A higher-order theory of emotional consciousness

Joseph E. LeDoux[a,b,1] and Richard Brown[c]

[a]Center for Neural Science, New York University, New York, NY 10003; [b]Emotional Brain Institute, Nathan Kline Institute, Orangeburg, NY 10962; and [c]Philosophy Program, LaGuardia Community College, The City University of New York, Long Island City, NY 10017

Contributed by Joseph E. LeDoux, January 14, 2017 (sent for review November 22, 2016; reviewed by Hakwan Lau and Dean Mobbs)

Emotional states of consciousness, or what are typically called emotional feelings, are traditionally viewed as being innately programmed in subcortical areas of the brain, and are often treated as different from cognitive states of consciousness, such as those related to the perception of external stimuli. We argue that conscious experiences, regardless of their content, arise from one system in the brain. In this view, what differs in emotional and nonemotional states are the kinds of inputs that are processed by a general cortical network of cognition, a network essential for conscious experiences. Although subcortical circuits are not directly responsible for conscious feelings, they provide nonconscious inputs that coalesce with other kinds of neural signals in the cognitive assembly of conscious emotional experiences. In building the case for this proposal, we defend a modified version of what is known as the higher-order theory of consciousness.

fear | amygdala | working memory | introspection | self

Much progress has been made in conceptualizing consciousness in recent years. This work has focused on the question of how we come to be aware of our sensory world, and has suggested that perceptual consciousness emerges via cognitive processing in cortical circuits that assemble conscious experiences in real-time. Emotional states of consciousness, on the other hand, have traditionally been viewed as involving innately programmed experiences that arise from subcortical circuits.

Our thesis is that the brain mechanisms that give rise to conscious emotional feelings are not fundamentally different from those that give rise to perceptual conscious experiences. Both, we propose, involve higher-order representations (HORs) of lower-order information by cortically based general networks of cognition (GNC). Thus, subcortical circuits are not responsible for feelings, but instead provide lower-order, nonconscious inputs that coalesce with other kinds of neural signals in the cognitive assembly of conscious emotional experiences by cortical circuits (the distinction between cortical and subcortical circuits is defined in *SI Appendix*, Box 1). Our theory goes beyond traditional higher-order theory (HOT), arguing that self-centered higher-order states are essential for emotional experiences.

Emotion, Consciousness, and the Brain

Detailed understanding of the emotional brain, and theorizing about it, is largely based on studies that fall under the heading of "fear." We will therefore focus on this body of work in discussing emotional consciousness. In light of this approach, we define "fear" as the conscious feeling one has when in danger. Although our conclusions may not apply equally well to all emotions, we maintain that the lessons from fear provide general principles that can at least be used as a starting point for theorizing about many emotions.

The Amygdala Fear Circuit View. Emotions like fear are often said to have been inherited from animal ancestors (1–6). These "basic emotions" are typically proposed to be wired into the brain's limbic system (7). Although the limbic system theory has been criticized extensively (6, 8–12), it still guides much research and theory in neuroscience. Fear, for example, is often said to be dependent upon a set of circuits that have as their hub the limbic area called the amygdala (5, 8, 13–19).

A great deal of research has shown that damage to the amygdala disrupts the ability of animals and people to respond to threats behaviorally and physiologically (13–15, 17, 19, 20). Furthermore, functional imaging studies in humans show that the amygdala is activated in the presence of threats (21–30). This work is often interpreted to mean that threats induce a state of fear by activating a fear circuit centered on the amygdala, and this same circuit controls the behavioral and physiological responses elicited by the threat; these responses are often called fear responses (5, 6, 8, 13, 17, 31) (Fig. 1*A*).

What is meant by "fear" varies among those who use it to account for behavioral and physiological responses to threats. For some, fear is a subjective state, a phenomenal experience elicited by danger. Darwin (1), for example, called emotions like fear states of mind inherited from animals. Mowrer (32) argued that rats freeze "by cause" of fear. Panksepp (5) noted that "fear is an aversive state of mind," the major driving force of which is a "subcortical FEAR system" (33). Others, like Perusini and Fanselow (31), agree that danger elicits fear and fear causes behavior, but they do not treat it as a subjective experience; for them, fear is a brain state that intervenes between threats and defensive behaviors. It is, in fact, common in behavioral neuroscience to construe animal and human behavior as being caused by so-called central states rather than by subjective experiences, while at the same time retaining the subjective state term (13, 16–18, 34). This approach allows animal research to be relevant to the human experience of fear, but leads to much confusion about what researchers mean when they use the term "fear" (20, 35).

Although we agree with those who argue that it is inappropriate to call upon subjective experiences to explain animal behavior, we maintain that subjective experience is a seminal part of human existence and is indispensable in the effort to understand human nature. A way is needed to conceive of behavioral responses to threats in animals and humans with similar constructs, but without attributing unmeasurable subjective states to animals, and without denying the role of subjective experience in humans. The notion of

Significance

Although emotions, or feelings, are the most significant events in our lives, there has been relatively little contact between theories of emotion and emerging theories of consciousness in cognitive science. In this paper we challenge the conventional view, which argues that emotions are innately programmed in subcortical circuits, and propose instead that emotions are higher-order states instantiated in cortical circuits. What differs in emotional and nonemotional experiences, we argue, is not that one originates subcortically and the other cortically, but instead the kinds of inputs processed by the cortical network. We offer modifications of higher-order theory, a leading theory of consciousness, to allow higher-order theory to account for self-awareness, and then extend this model to account for conscious emotional experiences.

Author contributions: J.E.L. and R.B. wrote the paper.

Reviewers: H.L., University of California, Los Angeles; and D.M., California Institute of Technology.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. Email: ledoux@cns.nyu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1619316114/-/DCSupplemental.

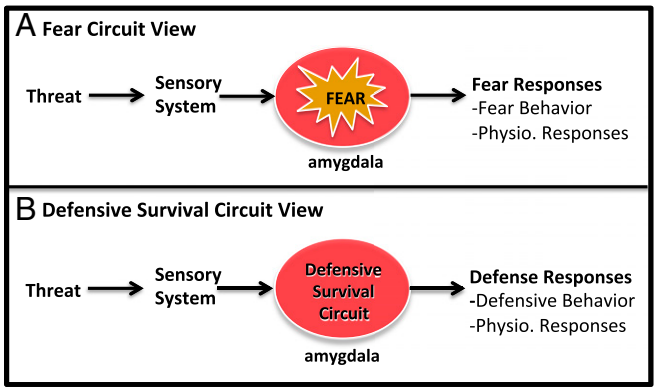

Fig. 1. Two views of amygdala contributions to threat processing. In the fear circuit view (*A*) the amygdala is responsible for both the subjective experience of fear and the control of so-called "fear responses." In the defensive survival circuit view (*B*) the amygdala controls defensive responses but is not responsible for subjective experiences elicited by threats.

evolutionarily conserved defensive survival circuits that account for the behavioral and physiological responses to threats, but that are not directly responsible for subjective experiences of fear, accomplishes this goal (20, 24, 35–37).

The Defensive Survival Circuit View. Defensive survival circuits are evolutionarily wired to detect and respond to innate threats and to respond to novel threats that have been learned about in the past (35, 36). As viewed here, defensive survival circuits indirectly contribute to the feeling of fear, but their activity does not constitute fear. The amygdala-centered circuit described above is an example of such a defensive survival circuit. Fig. 1*B* illustrates the defensive survival circuit view, relative to the fear circuit view, of amygdala contributions to threat processing.

That the defensive survival circuit is separate from the circuit that gives rise to the conscious experience of fear is suggested by several lines of evidence. First, it is well established that conscious feelings of fear and anxiety are poorly correlated with behavioral and physiological responses, such as those controlled by defensive survival circuits (38). If the same circuit was involved the correlation should be strong. Second, studies using subliminal stimulus presentation methods (e.g., backward masking or continuous flash suppression) to prevent or reduce awareness of visual stimuli show that visual threats activate the amygdala and elicit body responses despite the fact that participants deny seeing the stimulus (21–30). Under such conditions, participants do not report feeling fear, even when explicitly instructed to be introspective about what they are experiencing (39). Third, "blindsight" patients, who lack the ability to consciously see visual stimuli in a particular area of visual space, exhibit amygdala activation and physiological responses to visual threats presented in that part of space despite denying seeing the stimulus and without reporting fear (40–42) (see *SI Appendix*, Box 2 for a discussion of blindsight and other neurological patients that have contributed to consciousness research). Fourth, although damage to the amygdala interferes with bodily responses to threats, it does not interfere with conscious experience of emotions such as fear (43, 44). These findings all suggest that the amygdala-defensive survival circuits processes threats nonconsciously (8, 9, 20). This does not mean that defensive survival circuits play no role in conscious fear: they modulate the experience of fear, but are not directly responsible for the conscious experience itself.

How, then, does the conscious experiences of fear come about, if not as the product of an innate subcortical circuit? We argue below that fear results from the cognitive interpretation that you are in a dangerous situation, one in which physical or psychological harm may come to you. As such, an emotional experience like fear comes about much the same as any other conscious experience: as a result of processing by the GNC. However, these circuits process different inputs in emotional vs. nonemotional conscious experiences, and in different kinds of emotional experiences.

Consciousness in Contemporary Philosophy, Cognitive Science, and Neuroscience

In recent years empirical findings in cognitive science and neuroscience have helped reshape views of what consciousness is and how it comes about. In discussing this research we will emphasize consciousness as a subjective experience, as opposed to the condition of an organism simply being awake and responsive to sensory stimulation (45, 46).

