The contribution of nonhuman primate research to the understanding of emotion and cognition and its clinical relevance

Silvia Bernardi[a,b] and C. Daniel Salzman[a,b,c,d,1]

[a]Department of Psychiatry, Columbia University, New York, NY 10032; [b]Zuckerman Mind Brain Behavior Institute, Columbia University, New York, NY 10027; [c]Department of Neuroscience, Columbia University, New York, NY 10027; and [d]New York State Psychiatric Institute, New York, NY 10032

Edited by Robert H. Wurtz, National Institutes of Health, Bethesda, MD, and approved November 1, 2019 (received for review September 13, 2019)

Psychiatric disorders are often conceptualized as arising from dysfunctional interactions between neural systems mediating cognitive and emotional processes. Mechanistic insights into these interactions have been lacking in part because most work in emotions has occurred in rodents, often without concurrent manipulations of cognitive variables. Nonhuman primate (NHP) model systems provide a powerful platform for investigating interactions between cognitive operations and emotions due to NHPs' strong homology with humans in behavioral repertoire and brain anatomy. Recent electrophysiological studies in NHPs have delineated how neural signals in the amygdala, a brain structure linked to emotion, predict impending appetitive and aversive stimuli. In addition, abstract conceptual information has also been shown to be represented in the amygdala and in interconnected brain structures such as the hippocampus and prefrontal cortex. Flexible adjustments of emotional behavior require the ability to apply conceptual knowledge and generalize to different, often novel, situations, a hallmark example of interactions between cognitive and emotional processes. Elucidating the neural mechanisms that explain how the brain processes conceptual information in relation to emotional variables promises to provide important insights into the pathophysiology accounting for symptoms in neuropsychiatric disorders.

nonhuman primates | amygdala | emotion | prefrontal cortex | abstraction

The ability to do, think, and feel in a flexible manner is a fundamental feature of human behavior. This flexibility relies on the ability to generalize from knowledge of past experiences and to adapt instantly in novel ones. This type of generalization can occur by virtue of our capacity to abstract "shared features" in the environments we experience, whether these are directly observable or hidden, and to crystallize these features into "concepts." Subsequently, these concepts can serve as organizing principles to understand environments never experienced before. Difficulties in processes such as abstraction and generalization can thereby impact our interpretation of the environment, leading to psychiatric symptoms ranging from distortions of reality in psychotic disorders, where patients typically lack abstraction abilities, to disruptions in affective state or anxiety, where patients do not flexibly regulate their emotions.

One of the central applications of the ability to abstract and generalize is to assign predictive (rewarding or aversive) values to environmental stimuli in novel situations. Anticipatory emotional responses to sensory stimuli require accurate prediction of the positive or negative emotional outcome associated with a stimulus presentation. This prediction does not only rely on knowing the link between a sensory stimulus and reinforcement, as such links may differ depending upon the circumstances. In addition, the costs of actions required to acquire a reward or avoid an aversive stimulus may affect value assessments; these costs can also be context dependent. Adaptive emotional responses to sensory stimuli therefore rely on knowing how a stimulus, potential actions, and the overall current set of circumstances—or context—together predict a particular outcome. Of note, context here is defined broadly as the set of circumstances in which the subject operates, including internal (e.g., information about cognitive, affective, and homeostatic variables) and external (e.g., information about the environment) variables. As a result, a neural representation that represents the meaning of predicted reinforcement outcome must account for context-dependent effects.

Nonhuman primate (NHP) models offer a unique opportunity to explore high-level cognition because of the richness of their behavioral repertoire, among the closest to those of humans. NHPs also possess strong brain homology to humans. NHPs are perhaps unique among nonhuman mammals in possessing, like humans, a well-developed internal granular layer in prefrontal cortex (PFC) (1–3). In general, extensive elaboration, specialization, and differentiation evolved in primate PFC (4–8) compared to other animal models. Moreover, although some aspects of amygdala anatomy and neurochemistry are relatively conserved across species, anatomical studies have pointed out differences in the convergence of cortical inputs into the amygdala (9, 10). Stereological analyses showed significant differences in volume and neuronal density across the lateral, basal, and accessory basal nuclei in monkeys compared to rats (11). In addition, the primate amygdala possesses elaborated, bidirectional connectivity with medial and orbital PFC (12–19) and the visual cortex (20), a dominant sensory modality in primates. NHP models therefore provide an essential intermediate step for understanding fundamental mechanisms involving cortico-limbic operations that underlie interactions between cognitive and emotional processes. The development and testing of circuit-based therapeutics may ultimately rely on NHP models due to these critical anatomical homologies and to the mechanistic insight being provided by studies in NHPs (21).

In this article, we first review NHP studies that have contributed to the understanding of processes related to reinforcement outcome anticipation, which are fundamental to emotional behavior since emotions often adjust in anticipation of impending rewarding or aversive events. Next, we review studies aimed at

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Using Monkey Models to Understand and Develop Treatments for Human Brain Disorders," held January 7–8, 2019, at the Arnold and Mabel Beckman Center of the National Academies of Sciences and Engineering in Irvine, CA. NAS colloquia began in 1991 and have been published in PNAS since 1995. From February 2001 through May 2019 colloquia were supported by a generous gift from The Dame Jillian and Dr. Arthur M. Sackler Foundation for the Arts, Sciences, & Humanities, in memory of Dame Sackler's husband, Arthur M. Sackler. The complete program and video recordings of most presentations are available on the NAS website at http://www.nasonline.org/using-monkey-models.

Author contributions: S.B. and C.D.S. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Published under the PNAS license.

[1]To whom correspondence may be addressed. Email: cds2005@columbia.edu.

First published December 23, 2019.

understanding higher-level processes that can adjust representations of reinforcement expectation. These processes include the assessment of a current situation in terms of recent reward history, which can adjust how a particular stimulus is valued, and even higher-level operations that involve abstraction of hidden (not observable) variables used to flexibly assign meaning to stimuli.

Although the reviewed studies focus on neurophysiological mechanisms in behaving NHPs, these mechanisms have clinical applications. In particular, impairments in accurately anticipating reinforcement outcomes—whether due to impairments in emotional learning or in cognitive processes that help regulate emotions—can lead to difficulties that underlie many psychiatric disorders. For example, these impairments can lead to decision-making difficulties contributing to addictive behavior. Furthermore, difficulties in coordinating emotional responses flexibly lies at the core of several anxiety and mood disorders, and of disorders in which paranoia is prominent. Even some symptoms of personality disorders—where neurobiological insights into cause are lacking—may arise from the inability to flexibly generalize and regulate affective and cognitive processes.

The Role of Amygdalar Circuits in Outcome Anticipation

For many years, the amygdala had been studied during the acquisition and expression of defensive behaviors elicited by an aversive conditioned stimulus (CS) or unconditioned stimulus (US) (reviewed in ref. 22), largely in rodents. However, substantial evidence indicates that the amygdala is involved in learning the relationship between a CS and both appetitive or aversive USs, as well as in coordinating responses to these stimuli (22–27). The initial view that the amygdala was specialized for aversive processing may have resulted from studies that only studied negative emotional valence; it remained unclear whether the same amygdala neurons participated in processing both positive and negative emotional valences. Moreover, studies in rodents had almost exclusively utilized either auditory or olfactory CSs. In primates, vision is a dominant sensory modality, so investigating how the amygdala assigns emotional significance to visual stimuli was clearly important.

