

My Word

How does the non-conscious become conscious?

Joseph E. LeDoux

I have been thinking about consciousness for a long time. It all started when I was a graduate student in the 1970s working with Mike Gazzaniga on split-brain patients [1]. We did studies in which we coaxed the verbally silent right hemisphere to produce behavioral responses (stand up, laugh, and so on), and then asked the more loquacious left hemisphere why he did what he did. The left hemisphere had no idea why the responses occurred, but sure had answers — he stood up because he needed to stretch, and he laughed because we were funny. These fascinating observations set our minds a reeling. Drawing on social psychological principles of the day, such as cognitive dissonance and self-attribution, we came up with an account of why the left hemisphere, without a moment of hesitation, repeatedly confabulated perfectly reasonable stories about why he (the person) did what he did. We concluded that this was not a fluke, some strange consequence of the surgery, but instead a normal feature of the human brain — that our sense of who we are is a story we tell ourselves and others. And our behavior is a key part of that story.

Contemplating brain systems that might be responsible for non-consciously controlled behaviors, emotion systems came to mind. That is why, after completing my PhD, I turned to studies of how brains detect and respond to danger [2], work that I did in rodents. This research has been both successful and gratifying, but throughout I continued to ponder consciousness, and especially how emotions, like fear, might, in humans, result from narrations about non-consciously controlled behaviors such as those I was studying in rats [2,3].

Apropos of this, over the years I kept a watchful eye on the emerging field of cognitive neuroscience, and especially on studies that began to use new

methods, such as functional imaging, to shed light on how consciousness arises from brain networks. Most of this work has focused on visual consciousness [4–6]. The good news, I believe, is that the progress achieved can also help us understand emotional consciousness, especially if some adjustments are made. I don't mean that there's anything wrong with the science of vision. It is one of the most rigorous and advanced areas of brain research. The problem is instead that most studies of visual consciousness are lacking as a means for understanding the neural basis of complex real-life conscious experiences, including real-life visual experiences, and especially emotional experiences.

Conscious and non-conscious perception

Let us start with a clarification. There are all sorts of ideas about the 'C' word that have nothing to do with what I am interested in here. I am not concerned with whether electrons, rocks, or computers are conscious, nor even with whether, and if so how, bees, birds, or cats are conscious. My focus is on how human conscious experience comes about. And one thing we know is that what you are conscious of at any moment emerges from non-conscious processes in your brain. How, then, does the non-conscious become conscious? That is where visual research comes in handy.

A number of studies of visual consciousness have compared brain activity in situations in which people can give a verbal report of what they are seeing (can report on what they are consciously experiencing), as opposed to when, because of subliminal stimulus presentation, they cannot. In both the conscious and non-conscious conditions, areas of visual cortex are active. But when the participants can report about their experience, areas of prefrontal cortex known to be involved in higher cognitive functions are also active [4–7].

Different theories of consciousness explain these results in different ways [4,5,7]. Early sensory theories propose that visual cortex alone is sufficient for the experience, with prefrontal cortex just providing a kind of cognitive access that allows a verbal report about the experience. Cognitive theories, on

the other hand, say that visual cortex creates non-conscious representations that have to undergo additional processing via prefrontal cortex for a conscious experience to occur.

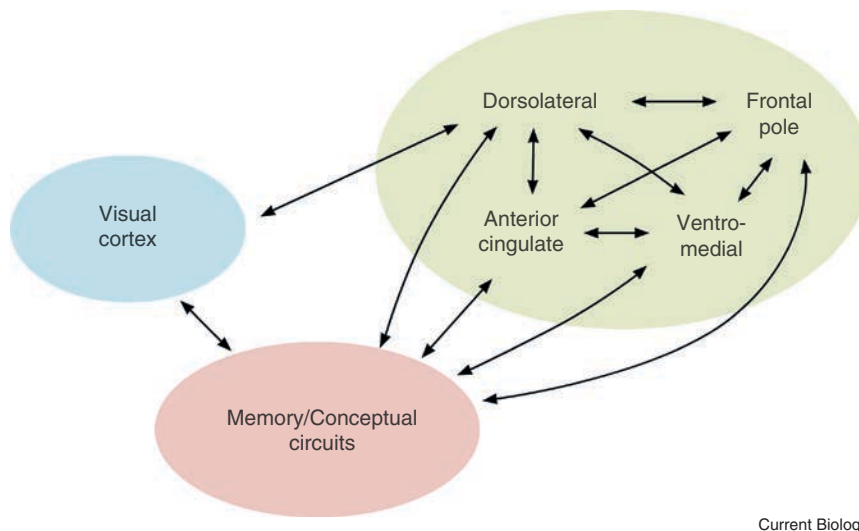
I believe that the evidence points to the need for the kind of additional processing that cognitive circuits enable. In particular, I favor the higher-order theory of consciousness, which assumes that prefrontal cortex actively re-represents the sensory cortex information and transforms the non-conscious sensory representation into a conscious experience [7–10].