Measuring Conscious Experiences. Essential to researching consciousness as subjective experience is some means of measuring internal states that cannot be observed by the scientist. The most common method is the use of verbal self-report (20, 47, 48). This allows researchers to distinguish conditions under which one is able to state when they experience a sensory event from when they do not. Verbal self-report depends on introspection, the ability to examine the content of one's mental states (49, 50). Introspection, in turn, is believed to involve such cognitive processes as attention, working memory, and metacognition (51–53), processes that are called upon in cognitive theories of consciousness (47, 54–57).

Nonverbal behavior is satisfactory for demonstrating that a human or other animal is conscious in the sense of being awake and responsive to stimuli, and for demonstrating cognitive capacities underlying working memory, attention, metacognition, problem-solving ability, and other indicators of intelligent behavior (52). However, because not all cognitive processing leads to conscious experience (52, 58–62), cognitive capacities indicated by nonverbal behavior alone are generally not sufficient to demonstrate conscious awareness (for further discussion of the measurement of consciousness through verbal and nonverbeal means, see *SI Appendix*, Box 3, and for discussion of nonconscious cognition, especially nonconscious working memory, see *SI Appendix*, Box 4).

Neural Correlates of Reportable Conscious Experiences. Evidence has mounted in recent years implicating specific brain circuits in introspectively reportable conscious experiences of visual stimuli in humans. For example, when self-reports of the stimulus are compromised by using subliminal stimulation procedures, such as masking (63), areas of the visual cortex (including primary and secondary areas) are functionally active, but when participants are able to consciously report seeing a visual stimulus, additional cortical areas become active (47, 54, 64–71). Most consistently implicated are various areas of the lateral and medial prefrontal cortex, but activations are also reported in the parietal cortex and insular cortex (Fig. 2). These cortical areas are, not surprisingly, components of the GNC (72–75). Related findings come from blindsight patients (76) who, because of damage to the visual cortex, are unable to report on the presence of visual stimuli in the part of visual space processed by the damaged area of the cortex, despite being able to respond nonverbally to the stimulus. When the stimulus is in the part of space they can see and report on, cortical areas of the GNC are activated, but when it is in the blind area these areas are not activated (65, 77). Although the imaging studies in healthy people and blindsight patients suggest correlates of consciousness, other studies show that disruption of activity in GNC areas, especially in the prefrontal or parietal cortex, impairs conscious awareness of visual stimuli (78–80).

Phenomenal Consciousness. A key idea that pervades discussions of consciousness is phenomenal experience. In the most general sense, so-called phenomenal consciousness is just the property of there being something that it is like for one, from one's point of view, to be in a particular state (81). When I consciously experience pain, or see red, there is something that it is like, and that something can only be known through experience. The specific phenomenal properties of an experience, sometimes called "qualia," are said to

be the specific aspects or contents of the conscious stream of experience (82). Explaining consciousness in the phenomenal sense is the (in)famous "hard problem" of consciousness (83). To address the hard problem is to provide an account of how it is that phenomenal consciousness emerges from brain activity: to explain why all of the information processing in the brain is not going on "in the dark," in the so-called cognitive unconscious (58).

First-Order vs. Higher-Order Theories of Consciousness. A key issue is whether introspection captures the nature of our phenomenally conscious experiences. Ned Block (84) has argued that introspection only reveals those aspects of consciousness to which we have cognitive access, what he calls "access consciousness." Phenomenal consciousness, according to Block, is a more fundamental level of experience that exists separately from and independent of cognitive access. The difference between phenomenal and access consciousness can be illustrated by considering first-order vs. higher-order theories of consciousness.

First-order theorists, such as Block, argue that processing related to a stimulus is all that is needed for there to be phenomenal consciousness of that stimulus (85–89). Conscious states, on these kinds of views, are states that make us aware of the external environment. Additional processes, such as attention, working memory, and metacognition, simply allow cognitive access to and introspection about the first-order state. In the case of visual stimuli, the first-order representation underlying phenomenal consciousness is usually said to involve the visual cortex, especially the secondary rather than primary visual cortex. (Fig. 3*A*). Cortical circuits, especially involving the prefrontal and parietal cortex, simply make possible cognitive (introspective) access to the phenomenal experience occurring in the visual cortex.

In contrast, David Rosenthal and other higher-order theorists argue that a first-order state resulting from stimulus-processing alone is not enough to make possible the conscious experience of a stimulus (90–93). In addition to having a representation of the external stimulus one also must be aware of this stimulus representation. This is made possible by a HOR, which makes the first-order state conscious (Fig. 3*B*). In other words, consciousness exists by virtue of the relation between the first- and higher-order states. Cognitive processes, such as attention, working memory, and metacognition are key to the conscious experience of the first-order state. In neural terms, the areas of the GNC, such as the prefrontal and parietal cortex, make conscious the sensory information represented in the secondary visual cortex. Varieties of HOT are described in *SI Appendix*, Box 7.

HOT has the advantage of appealing to a set of well-established cognitive operations underlying introspective access. In contrast, first-order theory is plagued by its appeal to a noncognitive kind of access that leads to experiences that can go undetected to the "conscious" person, and that are difficult to verify or falsify scientifically (20, 94–96). With this view, you can have phenomenal conscious experiences that you do not access (that you do not "know" exist) (88). Because it is truly hard to imagine what it might be like to have a conscious experience of a stimulus of which you are not aware, this key feature of first-order theory is contested (98–100) (debate on this topic is discussed further in *SI Appendix*, Box 5). Additionally, first-order theory is challenged by empirical findings (91, 101) and computer modeling (102). Block agrees that there must be some kind of introspective access (what he has called "awareness-access") when there is phenomenal consciousness (88, 89), but he also insists that this kind of access is not cognitive; what exactly noncognitive introspection might be is unclear (for further discussion of this point, see *SI Appendix*, Box 6).

Most cognitive theories call upon similar cognitive processes in accounting for conscious experience, but do so in somewhat different ways. Included are theories that emphasize attention and working memory (47, 55, 57, 59, 103–109), processing by a global workspace (56, 110, 111), or the interpretation of experience (112). A common thread that runs through various cognitive theories is that processing beyond the sensory cortex is required for conscious experience. In this sense, these other cognitive theories, although not explicitly recognized as HOTs, have a close affinity to the basic premise of HOT (for more on the relation between HOTs and other cognitive theories, see *SI Appendix*, Box 7).

Although the debate continues over first-order theory and HOT, our conclusion is that introspectively accessed states, which we view as phenomenally conscious experiences, are best described in terms of HORs that depend on the GNC, cortical circuits involving regions of the lateral and medial prefrontal cortex, posterior parietal cortex, and insular cortex (Fig. 2). These GNC have been implicated in attention, working memory, and metacognition (72–75), which—as discussed—are viewed as essential to consciousness in most cognitive theories. We are not suggesting that the brain areas included in the GNC are locations where consciousness literally occurs in a homuncular sense. Consciousness involves complex interactions between circuits in the GNC and sensory cortex, as well as other areas (especially those involved in memory). Furthermore, we do not mean to imply that these circuits function in a uniform way in conscious

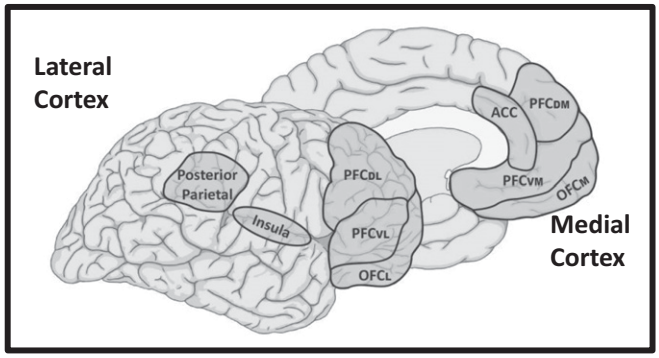

Fig. 2. GNC that contribute to conscious experiences. Functional imaging studies have implicated circuits spread across frontal and parietal areas in conscious experiences in humans. ACC, anterior cingulate cortex; OFCL, lateral orbital frontal cortex; OFCM, medial orbital frontal cortex; PFCDL, dorsolateral prefrontal cortex; PFCDM, dorsomedial prefrontal cortex; PFCVL, ventrolateral prefrontal cortex; PFCVM, ventromedial prefrontal cortex.

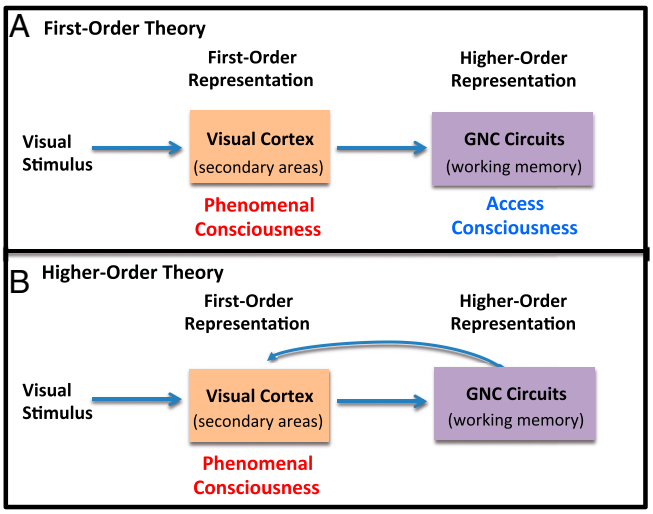

Fig. 3. First-order vs. higher-order theories of consciousness. Consciousness depends solely on sensory representations of stimuli in first-order theory (*A*) but depend on the representation of the lower-order information by circuits that underlie cognitive functions, such as working memory in HOT (*B*).

experiences; different subcircuits may well contribute to different aspects of consciousness (91, 113).

