

My Word Thoughtful feelings

Joseph E. LeDoux^{1,2,3}

I have researched how brains detect and respond to danger for most of my scientific career. Because the neural circuits involved are shared between humans and other mammals, information about how these circuits work in them is relevant to how they work in us. But there is disagreement about what this means, especially for human emotions. Resolving the issue is important, not only for understanding what emotions are, but also for developing effective treatments for emotional disorders. In a previous My Word debate, moderated by Leonard Mlodinow, Ralph Adolphs and Lisa Feldman Barrett argued their different views about emotions [1]. Here, I offer my take on this topic, and especially on emotional consciousness, building on a recent My Word I wrote about how conscious experiences emerge from non-conscious processing [2].

Darwin the psychologist

In his 1872 book, *The Expression of Emotions in Man and Animals*, Charles Darwin extended his theory of evolution to include emotions [3]. He adopted the common-sense notion of emotions as states of mind that cause us to respond in characteristic ways — for example, fear causes us to flee from danger. And because emotional states of mind enhanced the fitness of our mammalian ancestors, these mental states were selected for and passed on to us. It follows that we can use behaviors that typically occur in dangerous situations to know when humans and other mammals are feeling fear. Darwin astutely noted that cross-species transmission of mental states would have to take place by way of conserved features of the nervous system. Knowledge about the brain was limited in Darwin's day, and he had little to say about the brain regions or circuits that might be responsible for the inheritance of emotions. Fast-forward to the present.

The amygdala fear center

Modern research has helped fill in the neural part of Darwin's hypothesis. The brain area called the amygdala is often said to be the 'fear center' we've inherited from our animal ancestors. In the presence of danger, a biologically inherited feeling of fear is presumed to be unleashed within the amygdala, and this causes the expression of so-called fear behaviors (for review, see [4–6]). The fear center idea appears in countless scientific papers and books, and also in novels, plays, movies, and songs, and in popular writings about how to tame your fears, succeed as a stock trader, and lead a happier life. It is nothing short of a cultural meme.

Work by me and others [4–7] on how the amygdala detects and responds to danger in rats has added quite a bit of fuel to this memetic fire. But I actually don't think that feelings of fear are made by the amygdala, and I have made a concerted effort to clarify this in recent years [4–6]. Some of the

factors that underlie my position are these.

Because we often feel afraid when we are freezing to a snake or running from a bear, we, like Darwin, intuitively assume that fear causes the responses. But when we do this, are we confusing correlation with causation? A number of studies have, in fact, shown that fearful feelings and fear responses as not as tightly coupled as we imagine. But they should be if they are both products of the amygdala. Indeed, recent brain imaging findings show that amygdala activity is more strongly correlated with threat-elicited body responses than with subjective experiences of fear [8]. Moreover, when threatening stimuli are presented to people subliminally — for example, using quick exposures and other techniques — in brain imaging studies, the amygdala is activated and body responses are elicited. But the participants do not have any awareness of having seen the stimulus, and do not report feeling fear, which