Most research and scientific discussions about visual consciousness focus on the relation between visual cortex and one particular area of prefrontal cortex, the dorsolateral region. But I believe this view is too narrow. It might be sufficient to account for how highly artificial patterns of light forming dots or lines in laboratory studies are perceived, but it is lacking as an account of what goes on in real-life visual experiences. In particular, the role of memory in perception is missing. And once this limit is overcome, the way is paved for extending the model to many other kinds of experiences, including emotional experiences, since these are often triggered by visual or other sensory stimuli.

Memory and consciousness

Recognition of the identity of common objects requires more than just information about their visual properties. We don't innately know what a pencil, tuna salad sandwich, or a bicycle is. We have to learn what these are, and later use the memories we form to recognize them. For example, cherries and red marbles have some similar visual properties (both are reddish, roundish objects, often about the same size), but from experiences with them we come to know that they are distinct — one is a kind of fruit, and can be eaten, while the other is used in certain games, and is not edible. Memory is necessary to turn meaningless sensations into meaningful perceptions.

In light of this, it is notable that the dorsolateral prefrontal cortex not only receives visual and other sensory inputs, but also inputs from circuits that form and store long-term memories (Figure 1) [3,11], including both semantic and episodic memories



Current Biology

Figure 1. Visual and memory/conceptual inputs to prefrontal cortex.

Discussions of the neural mechanisms of consciousness often focus on the relation between visual cortex and dorsolateral prefrontal cortex. But memory/conceptual circuits (which include semantic and episodic memory circuits) should also be considered, since these receive visual inputs and connect with dorsolateral prefrontal cortex. Memory/conceptual circuits also connect with other prefrontal areas that, in turn, connect with dorsolateral prefrontal cortex (only some are shown). Visual inputs provide information about sensory properties of objects, while memory/conceptual inputs give the visual inputs meaning. Overall, inputs to the dorsolateral from memory circuits and other prefrontal areas outweigh inputs from visual cortex. Another area that may be important in consciousness is the frontal pole, which is interconnected with the dorsolateral area. It mainly receives inputs from memory/conceptual circuits and from other prefrontal areas, and it is considered the most conceptual of all brain areas. Of particular relevance is that the frontal pole has been implicated in self-introspection, and that it has a subregion, and cellular features, that are unique to the human brain. These anatomical considerations have important implications for understanding the non-conscious representations that contribute to conscious experiences. Specifically, as the non-conscious neural states that are immediately antecedent to conscious experience are, in principle, just one anatomical step removed from the conscious experience itself, such states, given the connections discussed, could include, or could simply be, prefrontal cortex states (see text for further discussion).

[12–14]. Semantic memory is about facts, such as the features and uses of objects. But as the cherry/marble example shows, object recognition requires knowledge both of what an object is and what it isn't. In other words, it also utilizes conceptual knowledge, a complex form of semantic memory [15]. For the most part, however, we experience and remember our lives as complex episodes rather than encounters with isolated stimuli. Episodes include facts and concepts, but in the context of personal experiences [12,13,16,17]. They are marked by what happened to you, and where and when it happened [16]. It is you, and only you, who knows what it is like to have the experiences you have.

The relation of memory circuits to prefrontal cortex

The circuits underlying each kind of memory have multiple components,

some of which directly connect with the dorsolateral region of prefrontal cortex, while others connect with it indirectly by way of other prefrontal areas, including the anterior cingulate, ventromedial, orbital, and other areas. Only some of these regions and the complex relations between them are shown in Figure 1.