Emotions in Light of the First- vs. Higher-Order Debate

Panksepp's (5, 114) emotion theory described above can be reconceived as a first-order emotion theory. In his view, core phenomenal states of emotional consciousness, such as feelings of fear, are innate experiences that arise in humans and other mammals from evolutionarily conserved subcortical circuits. These states are described as "implicit procedural (perhaps truly unconscious), sensory-perceptual and affective states" (115, 116). Although they "lack reflective awareness," they nevertheless "give us a specific feeling." Then, through cognitive processing by cortical circuits (presumably the GNC), the states can be accessed and introspectively experienced. There is a striking similarity between Panksepp's theory and Block's distinction between phenomenal and access consciousness. Like Block, Panksepp calls upon conscious states that the organism is unaware of but cannot introspect or talk about (they "give us a certain feeling" but "lack reflective awareness" and "are perhaps truly unconscious").

Similar to Panksepp, Antonio Damasio builds on the idea that subcortical circuits (what he calls "emotional action" systems) control innate behaviors and related physiological responses. Unlike Panksepp, though, Damasio assumes that these subcortical action systems operate nonconsciously. For him, basic/core phenomenal feelings result when feedback from the body responses is represented in "body sensing" areas of the brain to create emotion-specific "body states" or "somatic markers" (6, 117). Initially, Damasio emphasized the importance of body sensing areas of the cortex in giving rise to feelings, but more recently he has revised his view, arguing that core feelings are products of subcortical circuits that receive primary sensory signals from the body (118). The subcortical sensory representations are genuine first-order sensory states that, in the theory, account for phenomenal experiences of basic emotions. Then, through cognitive processing in cortical areas, including the insular, somatosensory, and various prefrontal regions, introspectively accessible experiences emerge. Findings from patients with autonomic failure and alexithymia (119), although superficially supportive, primarily show quantitative changes in subjective experience, more consistent with a modulatory role of body feedback.

The subcortical representations proposed by Pankesepp and Damasio clearly occur in humans and other mammals. However, the evidence that these actually give rise to phenomenally experienced feelings that exist independent of introspective access is, to us, not convincing (20). Given that the arguments both theorists make about animal consciousness rest on the similarity of the circuitry in animals and humans, the weakness of the evidence about these subcortical circuits supporting phenomenal consciousness in humans, and the inability to directly measure consciousness in animals, questions the value of the animal states as a way of understanding the brain mechanisms of human emotional consciousness.

In our opinion, the subcortical circuits proposed by Panksepp and Damasio are better interpreted as contributing nonconscious first-order representations that indirectly influence the higher-order assembly of conscious feelings by the GNC. We develop a HOT of emotion below.

A Modified Higher-Order Theory

Traditional HOT needs to be modified before using it as the foundation for a theory of emotional consciousness. Specifically needed are changes to its treatment of introspection, of higher-order states that lack an external source, and of the self.

Introspection and the Higher-Order Account of Phenomenal Experience. As noted above, in general, first-order and higher-order theories seek to explain the same thing: the phenomenology of experience, the essence of what it is like to have an experience. However, Rosenthal, the leading higher-order theorist, does not accept Block's distinction between phenomenal and access consciousness; for him there is just consciousness. The reason Rosenthal avoids phenomenal consciousness hinges on his construal of it as tacitly committing one to a first-order view (*SI Appendix*, Box 8).

According to HOT, one is not typically conscious of the higher-order state itself, but instead is conscious of the first-order state by virtue of the HOR of it. To be aware of the higher-order state (to be conscious that you are in that state) requires yet another HOR. Higher-order theorists typically reserve the term "introspection" for this additional level of representations (90). This amounts to the claim that introspection consists in a conscious higher-order state (i.e., a third-order state, or a HOR of the initial HOR). For example, Rosenthal uses introspection to refer to situations when one is attentively and deliberately focused on one's conscious experiences. He argues that this additional state (the HOR of the HOR) is considerably less common than simply noticing one's experiences, and thus that introspection is not a key part of normal, everyday consciousness (90).

We propose a more inclusive view of introspection, in which the term indicates the process by which phenomenally experienced states result. Following Armstrong (120), we argue that introspection can involve either passive noticing (as, for example, in the case of consciously seeing a ripe strawberry on the counter) or active scrutinizing (as in the case of deliberate focused attention to our conscious experience of the ripe strawberry). Both kinds of introspection, in our view, lead to phenomenal experience. Thus, as we use the term "introspection," the HOR that is responsible for consciousness on the traditional HOT would count as passive introspectively noticing one's first-order states, and the second HOR would count as introspectively scrutinizing the first HOR. This notion of a HOR of a HOR is a part of our modified HOT, described next.

Accounting for HORs of Absent Stimuli: HOROR Theory. A key criticism of HOT is that it relies on the existence of a relation between the higher-order state and the state which it represents (121). That is, the first-order state is said to be transformed into a phenomenally conscious state by virtue of it being represented by the higher-order state (92, 122, 123). This is depicted in Fig. 3*B* by the arrow between the higher-order and lower-order state; in other words, the higher-order state makes conscious the lower-order state. This idea roughly preserves the perception-like nature of higher-order awareness. However, it is possible to have the experience as of seeing something without it being there to be seen, as in the case of hallucinations or in dreams. In a somewhat similar way it might be possible to have higher-order awareness in the absence of a first-order representation. In fact, there seems to be empirical confirmation of this (see discussion of Charles Bonnet Syndrome below and in *SI Appendix*, Box 9). To account for this limitation a version that does not depend on a relation between the higher-order state and a sensory representation has been proposed (96, 124). This revised HOT involves a HOR of a representation, and is thus called HOROR.

HOROR theory argues that phenomenal consciousness does not reflect a sensory state (as proposed by first-order theory) or the relation between a sensory state and a higher-order cognitive state of working memory (as proposed by traditional HOT) (96, 100, 124). Instead, HOROR posits that phenomenal consciousness consists of having the appropriate HOR of lower-order information, where lower-order does not necessarily mean sensory, but instead refers to a prior higher-order state that is rerepresented (Fig. 4*A*). This second HOR is thought-like and, in virtue of this, instantiates the phenomenal, introspectively accessed experience of the external sensory stimulus. That is, to have a phenomenal experience is to be introspectively aware of a nonconscious HOR. This introspective awareness, in ordinary circumstances, will be the passive noticing discussed earlier. This passive kind of noticing, which we postulate is responsible for the existence of phenomenal consciousness, differs from the active scrutinizing of one's conscious experience that requires deliberate attentive focus on one's phenomenal consciousness. Active introspection

requires an additional layer of HOR (and thus a HOR of a HOROR).

HOROR theory is supported by empirical cases where phenomenal consciousness plausibly outstrips first-order representations. One example, as mentioned, occurs in Charles Bonnet Syndrome, and another involves empirical findings of inattentional inflation in healthy humans (91, 96). In these cases, what it is like for the subjects is better tracked by the higher-order states. Specifically, these cases suggest that there is more that is phenomenally experienced than can be accounted for in terms of first-order representations. For example, in the extreme case of the rare form of Charles Bonnet Syndrome subjects have severe damage to their visual cortex (and thus presumably lack first-order representations, or at least have sparse first-order representations) and yet have rich and vivid visual experiences in hallucinations.

If HOROR is correct, the nonconscious events that contribute to consciousness are not sensory cortex events but instead are HORs of sensory events (the first HOR). HOROR thus depends on the existence of nonconscious HORs. Our hypothesis is that the nonconscious HOR, and the HOROR that instantiates the awareness of ourselves as being in the HOR, both occur within working memory. Although some have questioned the capacity of nonconscious working memory, and thus its ability to support consciousness (88, 125, 126), the plausibility of our hypothesis is supported by results from recent studies indicating that nonconscious processing within working memory is more robust than previously observed (79, 101, 108, 127) (for additional discussion of nonconscious working memory, see *SI Appendix*, Box 4).

Given that the conscious experience of a stimulus presumably depends on object memories, which are stored in the medial temporal lobe memory systems (128–130), the lower-order representations that underlie consciousness could involve memory rather than—or in addition to—sensory representations. That is, so-called explicit memories, until retrieved into working memory, are in a nonconscious state (20). The relation of memory to consciousness deserves more attention than it currently receives.

Refining the Role of the Self in HOT. In traditional HOT (90, 122, 131, 132), the content of the HOR is talked about as "I thoughts" (133) and is postulated to have something like the content "I am in state *x*." For example, if one were consciously seeing red then the higher-order state would have the content "I am seeing red." It is important to note that in this view the self is implied to be represented by the reference to "I" in the propositional statement. However, this representation is not of the conscious self (113) and can be construed as being implicit or nonconscious (134). The higher-order state thus represents the first-order state as belonging to oneself, but in an extremely thin sense that does not invoke the conscious self or self-consciousness. Thus, a distinction between self and other that includes body sensations or mental states, but that does not explicitly represent the conscious self, is not a state of self-awareness. Because of this, and to avoid confusion, we phrase the thesis above as simply "being aware of the state" rather than "I am aware of myself as being in the state."

When a higher-order state includes information about oneself, it becomes possible for there being something that it is like for "you" to be in that state. This is what we call "self-HOROR," a HOROR that includes information about the self. Tulving refers to experiences that include the self as "autonoetic consciousness," and experiences that do not as "noetic consciousness" (135–137). The presence of the self in the nonconscious representation allows for an autonoetic conscious (a self-HOROR) state to result from the re-representation (Fig. 4*B*).

Tulving proposed that noetic and autonoetic consciousness are associated with two classes of explicit or conscious memory, called semantic and episodic memory (135–137). Semantic memories consist of factual knowledge and are experienced as noetic states of consciousness; episodic memories are about facts anchored in space and time that involve the self, and are experienced as autonoetic states of consciousness (138). The involvement of the self makes episodic memories personal, that is, autobiographical. With autonoetic consciousness, one can engage in mental time travel to remember the past and imagine the future self.