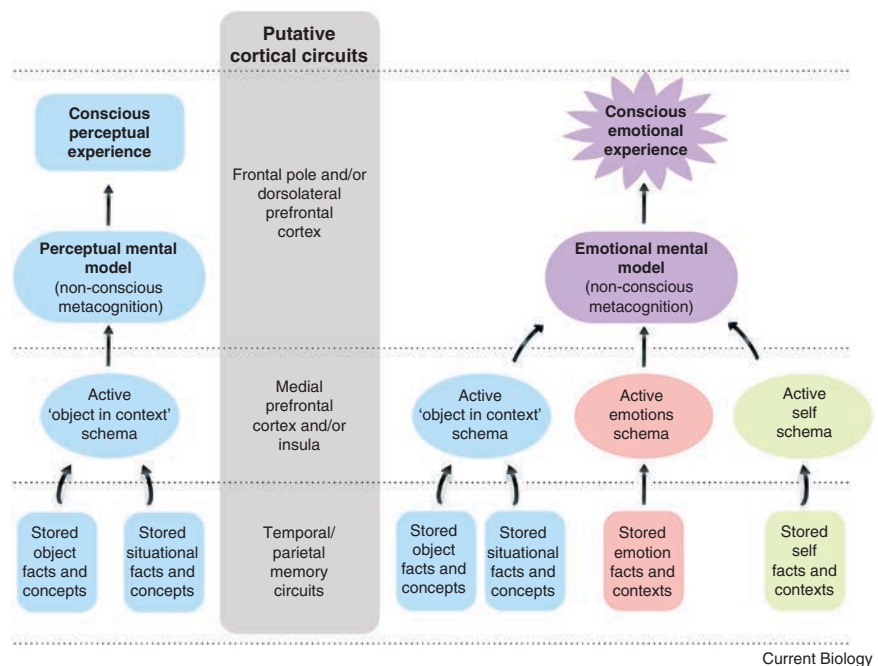


Figure 1. Memories, schemas, mental models, and conscious experiences.

Conscious perceptual and emotional experiences are both proposed to depend on antecedent non-conscious states, including memories, schemas, and mental models (i.e. metacognitions). But emotions have antecedent components lacking in non-emotional experiences. The states shown represent neural events that follow processing by sensory cortex. Brain and body states that are elicited in some instances of some emotions are proposed to affect an emotional experience by influencing the content of the active emotion schema. Although the arrows depict information flow from lower to higher levels, each level connects with its immediate antecedent level, and in most cases, with other lower levels as well.

they should do if amygdala activity is what makes fear. And while body responses to threats are disrupted in people who have suffered damage to the amygdala, they can sometimes report feeling fear — which should not happen if the amygdala makes fear.

Together, these findings suggest that the amygdala is indeed wired by evolution to detect and produce body responses to certain kinds of threats, but that it is not required to feel fear. For these and other reasons, I have argued that the amygdala's role in detecting and responding to threats is more appropriately considered in terms of a non-conscious defensive survival circuit than a conscious fear circuit [4,5].

The amygdala has such a prominent presence in discussions about fear because most research has focused on predatory defense responses, which depend, in part, on amygdala circuits. But the fact is, fear of bodily harm can result from many other kinds of events in life besides predators. Lack of food or water can cause us to fear starvation or dehydration; extreme low temperatures can cause fear of death by hypothermia; the news that you have a life-threatening disease elicits fear, as does the mere possibility of illness during an outbreak of a contagious virus. Additionally, fear can result from political instability, economic loss, social abuse, or existential concerns. Not all of these fear triggers depend on amygdala circuits, and even for those that do, the amygdala plays an ancillary rather than an essential role in fear. Also worth noting is that the amygdala is not the only brain area involved in defensive behavior, and that it contributes to a variety of behavioral and cognitive functions that have nothing to do with danger.

A cognitive approach to emotion

My alternative to the standard fear center hypothesis is this. When one faces certain kinds of danger, defensive responses like fleeing or freezing can co-occur with feelings of fear, not because the neural processes underlying these two kinds of events are intimately entwined in the amygdala, but instead because the events have the same starting point — a threatening stimulus that enters the

brain through the visual (or some other sensory) system. From there, the paths underlying the responses and feelings diverge.

Visual connections to the amygdala trigger innate behavioral responses, much as the Darwinian logic suggests. The conscious feeling of fear, on the other hand, results from visual connections to cognitive circuits of the prefrontal cortex, where diverse kinds of information are integrated in the process of creating the conscious feeling of fear [5–8]. While the effects of amygdala activation can indirectly affect the resultant feeling, they do not themselves make the feeling. To feel fear, all that is needed is a cognitive interpretation (a belief) that you are in danger [5,6].

The notion that emotion involves cognition is counterintuitive to some. Cognition is about thinking; emotion is about feeling. Why would we need cognition to feel fear? But whether an idea is consistent with our intuitions is irrelevant — we wouldn't need research if commonsense were sufficient to understand nature. Research, in fact, has made the cognitive approach to emotion a leading perspective today [9–11].

