Padoa-Schioppa C, Assad JA. The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. Nat Neurosci. 2008; 11:95–102. [PubMed: 18066060]

Pare D, Quirk GJ, Ledoux JE. New vistas on amygdala networks in conditioned fear. J Neurophysiol. 2004; 92:1–9. [PubMed: 15212433]

Pare D, Royer S, Smith Y, Lang EJ. Contextual inhibitory gating of impulse traffic in the intra-amygdaloid network. Ann N Y Acad Sci. 2003; 985:78–91. [PubMed: 12724150]

Parkinson J, Crofts HS, McGuigan M, Tomic DL, Everitt BJ, Roberts AC. The role of the primate amygdala in conditioned reinforcement. J Neurosci. 2001; 21:7770–80. [PubMed: 11567067]

Paton J, Belova M, Morrison S, Salzman C. The primate amygdala represents the positive and negative value of visual stimuli during learning. Nature. 2006; 439:865–70. [PubMed: 16482160]

Pearce J, Hall G. A model for Pavlovian conditioning: variations in the effectiveness of conditioned but not unconditioned stimuli. Psychol Rev. 1980; 87:532–52. [PubMed: 7443916]

Pessoa L. On the relationship between emotion and cognition. Nat Rev Neurosci. 2008; 9:148–58. [PubMed: 18209732]

Petrides, M.; Pandya, D. Comparative architectonic analysis of the human and macaque frontal cortex. In: Boller, F.; Grafman, J., editors. Handbook of Neuropsychology. New York: Elsevier; 1994. p. 17-57.

Phelps EA, LeDoux JE. Contributions of the amygdala to emotion processing: from animal models to human behavior. Neuron. 2005; 48:175–87. [PubMed: 16242399]

Pitkanen A, Amaral DG. Organization of the intrinsic connections of the monkey amygdaloid complex: projections originating in the lateral nucleus. J Comp Neurol. 1998; 398:431–58. [PubMed: 9714153]

Platt ML, Glimcher PW. Neural correlates of decision variables in parietal cortex. Nature. 1999; 400:233–38. [PubMed: 10421364]

Preuss TM. Do rats have prefrontal cortex? The Rose-Woolsey-Akert program reconsidered. J Cogn Neurosci. 1995; 7:1–24.

Price JL. Definition of the orbital cortex in relation to specific connections with limbic and visceral structures and other cortical regions. Ann N Y Acad Sci. 2007; 1121:54–71. [PubMed: 17698999]

Pritchard TC, Schwartz GJ, Scott TR. Taste in the medial orbitofrontal cortex of the macaque. Ann N Y Acad Sci. 2007; 1121:121–35. [PubMed: 17698994]

Quirk GJ, Mueller D. Neural mechanisms of extinction learning and retrieval. Neuropsychopharmacology. 2008; 33:56–72. [PubMed: 17882236]

Quirk GJ, Russo GK, Barron JL, Lebron K. The role of ventromedial prefrontal cortex in the recovery of extinguished fear. J Neurosci. 2000; 20:6225–31. [PubMed: 10934272]

Rangel A, Camerer C, Montague PR. A framework for studying the neurobiology of value-based decision making. Nat Rev Neurosci. 2008; 9:545–56. [PubMed: 18545266]

Rigotti M, Ben Dayan, Rubin D, Morrison SE, Salzman CD, Fusi S. Attractor concretion as a mechanism for the formation of context representations. NeuroImage. 2010a In press. 10.1016/j.neuroimage.2010.01.047

Rigotti M, Ben-Dayan Rubin D, Wang XJ, Fusi S. The importance of the diversity in neural responses in context-dependent tasks. Submitted. 2010b

Robbins T, Arnsten A. The neuropsychopharmacology of fronto-executive function: monoaminergic modulation. Annu Rev Neurosci. 2009; 32:267–87. [PubMed: 19555290]

Roesch MR, Olson CR. Neuronal activity related to reward value and motivation in primate frontal cortex. Science. 2004; 304:307–10. [PubMed: 15073380]

Rolls, E.; Deco, G. Computational Neuroscience of Vision. Oxford: Oxford Univ. Press; 2002.