Other animals have factual elements of episodic memory (the ability to form memories about what, where, and when some event occurred) (139, 140). Whether they can engage in self-referential conscious thinking (141), and thus have self-awareness, is less certain (20, 142, 143). Self-awareness is viewed by many as a uniquely human experience (112, 135, 136, 144–147).

Klein (148) draws an important distinction between episodic memory (memory of what happened, where it happened, and when it happened) and autonoetic consciousness (the awareness that the facts about what, where, and when happened to you). If so, episodic memory can be thought of as simply a complex form of semantic memory rather than as memory about the self. Only when the self is explicitly integrated into the episode does autonoetic awareness occur. In this sense, animals may have episodic (what, where, when) memory, and maybe even noetic awareness, but may nevertheless lack autonoetic awareness.

Self-HOROR depends on self-knowledge, which includes both explicit (consciously accessible) and implicit (not consciously accessible) memory (9). Our focus here is on information about the self that has the potential for conscious access, or at least for influencing conscious experience, and thus that falls in the domain of explicit memory. Explicit memory is acquired and retrieved via the medial temporal lobe memory system (128–130). Such memories are in effect nonconscious until they are retrieved into working memory and rendered conscious (20).

We view the self as a set of autobiographical memories about who you are and what has happened to you in your life, and how you think, act, and feel in particular situations (149, 150). Such bodies of information are called schema (151–153). As part of self-HOROR theory, we thus propose that autobiographical self-schema (154, 155) contribute to conscious states in which the self is involved.

The cognitive functions (working memory, attention, metacognition, and so forth) and neural circuits (especially prefrontal circuits) that underlie HORs are also required for conscious experiences involving the self, including emotional states involving the self (145, 156–162). These functions integrate nonconscious factual (what) and episodic context (where and when) information with nonconscious representations of autobiographical

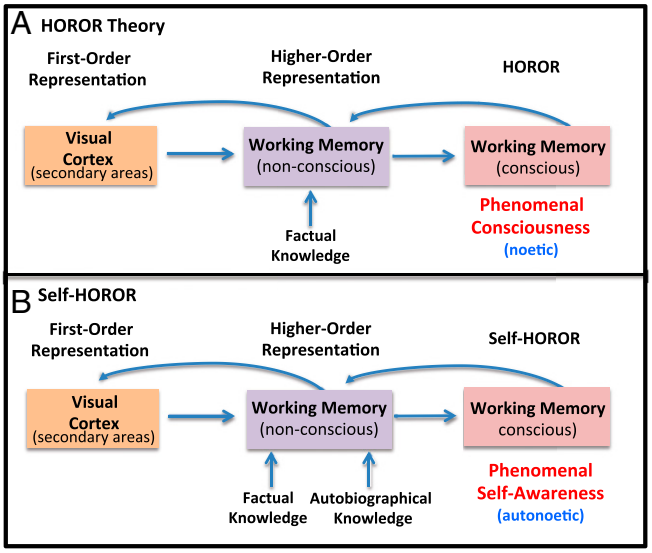

Fig. 4. HOROR theory. (*A*) HOROR is a variant on HOT (see Fig. 3 for HOT). In it, the first HOR is nonconscious and must be rerepresented by another higher-order state. This creates a noetic (factual) state of consciousness. (*B*) However, if the self is part of the nonconscious representation then the HOROR results in an autonoetic (self-aware) state of consciousness.

information into a nonconscious HOR that is rerepresented as an autonoetic HOROR (a self-HOROR) that supports a conscious experience of the event as happening to you.

We propose that these various representations involve interactions between the medial temporal lobe memory system and areas of the GNC in creating the nonconscious representations that are the basis of the self-HOROR within the GNC. Earlier we noted that circuits within the GNC should not be thought of as making unified contributions to consciousness, and indeed evidence exists for a distinction within prefrontal cortex circuits for self-referential (autonoetic) conscious experiences as opposed to experiences of external stimuli (91, 113).

Self-representations have been described as integrated hubs for information processing, a kind of "associative glue" (156). This notion becomes especially important in emotional states that, we argue, crucially depend on representation of the self as part of the higher-order state that constitutes the felt experience.

A Higher-Order Theory of Emotional Consciousness

The modifications of HOT just described allow us to view the phenomenal states we call emotions as HORs. Particularly important to our HOT of emotional consciousness (HOTEC) is the notion that emotions depend on the self. Without the self there is no fear or love or joy. If some event is not affecting you, then it is not producing an emotion. When your friend or child suffers you feel it because they are part of you. When the suffering of people you don't know affects you emotionally, it is because you empathize with them (put yourself in their place, feel their pain): no you, no emotion. The self is, as noted above, the glue that ties such multidimensional integrated representations together (156).

First-order theories of emotion require two circuits of consciousness: a subcortical circuit for first-order phenomenal conscious feeling (one that is not available to introspection and yet is supposedly consciously experienced) and a cortical circuit for higher-order, introspectively experienced conscious feelings. In HOTEC only one circuit of consciousness is required: the GNC. That GNC circuits contribute to conscious emotions, and to representations of the self in emotions, is indicated by studies showing activation of prefrontal areas (163) and differences in prefrontal morphology (164) in relation to fearful or other emotional experiences, as well as studies showing activation of prefrontal areas in self-judgments about emotions (161) and deficits in self-consciousness in emotions in people with damage to fronto-temporal areas (162).

An important prediction of our theory is that damage to first-order subcortical circuits, such as a defensive survival circuit or body-sensing circuits, should mute but not eliminate feelings of fear. Evidence consistent with exists (43, 44). Relevant research questions and predictions are summarized in *SI Appendix*, Box 10.

If the circuits that give rise to cognitive states of awareness (the GNC) also give rise to emotional self-awareness, how can we distinguish emotional from nonemotional cognitive states of awareness? One of us (J.E.L.) has long argued that emotional states of consciousness depend on the same fundamental neural mechanisms as any other state of consciousness, but that the inputs processed are different (8, 9, 20, 165, 166). Below, we incorporate this notion into a HOT of emotion, using the feeling of fear resulting from an encounter with a visual threat as an example.

A visual threat, say a snake at your feet, is processed in two sets of circuits in parallel in the process of giving rise to a feeling of fear. Cortical circuits involving the visual cortex, memory systems of the medial temporal lobe, and the cognitive systems related to the GNC are engaged in the process of representing the stimulus and ultimately experiencing it (Fig. 5*A*). At the same time, subcortical defensive survival circuits centered on the amygdala control innate behavioral and physiological responses that help the organism adapt to the situation (Fig. 5*B*). Both sets of circuits are far more complex that the simplified versions shown for illustrative purposes.

Stimulus-processing streams beginning in the retina and continuing through the various stages of the visual cortex create a representation of the snake. Then, through connections from the visual cortex to the medial temporal lobe and GNC this representation is integrated with long-term semantic memories, allowing the factual information about snakes and their potential for causing harm to be rerepresented nonconsciously in working memory (a HOR). This nonconscious information is then the basis for a further rerepresentation (a HOROR) that allows a noetic conscious experience of the facts of the situation, an awareness that a potentially dangerous situation is unfolding. Retrieval of self-schema into the representation allows the representation to constitute an autonoeitc experience in which you are a part (a self-HOROR).

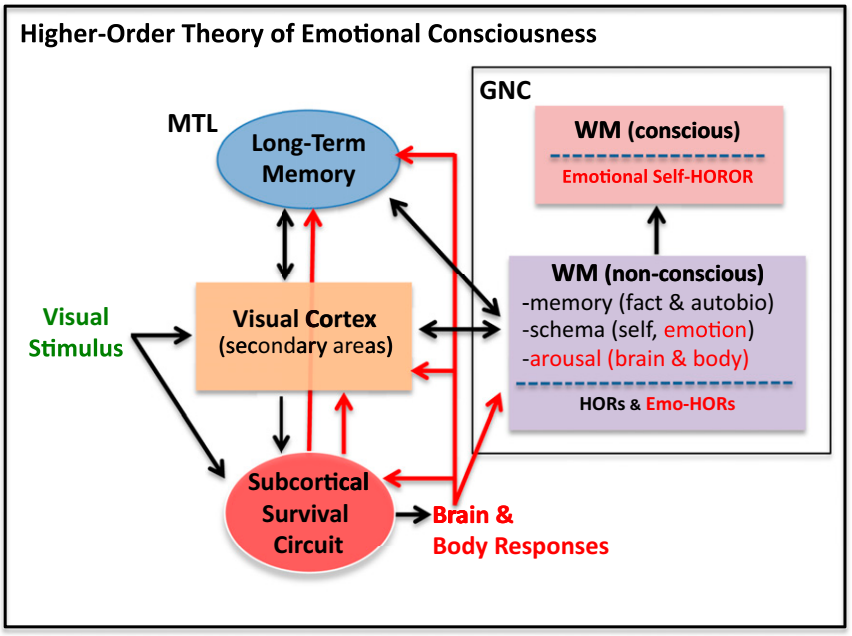

Fig. 5. HOTEC extends HOT and HORROR theory to account for the self. In HOTEC, the difference between an emotional and nonemotional state of consciousness is accounted for by the kinds of inputs processed by the GNC. Red lines show circuit interactions that are especially important in emotional states. Red text indicates states/events that occur during emotional but not nonemotional experiences. See main text for additional details. Emo-HOR, emotional higher-order representation; MTL, medial temporal lobe; WM, working memory.