Cognition and consciousness

But let's dig deeper. If I am correct, understanding the cognitive basis of emotional experience is a subset of the problem of understanding the cognitive foundations of consciousness. What you are conscious of at any one moment is believed to reflect the contents of working memory, a mental workspace that underlies the control of thought and action [12]. Most research on working memory has used visual stimuli and has focused on interactions between areas of visual cortex and prefrontal cortex. The key area of prefrontal cortex involved is the dorsolateral region [13], which is also the main prefrontal region implicated in visual consciousness [14,15].

In a previous My Word [2], I suggested that the dorsolateral prefrontal cortex integrates inputs from visual cortex and temporal lobe memory circuits to transform meaningless sensations into meaningful perceptual experiences of objects in complex contexts. I also noted that the dorsolateral

area receives indirect inputs from memory circuits by way of other prefrontal regions, providing additional conceptual tools for understanding the stimuli we encounter in life. For example, the ventromedial prefrontal area, which connects with the dorsolateral region, receives inputs from memory circuits and has been implicated in processing schemas [16]. These are collections of semantic memories about recurring objects and situations. The schema components active in the moment serve as non-conscious conceptual templates for understanding present stimuli in the context in which they are embedded, and in relation to the person's needs and goals.

Another prefrontal region of interest in relation to conceptualization capacities is the frontal pole (also known as polar, anterior, or rostral prefrontal cortex) [2,5,15,17–19]. The lateral part of the frontal pole is considered a unique human specialization and is especially important in higher cognition. It has minimal if any sensory inputs, and instead is extensively connected with memory circuits in the temporal and parietal lobes, and with other prefrontal areas that are also connected with memory circuits, including the schema-forming ventromedial area. Consistent with these connections, the lateral frontal pole has the greatest conceptual prowess of any brain area; it has been implicated in hierarchical relational reasoning, stimulus independent thought, subjective metacognition, mentalizing and recollecting about one's self, prospective memory, and introspection.

I propose that the dorsolateral region and the lateral frontal pole integrate non-conscious object and context schemas to form mental models (Figure 1). Here, a mental model is conceived of as a non-conscious (implicit) metacognitive representation that conceptualizes the present situation, predicts future outcomes, and exerts top-down control over other cognitive processes. It is noteworthy that the dorsolateral region and the lateral frontal pole have been proposed to be components of a higher-order network that underlies perceptual consciousness [2,5,15]. How non-conscious memories, schemas, and

mental models might contribute to conscious perceptions is depicted in Figure 1 (other cortical circuits that may also contribute are discussed elsewhere [2]).

Emotional and self schemas

What about emotion? The core of my idea is that regardless of whether an experience is an emotional or non-emotional one, the same general cognitive processes and circuits are involved [2,5,6]. The difference, I suggest, is that the cognitive circuits work with different information in emotional *versus* non-emotional situations.

When a threatening stimulus, say a snake, is encountered, visual and memory circuits, including perceptual schema circuits, will be engaged to identify what the object is in light of the situational context. These memories contribute to the non-conscious mental model that underlies conscious emotional experience, but two additional kinds of memory schemas are also important [5,6] (Figure 1).

The first is an emotion schema. For example, your ‘fear schema’ is the collection of memories that you have accumulated about things and situations that you have come to know of as dangerous, what typically happens in danger, and how people typically act. But a fear schema is not simply a perceptual ‘object-in-context’ representation focused on dangerous stimuli. It defines the conceptual space of an emotional experience by providing prescriptive information — that fear is what people feel when in danger [10]. While the amygdala is activated in some forms of danger, its effects on conscious fear experiences are mediated by their impact on the active fear schema.

The other kind of schema that contributes to an emotional experience is a ‘self-schema’ — the collection of memories you have accumulated about yourself. These are autobiographical memories that include semantic information about you, but also episodic memories about your personal relationship to various kinds of experiences that you have had in your life — for example, how you typically act and feel when in danger. Most important, your self-schema makes an experience feel as if it is happening to

you. Self-schema can also be involved in non-emotional situations, but they play an especially important role in emotions. For example, to feel fear in a dangerous situation your fear schema has to define the situation as fearful, and you have to be aware that it is you who is in harm’s way. In other words, you have to be part of the subject matter of the dangerous experience in order to feel fear.