Romanski LM, Bates JF, Goldman-Rakic PS. Auditory belt and parabelt projections to the prefrontal cortex in the rhesus monkey. J Comp Neurol. 1999; 403:141–57. [PubMed: 9886040]

Rudebeck PH, Bannerman DM, Rushworth MF. The contribution of distinct subregions of the ventromedial frontal cortex to emotion, social behavior, and decision making. Cogn Affect Behav Neurosci. 2008; 8:485–97. [PubMed: 19033243]

Rushworth MF, Behrens TE. Choice, uncertainty and value in prefrontal and cingulate cortex. Nat Neurosci. 2008; 11:389–97. [PubMed: 18368045]

Russell JA. A circumplex model of affect. J Pers Soc Psychol. 1980; 39:1161–78.

Saddoris MP, Gallagher M, Schoenbaum G. Rapid associative encoding in basolateral amygdala depends on connections with orbitofrontal cortex. Neuron. 2005; 46:321–31. [PubMed: 15848809]

Salzman CD, Belova MA, Paton JJ. Beetles, boxes and brain cells: neural mechanisms underlying valuation and learning. Curr Opin Neurobiol. 2005; 15:721–29. [PubMed: 16271457]

Salzman CD, Paton JJ, Belova MA, Morrison SE. Flexible neural representations of value in the primate brain. Ann N Y Acad Sci. 2007; 1121:336–54. [PubMed: 17872400]

Samejima K, Ueda Y, Doya K, Kimura M. Representation of action-specific reward values in the striatum. Science. 2005; 310:1337–40. [PubMed: 16311337]

Sanghera MK, Rolls ET, Roper-Hall A. Visual responses of neurons in the dorsolateral amygdala of the alert monkey. Exp Neurol. 1979; 63:610–26. [PubMed: 428486]

Schoenbaum G, Setlow B, Saddoris MP, Gallagher M. Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. Neuron. 2003; 39:855–67. [PubMed: 12948451]

Seymour B, O'Doherty JP, Dayan P, Koltzenburg M, Jones AK, et al. Temporal difference models describe higher-order learning in humans. Nature. 2004; 429:664–67. [PubMed: 15190354]

Spiegler BJ, Mishkin M. Evidence for the sequential participation of inferior temporal cortex and amygdala in the acquisition of stimulus-reward associations. Behav Brain Res. 1981; 3:303–17. [PubMed: 7306385]

Stefanacci L, Amaral DG. Topographic organization of cortical inputs to the lateral nucleus of the macaque monkey amygdala: a retrograde tracing study. J Comp Neurol. 2000; 421:52–79. [PubMed: 10813772]

Stefanacci L, Amaral DG. Some observations on cortical inputs to the macaque monkey amygdala: an anterograde tracing study. J Comp Neurol. 2002; 451:301–23. [PubMed: 12210126]

Stefanacci L, Suzuki WA, Amaral DG. Organization of connections between the amygdaloid complex and the perirhinal and parahippocampal cortices in macaque monkeys. J Comp Neurol. 1996; 375:552–82. [PubMed: 8930786]

Stuss DT, Levine B, Alexander MP, Hong J, Palumbo C, et al. Wisconsin Card Sorting Test performance in patients with focal frontal and posterior brain damage: effects of lesion location and test structure on separable cognitive processes. Neuropsychologia. 2000; 38:388–402. [PubMed: 10683390]

Sugase-Miyamoto Y, Richmond BJ. Neuronal signals in the monkey basolateral amygdala during reward schedules. J Neurosci. 2005; 25:11071–83. [PubMed: 16319307]

Sugrue LP, Corrado GS, Newsome WT. Matching behavior and the representation of value in the parietal cortex. Science. 2004; 304:1782–87. [PubMed: 15205529]

Susillo D, Abbott L. Generating coherent patterns of activity from chaotic neural networks. Neuron. 2009; 63:544–57. [PubMed: 19709635]

Sutton, R.; Barto, A. Reinforcement Learning. Cambridge, MA: MIT Press; 1998.