However, although self-HOROR is necessary it is not sufficient for an emotional experience. Self-HOROR can be purely cognitive (your personal knowledge that there is a pencil present and that it is you that is looking at this pencil) or emotional (there is a snake present and you are afraid it may harm you). What accounts for this difference? This takes us to the second set of circuits that contributes to fear: defensive survival circuits.

Activation of defensive circuits during an emotional experience results in various consequences that provide additional kinds of inputs to working memory, thus altering processing in the in the GNC in ways that do not occur in nonemotional situations (20) (Fig. 5*B*). For one thing, outputs of the amygdala activate arousal systems that release neuromodulators that affect information processing widely in the brain, including in the GNC. Amygdala outputs also trigger behavioral and physiological responses in the body that help cope with the dangerous situation; feedback to the brain from the amygdala-triggered body responses (as in Damasio) also change processing in many relevant circuits, including the GNC, arousal circuits, and survival circuits, creating loops that sustain the reaction. In addition, the amygdala has direct connections with the sensory cortex, allowing bottom-up attentional control over sensory processing and memory retrieval, and also has direct connections with the GNC. These various effects on the GNC may well influence top-down attentional control over sensory processing, memory retrieval, and other cognitive functions. The GNC is also important in regulation over time of the overall brain and body state that results from survival circuit activation (21, 167, 168).

Unlike in traditional HOT, where there is somewhat of a hierarchical, or serial, relation between lower- and higher-order cortical circuits, in HOTEC the amygdala represents an additional lower-order circuit that can be activated in parallel with, and even independent of, the GNC. For example, subcortical inputs from visual or auditory areas of the thalamus can activate the amygdala before and independent of the GNC (8, 20). This finding suggests that direct amygdala activation, say by deep brain stimulation, might produce subjective experiences related to fear. Some claim that findings from such stimulations show that fear experiences are directly encoded in the amygdala (5). However, the subjective consequences of brain stimulation are not necessarily encoded by the neurons stimulated (20, 169). Moreover, the vague nature of the verbal reports in these studies (170) are less consistent with induction of specific emotional experiences than with the possibility that activation of body or brain arousal by amygdala outputs could, along the lines proposed by Schachter (171), induce a distressing state of dissonance that is cognitively interpreted and labeled as fear or anxiety (20). This information, together with the observation that amygdala damage (see above) does not eliminate fear experiences, suggest that fear can exist independent of the lower-order inputs. Whether such experiences are different, either quantitatively or qualitatively, because of the absence of the consequences of amygdala activation, is not known.

One additional factor needs to be considered. An emotion schema is a collection of information about a particular emotion (20, 172–174). A fear schema, for example, would include factual information (semantic memories) about harmful objects and situations, and about behavioral and body responses that occur in such situations. Thus, if you find yourself in a situation in which a harmful stimulus is present (a threat), and notice, through self-monitoring, that you are freezing and your heart is racing, these factors will likely match facts associated with fear in the fear schema and, through pattern completion, activate the fear schema. The schema will also include factual memories about how to cope with danger, and episodic memories about how you cope with such situations, which will bias the particular thoughts and actions used to cope. Emotion schema are learned in childhood and used to categorize situations as one goes through life. As one becomes more emotionally experienced, the states become more differentiated: fright comes to be distinguished from startle, panic, dread, and anxiety. When an emotion schema is present as part of a HOR, it biases the content of the experience that the HOR will support. Thus, an autonoetic emotional experience of fear is based on an emotional HOROR that includes the self (a fear self-HOROR). Of particular note is that the presence of a vague threat or physiological arousal, as noted above in relation to brain stimulation, may be sufficient to induce pattern completion of the fear schema.

Tulving argued that autonoetic consciousness is an exclusive feature of the human brain (135). Other animals could, in principle, experience noetic states about being in danger. However, because such states lack the involvement of the self, as a result of the absence of autonoetic awareness, the states would not, in our view, be emotions.

Another unique feature of the human brain that is relevant to self-awareness is natural language. Language organizes experiences into categories and shapes thought (175, 176). Words related to various emotions are an important part of the emotion schema stored in memory. More than three-dozen words exist in English to characterize fear-related experiences (177). It has long been thought that language influences experience (178), including emotional experience (179, 180). Language also allows symbolic representation of the experience of emotions without the actual exposure to the stimuli that normally elicit these emotions (181). Damasio's (6, 117) notion of "as-if loops" and Rolls's "if-then syntactic thoughts" (169) are consistent with this idea. Although the ability to imagine emotions is useful, this can also become a vehicle for excessive rumination, worry, and obsessions. If self-awareness and emotion can exist without language, these are surely different when the organism has language.

We have emphasized here how defensive circuits can contribute to conscious emotional feelings of fear. This makes it easy to slip into the idea that a defensive survival circuit is a fear circuit. However, the feeling of fear can occur in response to activity in other survival circuits as well (e.g., energy, fluid balance, thermoregulatory circuits); in relevant circumstances, you can fear dying of starvation, dehydration, or freezing to death. We call these fears (or "anxieties") because they are triggered by specific stimuli and interpreted in terms of stored schemas related to danger and harm to well-being. An emotional experience results from the cognitive representation of situations in which you find yourself, in light of what you know about such situations from past experiences that have provided you with factual knowledge and personal memories.

One implication of our view is that emotions can never be unconscious. Responses controlled by subcortical survival circuits that operate nonconsciously sometimes occur in conjunction with emotional feelings but are not emotions. An emotion is the conscious experience that occurs when you are aware that you are in particular kind of situation that you have come, through your experiences, to think of as a fearful situation. If you are not aware that you are afraid, you are not afraid; if you are not afraid, you aren't feeling fear. Another implication is that you can never be mistaken about what emotion you are feeling. The emotion is the experience you are having: if you are feeling afraid but someone tells you that they think you were angry or jealous, they may be accurate about why feeling angry or jealous might have been appropriate, given behaviors you expressed in the situation, but they would be wrong about what you actually experienced.

An important question to consider is the function of fear and other states of emotional awareness. Our proposal that emotions are cognitive states is consistent with the idea that once they are assembled in the GNC they can contribute to decision making (6, 117, 169), as well as to imaginations about one's future self and the emotions it may experience, and about decisions and actions one's future self might take when these emotions occur. This notion overlaps with a proposal by Mobbs and colleagues (24, 164). Emotion schema, built up by past emotions, would provide a context and set of constraints for such anticipated emotions. In the short-term, anticipated emotions might, like the GNC itself, play a role in top-down modulation of perceptual and memory processing, but also processing in subcortical survival circuits that contribute to the initial assembly of the emotional state. Considerable evidence

shows that top-down cognitive modulation of survival circuits occurs (21, 167, 168), and presumably emotional schema within the GNC, could similarly modulate survival circuit activity.

Relation of HOTEC to Other Theories of Emotion

A key aspect of our HOTEC is the HOR of the self; simply put, no self, no emotion. HOROR, and especially self-HOROR, make possible a HOT of emotion in which self-awareness is a key part of the experience. In the case of fear, the awareness that it is you that is in danger is key to the experience of fear. You may also fear that harm will come to others in such a situation but, as argued above, such an experience is only an emotional experience because of your direct or empathic relation to these people.

One advantage of our theory is that the conscious experience of all emotions (basic and secondary) (2–5), and emotional and nonemotional states of consciousness, are all accounted for by one system (the GNC). As such, elements of cognitive theories of consciousness by necessity contribute to HOTEC. Included implicitly or explicitly are cognitive processes that are key to other theories of consciousness, such as working memory (54, 55, 57, 103, 104, 182), attention amplification (105, 109), and reentrant processing (87).

Our theory of emotion, which has been in the making since the 1970s (8, 9, 20, 35, 166, 183), shares some elements with other cognitive theories of emotion, such as those that emphasize processes that give rise to syntactic thoughts (169), or that appraise (184–186), interpret (112), attribute (171, 187), and construct (188–192) emotional experiences. Because these cognitive theories of emotion depend on the rerepresentation of lower-order information, they are higher-order in nature.

Conclusion

In this paper we have aimed to redirect attention from subcortical to cortical circuits in the effort to understand emotional consciousness. In doing so, we built upon contemporary theorizing about perceptual consciousness in philosophy, cognitive science, and neuroscience, and especially on the debate between first- and higher-order theories, in an effort to account for how feelings arise as a result of introspective awareness of internal information.

Using the emotion of fear to illustrate our views, we argue that in the presence of a threat different circuits underlie the conscious feelings of fear and the behavioral responses and physiological responses that also occur. The experience of fear, the conscious emotional feeling we propose, results when a first-order representation of the threat enters into a HOR, along with relevant long-term memories—including emotion schema—that are retrieved. This initial HOR involving the threat and the relevant memories occurs nonconsciously. Then, a HOROR allows for the conscious noetic experience of the stimulus as dangerous. However, to have the emotional autonoetic experience of fear, the self must be included in the HOROR. In typical instances of fear, these representations are supplemented by the consequences of activation of subcortical survival circuits (not just defensive circuits but any circuit that indicates a threat to well-being). However, as noted earlier, fear can occur when the defensive survival circuit is damaged. Furthermore, existential fear/anxiety about the meaningless of life or the eventuality of death may not engage survival circuits at all. Our theory can thus potentially account for all forms of fear: those accompanied by brain arousal and bodily responses and those that are purely cognitive and even existential. Although we have focused on fear, we believe that the basic principles involved can be leveraged to understand other emotions as well.