Paton et al. (28) exploited monkeys' visual capabilities and employed a trace-conditioning task containing a contingency reversal for 2 visual CSs to establish that the amygdala appears to contain distinct appetitive and aversive neural systems. In this task, each CS was initially paired with an appetitive (liquid reward) or aversive (air puff) outcome, respectively (Fig. 1 *A* and *B*). After monkeys learned these associations, reinforcement contingencies reversed. Monkeys learned the new associations. Anticipatory licking and blinking reflected the monkeys' subjective valuation of impending USs and closely tracked the objective reinforcement to be received on every trial; thus, in this task, subjective valuation and objective reinforcement associated with stimuli were highly correlated. Individual amygdala neurons often responded preferentially to either positively or negatively conditioned CSs both before and after the reversal in contingencies (Fig. 1*C*). Such neurons were defined as encoding the positive or negative value of a CS based on the differential response to a CS when paired with rewarding or aversive USs. Value-coding neurons in the amygdala updated their response to a CS fast enough to account for behavioral learning of the reversed reinforced contingencies. Monkeys capacity to learn and reverse stimulus–outcome contingencies rapidly, just like humans can, was critical to the success of these experiments; behavioral paradigms that achieve such rapid learning and reversal are not easily realized in rodents.

Many value-coding neurons actually responded to stimuli of both valences but consistently responded more strongly to stimuli associated with either positive or negative valence (Fig. 1*C*). Nonetheless, analyses of simultaneously recorded neurons indicated that a greater frequency of short-latency peaks in shuffle-corrected cross-correlograms were observed for pairs of neurons sharing the same response selectivity compared to other pairs (29). These data provided physiological evidence for distinct and preferentially connected appetitive and aversive amygdalar circuits.

How does the amygdala acquire and update knowledge of the valence of stimuli? Error signals reflecting the difference between expected and received reinforcement have long been thought to be important for such learning (30, 31). Neurophysiological reflections of prediction errors have been shown in different brain areas, including reward prediction errors (RPEs) carried by midbrain dopamine neurons (32, 33), and as well as lateral habenula medial PFC (34–36), striatum (37–39), globus pallidus (40), and lateral habenula (41). Neurons in this last nucleus in particular encode prediction errors with the opposite valence compared to dopaminergic neurons, responding most strongly when an aversive stimulus is unexpected (41).

Given that the amygdala represents expected appetitive and aversive outcomes, an important question concerns whether evidence links RPEs and value representations in the amygdala. To address this question, Belova et al. (42) showed that neural

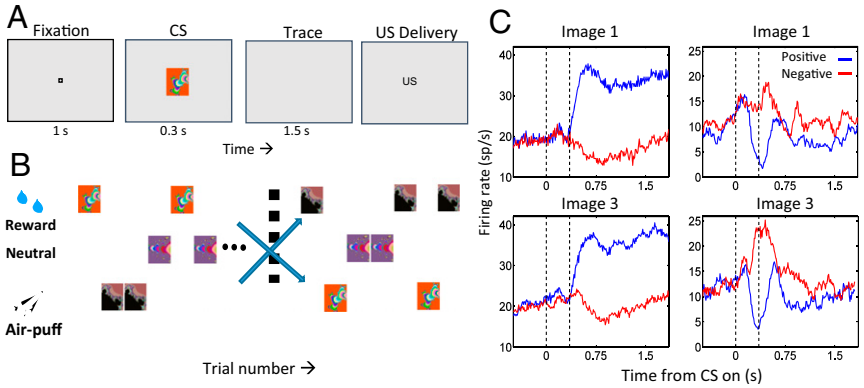

Fig. 1. Amygdala neurons represent the positive and negative valence of CSs. (*A*) Sequence of a trace-conditioning trial. Monkeys centered gaze at a fixation point for 1 s and viewed a fractal image for 300 ms. US delivery followed a 1,500-ms trace interval. (*B*) Task structure. Positive images, liquid reward; negative images, aversive air puff; nonreinforced images, no US. After monkeys learned initial contingencies, image reinforcement contingencies switched without warning. (*C*) PSTHs for 2 neurons. Reward trials, blue; air puff trials, red. Image 1 was initially rewarded, then paired with air puff after reversal; image 3, opposite contingencies. (*Left* column) Positive value-coding neuron, responding more strongly to both images when rewarded. (*Right* column) Negative value-coding neuron, responding most strongly when air puff follows each image presentation. Reprinted with permission from ref. 28.

responses to reinforcement in the amygdala were often stronger when a rewarding or aversive US occurs unexpectedly, such as immediately after a reversal in CS–US contingencies, or at a random time (42). These responses differed from classic RPE signals because they lacked a prototypical phasic, short-latency response emblematic of RPEs, as amygdala responses were typically sustained. Moreover, since CSs predicted reinforcement with 80% probability, the authors were able to examine neural responses upon reinforcement omission. Response modulation, a signature of RPE signals, was not observed. Nonetheless, there was a significant statistical association between whether expectation modulated responses to a US and whether the same cell also encoded the value of a CS, suggesting a link between these response properties. Interestingly, the basal forebrain, a structure bidirectionally connected with the amygdala, also contains neurons in which expectation modulates responses to USs without registering reinforcement omission (43).

Amygdala neurons also exhibited a diversity in response profiles to unexpected USs. Some neurons enhanced responses to an unexpected reward or to an aversive stimulus, but not both, which are valence-specific modulations. These neural responses are more analogous to signals observed in dopamine neurons or the habenula, respectively (32, 41), without characteristic phasic responses and modulation by reinforcement omission. Other neurons modulated responses to both valences of unexpected USs, a valence-nonspecific profile similar to that observed in rodents (44). The amygdala projects to brain structures involved in autonomic responses (45). Since autonomic arousal can occur in relation to both positive and negative emotional states of high intensity, neurons that respond to unexpected reinforcement in a valence-nonspecific manner might mediate such autonomic arousal, or attention that can be amplified in relation to both emotional valences.

A key concept in reinforcement learning models is that values are assigned to states (31). Neurophysiological evidence indicates that amygdala neurons represent not just reinforcement expected upon encountering a stimulus, but instead represent the value of a state. An animal experiences a variety of states even in a simple conditioning protocol (46). Specifically, amygdala neurons' responses to a fixation point, which monkeys voluntarily foveate to begin a trial, reflects the mildly positive value that this stimulus possesses to motivated subjects (46). Neurons belonging to the appetitive system (those that respond more strongly to a CS when it predicts a reward than when the same CS predicts an aversive stimulus) tend to increase their firing rate in response to a fixation point, but neurons belonging to the aversive system tend to decrease their firing rate to the same fixation point. Moreover, neurons in the appetitive system tend to fire more strongly to rewarding USs, and neurons in the aversive system have the opposite response profile (46), a finding confirmed in rodents at the population level (47). These findings raise the possibility that the amygdala represents the motivational significance of many types of stimuli. Indeed, one recent result suggests that neural ensembles in the amygdala that encode the value of CSs also encode information about the hierarchical rank of conspecifics within a social group (48). Hierarchical status is social information that can influence many emotional and cognitive operations.