Suzuki WA, Amaral DG. Perirhinal and parahippocampal cortices of the macaque monkey: cortical afferents. J Comp Neurol. 1994; 350:497–533. [PubMed: 7890828]

Tremblay L, Schultz W. Relative reward preference in primate orbitofrontal cortex. Nature. 1999; 398:704–8. [PubMed: 10227292]

von der Malsburg C. The what and why of binding: the modeler's perspective. Neuron. 1999; 24:95–104. [PubMed: 10677030]

Wallis JD. Orbitofrontal cortex and its contribution to decision-making. Annu Rev Neurosci. 2007; 30:31–56. [PubMed: 17417936]

Wallis JD, Anderson KC, Miller EK. Single neurons in prefrontal cortex encode abstract rules. Nature. 2001; 411:953–56. [PubMed: 11418860]

Wallis JD, Miller EK. Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. Eur J Neurosci. 2003; 18:2069–81. [PubMed: 14622240]

Wang XJ. Probabilistic decision making by slow reverberation in cortical circuits. Neuron. 2002; 36:955–68. [PubMed: 12467598]

Weiskrantz L. Behavioral changes associated with ablation of the amygdaloid complex in monkeys. J Comp Neurol. 1956; 49:381–91.

Wellman LL, Gale K, Malkova L. GABAA-mediated inhibition of basolateral amygdala blocks reward devaluation in macaques. J Neurosci. 2005; 25:4577–86. [PubMed: 15872105]

Wilson FA, Rolls ET. The primate amygdala and reinforcement: a dissociation between rule-based and associatively-mediated memory revealed in neuronal activity. Neuroscience. 2005; 133:1061–72. [PubMed: 15964491]

Wise SP. Forward frontal fields: phylogeny and fundamental function. Trends Neurosci. 2008; 31:599–608. [PubMed: 18835649]

Wittgenstein, W. Philosophical Investigations. Oxford: Blackwell; 1958.

Yakovlev V, Fusi S, Berman E, Zohary E. Inter-trial neuronal activity in inferior temporal cortex: a putative vehicle to generate long-term visual associations. Nat Neurosci. 1998; 1:310–17. [PubMed: 10195165]

Glossary

PFC	prefrontal cortex
ACC	anterior cingulate cortex
OFC	orbitofrontal cortex
CS	conditioned stimulus
US	unconditioned stimulus