The relation of self to consciousness is complex. Conscious experience requires some involvement of the self — this makes it possible for you to know when you are consciously experiencing something: knowing that you are looking at an apple, for example. But emotions require that you also know that it is you that is having the experience — that it is you that is going to be bitten by the snake. The distinction can perhaps be understood as one between a noetic self (a factual semantic self of the moment) and an auto-noetic self (a reflective self with a personal past and future) [5,6,20]. Not all conscious experiences that involve auto-noetic self-awareness are emotional experiences. But one’s auto-noetic self — as characterized by one’s momentarily active self-schema — is always part of an emotional experience [5,6].

The brain circuits underlying fear and self-schemas are not as well understood as the circuits involved in perceptual schema. But clues are available. Several prefrontal areas have been implicated in emotional processing, including medial prefrontal areas (ventromedial, orbital, and anterior cingulate) and the insula cortex [5]. Each of these areas receives inputs from memory circuits, and from the amygdala and other subcortical circuits involved in body homeostasis. The same areas (and others) have been implicated in aspects of self-processing, including the sense of ownership of mental states [5]. Circuits involving medial prefrontal and insula cortical areas conceivably contribute to the construction of emotion and self-schemas.

Mental models and emotional experiences

As in conscious perception, conscious emotion requires another step. Momentarily active emotion and self-

schemas have to be integrated to form an emotional mental model, a non-conscious representation that shapes the content of conscious emotional experiences (Figure 1, right).

Each of the putative emotion and self-schema circuits mentioned above interacts with both the dorsolateral prefrontal area and the frontal pole, making these excellent candidate sites for the emotional mental model. For emotional experiences, however, the frontal pole may be especially important. In addition to its roles in higher cognition mentioned above [17–19], the frontal pole is also involved in various aspects of emotional regulation and self-processing (reviewed in [19,21]). It seems well suited for conceptualizing the relationship of one’s self to emotional situations.

According to the higher-order theory of consciousness, conscious experiences depend on higher-order mental states [5,6,15], especially metacognitive states [22]. Higher-order states, though, are not conscious; they are the penultimate non-conscious states that precede conscious experiences. The perceptual and emotional mental models depicted in Figure 1 are in effect non-conscious metacognitive higher-order states. While this leaves open the question of how a conscious experience itself comes about, identification of the neural circuits instantiating the mental models would, if this idea is correct, leave us just one step removed from the neural basis of the conscious experiences themselves.

Emotions, words and experiences

Emotion words categorize emotional experiences and provide conceptual anchors that help us understand and remember our experiences [5,9,10,23]. These labels are not required to feel emotionally aroused, but are required to feel the emotion named by the label. A distressed young child, lacking specific emotion words cannot not experience herself as being in a state that an older child experiences as fear when her mental model, drawing upon her emotion and self schemas, conceptualizes her experience that way.

But even in adults the non-conscious underpinnings of emotions are not always precise enough to produce an

experience that is clearly identified with a common emotion word. One may feel uncomfortable, concerned, or distressed in a situation, and not progress to something more specific. But as the situation unfolds and more information is collected, it is also possible that a vague feeling may turn into one labeled and experienced as fear, which might, with additional information, morph into anger or jealousy, or to relief.

To an observer who only has access to your outward behavior, it may seem you are not feeling what you are experiencing. But if fear is what you feel in the moment of the experience, you feel it, regardless of what it looks like from the outside [5]. Your first-person knowledge of *why* you are feeling what you are feeling is less reliable, however, than the fact that you *are* feeling that way. Also, your memory of what you experienced in the past is less reliable than what you consciously feel in the moment.

Emotions are personal and cultural, not universal, states

If each person has unique emotion schemas and self-schemas, and thus unique emotional mental models, emotional experiences must differ across individuals. This idea, that emotions are personalized, goes against Darwin's conclusion that emotions are similar in all humans, a conclusion based largely on his observation that facial expressions of emotion are similar around the world [3]. But research has questioned the idea that emotions are expressed in rigid universal ways [9,23]. In fact, cultural differences in how emotions are expressed and experienced are well established [23,24].