The same expected reward can have different motivational meanings (and thereby different state values) depending upon internal variables, such as knowledge of the magnitude of other recently received rewards. To study the neurophysiology mediating this assignment of motivational meaning, the activity of neurons in the amygdala and orbitofrontal cortex (OFC) (Brodmann area 13/13m) of monkeys were recorded during a Pavlovian task in which the relative amount of liquid reward associated with one CS was manipulated by changing the reward amount associated with a second CS (49). Anticipatory licking tracked relative reward magnitude, implying that monkeys integrated information about recent rewards—a process that changes an internal state variable—to adjust the subjective meaning of a CS. Upon changes in relative reward, neural responses in the amygdala and OFC also updated, despite the fact that the US associated with the CS had not changed (Fig. 2) (49). These results tie neural response properties in the amygdala and OFC to state value as manipulated by recent reward history. Strikingly, neural responses to reward-predictive cues updated more rapidly in OFC than amygdala, and activity in OFC but not the amygdala was correlated with recent reward history (49). These results highlight a distinction between the amygdala and OFC in assessing reward history to adjust an internal state variable to support flexible assignment of motivational meaning to sensory cues.

Other studies have reported neural signals in amygdala that reflect expected reward value over long timescales during planning (50). Grabenhorst et al. (51) reported prospective amygdala signaling of immediate, current-trial choices and self-determined choice sequences that determined distant rewards. Thus, monkeys could implement internal plans to save reward consumption until the end of a sequence of trials, increasing overall reward since deferring reward resulted in reward increased by an "interest rate" (50). This response profile could reflect plans up to 2 min in the absence of external cues (52). Activity in the amygdala is thereby not limited to the assignment of valence to stimuli about to be received, but also reflects planning that determines long-term cumulative reward. Many aspects of these results may be viewed as an extension of the notion that the amygdala encodes state value, as signals related to planning that results in increased reward may reflect the value of the state in part defined by an internal plan to receive more reward later. The mechanisms of this type of cognitive influence on amygdala neural activity is ripe for investigation.

Cortico-Amygdalar Mechanisms and Their Role in the Regulation of Emotions

The amygdala is anatomically interconnected with a broad range of cortical and subcortical brain structures likely responsible for its role in emotional expression and for interactions between cognitive and emotional processes (18, 45). Within PFC, one prominent bidirectional anatomical pathway lies between the amygdala and OFC (12, 15–17). Both amygdala and OFC represent the motivational significance of sensory stimuli (5, 53–56), and they are thought to play an important role in value-based decision making (5, 22, 24, 26, 50–52, 57, 58). Neurons in OFC track positive and negative value in a consistent manner across the different sensory events in a conditioning trial, including the fixation point, CS, and US presentations (26, 54, 59). Moreover, OFC neural responses are correlated with monkeys' behavioral use of information about both rewarding and aversive CSs (60, 61).

OFC has long been conceptualized as a more cognitive brain area than the amygdala (62), and it has been proposed as a key structure in mediating reversal learning—a task often used to study flexible updating of reinforcement expectation—by some (but not all) studies (63–70). Leveraging the richness of bidirectional amygdala–cortical connectivity of NHP models, the physiological role of the amygdala and OFC in reversal learning was recently studied (28, 61). In principle, flexible adjustment of reinforcement expectation during reversal learning can involve 2 types of mechanisms, a trial-by-trial learning process or an inferential mechanism. In trial-by-trial learning, each CS–US (or action–US) association is learned independently upon reversal. Although behavioral training can lead to faster learning rates, learning about each CS–US is a distinct process. In reversal learning mediated by inference, knowledge of the structure of the task—whereby all CS–US contingencies switch simultaneously—can enable a subject to adjust their responses to the first appearance of a CS after a reversal if a different CS has already been experienced after the reversal.

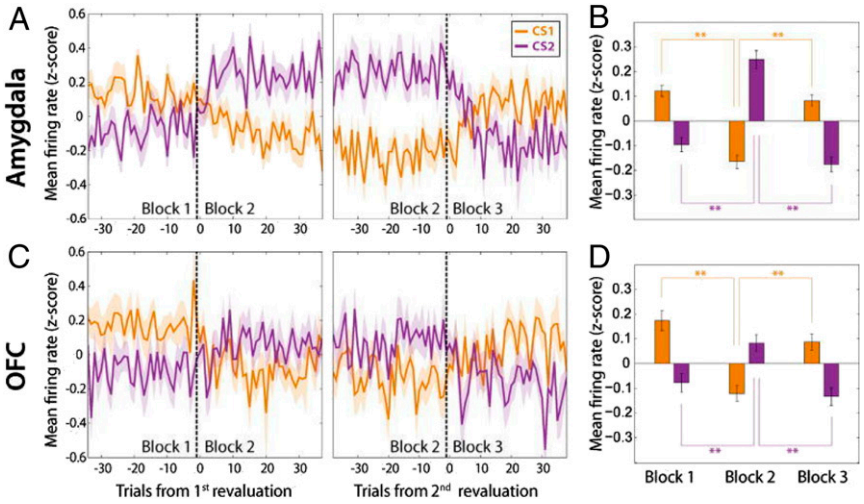

Fig. 2. Neurons in the amygdala and OFC represent the relative amount of reward associated with a CS. (*A*) Population average firing rate plotted as a function of trial number for amygdala neurons that responded selectively to the amount of expected reward. (*Left*) Activity changes in relation to a revaluation of CS1 (orange), which occurs by increasing the reward amount associated with CS2 (purple). (*Right*) Activity changes in relation to a second revaluation of CS1, which now occurs by decreasing the reward amount associated with CS2. (*B*) Average firing rate of neurons in *A* for each block. $^{**}P < 0.01$ (Wilcoxon sign-rank test). (*C* and *D*) Same as *A* and *B*, except for neurons recorded in OFC. Adapted from ref. 49, with permission from Elsevier.

Morrison et al. (61) examined OFC (area 13/13m) and amygdala neurophysiological responses using a reversal-learning study and showed that appetitive and aversive networks in OFC and amygdala exhibit different learning rates in the 2 brain areas. For positive cells, changes in OFC neural activity after reversal occurred more rapidly than those in positive cells in the amygdala; for negative cells, the aversive network in the amygdala learned more rapidly than in OFC (Fig. 3). In each case, the faster-changing area was completing its transition around the time of the onset of changes in behavior. The findings suggest that distinct sequences of neural processing lead to the updating of representations within appetitive and aversive networks within these structures. Perhaps the faster learning response of the aversive network in the amygdala reflects the preservation across evolution of an aversive system that learns very quickly in order to avoid threats (71).