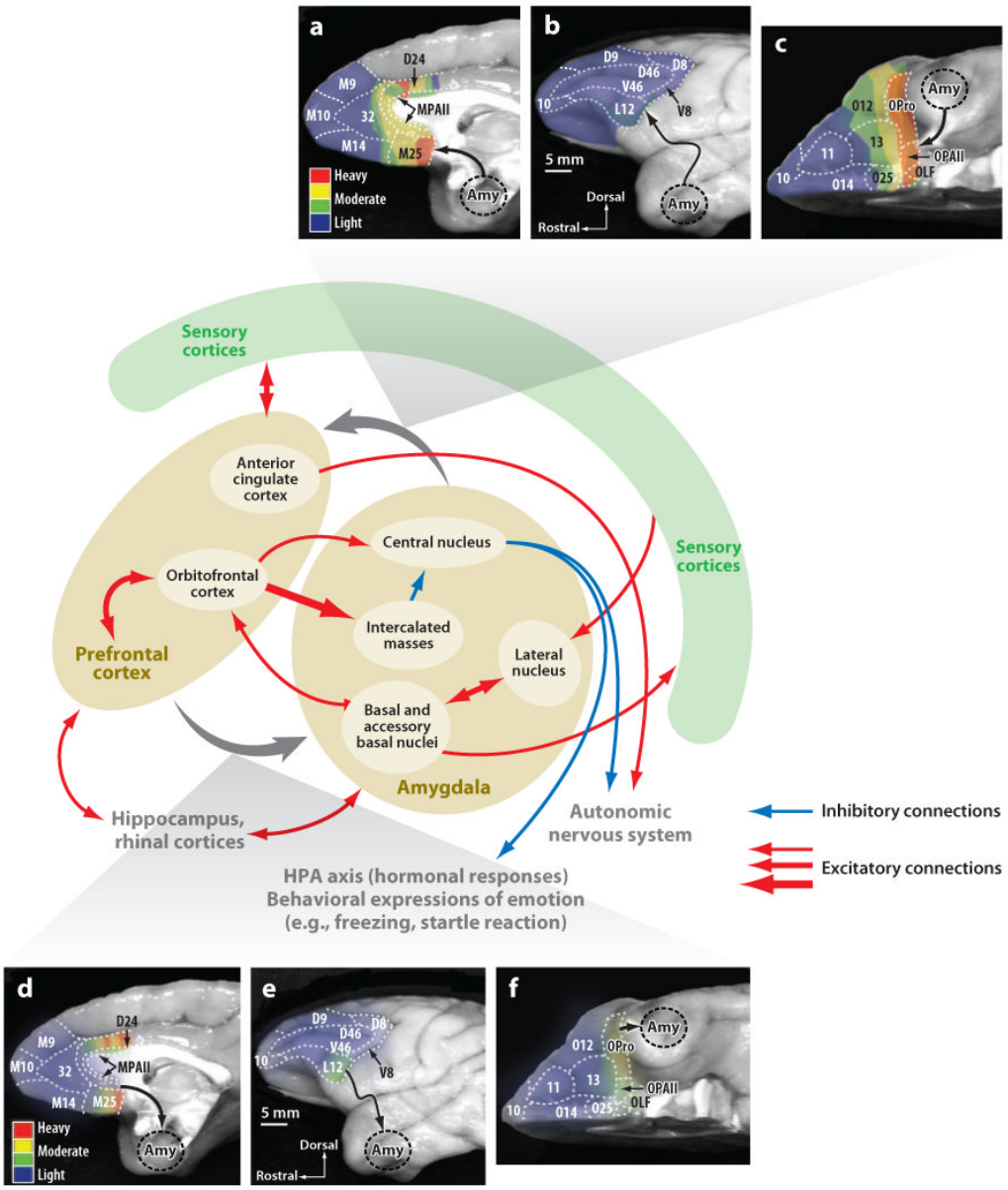

Figure 1.

Overview of anatomical connections of the amygdala and the prefrontal cortex (PFC). Schematic showing some (but not all) the main projections of the amygdala and the PFC. The interconnections of the amygdala and the PFC (and especially the OFC) are emphasized. (*a–c*) Summary of projections from the amygdala to the PFC (density of projections is color coded). (*d–f*) Summary of projections from the PFC to the amygdala (projection density is color coded). The complex circuitry between the amygdala and the OFC is also highlighted (*red arrows connect the structures*). Medial amygdala nuclei not shown. Many additional connections of both amygdala and PFC are not shown. Panels *a–f* were adapted with permission from figures 5 and 6 of Ghashghaei et al. (2007).