1. Darwin C (1872) *The Expression of the Emotions in Man and Animals* (Fontana, London).
2. Tomkins SS (1962) *Affect, Imagery, Consciousness* (Springer, New York).
3. Ekman P (1992) An argument for basic emotions. *Cogn Emotion* 6:169–200.
4. Izard CE (1992) Basic emotions, relations among emotions, and emotion-cognition relations. *Psychol Rev* 99(3):561–565.
5. Panksepp J (1998) *Affective Neuroscience* (Oxford Univ Press, New York).
6. Damasio A (1994) *Descarte's Error: Emotion, Reason, and the Human Brain* (Gosset/Putnam, New York).
7. MacLean PD (1952) Some psychiatric implications of physiological studies on frontotemporal portion of limbic system (visceral brain). *Electroencephalogr Clin Neurophysiol* 4(4):407–418.
8. LeDoux JE (1996) *The Emotional Brain* (Simon and Schuster, New York).
9. LeDoux JE (2002) *Synaptic Self: How Our Brains Become Who We Are* (Viking, New York).
10. LeDoux JE (1991) Emotion and the limbic system concept. *Concepts Neurosci* 2: 169–199.
11. Swanson LW (1983) The hippocampus and the concept of the limbic system. *Neurobiology of the Hippocampus*, ed Seifert W (Academic, London), pp 3–19.
12. Kötter R, Meyer N (1992) The limbic system: A review of its empirical foundation. *Behav Brain Res* 52(2):105–127.
13. Davis M (1992) The role of the amygdala in conditioned fear. *The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction*, ed Aggleton JP (Wiley-Liss, New York), pp 255–306.
14. Fanselow MS, Poulos AM (2005) The neuroscience of mammalian associative learning. *Annu Rev Psychol* 56:207–234.
15. Maren S (2005) Synaptic mechanisms of associative memory in the amygdala. *Neuron* 47(6):783–786.
16. Lang PJ, Davis M (2006) Emotion, motivation, and the brain: Reflex foundations in animal and human research. *Prog Brain Res* 156:3–29.
17. Adolphs R (2013) The biology of fear. *Curr Biol* 23(2):R79–R93.
18. Rosen JB, Schulkin J (1998) From normal fear to pathological anxiety. *Psychol Rev* 105(2):325–350.
19. Gross CT, Canteras NS (2012) The many paths to fear. *Nat Rev Neurosci* 13(9): 651–658.
20. LeDoux JE (2015) *Anxious: Using the Brain to Understand and Treat Fear and Anxiety* (Viking, New York).
21. Phelps EA (2006) Emotion and cognition: Insights from studies of the human amygdala. *Annu Rev Psychol* 57:27–53.
22. Dolan RJ, Vuilleumier P (2003) Amygdala automaticity in emotional processing. *Ann N Y Acad Sci* 985:348–355.
23. Taylor JM, Whalen PJ (2015) Neuroimaging and anxiety: The neural substrates of pathological and non-pathological anxiety. *Curr Psychiatry Rep* 17(6):49.
24. Mobbs D, Hagan CC, Dalgleish T, Silston B, Prévost C (2015) The ecology of human fear: Survival optimization and the nervous system. *Front Neurosci* 9:55.
25. Mineka S, Ohman A (2002) Phobias and preparedness: The selective, automatic, and encapsulated nature of fear. *Biol Psychiatry* 52(10):927–937.
26. Vuilleumier P, Pourtois G (2007) Distributed and interactive brain mechanisms during emotion face perception: Evidence from functional neuroimaging. *Neuropsychologia* 45(1):174–194.
27. Luo Q, et al. (2010) Emotional automaticity is a matter of timing. *J Neurosci* 30(17): 5825–5829.
28. Morris JS, Ohman A, Dolan RJ (1999) A subcortical pathway to the right amygdala mediating "unseen" fear. *Proc Natl Acad Sci USA* 96(4):1680–1685.
29. Williams LM, et al. (2006) Mode of functional connectivity in amygdala pathways dissociates level of awareness for signals of fear. *J Neurosci* 26(36):9264–9271.
30. Hariri AR, Tessitore A, Mattay VS, Fera F, Weinberger DR (2002) The amygdala response to emotional stimuli: A comparison of faces and scenes. *Neuroimage* 17(1): 317–323.
31. Perusini JN, Fanselow MS (2015) Neurobehavioral perspectives on the distinction between fear and anxiety. *Learn Mem* 22(9):417–425.
32. Mowrer OH (1960) *Learning Theory and Behavior* (Wiley, New York).
33. Panksepp J, Fuchs T, Iacabucci P (2011) The basic neuroscience of emotional experiences in mammals: The case of subcortical FEAR circuitry and implications for clinical anxiety. *Appl Anim Behav Sci* 129:1–17.
34. Bolles RC, Fanselow MS (1980) A perceptual-defensive-recuperative model of fear and pain. *Behav Brain Sci* 3:291–323.
35. LeDoux JE (2014) Coming to terms with fear. *Proc Natl Acad Sci USA* 111(8): 2871–2878.
36. LeDoux J (2012) Rethinking the emotional brain. *Neuron* 73(4):653–676.
37. LeDoux JE, Pine DS (2016) Using neuroscience to help understand fear and anxiety: A two-system framework. *Am J Psychiatry* 173(11):1083–1093.
38. Rachman S, Hodgson R (1974) I. Synchrony and desynchrony in fear and avoidance. *Behav Res Ther* 12(4):311–318.
39. Bornemann B, Winkielman P, van der Meer E (2012) Can you feel what you do not see? Using internal feedback to detect briefly presented emotional stimuli. *Int J Psychophysiol* 85(1):116–124.
40. Morris JS, DeGelder B, Weiskrantz L, Dolan RJ (2001) Differential extrageniculostriate and amygdala responses to presentation of emotional faces in a cortically blind field. *Brain* 124(Pt 6):1241–1252.
41. Tamietto M, de Gelder B (2010) Neural bases of the non-conscious perception of emotional signals. *Nat Rev Neurosci* 11(10):697–709.
42. Bertini C, Cecere R, Làdavas E (2013) I am blind, but I "see" fear. *Cortex* 49(4): 985–993.
43. Feinstein JS, et al. (2013) Fear and panic in humans with bilateral amygdala damage. *Nat Neurosci* 16(3):270–272.
44. Anderson AK, Phelps EA (2002) Is the human amygdala critical for the subjective experience of emotion? Evidence of intact dispositional affect in patients with amygdala lesions. *J Cogn Neurosci* 14(5):709–720.