Neural Representations of Conceptual Information and Their Potential Role in Flexible Cognitive and Emotional Behavior

In the studies described until now, NHPs were learning and relearning contingencies in different situations. In real life, this is

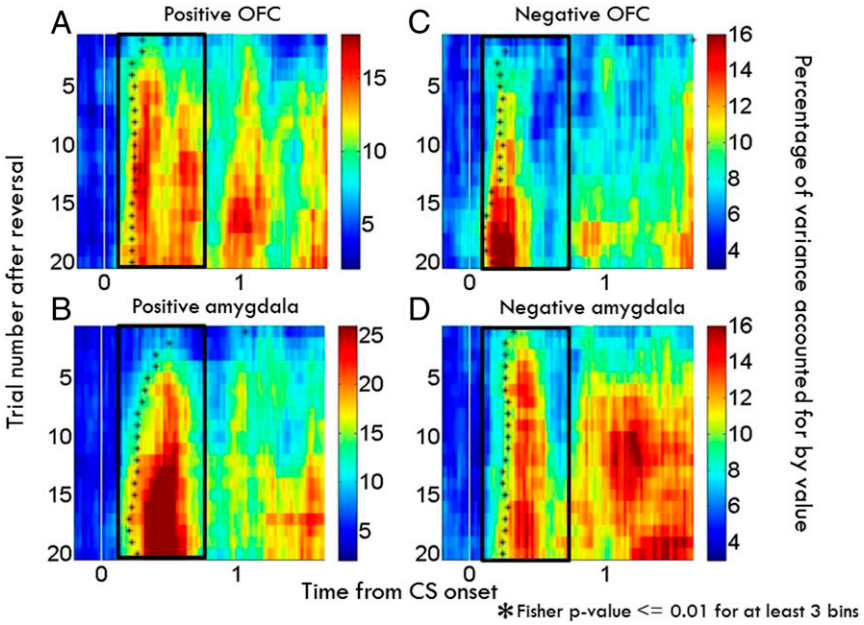

Fig. 3. Neurons within appetitive and aversive networks in amygdala and OFC update responses to changed reinforcement contingencies at different rates. Time course of changes in value-related signals in amygdala and OFC plotted as a function of time and trial number relative to reversal (*A–D*). For each bin, an index computed for each cell the proportion of variance accounted for by image value divided by the total variance using a 2-way ANOVA (61), and this index was averaged across populations. (*A* and *B*) Average contribution of image value in positive value-coding neurons in OFC (*A*) and amygdala (*B*). (*C* and *D*) Same as *A* and *B*, except for negative value-coding cells. Black asterisks, time when the contribution-of-value index becomes significant (asterisks placed in center of the first of at least 3 consecutive significant bins; Fisher $P < 0.01$). Bin size, 200 ms. Bin steps, 20 ms. Adapted from ref. 61, with permission from Elsevier.

the equivalent of learning that some sources of heat can burn if touched. Of course, humans can abstract a concept, "heat," and apply it flexibly to different sources of heat to not get burned every time they get close to a stove, yet still expect pleasure from a hot shower. In this example, both an abstract concept (heat), and 2 rules (heat can burn, or be comforting), can be used to adjust behavior, even upon seeing a stove or other source of heat never seen before. Subjects do not need to relearn the contingencies of the newly observed stove to adjust responses. Instead, subjects can generalize to the new situation, or context, and avoid touching the stove.

Investigation of context-dependent changes in contingencies has occurred in rodents, typically employing contexts cued by sensory stimuli, such as through using different textures in 2 chambers (72), but providing an exhaustive review of these studies is beyond the scope of this review. Overall, in those studies, explicit (observable) contextual cues signal the rules governing the relationship between the environment and reinforcement. This bears similarity to the cases where each shower is labeled with "comforting heat" and each stove is labeled with "do not touch, risk of burning."

The ability to adjust behavior flexibly in a context-dependent manner—yet when the contexts are not cued—requires that the brain represent hidden variables, such as rules. NHPs provide an important platform for studying context-dependent behavior when contexts are described by hidden variables. The encoding of rules in PFC has been extensively investigated in NHP models (73–75), but in similar rodent studies, experiments have rarely used an uncued rule. Moreover, investigation of the role of PFC in creating neural representations that can support generalization to new situations has not occurred. Generalization to new situations is a critical element of human cognition, including the ability to anticipate outcomes in novel situations and consequently regulate emotional responses. Most studies of emotion have focused on learning mechanisms (e.g., fear learning, or extinction), but mechanisms by which the brain constructs representations of conceptual information that can be used to regulate emotions have remained largely unstudied.

Saez et al. (76) recently investigated the nature of conceptual information and its use in the process of stimulus–value updating in changing contexts. Here, monkeys performed a serial reversal-learning task in which 2 CS–US pairs switched contingencies many times every experiment. Each day, monkeys would be presented with 2 novel CSs and would learn which one was paired with a reward US and which with nothing. The CSs then switched contingencies many times in experiments, with each set of CS–US pairs comprising a "task set."

Unlike in Morrison et al. (61), in this study, monkeys' behavior suggested they employed inference to update reinforcement expectation, as assayed by anticipatory licking. Monkeys adjusted their licking in a switch-like manner after recognizing that reinforcement contingencies had changed. As soon as they experienced that one CS had switched contingencies, they inferred that the second CS had also changed contingencies and adjusted their licking. This behavior was consistent with the possibility that upon experiencing one switched CS–US contingency, monkeys can generalize their knowledge of task structure (i.e., the task sets that defined 2 contexts) and apply it to the other CS. We refer to each task set as a "context," although it should be reiterated that context is an implicit, not perceptually cued, rule. This type of inference has not, to our knowledge, been reported in rodents during serial reversal-learning tasks, and it is likely mediated by the extensive elaboration of the PFC (and perhaps amygdala) in monkeys. These results suggest that the monkeys acquired and retained a "mental map," a representation of the rules of the task. It is important to keep in mind that this is a mental map of the rules governing environmental contingencies; therefore, context is an implicit variable, not perceptually cued, as opposed to a mental map of locations and the concrete external environment of the agent.

During performance of this task, Saez et al. (76) recorded single units from the amygdala, the anterior cingulate cortex (ACC) (another part of PFC bidirectionally connected with the amygdala) and the OFC (Brodmann areas 13, 13m). They used a linear decoder to analyze task-relevant signals as a function of time by considering populations of neurons collectively (Fig. 4). Information about the map—the task set—was encoded not only in ACC and OFC but also in the amygdala (76). This signal reflected a process of abstraction (76). Strikingly, errors in behavior (licking when no reward was going to be delivered, or not licking when reward was going to occur) were correlated with the failure to maintain upon stimulus appearance a representation of the context in the amygdala (76).