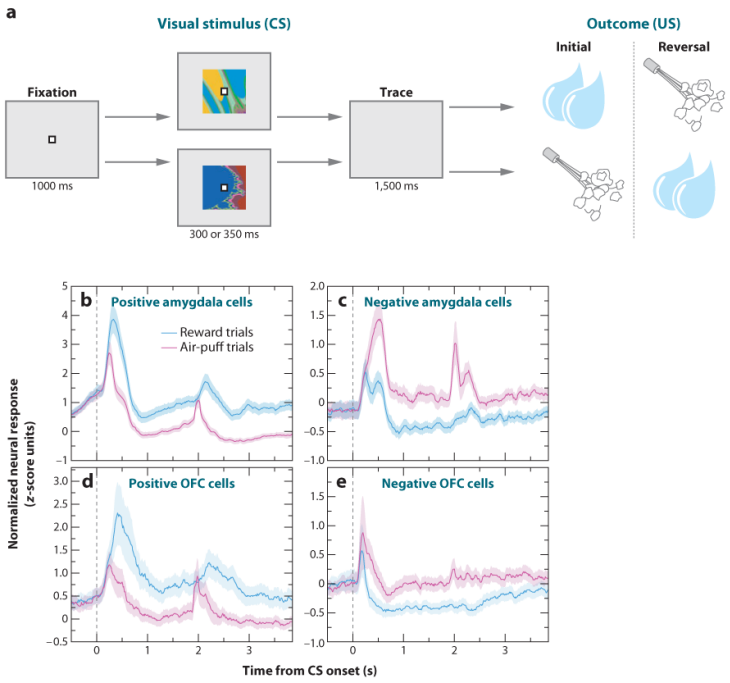

Figure 2.

Neural representation of positive and negative valence in the amygdala and the OFC. (*a*) Trace-conditioning task involving both appetitive and aversive conditioning. Monkeys first centered gaze at a fixation point. Each experiment used novel abstract images as conditioned stimuli (CS). After fixating for 1 s, monkeys viewed a CS briefly, and following a 1.5-ms trace interval, unconditioned stimulus (US) delivery occurred. One CS predicted liquid reward, and a second CS predicted an aversive air puff directed at the face. After monkeys learned these initial associations, as indicated by anticipatory licking and blinking, the reinforcement contingencies were reversed. A third CS appeared on one-third of the trials, and it predicted either nothing or a much smaller reward throughout the experiment (not depicted in the figure). (*b–e*) Normalized and averaged population peri-stimulus time histograms (PSTHs) for positive and negative encoding amygdala (*b*,*c*) and OFC (*d*,*e*) neurons.

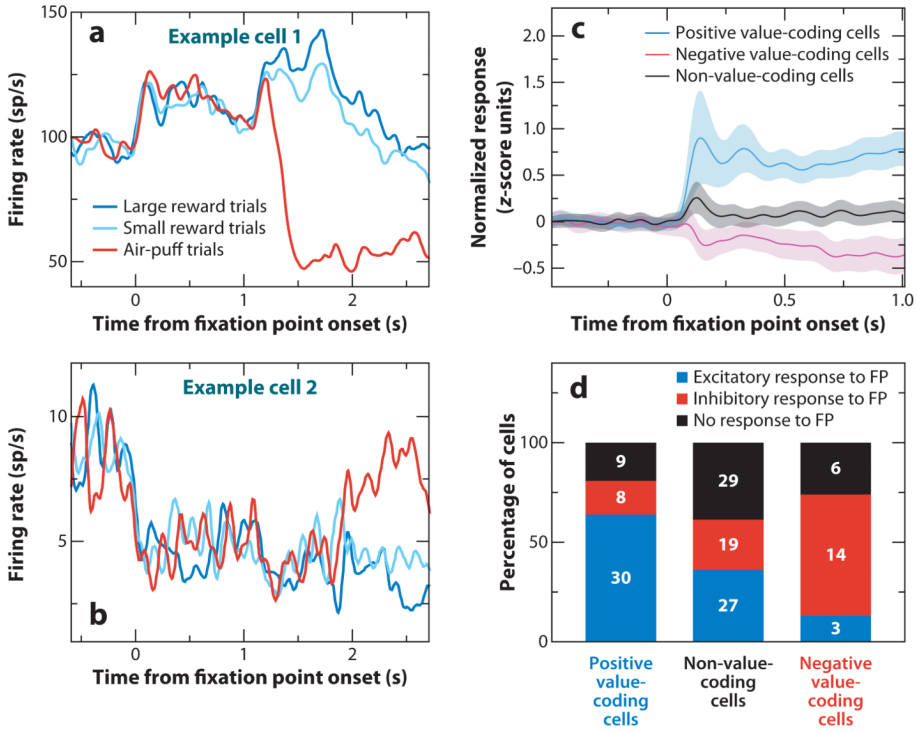

Figure 3.