45. Manson N (2000) State consciousness and creature consciousness: A real distinction. *Philos Psychol* 13(3):405–410.
46. Rosenthal DM (1986) Two concepts of consciousness. *Philos Stud* 49:329–359.
47. Frith C, Perry R, Lumer E (1999) The neural correlates of conscious experience: An experimental framework. *Trends Cogn Sci* 3(3):105–114.
48. Jack AI, Shallice T (2001) Introspective physicalism as an approach to the science of consciousness. *Cognition* 79(1-2):161–196.
49. Overgaard M (2003) On the theoretical and methodological foundations for a science of consciousness. *Bulletin fra Forum For Antropologisk Psykologi* 13:6–31.
50. Jack AI, Roepstorff A (2003) Why trust the subject? *J Conscious Stud* 10(9-10):v–xx.
51. Terrace H, Metcalfe J (2004) *The Missing Link in Cognition: Origins of Self-Reflective Consciousness* (Oxford Univ Press, New York).
52. Smith JD (2009) The study of animal metacognition. *Trends Cogn Sci* 13(9):389–396.
53. Overgaard M, Sandberg K (2014) Kinds of access: Different methods for report reveal different kinds of metacognitive access. *The Cognitive Neuroscience of Metacognition*, eds Fleming SM, Frith CD (Springer, Berlin), pp 67–86.
54. Dehaene S (2014) *Consciousness and the Brain: Deciphering How the Brain Codes Our Thoughts* (Penguin, New York).
55. Shallice T (1988) Information processing models of consciousness. *Consciousness in Contemporary Science*, eds Marcel A, Bisiach E (Oxford Univ Press, Oxford), pp 305–333.
56. McGovern K, Baars BJ (2007) Cognitive theories of consciousness. *The Cambridge Handbook of Consciousness*, eds Zelazo PD, Moscovitch M, Thompson E (Cambridge Univ Press, New York), pp 177–205.
57. Baars BJ, Franklin S (2003) How conscious experience and working memory interact. *Trends Cogn Sci* 7(4):166–172.
58. Kihlstrom JF (1987) The cognitive unconscious. *Science* 237(4821):1445–1452.
59. Jacobs C, Silvanto J (2015) How is working memory content consciously experienced? The 'conscious copy' model of WM introspection. *Neurosci Biobehav Rev* 55:510–519.
60. Jacob J, Jacobs C, Silvanto J (2015) Attention, working memory, and phenomenal experience of WM content: memory levels determined by different types of top-down modulation. *Front Psychol* 6:1603.
61. Bergström F, Eriksson J (2014) Maintenance of non-consciously presented information engages the prefrontal cortex. *Front Hum Neurosci* 8:938.
62. Kiefer M (2012) Executive control over unconscious cognition: Attentional sensitization of unconscious information processing. *Front Hum Neurosci* 6:61.
63. Breitmeyer BG, Ogmen H (2006) *Visual Masking: Time Slices Through Conscious and Unconscious Vision* (Oxford Univ Press, Oxford).
64. Rees G, Frith C (2007) Methodologies for identifying the neural correlates of consciousness. *A Companion to Consciousness*, eds Velmans M, Schneider S (Blackwell, Oxford).
65. Lau HC, Passingham RE (2006) Relative blindsight in normal observers and the neural correlate of visual consciousness. *Proc Natl Acad Sci USA* 103(49):18763–18768.
66. Gaillard R, et al. (2009) Converging intracranial markers of conscious access. *PLoS Biol* 7(3):e61.
67. Dehaene S, Changeux JP, Naccache L, Sackur J, Sergent C (2006) Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends Cogn Sci* 10(5): 204–211.
68. Naccache L, Blandin E, Dehaene S (2002) Unconscious masked priming depends on temporal attention. *Psychol Sci* 13(5):416–424.
69. Craig AD (2009) How do you feel—now? The anterior insula and human awareness. *Nat Rev Neurosci* 10(1):59–70.
70. Craig AD (2010) The sentient self. *Brain Struct Funct* 214(5-6):563–577.
71. Frith C, Dolan R (1996) The role of the prefrontal cortex in higher cognitive functions. *Brain Res Cogn Brain Res* 5(1-2):175–181.
72. Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 24:167–202.
73. Goldman-Rakic PS (1999) Working memory, neural basis. *MIT Encyclopedia of Cognitive Sciences*, eds Wilson RA, Keil FC (MIT Press, Cambridge, MA).
74. Fuster JM (2003) *Cortex and Mind: Unifying Cognition* (Oxford Univ Press, Oxford).
75. Fleming SM, Huijgen J, Dolan RJ (2012) Prefrontal contributions to metacognition in perceptual decision making. *J Neurosci* 32(18):6117–6125.
76. Weiskrantz L (1997) *Consciousness Lost and Found: A Neuropsychological Exploration* (Oxford Univ Press, New York).
77. Persaud N, et al. (2011) Awareness-related activity in prefrontal and parietal cortices in blindsight reflects more than superior visual performance. *Neuroimage* 58(2): 605–611.
78. Vuilleumier P, et al. (2008) Abnormal attentional modulation of retinotopic cortex in parietal patients with spatial neglect. *Curr Biol* 18(19):1525–1529.
79. Del Cul A, Dehaene S, Reyes P, Bravo E, Slachevsky A (2009) Causal role of prefrontal cortex in the threshold for access to consciousness. *Brain* 132(Pt 9):2531–2540.
80. Pascual-Leone A, Walsh V (2001) Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science* 292(5516):510–512.
81. Carruthers P (2016) Higher-order theories of consciousness. *The Stanford Encyclopedia of Philosphy*, ed Zalta EN (Metaphysics Research Laboratory, Stanford Univ, Stanford, CA).
82. Nagel T (1974) What is it like to be a bat? *Philos Rev* 83:4435–4450.
83. Chalmers D (1996) *The Conscious Mind* (Oxford Univ Press, New York).
84. Block N (1995) How many concepts of consciousness? *Behav Brain Sci* 18(2):272–284.
85. Dretske F (1995) *Naturalizing the Mind* (MIT Press, Cambridge, MA).
86. Tye M (2000) *Consciousness, Color, and Content* (MIT Press, Cambridge, MA).
87. Lamme VAF (2005) Independent neural definitions of visual awareness and attention. *Cognitive Penetrability of Perception: Attention, Action, Strategies, and Bottom-Up Constraints*, ed Raftopoulos A (Nova Science, New York), pp 171–191.
88. Block N (2011) Perceptual consciousness overflows cognitive access. *Trends Cogn Sci* 15(12):567–575.
89. Block N (2007) Consciousness, accessibility, and the mesh between psychology and neuroscience. *Behav Brain Sci* 30(5-6):481–499, discussion 499–548.
90. Rosenthal DM (2005) *Consciousness and Mind* (Oxford Univ Press, Oxford).
91. Lau H, Rosenthal D (2011) Empirical support for higher-order theories of conscious awareness. *Trends Cogn Sci* 15(8):365–373.
92. Carruthers P (2005) *Consciousness: Essays from a High-Order Perspective* (Oxford Univ Press, Oxford).
93. Cleeremans A, Timmermans B, Pasquali A (2007) Consciousness and metarepresentation: A computational sketch. *Neural Netw* 20(9):1032–1039.
94. Cohen MA, Dennett DC (2011) Consciousness cannot be separated from function. *Trends Cogn Sci* 15(8):358–364.
95. Brown R (2012) The myth of phenomenological overflow. *Conscious Cogn* 21(2): 599–604.
96. Brown R (2012) The brain and its states. *Being in Time: Dynamical Models of Phenomenal Experience*, eds Edelman S, Fekete T, Zach N (John Benjamins, Philadelphia), pp 211–238.
97. Kouider S, de Gardelle V, Sackur J, Dupoux E (2010) How rich is consciousness? The partial awareness hypothesis. *Trends Cogn Sci* 14(7):301–307.
98. Stazicker J (2011) Attention, visual consciousness and indeterminacy. *Mind Lang* 26(2):156–184.
99. Phillips IB (2011) Perception and iconic memory: What Sperling doesn't show. *Mind Lang* 26(4):381–411.
100. Brown R (2014) Consciousness doesn't overflow cognition. *Front Psychol* 5:1399.
101. Lau HC, Passingham RE (2007) Unconscious activation of the cognitive control system in the human prefrontal cortex. *J Neurosci* 27(21):5805–5811.
102. Maniscalco B, Lau H (2016) The signal processing architecture underlying subjective reports of sensory awareness. *Neurosci Conscious*, 10.1093/nc/niw002.
103. Baddeley A (2000) The episodic buffer: A new component of working memory? *Trends Cogn Sci* 4(11):417–423.
104. Schacter DL (1989) On the relation between memory and consciousness: Dissociable interactions and conscious experience. *Varieties of Memory and Consciousness: Essays in Honour of Endel Tulving*, eds Roediger HLI, Craik FIM (Lawrence Erlbaum Associates, Hillsdale, NJ), pp 355–389.
105. Prinz JJ (2012) *The Conscious Brain: How Attention Engenders Experience* (Oxford Univ Press, New York).
106. Crick F, Koch C (2003) A framework for consciousness. *Nat Neurosci* 6(2):119–126.
107. Cohen MA, Cavanagh P, Chun MM, Nakayama K (2012) The attentional requirements of consciousness. *Trends Cogn Sci* 16(8):411–417.
108. Bor D, Seth AK (2012) Consciousness and the prefrontal parietal network: Insights from attention, working memory, and chunking. *Front Psychol* 3:63.
109. Koch C (2004) *The Quest for Consciousness: A Neurobiological Approcah* (Roberts and Co., Denver).
110. Baars BJ (2005) Global workspace theory of consciousness: Toward a cognitive neuroscience of human experience. *Prog Brain Res* 150:45–53.
111. Baars BJ (1988) *A Cognitive Theory of Consciousness* (Cambridge Univ Press, New York).
112. Gazzaniga MS (2008) *Human: The Science Behind What Makes us Unique* (Ecco, New York).
113. Lau H, Rosenthal D (2011) The higher-order view does not require consciously self-directed introspection: Response to Malach. *Trends Cogn Sci* 15(11):508–509.
114. Panksepp J (2012) *The Archaeology of Mind: Neuroevolutionary Origins of Human Emotion* (WW Norton & Company, New York).
115. Vandekerckhove M, Panksepp J (2009) The flow of anoetic to noetic and autonoetic consciousness: A vision of unknowing (anoetic) and knowing (noetic) consciousness in the remembrance of things past and imagined futures. *Conscious Cogn* 18(4): 1018–1028.
116. Vandekerckhove M, Panksepp J (2011) A neurocognitive theory of higher mental emergence: From anoetic affective experiences to noetic knowledge and autonoetic awareness. *Neurosci Biobehav Rev* 35(9):2017–2025.
117. Damasio AR (1999) *The Feeling of What Happens: Body and Emotion in the Making of Consciousness* (Harcourt Brace, New York).
118. Damasio A, Carvalho GB (2013) The nature of feelings: Evolutionary and neurobiological origins. *Nat Rev Neurosci* 14(2):143–152.
119. Bogdanov VB, et al. (2013) Alexithymia and empathy predict changes in autonomic arousal during affective stimulation. *Cogn Behav Neurol* 26(3):121–132.
120. Armstrong DM (1981) *What Is Consciousness? The Nature of Mind*, ed Heil J (Cornell Univ Press, Ithaca, New York).
121. Block N (2011) The higher-order approach to consciousness is defunct. *Analysis* 71(3): 419–431.
122. Gennaro RJ (2011) *The Consciousness Paradox* (MIT Press, Cambridge, MA).
123. Kriegel U (2009) *Subjective Consciousness: A Self-Representational Theory* (Oxford Univ Press, Oxford).
124. Brown R (2015) The HOROR theory of phenomenal consciousness. *Philos Stud* 172(7): 1783–1794.
125. Peters MA, Lau H (2015) Human observers have optimal introspective access to perceptual processes even for visually masked stimuli. *eLife* 4:e09651.
126. Stein T, Kaiser D, Hesselmann G (2016) Can working memory be non-conscious? *Neurosci Conscious*, 10.1093/nc/niv011.
127. Soto D, Silvanto J (2014) Reappraising the relationship between working memory and conscious awareness. *Trends Cogn Sci* 18(10):520–525.
128. Squire L (1987) *Memory and Brain* (Oxford Univ Press, New York).
129. Eichenbaum H (2002) *The Cognitive Neuroscience of Memory* (Oxford Univ Press, New York).