The data reported by Saez et al. (76) indicate that inference can be used by monkeys and that neural activity that reflects this process is represented not only in PFC, but also in amygdala. Inference is a cognitive process that can occur by knowing concepts. Concepts, in turn, may be thought of as arising from a process of abstraction, a process that finds features, whether explicit or hidden, shared by instances. The process of abstraction is an active area of research in reinforcement learning, as it provides a solution for the notorious "curse of dimensionality," i.e., the exponential growth of the solution space required to encode all states of the environment (77). In real life, the abstraction of the concept of heat obviates the need to touch every heat source to discover its burning properties. However, a clear account of how the brain may represent abstract variables has remained elusive, and this is a critical issue for understanding many cognitive functions, including the regulation of emotion.

Concepts ("abstract" variables) can correspond to hidden variables not directly observable in the environment, as well as to explicit variables, such as a perceptual category, like color. Moreover, the same stimulus or event may be described by multiple abstract variables, allowing generalization to different types of situations. Returning to our previous example, the concept of heat includes an understanding that heat can burn, as well as the fact that heat can make you sweat. Thus, one can link "heat" to these 2 abstract variables (and more). With the goal of elucidating how the brain represents abstract information, Bernardi et al. (78) employed a task in which monkeys performed a more complex serial reversal-learning task; this task included an operant action, and switching between 2 uncued task sets (Fig. 5). Each set had 4 trial conditions containing different stimulus–response–outcome mappings. The stimuli, responses, and reinforcement outcomes were explicit variables (observable), while the 2 task sets were described by a hidden variable.

This task engaged a series of cognitive operations, including perceiving a stimulus, making a decision as to which action to execute, and then expecting and sensing reward to determine whether an error occurred so as to update the decision process on the next trial. Monkeys exhibited inference, as they would adjust their behavior after a reversal on the first appearance of a trial condition if they had already experienced other switched types of trial conditions.

Bernardi et al. (78) recorded neural activity in anterior hippocampus (HPC), dorsolateral PFC (DLPFC), and ACC. The HPC was targeted in this study because of its role in generating episodic associative memories, as well as its role in processes of abstraction suggested by imaging studies (79, 80). PFC areas were targeted because encoding of rules has been reported (73, 74). Moreover, a recent study observed that different oscillatory dynamics might reflect different functional roles for categorization based on bottom-up features or more abstract concepts, with ventrolateral PFC gamma oscillations more engaged for lower-level abstraction and DLPFC beta for higher-level abstraction (81).

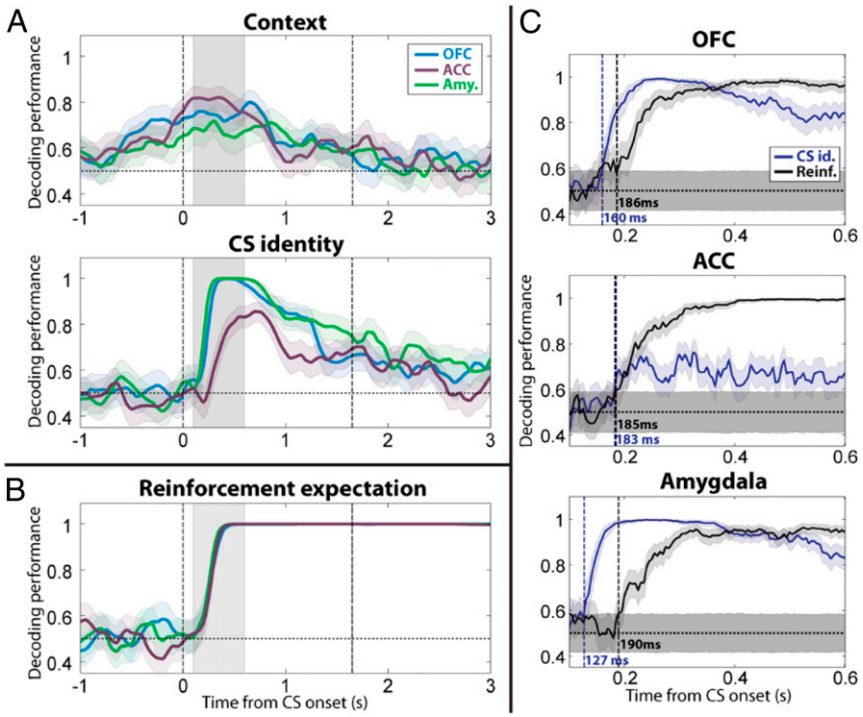

Fig. 4. Neural ensembles in amygdala, OFC, and ACC encode uncued contexts, stimulus identity, and expected reinforcement. (*A*) Decoding performance for context and CS identity in amygdala, OFC, and ACC (250-ms sliding window, 50-ms steps). Blue, OFC; purple, ACC; green, amygdala. Shaded areas, 95% confidence intervals (bootstrap). Vertical dashed lines, CS onset and earliest possible US onset. (*B*) Decoding performance for reinforcement expectation plotted vs. time relative to image onset. (*C*) Timing of onset of CS identity (blue) and reinforcement expectation (black) signals in OFC, ACC, and amygdala (50-ms sliding window, 5-ms steps for 500-ms window shown in *A* and *B* by gray shading). Vertical dashed lines (and labels) indicate the first time bin where decoding performance is significantly above chance level and remains there for 10 time bins. Shaded areas, 95% confidence intervals around chance (shuffle). Adapted from ref. 76, with permission from Elsevier.

Ponsen et al. (82) define abstraction as an operation that changes the representation of an object by hiding or removing less critical details while preserving desirable properties, a process that could enable generalization upon encountering other similar but novel objects. Bernardi et al. (78) sought to determine whether neural ensembles represented variables in a manner reflecting abstraction. They therefore did not only ask whether a neural ensemble represents information. Instead, they also determined whether the format of a representation would support generalization to novel situations. A variable was defined as being represented in an abstract format if a linear decoder could be trained to classify it on a subset of trial conditions (e.g., only on trial types A and B), and then be used to correctly classify the variable on held-out trial conditions (e.g., trial conditions C and D). The decoder's ability to classify task conditions not previously "experienced" by the decoder was used as a quantitative index as to whether the representation of a variable was generalizable and hence in an abstract format. Note that this use of a decoder differed from using a decoder to determine whether a neural ensemble simply encodes a variable, which traditionally entails training and testing a decoder on the same types of trial conditions.

Neural ensembles in DLPFC, ACC, and HPC were observed to represent multiple variables in an abstract format simultaneously during performance of this serial reversal-learning task. Prior to the stimulus presentation of a trial, in DLPFC and ACC, the recently performed action reflecting a decision, the reinforcement recently received, and the hidden variable context were all represented in abstract format. The HPC encoded the context and value of the recent trial in an abstract format, but not the action, which was represented but not in an abstract format.