Although the dorsolateral region receives much attention in terms of cognitive functions, another prefrontal area that is also important is the frontal pole, the forward-most area of the human brain [3,7,8,18,19]. It is said to engage in the highest levels of abstract conceptualization of any brain area. Not surprisingly, then, it has minimal if any sensory inputs, and instead is mainly connected with other prefrontal areas and with memory/conceptual circuits. Part of its allure is that it has unique properties only found in the human brain [18,19].

Sensory inputs to prefrontal cortex provide information about the physical properties of stimuli in the external world. The memory inputs function as schema that add conceptual and personal meaning. Together, they allow prefrontal cortex to build mental models that simulate what might be present, despite not having all of the visual details that are represented in visual cortex itself [3,8]. Such schema-based models are used in top-down control over the activities of lower-level sensory, memory, and other processors, not only predicting what a stimulus is, or might be, but also ascertaining how it relates to you, and how you might act, or not act, in its presence.

The inclusion of memory and conceptual understanding in the hierarchical interface between the sensory world and higher cognition provides a richer account of how sensations ultimately impact thoughts and actions. But because cognitive processing can be conscious or non-conscious, this still leaves us with the question of how non-conscious representations come to be consciously experienced.

Working memory and consciousness

The cognitive functions of prefrontal cortex are often discussed in terms of working memory, a mental workspace involved in the control of thought and action [20–22]. This differs from long-term memory in that it only temporarily stores information.

Working memory uses executive control functions, like attention, to select, monitor, temporarily maintain, and integrate diverse kinds of information from specialized processors. The specialized processors, which involve circuits underlying sensory, memory, verbal, and other functions, are mostly located in posterior cortical areas, each of which is reciprocally connected with prefrontal areas [3,11–14]. Executive functions, on the other hand, are mostly associated with prefrontal cortex itself [21] (but with parietal cortex also contributing).

Working memory is relevant here because the information it represents in-the-moment is generally said to be what we are conscious of [20,22,23]. The traditional view of this relationship is that when information enters working memory, we are conscious of it. But

more recently it has become apparent that not all information in working memory is consciously experienced. In other words, information can be represented in working memory and used in thought and action at a non-conscious level [23,24]. Some non-conscious aspects of working memory, like temporary storage of memory, involve various low-level (sensory) and intermediate-level (memory) processors that interact with the prefrontal network [3,25]. But prefrontal areas have also been implicated in non-conscious working memory [3,23,24].

The conscious aspects of working memory are said depend on what is called the episodic buffer [20], a hypothetical process that integrates information from lower-level processors to produce higher-level representations that give coherence to our conscious experiences. For example, when experiencing a musical performance, the music, musicians, and space are seamlessly integrated into the representations; you may focus on one instrument at times, but that does not completely isolate it from its context. Particularly relevant here is the suggestion that the cognitive functions attributed to the episodic buffer may depend on the prefrontal cortex, and especially the frontal pole [26]. Consistent with this are results implicating the frontal pole in the ability to introspect about one's self [27]. Prefrontal areas, especially the frontal pole and the dorsolateral area, are well suited to play important, if not crucial, roles in complex conceptions, and perhaps conscious experience itself [3,7,8].

Higher-order experience in light of non-conscious working memory

Higher-order theory, as noted, has traditionally assumed that the penultimate non-conscious state that is re-represented by prefrontal cortex is a sensory cortex state, and that the re-representation does something to make this lower-level state conscious [7–10]. But the thrust of the discussion above is that sensory states might not be the only, or even the main, lower-level states that are antecedent to conscious experiences.

As we have seen, dorsolateral prefrontal cortex and the frontal pole receive relatively more information

about sensory stimuli from hierarchically organized memory and conceptual processing circuits, and from other prefrontal areas that also receive memory and conceptual inputs, than from visual cortex. Memory-tinged prefrontal representations may constitute, or contribute to, non-conscious working memory and its ability to influence thought and behavior without conscious participation. But, in addition, non-conscious prefrontal representations that fuse sensation and schema-based memory to make visual stimuli meaningful might also constitute penultimate states that are antecedent to conscious experiences [3,8]. And in some situations, especially involving top-down mental modeling, sensory states may not be needed at all. Better understanding of the penultimate non-conscious state (or states) in various kinds of conscious experiences is crucial, as conscious experience is always preceded by such non-conscious events.