My view is that situations, rather than emotional feelings, are universal [5]. For example, because danger is a condition of life [5], all people, regardless of their cultural background, have a conception of danger, and thus experience something they name with a word that translates into what English speaking folks call fear.

Emotion words anchor the unique experiences an individual has to others in the same culture, and via translation, to related experiences of individuals in other cultures. But because an individual's emotion and self-schemas

and mental models are embedded in cultural narratives, people in different cultures experience danger in different ways. That emotion words can be translated across languages does not mean that members of different cultures have the same experiences. Though useful and important, semantic translations distort cultural differences in underlying meaning.

Implications for mental health

An important consequence of the conflation of the neural basis of conscious feelings and body responses is the failure of the pharmaceutical industry to find new, effective medications for treating problems related to fear and anxiety [25]. Many such studies have used behavioral responses in animals to test possible treatments, assuming that a drug that decreases defensive behaviors, such as freezing or escape, in rats or mice will be an effective treatments for fear anxiety disorders in humans. The results from decades of research have been so disappointing that the efforts have been scaled-back.

But the problem is not with the research. It with the widespread assumption that disordered fear and anxiety are problems that stem from pathophysiological alterations in circuits that we have inherited from our mammalian ancestors. To assess whether the circuits have been changed by treatments (including pharmaceutical and behavioral/cognitive ones), objectively observable/measurable symptoms (behavioral and physiological responses) have generally been preferred over subjective experiences, as reflected in self-reports. The latter have been viewed as less reliable, in part because they sometimes paint a different picture than the objective measures. But the thrust of what I have argued here is that subjective feelings are not an optional, less-reliable measure of fear or anxiety: they are the fear or anxiety that the person experiences.

Clearly, behavioral/physiological symptoms, which can be studied in animals, contribute to both normal and pathological feelings of fear and anxiety [5]. But assuming that when these objective symptoms are changed subjective/conscious symptoms will come along for the ride is not the

solution. The conscious part, the fear or the anxiety itself, has to be addressed, and addressed directly; otherwise the person will continue to feel fearful or anxious. By the same token, changing one's conscious symptoms will not necessarily make the objective ones go away. In other words, the constellation of symptoms that occur in a fear or anxiety disorder are best viewed as reflecting a federation of neural systems that must each be addressed in the treatment process [26].

Treatment providers presumably want their clients/patients to feel better. But the dominance of the medical disease model, and its emphasis on objective symptoms, has, for decades, sidelined the important subjective experience [26]. The good news is twofold: consciousness is currently a thriving area of scientific research (see <https://theassc.org/>), and there is growing recognition by clinicians that improved subjective well-being does not simply follow from behavioral/physiological symptom reduction alone. As we go forward, the science of consciousness has the potential to provide novel insights into psychological disorders [6].

Our emotions define our lives and our well-being. So long as we misconstrue what they are, we will be hampered in our efforts to use research to find more effective ways to relieve emotional suffering. In the end, our understanding emotion in the brain is only as good as our conceptualization of what an emotion is. If we don't know what we are looking for, we will surely fail to find it, and will continue to mislead others in the process.

REFERENCES

1. Adolphs, R., Mlodinow, L., and Barrett, L.F. (2019). What is an emotion? *Curr. Biol.* 29, R1060–R1064.
2. LeDoux, J.E. (2020). My Word: How does the non-conscious become conscious? *Curr. Biol.* 30, R1–R4.
3. Darwin, C. (1872). *The expression of the emotions in man and animals*, (London: Fontana Press).
4. LeDoux, J. (2012). Rethinking the emotional brain. *Neuron* 73, 653–676.
5. LeDoux, J. (2019). *The Deep History of Ourselves: The four-billion-year story of how we got conscious brains*, (New York: Viking).
6. LeDoux, J.E., and Brown, R. (2017). A higher-order theory of emotional consciousness. *Proc. Natl. Acad. Sci. USA* 114, E2016–E2025.
7. Perusini, J.N. and Fanselow, M.S. (2015) Neurobehavioral perspectives on the distinction