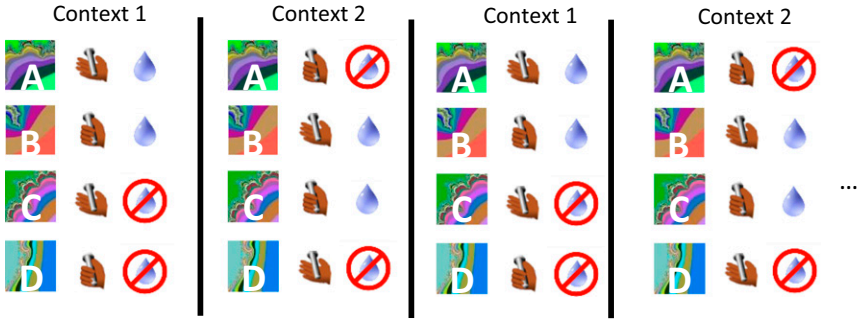

Fig. 5. Serial reversal-learning task in which 2 stimulus–response–outcome task sets define 2 contexts. Monkeys hold a bar and then fixate to begin a trial. Upon viewing an image, monkeys must hold or release a bar to perform correctly. Two stimuli in each context were rewarded after correct decisions (context 1: A, C; context 2: B, C).

The content and format of representations of these variables changed as task events engaged the cognitive operations needed to perform the task. After an image appeared, neural representations of the planned action and the expected reward occur more rapidly in DLPFC and ACC than in HPC, suggesting that these prefrontal areas may play a more prominent role in the decision process. Notably, value and action were in abstract format in all brain areas shortly after image onset, but context was not abstract in the DLPFC despite being decodable. In ACC, context was only weakly abstract despite being strongly decodable. In order to make a correct decision on this task, information about context must be "mixed" nonlinearly with information about the stimulus. Thus, a brain area mediating the decision process may not maintain the representation of context in an abstract format after a stimulus appearance. Both the rapid emergence of neural signals reflecting the action and value of the current trial, and the fact that the representation of context becomes less abstract, suggests a primary role for the PFC, especially DLPFC, in decision making on this task. Nonetheless, prior to the presentation of the stimulus on the next trial, the representation of the hidden variable context evolves into an abstract format in all 3 brain structures. These results highlight how the format in which a variable is represented can distinguish between the coding properties of brain areas, even when the content of information represented in those areas is similar.

Conclusion

Neuropsychiatric disorders are not merely disorders of emotion. Instead, dysfunction in cognitive operations—operations that form, encode, and utilize representations of conceptual information—play a vital role in mechanisms that regulate or update emotional responses to environmental events. Electrophysiological studies in NHPs have now described how appetitive and aversive amygdalar networks are updated during learning, and how amygdala signals are related to neural signals in the OFC upon reversals in contingencies. More recent studies have provided insights into how abstract variables are represented in the brain, a function critical to being able to adjust emotional and cognitive behavior upon encountering new situations. Future studies must delineate how these representations of abstract variables are created and utilized to confer such flexibility.

Efforts to understand the neural underpinnings of how abstract variables are represented and utilized are vital to the understanding of psychiatric morbidity. In humans, disruption in the neural mechanisms underlying the attribution of value, and therefore the modulation of anticipatory emotions and decision making, often leads to maladaptive patterns of behavior. For example, addiction is commonly understood as resulting from disruption in circuitry that normally processes reward-predictive stimuli and that must regulate behavioral responses to such stimuli, a process that must utilize conceptual information to generalize appropriately (71, 83). Deficits in contextual processing—even when contexts are not explicitly cued—can lead to interpersonal sensitivity with failures in emotional regulation such as that observed in social anxiety and borderline personality disorders (84) and schizophrenia (71, 85). Moreover, the inability to contextualize how stimuli and actions are related to aversive events can lead to posttraumatic stress disorder (86). The development of transformative treatment strategies for a range of psychiatric disorders likely relies on our acquiring a much more detailed understanding of neural circuitry dedicated to processing rewarding and aversive information, and their interactions with circuitries responsible for cognitive functions (71, 87). Because of brain homology and the richness of its behavioral repertoire, NHP models likely will prove to be a critical platform for developing new treatments. This model has already proven powerful for treatments of movement disorders, such as Parkinson's disease, by using brain stimulation informed by neurophysiological studies. Persistence and creativity in experimental approaches promise the same rewards for helping cure the complex symptomatology of neuropsychiatric illness.

Data Availability. The manuscript contains no original data.

ACKNOWLEDGMENTS. C.D.S. received support by the Simons Foundation, National Institute of Mental Health (NIMH) R21MH116348, R01MH082017), and National Institute on Drug Abuse (NIDA), (R21DA045989). S.B. received support from NIMH (1K08MH115365, T32MH015144 and R25MH086466), and from the American Psychiatric Association, the Leon Levy Foundation, and the Brain & Behavior Research Foundation young investigator fellowships.

1. D. Ongür, J. L. Price, The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans. *Cereb. Cortex* **10**, 206–219 (2000).
2. M. Petrides, D. N. Pandya, Dorsolateral prefrontal cortex: Comparative cytoarchitectonic analysis in the human and the macaque brain and corticocortical connection patterns. *Eur. J. Neurosci.* **11**, 1011–1036 (1999).
3. M. Petrides, D. N. Pandya, Comparative cytoarchitectonic analysis of the human and the macaque ventrolateral prefrontal cortex and corticocortical connection patterns in the monkey. *Eur. J. Neurosci.* **16**, 291–310 (2002).
4. S. Bouret, B. J. Richmond, Ventromedial and orbital prefrontal neurons differentially encode internally and externally driven motivational values in monkeys. *J. Neurosci.* **30**, 8591–8601 (2010).
5. X. Cai, C. Padoa-Schioppa, Contributions of orbitofrontal and lateral prefrontal cortices to economic choice and the good-to-action transformation. *Neuron* **81**, 1140–1151 (2014).
6. P. S. Goldman-Rakic, A. R. Cools, K. Srivastava, The prefrontal landscape: Implications of functional architecture for understanding human mentation and the central executive. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **351**, 1445–1453 (1996).
7. R. Levy, P. S. Goldman-Rakic, Segregation of working memory functions within the dorsolateral prefrontal cortex. *Exp. Brain Res.* **133**, 23–32 (2000).
8. S. P. Wise, Forward frontal fields: Phylogeny and fundamental function. *Trends Neurosci.* **31**, 599–608 (2008).
9. A. J. McDonald, F. Mascagni, Immunohistochemical localization of the β2 and β3 subunits of the GABAA receptor in the basolateral amygdala of the rat and monkey. *Neuroscience* **75**, 407–419 (1996).
10. A. J. McDonald, Cortical pathways to the mammalian amygdala. *Prog Neurobiol.* **55**, 257–332 (1998).
11. L. J. Chareyron, P. Banta Lavenex, D. G. Amaral, P. Lavenex, Stereological analysis of the rat and monkey amygdala. *J. Comp. Neurol.* **519**, 3218–3239 (2011).
12. J. L. Price, Comparative aspects of amygdala connectivity. *Ann. N. Y. Acad. Sci.* **985**, 50–58 (2006).
13. J. Price, R. Russchen, D. Amaral, "The limbic region. II. The amygdaloid complex" in *Handbook of Chemical Neuroanatomy, Vol. 5, Integrated Systems of the CNS, Part I, Hypothalamus, Hippocampus, Amygdala, Retina*, A. Bjorklund, T. Hokfelt, L. W. Swanson, Eds. (Elsevier, 1987), pp. 279–388.
14. D. G. Amaral, J. L. Price, Amygdalo-cortical projections in the monkey (*Macaca fascicularis*). *J. Comp. Neurol.* **230**, 465–496 (1984).
15. S. T. Carmichael, J. L. Price, Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. *J. Comp. Neurol.* **363**, 615–641 (1995).
16. H. T. Ghashghaei, H. Barbas, Pathways for emotion: Interactions of prefrontal and anterior temporal pathways in the amygdala of the rhesus monkey. *Neuroscience* **115**, 1261–1279 (2002).
17. H. T. Ghashghaei, C. C. Hilgetag, H. Barbas, Sequence of information processing for emotions based on the anatomic dialogue between prefrontal cortex and amygdala. *Neuroimage* **34**, 905–923 (2007).
18. L. Stefanacci, D. G. Amaral, Some observations on cortical inputs to the macaque monkey amygdala: An anterograde tracing study. *J. Comp. Neurol.* **451**, 301–323 (2002).
19. L. Stefanacci, W. A. Suzuki, D. G. Amaral, Organization of connections between the amygdaloid complex and the perirhinal and parahippocampal cortices in macaque monkeys. *J. Comp. Neurol.* **375**, 552–582 (1996).
20. D. G. Amaral, H. Behniea, J. L. Kelly, Topographic organization of projections from the amygdala to the visual cortex in the macaque monkey. *Neuroscience* **118**, 1099–1120 (2003).
21. S. N. Haber, B. Knutson, The reward circuit: Linking primate anatomy and human imaging. *Neuropsychopharmacology* **35**, 4–26 (2010).
22. J. E. LeDoux, Emotion circuits in the brain. *Annu. Rev. Neurosci.* **23**, 155–184 (2000).
23. M. G. Baxter, E. A. Murray, The amygdala and reward. *Nat. Rev. Neurosci.* **3**, 563–573 (2002).
24. F. Gore, E. C. Schwartz, C. D. Salzman, Manipulating neural activity in physiologically classified neurons: Triumphs and challenges. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20140216 (2015).