Amygdala neurons track state value during the fixation interval. (*a*,*b*) PSTHs aligned on fixation point onset from two example amygdala neurons (*a*, positive encoding; *b*, negative encoding) revealing responses to the fixation point consistent with their encoding state value. (*c*) Averaged and normalized responses to the fixation point for positive, negative, and nonvalue-coding amygdala neurons. (*d*) Histograms showing the number of cells that increased, decreased, or did not change their firing rates as a function of which valence the neuron encoded. Note that the fixation point may be understood as a mildly positive stimulus, so positive neurons tend to increase their response to it and negative neurons decrease their response. Adapted with permission from Belova et al. (2008, figure 2).

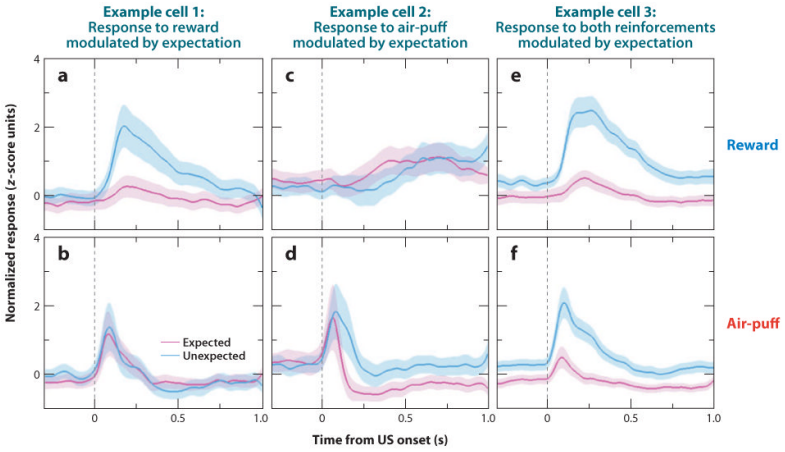

Figure 4.

Valence-specific and valence-nonspecific encoding in the amgydala. (*a–f*). Normalized and averaged neural responses to reinforcement when it was expected (*magenta*) and unexpected (*cyan*) for reward (*a,c,e*) and air puff (*b,d, f*). Expectation modulated reinforcement responses for only one valence of reinforcement in some cells (*a–d*) but modulated reinforcement responses for both valences in many cells (*e,f*). These responses are consistent with a role of the amygdala in valence-specific processes, as well as valence-nonspecific processes, such as attention, arousal, and motivation. Adapted with permission from Belova et al. (2007, figure 3).

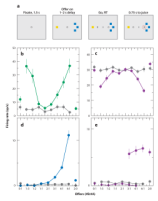

Figure 5.

OFC neural responses during economic decision-making. (*a*) Behavioral task. Monkeys centered gaze at a fixation point and then viewed two visual tokens that indicate the type and quantity of juice reward being offered for potential saccades to each location (*tokens, yellow and blue squares*). After fixation point extinction, the monkey is free to choose which reward it wants by making a saccade to one of the targets. The amounts of juices offered of each type are titrated against each other to develop a full psychometric characterization of the monkey's preferences as a function of the two juice types offered. (*b–e*) Activity of four neurons revealing different types of response profiles. X-axis shows the quantity of each offer type. Chosen value neurons increased (*b*) or decreased (*c*) their firing when the value of their chosen option increased. Offer value neurons (*d*) increased their firing when the value of one of the juices offered increased. Juice neurons (*e*) increased their firing for trials with a particular juice type offered, independent of the amount of juice offered. Adapted from Padoa-Schioppa & Assad (2006, figures 1 and 3) with permission.