- between fear and anxiety. *Learn. Mem.* **22**, 417–25.
8. Taschereau-Dumouchel, V., Kawato, M., and Lau, H. (2019). Multivoxel pattern analysis reveals dissociations between subjective fear and its physiological correlates. *Mol. Psych.* **Oct 29**. doi: 10.1038/s41380-019-0520-3. [Epub ahead of print].
 9. Barrett L.F. and Russell J.A. (eds.) (2015) *The Psychological Construction of Emotion*. (New York: Guilford Press).
 10. Clore, G.L., and Ortony, A. (2013). Psychological construction in the OCC model of emotion. *Emot. Rev.* **5**, 335–343.
 11. Scherer, K.R. (2019). Studying appraisal-driven emotion processes: taking stock and moving to the future. *Cogn. Emot.* **33**, 31–40.
 12. Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends Cogn. Sci.* **4**, 417–423.
 13. Miller, E.K. (2013). The 'working' of working memory. *Dialogues Clin. Neurosci.* **15**, 411–418.
 14. Dehaene, S., Lau, H., and Kouider, S. (2017). What is consciousness, and could machines have it? *Science* **358**, 486–492.
 15. Lau, H., and Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends Cogn. Sci.* **15**, 365–373.
 16. Gilboa, A., and Marlatte, H. (2017). Neurobiology of schemas and schema-mediated memory. *Trends Cogn. Sci.* **21**, 618–631.
 17. Koechlin, E., and Hyafil, A. (2007). Anterior prefrontal function and the limits of human decision-making. *Science* **318**, 594–598.
 18. Burgess, P.W., Dumontheil, I., and Gilbert, S.J. (2007). The gateway hypothesis of rostral prefrontal cortex (area 10) function. *Trends Cogn. Sci.* **11**, 290–298.
 19. Bramson, B., Folloni, D., Verhagen, L., Hartogsveld, B., Mars, R.B., Toni, I., and Roelofs, K. (2020). Human lateral Frontal Pole contributes to control over emotional approach-avoidance actions. *J. Neurosci.* Feb 6. <https://doi.org/10.1523/JNEUROSCI.2048-19.2020>. [Epub ahead of print].
 20. Tulving, E. (2005). Episodic memory and autonoesis: Uniquely human? In *The Missing Link in Cognition*, H.S. Terrace and J. Metcalfe, eds. (New York: Oxford University Press), pp. 4–56.
 21. Koch, S.B.J., Mars, R.B., Toni, I., and Roelofs, K. (2018). Emotional control, reappraised. *Neurosci. Biobehav. Rev.* **95**, 528–534.
 22. Lau, H. (2019). Consciousness, Metacognition, & Perceptual Reality Monitoring. <https://doi.org/10.31234/osf.io/ckbyf>
 23. Jackson, J.C., Watts, J., Henry, T.R., List, J.M., Forkel, R., Mucha, P.J., Greenhill, S.J., Gray, R.D., and Lindquist, K.A. (2019). Emotion semantics show both cultural variation and universal structure. *Science* **366**, 1517–1522.
 24. Markus, H.R., and Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychol. Rev.* **98**, 224–253.
 25. Hyman, S.E. (2012). Revolution stalled. *Sci. Transl. Med.* **4**, 155cm111.
 26. LeDoux, J., Brown, R., Pine, D.S., and Hofmann, S.G. (2018). Know Thyself: Well-Being and Subjective Experience. In *Cerebrum*. (New York: The Dana Foundation). <https://www.dana.org/article/know-thyself-well-being-and-subjective-experience/>

¹Center for Neural Science and Department of Psychology, New York University, New York, NY 1003, USA. ²Emotional Brain Institute, Nathan Kline Institute, Orangeburg, NY 10962, USA. ³Department of Psychiatry, and Department of Child and Adolescent Psychiatry, New York University Langone Medical School, New York, NY 1003, USA. E-mail: ledoux@cns.nyu.edu.