25. K. M. Tye, G. D. Stuber, B. de Ridder, A. Bonci, P. H. Janak, Rapid strengthening of thalamo-amygdala synapses mediates cue-reward learning. *Nature* **453**, 1253–1257 (2008).
26. S. E. Morrison, C. D. Salzman, Re-valuing the amygdala. *Curr. Opin. Neurobiol.* **20**, 221–230 (2010).
27. J. Resnik, R. Paz, Fear generalization in the primate amygdala. *Nat. Neurosci.* **18**, 188–190 (2015).
28. J. J. Paton, M. A. Belova, S. E. Morrison, C. D. Salzman, The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature* **439**, 865–870 (2006).
29. W. Zhang *et al*., Functional circuits and anatomical distribution of response properties in the primate amygdala. *J. Neurosci.* **33**, 722–733 (2013).
30. R. A. Rescorla, A. R. Wagner, "A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement" in *Classical Conditioning II: Current Research and Theory*, A. H. Black, W. F. Prokasy, Eds. (Appleton Century Crofts, 1972), pp. 64–99.
31. R. S. Sutton, A. G. Barto, *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, MA, 1998).
32. W. Schultz, P. Dayan, P. R. Montague, A neural substrate of prediction and reward. *Science* **275**, 1593–1599 (1997).
33. J. R. Hollerman, W. Schultz, Dopamine neurons report an error in the temporal prediction of reward during learning. *Nat. Neurosci.* **1**, 304–309 (1998).
34. H. Seo, D. Lee, Temporal filtering of reward signals in the dorsal anterior cingulate cortex during a mixed-strategy game. *J. Neurosci.* **27**, 8366–8377 (2007).
35. M. Matsumoto, K. Matsumoto, H. Abe, K. Tanaka, Medial prefrontal cell activity signaling prediction errors of action values. *Nat. Neurosci.* **10**, 647–656 (2007).
36. S. W. Kennerley, T. E. J. Behrens, J. D. Wallis, Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nat. Neurosci.* **14**, 1581–1589 (2011).
37. K. Oyama, I. Hernádi, T. Iijima, K. Tsutsui, Reward prediction error coding in dorsal striatal neurons. *J. Neurosci.* **30**, 11447–11457 (2010).
38. H. Kim, J. H. Sul, N. Huh, D. Lee, M. W. Jung, Role of striatum in updating values of chosen actions. *J. Neurosci.* **29**, 14701–14712 (2009).
39. P. Apicella, Leading tonically active neurons of the striatum from reward detection to context recognition. *Trends Neurosci.* **30**, 299–306 (2007).
40. S. Hong, O. Hikosaka, The globus pallidus sends reward-related signals to the lateral habenula. *Neuron* **60**, 720–729 (2008).
41. M. Matsumoto, O. Hikosaka, Lateral habenula as a source of negative reward signals in dopamine neurons. *Nature* **447**, 1111–1115 (2007).
42. M. A. Belova, J. J. Paton, S. E. Morrison, C. D. Salzman, Expectation modulates neural responses to pleasant and aversive stimuli in primate amygdala. *Neuron* **55**, 970–984 (2007).
43. K. Zhang, C. D. Chen, I. E. Monosov, Novelty, salience, and surprise timing are signaled by neurons in the basal forebrain. *Curr. Biol.* **29**, 134–142.e3 (2019).
44. S. J. Shabel, P. H. Janak, Substantial similarity in amygdala neuronal activity during conditioned appetitive and aversive emotional arousal. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 15031–15036 (2009).
45. D. Amaral, J. Price, A. Pitkanen, S. Carmichael, *The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction*, J. Aggleton, Ed. (Wiley-Liss, 1992), pp. 1–66.
46. M. A. Belova, J. J. Paton, C. D. Salzman, Moment-to-moment tracking of state value in the amygdala. *J. Neurosci.* **28**, 10023–10030 (2008).
47. B. F. Grewe *et al*., Neural ensemble dynamics underlying a long-term associative memory. *Nature* **543**, 670–675 (2017).
48. J. Munuera, M. Rigotti, C. D. Salzman, Shared neural coding for social hierarchy and reward value in primate amygdala. *Nat. Neurosci.* **21**, 415–423 (2018).
49. R. A. Saez, A. Saez, J. J. Paton, B. Lau, C. D. Salzman, Distinct roles for the amygdala and orbitofrontal cortex in representing the relative amount of expected reward. *Neuron* **95**, 70–77.e3 (2017).
50. I. Hernádi, F. Grabenhorst, W. Schultz, Planning activity for internally generated reward goals in monkey amygdala neurons. *Nat. Neurosci.* **18**, 461–469 (2015).
51. F. Grabenhorst, I. Hernádi, W. Schultz, Prediction of economic choice by primate amygdala neurons. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 18950–18955 (2012).
52. F. Grabenhorst, I. Hernadi, W. Schultz, Primate amygdala neurons evaluate the progress of self-defined economic choice sequences. *eLife* **5**, e18731 (2016).
53. J. D. Wallis, Cross-species studies of orbitofrontal cortex and value-based decision-making. *Nat. Neurosci.* **15**, 13–19 (2011).
54. S. E. Morrison, C. D. Salzman, Representations of appetitive and aversive information in the primate orbitofrontal cortex. *Ann. N. Y. Acad. Sci.* **1239**, 59–70 (2011).
55. W. Schultz, Neuronal reward and decision signals: From theories to data. *Physiol. Rev.* **95**, 853–951 (2015).
56. T. Hosokawa, S. W. Kennerley, J. Sloan, J. D. Wallis, Single-neuron mechanisms underlying cost-benefit analysis in frontal cortex. *J. Neurosci.* **33**, 17385–17397 (2013).
57. C. Padoa-Schioppa, J. A. Assad, Neurons in the orbitofrontal cortex encode economic value. *Nature* **441**, 223–226 (2006).