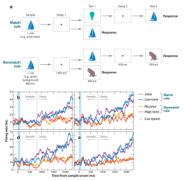

Figure 6.
Single neurons encode rules in PFC. (*a*) Behavioral task. Monkeys grasped a lever to initiate a trial. They then had to center gaze at a fixation point while viewing a sample object, wait during a brief delay, and then view a test object. Two types of trials are depicted (*double horizontal arrows*). On match rule trials, monkeys had to release the lever if the test object matched the sample object. On nonmatch rule trials, monkeys had to release the lever if the test object did not match the sample. Otherwise, they had to hold the lever until a third object appeared that always required lever release. The rules in effect varied trial-by-trial by virtue of a different sensory cue (e.g., tones or juice) presented during viewing of the sample object. (*b*,*c*) PFC neurons encoding match (*b*) or nonmatch (*c*) rules. Activity was higher in relation to the rule in effect regardless of the stimuli shown. Adapted with permission from Wallis et al. (2001, figures 1 and 2).

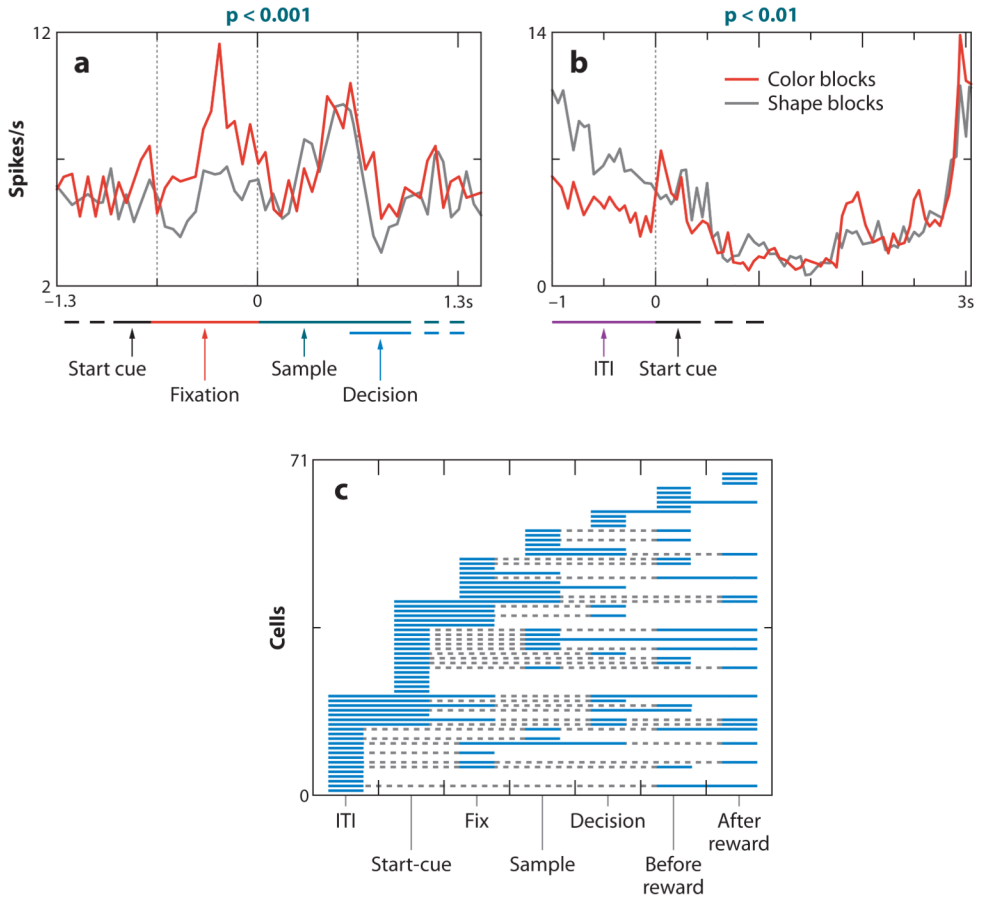

Figure 7.