130. Suzuki WA, Amaral DG (2004) Functional neuroanatomy of the medial temporal lobe memory system. *Cortex* 40(1):220–222.
131. Weisberg J (2011) Misrepresenting consciousness. *Philos Stud* 154(3):409–433.
132. Rosenthal D, Weisberg J (2008) Higher-order theories of consciousness. *Scholarpedia* 3(5):4407.
133. Bennett J (1988) Thoughtful brutes. *Proceedings and Addresses of the American Philosophical Association* 62:197–210.
134. Rosenthal D (2012) Awareness and identification of self. *Consciousness and the Self: New Essays*, eds Liu J, Perry J (Cambridge Univ Press, New York), pp 22–50.
135. Tulving E (2005) Episodic memory and autonoesis: Uniquely human? *The Missing Link in Cognition*, eds Terrace HS, Metcalfe J (Oxford Univ Press, New York), pp 4–56.
136. Tulving E (2001) The origin of autonoesis in episodic memory. *The Nature of Remembering: Essays in Honor of Robert G. Crowder*, eds Roediger HL, Nairne JS, Neath I, Suprenant AM (American Psychololgy Association, Washington, DC), pp 17–34.
137. Tulving E (1983) *Memory and Consciousness* (Clarendon Press, Oxford).
138. Schacter DL (1985) Multiple forms of memory in humans and animals. *Memory Systems of the Brain: Animal and Human Cognitive Processes*, eds Weinberger NM, McGaugh JL, Lynch G (Guilford Publications, New York), pp 351–379.
139. Clayton NS, Dickinson A (1998) Episodic-like memory during cache recovery by scrub jays. *Nature* 395(6699):272–274.
140. Ergorul C, Eichenbaum H (2004) The hippocampus and memory for "what," "where," and "when". *Learn Mem* 11(4):397–405.
141. Block N (2009) Comparing the major theories of consciousness. *The Cognitive Neurosciences IV*, ed Gazzaniga MS (MIT Press, Cambridge, MA), 4th Ed, pp 1111–1122.
142. Clayton NS, Bussey TJ, Dickinson A (2003) Can animals recall the past and plan for the future? *Nat Rev Neurosci* 4(8):685–691.
143. Suddendorf T, Corballis MC (2010) Behavioural evidence for mental time travel in nonhuman animals. *Behav Brain Res* 215(2):292–298.
144. Lewis M (2013) *The Rise of Consciousness and the Development of Emotional Life* (Guilford Press, New York).
145. Lou HC, Changeux JP, Rosenstand A (April 11, 2016) Towards a cognitive neuroscience of self-awareness. *Neurosci Biobehav Rev*, 10.1016/j.neubiorev.2016.04.004.
146. Werning M (2010) Descartes discarded? Introspective self-awareness and the problems of transparency and compositionality. *Conscious Cogn* 19(3):751–761.
147. Kihlstrom JF, Klein SB (1997) Self-knowledge and self-awareness. *Ann N Y Acad Sci* 818:4–17.
148. Klein SB (2013) Making the case that episodic recollection is attributable to operations occurring at retrieval rather than to content stored in a dedicated subsystem of long-term memory. *Front Behav Neurosci* 7:3.
149. Conway MA (2005) Memory and the self. *J Mem Lang* 53(4):594–628.
150. Marsh EJ, Roediger HL (2013) Episodic and autobiographical memory. *Handbook of Psychology: Volume 4, Experimental Psychology*, eds Healy AF, Proctor RW (John Wiley & Sons, New York), 2nd Ed, pp 472–494.
151. Piaget J (1971) *Biology and Knowledge* (Edinburgh Univ Press, Edinburgh).
152. Mandler JM (1984) *Stories, Scripts, and Scenes: Aspects of Schema Theory* (Lawrence Erlbaum Associates, Hillsdale, NJ).
153. O'Sullivan CS, Durso FT (1984) Effect of schema-incongruent information on memory for stereotypical attributes. *J Pers Soc Psychol* 47(1):55–70.
154. Cox WT, Abramson LY, Devine PG, Hollon SD (2012) Stereotypes, prejudice, and depression: The integrated perspective. *Perspect Psychol Sci* 7(5):427–449.
155. Petersen LE, Stahlberg D, Dauenheimer D (2000) Effects of self-schema elaboration on affective and cognitive reactions to self-relevant information. *Genet Soc Gen Psychol Monogr* 126(1):25–42.
156. Sui J, Humphreys GW (2015) The integrative self: How self-reference integrates perception and memory. *Trends Cogn Sci* 19(12):719–728.
157. Frith U, Happé F (1999) Theory of mind and self-consciousness: What is it like to be autistic? *Mind Lang* 14(1):82–89.
158. Frith CD (2012) The role of metacognition in human social interactions. *Philos Trans R Soc Lond B Biol Sci* 367(1599):2213–2223.
159. Fleming SM, Dolan RJ, Frith CD (2012) Metacognition: Computation, biology and function. *Philos Trans R Soc Lond B Biol Sci* 367(1594):1280–1286.
160. Wheeler MA, Stuss DT (2003) Remembering and knowing in patients with frontal lobe injuries. *Cortex* 39(4-5):827–846.
161. Ochsner KN, et al. (2004) Reflecting upon feelings: An fMRI study of neural systems supporting the attribution of emotion to self and other. *J Cogn Neurosci* 16(10): 1746–1772.
162. Sturm VE, Rosen HJ, Allison S, Miller BL, Levenson RW (2006) Self-conscious emotion deficits in frontotemporal lobar degeneration. *Brain* 129(Pt 9):2508–2516.
163. Williams LM, et al. (2006) Amygdala-prefrontal dissociation of subliminal and supraliminal fear. *Hum Brain Mapp* 27(8):652–661.
164. Koizumi A, Mobbs D, Lau H (2016) Is fear perception special? Evidence at the level of decision-making and subjective confidence. *Soc Cogn Affect Neurosci* 11(11): 1772–1782.
165. LeDoux JE (2008) Emotional colouration of consciousness: How feelings come about. *Frontiers of Consciousness: Chichele Lectures*, eds Weiskrantz L, Davies M (Oxford Univ Press, Oxford), pp 69–130.
166. LeDoux JE (1984) Cognition and emotion: Processing functions and brain systems. *Handbook of Cognitive Neuroscience*, ed Gazzaniga MS (Plenum, New York), pp 357–368.
167. Buhle JT, et al. (2014) Cognitive reappraisal of emotion: A meta-analysis of human neuroimaging studies. *Cereb Cortex* 24(11):2981–2990.
168. Schiller D, Delgado MR (2010) Overlapping neural systems mediating extinction, reversal and regulation of fear. *Trends Cogn Sci* 14(6):268–276.
169. Rolls ET (2008) Emotion, higher-order syntactic thoughts, and consciousness. *Frontiers of Consciousness: Chichele Lectures*, eds Weiskrantz L, Davies M (Oxford Univ Press, Oxford), pp 131–167.
170. Berridge KC, Kringelbach ML (2011) Building a neuroscience of pleasure and well-being. *Psychol Well Being* 1(1):1–3.
171. Schachter S (1975) Cognition and centralist-peripheralist controversies in motivation and emotion. *Handbook of Psychobiology*, eds Gazzaniga MS, Blakemore CB (Academic, New York), pp 529–564.
172. Izard CE (2007) Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspect Psychol Sci* 2(3):260–280.
173. Lang PJ (1979) Presidential address, 1978. A bio-informational theory of emotional imagery. *Psychophysiology* 16(6):495–512.
174. Beck AT, Freeman A (1990) *Cognitive Therapy of Personality Disorders* (Guilford Press, New York).
175. Bowerman M, Levinson SC, eds (2001) *Language Acquisition and Conceptual Development* (Cambridge Univ Press, Cambridge, UK).
176. Gentner D, Goldin-Meadow S, eds (2003) *Language in Mind: Advances in the Study of Language and Thought* (MIT Press, Cambridge, MA).
177. Marks I (1987) *Fears, Phobias, and Rituals: Panic, Anxiety and Their Disorders* (Oxford Univ Press, New York).
178. Whorf BL (1956) *Language, thought, and Reality* (Technology Press of MIT, Cambridge, MA).
179. Wierzbicka A (1994) Emotion, language, and cultural scripts. *Emotion and Culture: Empirical Studies of Mutual Influence*, eds Kitayama S, Markus HR (American Psychological Association, Washington, DC), pp 133–196.
180. Russell JA (1991) Natural language concepts of emotion. *Perspectives in Personality*, eds Hogan R, Jones WH, Stewart AJ, Healy JM, Ozer DJ (Jessica Kingsley, London), Vol 3.
181. Forsyth JP, Eifert GH (1996) The language of feeling and the feeling of anxiety: Contributions of the behaviorisms toward understanding the function-altering effects of language. *Psychol Rec* 46:607–649.
182. Schacter DL, Buckner RL, Koutstaal W (1998) Memory, consciousness and neuroimaging. *Philos Trans R Soc Lond B Biol Sci* 353(1377):1861–1878.
183. Gazzaniga MS, LeDoux JE (1978) *The Integrated Mind* (Plenum, New York).
184. Lazarus RS (1991) Cognition and motivation in emotion. *Am Psychol* 46(4):352–367.
185. Scherer K (2000) Emotions as episodes of subsystem synchronization driven by nonlinear appraisal processes. *Emotion, Development, and Self-Organization: Dynamic Systems Approaches to Emotional Development*, eds Lewis M, Granic I (Cambridge Univ Press, New York), pp 70–99.
186. Ortony A, Clore GL (1989) Emotions, moods, and conscious awareness. *Cogn Emotion* 3:125–137.
187. Mandler G (1975) *Mind and Emotion* (Wiley, New York).
188. Boiger M, Mesquita B (2012) The construction of emotion in interactions, relationships, and cultures. *Emot Rev* 4(3):221–229.
189. Averill JR (1980) A constructivist view of emotion. *Emotion: Theory, Research and Experience, Volume I. Theories of Emotiojn*, eds Plutchik R, Kellerman H (Academic, New York), pp 305–339.
190. Harre R, ed (1986) *The Social Construction of Emotions* (Basal Blackwell, Oxford).
191. Barrett LF (2017) *How Emotions Are Made: The Secret Life of the Brain* (Houghton Mifflin Harcourt, New York).
192. Barrett LF, Russell JA, eds (2015) *The Psychological Construction of Emotion* (Guilford Press, New York).