58. C. D. Salzman, S. Fusi, Emotion, cognition, and mental state representation in amygdala and prefrontal cortex. *Annu. Rev. Neurosci.* **33**, 173–202 (2010).
59. L. Tremblay, W. Schultz, Relative reward preference in primate orbitofrontal cortex. *Nature* **398**, 704–708 (1999).
60. S. E. Morrison, C. D. Salzman, The convergence of information about rewarding and aversive stimuli in single neurons. *J. Neurosci.* **29**, 11471–11483 (2009).
61. S. E. Morrison, A. Saez, B. Lau, C. D. Salzman, Different time courses for learning-related changes in amygdala and orbitofrontal cortex. *Neuron* **71**, 1127–1140 (2011).
62. R. C. Wilson, Y. K. Takahashi, G. Schoenbaum, Y. Niv, Orbitofrontal cortex as a cognitive map of task space. *Neuron* **81**, 267–279 (2014).
63. P. H. H. Rudebeck, E. A. A. Murray, The orbitofrontal oracle: Cortical mechanisms for the prediction and evaluation of specific behavioral outcomes. *Neuron* **84**, 1143–1156 (2014).
64. L. K. Fellows, M. J. Farah, Ventromedial frontal cortex mediates affective shifting in humans: Evidence from a reversal learning paradigm. *Brain* **126**, 1830–1837 (2003).
65. M. F. S. Rushworth, M. P. Noonan, E. D. Boorman, M. E. Walton, T. E. Behrens, Frontal cortex and reward-guided learning and decision-making. *Neuron* **70**, 1054–1069 (2011).
66. B. K. H. Chau *et al*., Contrasting roles for orbitofrontal cortex and amygdala in credit assignment and learning in macaques. *Neuron* **87**, 1106–1118 (2015).
67. P. H. Rudebeck *et al*., Frontal cortex subregions play distinct roles in choices between actions and stimuli. *J. Neurosci.* **28**, 13775–13785 (2008).
68. A. Izquierdo, R. K. Suda, E. A. Murray, Bilateral orbital prefrontal cortex lesions in rhesus monkeys disrupt choices guided by both reward value and reward contingency. *J. Neurosci.* **24**, 7540–7548 (2004).
69. P. H. Rudebeck, R. C. Saunders, A. T. Prescott, L. S. Chau, E. A. Murray, Prefrontal mechanisms of behavioral flexibility, emotion regulation and value updating. *Nat. Neurosci.* **16**, 1140–1145 (2013).
70. G. Schoenbaum, B. Setlow, M. P. Saddoris, M. Gallagher, Encoding predicted outcome and acquired value in orbitofrontal cortex during cue sampling depends upon input from basolateral amygdala. *Neuron* **39**, 855–867 (2003).
71. S. Bernardi, D. Salzman, "Appetitive and aversive systems in the amygdala" in *Decision Neuroscience: An Integrative Perspective*, J.-C. Dreher, L. Tremblay, Eds. (Academic Press, 2016), pp. 33–46.
72. S. McKenzie *et al*., Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas. *Neuron* **83**, 202–215 (2014).
73. J. D. Wallis, K. C. Anderson, E. K. Miller, Single neurons in prefrontal cortex encode abstract rules. *Nature* **411**, 953–956 (2001).
74. M. J. Buckley *et al*., Dissociable components of rule-guided behavior depend on distinct medial and prefrontal regions. *Science* **325**, 52–58 (2009).
75. M. G. Stokes *et al*., Dynamic coding for cognitive control in prefrontal cortex. *Neuron* **78**, 364–375 (2013).
76. A. Saez, M. Rigotti, S. Ostojic, S. Fusi, C. D. Salzman, Abstract context representations in primate amygdala and prefrontal cortex. *Neuron* **87**, 869–881 (2015).
77. R. Bellman, *Dynamic Programming* (Princeton University Press, 1957).
78. S. Bernardi *et al*., The geometry of abstraction in hippocampus and pre-frontal cortex. https://www.biorxiv.org/content/10.1101/408633v3 (4 October 2019).
79. D. Kumaran, J. J. Summerfield, D. Hassabis, E. A. Maguire, Tracking the emergence of conceptual knowledge during human decision making. *Neuron* **63**, 889–901 (2009).
80. A. C. Schapiro, T. T. Rogers, N. I. Cordova, N. B. Turk-Browne, M. M. Botvinick, Neural representations of events arise from temporal community structure. *Nat. Neurosci.* **16**, 486–492 (2013).
81. A. Wutz, R. Loonis, J. E. Roy, J. A. Donoghue, E. K. Miller, Different levels of category abstraction by different dynamics in different prefrontal areas. *Neuron* **97**, 716–726.e8 (2018).
82. M. Ponsen, M. E. Taylor, K. Tuyls, "Abstraction and generalization in reinforcement learning: A summary and framework" in *Adaptive and Learning Agents*, (Springer, Berlin, Heidelberg, 2010), pp. 1–32.
83. N. D. Volkow, G.-J. Wang, J. S. Fowler, D. Tomasi, F. Telang, Addiction: Beyond dopamine reward circuitry. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 15037–15042 (2011).
84. Y. Schaffer, O. Barak, Y. Rassovsky, Social perception in borderline personality disorder: The role of context. *J. Pers. Disord.* **29**, 275–288 (2013).
85. A. W. MacDonald, , 3rd, M. F. Pogue-Geile, M. K. Johnson, C. S. Carter, A specific deficit in context processing in the unaffected siblings of patients with schizophrenia. *Arch. Gen. Psychiatry* **60**, 57–65 (2003).
86. M. R. Milad *et al*., Neurobiological basis of failure to recall extinction memory in posttraumatic stress disorder. *Biol. Psychiatry* **66**, 1075–1082 (2009).
87. S. Ramirez *et al*., Activating positive memory engrams suppresses depression-like behaviour. *Nature* **522**, 335–339 (2015).