PFC neurons encode rules in effect across time within a trial. Monkeys performed a task in which they had to match either the shape or color of two simultaneously presented objects with a sample object viewed earlier in the trial. Monkeys learned by trial and error whether a shape or color rule was in effect within a block of trials, and block switches were uncued to the monkey. (*a*,*b*) Two PFC cells that fired differentially depending on the rule in effect; activity differences emerged during the fixation (*a*) and intertrial intervals (ITI) (*b*). Activity is aligned on a start cue, which occurs before fixation on every trial. During the sample interval, one stimulus is presented over the fovea. During the decision interval, two stimuli are presented to the left and right; one matched the sample stimulus in color, and the other matched in shape. The correct choice can be chosen only if one has learned the rule in effect for the current block. (*c*) Distribution of activity differences between shape and color rules for each cell studied in each time interval of a trial. Each line corresponds to a single cell, and the solid parts of a line indicate when the cell fired differentially between color and shape blocks. Encoding of rules occurred in all time epochs, indicating that PFC neurons encode the rule in effect across time within a trial. Adapted with permission from Mansouri et al. (2006, figures 2 and 3).

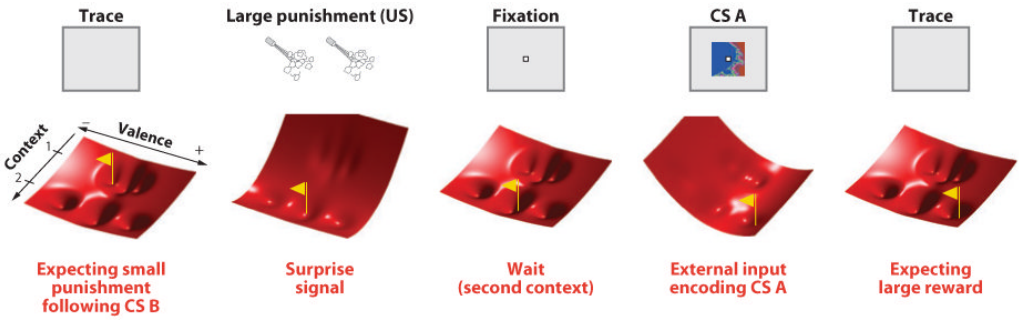

Figure 8.

The dynamics of context-dependent values. We consider a hypothetical variation of the experiment by Paton et al. (2006) in which there are two contexts. In the first context, CS A and B predict small rewards and punishments, respectively. In the second context, CS A and B are associated with large rewards and punishments. Six mental states now correspond to six valleys in the energy landscape. In panel 1, we consider the first trial of context 2, immediately after switching from context 1. Stimulus B has just been presented (not shown), and it is believed to predict small punishment. However a large punishment is delivered (not shown), and a surprise signal tilts the energy function (panel 2), inducing a transition to the neutral mental state of context 2 at the beginning of the next trial (panel 3; we assume for this example that the fixation interval has a neutral value). Now the system has already registered that it is in context 2. Consequently, the appearance of CS A tilts the energy landscape, and the mental state settles at a large positive value (panel 4). After CS disappearance, the network relaxes into the high positive value mental state of context 2. Thus the network does not need to relearn that CS A predicts a large reward because the network has already formed a representation for all the mental states contained within this simple experiment. Just knowing that the context has changed is sufficient for subjects to make an accurate prediction about impending reinforcement. For a detailed attractor model implementing this form of context dependency see Rigotti et al. (2010b).