Clearly, challenges abound. I have emphasized certain prefrontal components over others for illustrative purposes, but much remains unknown. The prefrontal network is complex and dynamic. Its representations constantly change from moment to moment as the situation changes. A given conscious experience, depending on what it is an experience of (for example, a meaningless versus a threatening stimulus) likely involves somewhat different lower-order circuits, and perhaps even somewhat different components of the higher-order network. As a result of the redundancy of the inputs and the representations they engender, damage to any one member of the prefrontal team may simply shift the burden to its partners, which may, under some or even many circumstances, be able to carry on. For example, the frontal pole confers conceptual advantages, but except in the most taxing situations, may not be absolutely necessary. And in the absence of the entire prefrontal cortex, posterior conceptual circuits, perhaps with the aid of parietal executive functions, may be able to pull off some degree of higher-level representation sufficient to have some kind of conscious experience, especially of simple perceptual stimuli such as those often used in visual research.

Wrapping Up

The model presented is not meant as a grand solution to the problem of consciousness. Instead, it is a suggestion about how the explanatory scope of consciousness research might be enriched by recognizing the complex nature of the connectivity between sensory and memory circuits, and between these and higher cognitive circuits that mostly involve prefrontal cortex. A virtue of the model is that it is potentially applicable to any and all kinds of experiences, whether they involve external stimuli, body states, inner thoughts, or emotions. All are viewed as being rendered conscious through memory-informed conceptualizations that create non-conscious working memory states, which, in turn, are antecedent to conscious experiences.

I started this piece reminiscing about my graduate studies that implicated silent narrations in consciousness. Like schema, narratives are based on prior knowledge and depend on prefrontal cortex [28]. The framework presented here might therefore even provide a vantage point from which to approach the question of how narrations contribute to our sense of who we are, an issue that has kept me thinking about consciousness for more than four decades.

REFERENCES

1. Gazzaniga, M.S., and LeDoux, J.E. (1978). *The Integrated Mind*, (New York: Plenum).
2. LeDoux, J.E. (1996). *The Emotional Brain*, (New York: Simon and Schuster).
3. LeDoux, J. (2019). *The Deep History of Ourselves: The four-billion-year story of how we got conscious brains*, (New York: Viking).
4. Block, N. (2009). Comparing Theories of Consciousness, in Michael Gazzaniga (ed.) *The Cognitive Neurosciences IV*, MIT Press, 1111–1122.
5. Dehaene, S., Lau, H., and Kouider, S. (2017). What is consciousness, and could machines have it? *Science* 358, 486–492.
6. Shea, N., and Frith C.D. (2019). The global workspace needs metacognition. *Trends Cogn. Sci.* 23, 560–571.
7. Lau, H., and Rosenthal, D. (2011). Empirical support for higher-order theories of conscious awareness. *Trends Cogn. Sci.* 15, 365–373.
8. Brown, R., Lau, H., and LeDoux, J.E. (2019). Understanding the Higher-Order Approach to Consciousness. *Trends Cogn. Sci.* 23, 754–768.
9. Rosenthal, D.M. (2005). *Consciousness and mind*, (Oxford: Oxford University Press).
10. Cleeremans, A., Achoui, D., Beaury, A., Keuninckx, L., Martin, J.R., Munoz-Moldes, S., Vuillaume, L., and de Heering, A. (2019). Learning to be conscious. *Trends Cogn. Sci.* <https://doi.org/10.1016/j.tics.2019.11.011>. [Epub ahead of print].

11. Yeterian, E.H., Pandya, D.N., Tomaiuolo, F., and Petrides, M. (2012). The cortical connectivity of the prefrontal cortex in the monkey brain. *Cortex* 48, 58–81.
12. Beran, M.J., Menzel, C.R., Parrish, A.E., Perdue, B.M., Sayers, K., Smith, J.D., and Washburn, D.A. (2016). Primate cognition: attention, episodic memory, prospective memory, self-control, and metacognition as examples of cognitive control in nonhuman primates. *Wiley Interdiscip Rev Cogn Sci.* 7, 294–316.
13. Murray, E.A., Wise, S.P., and Graham, K.S. (2017). *The Evolution of Memory Systems: Ancestors, anatomy, and adaptations*, (Oxford: Oxford University Press).
14. Libby, L.A., Hannula D.E., Ranganath, C. (2014). Medial temporal lobe coding of item and spatial information during relational binding in working memory. *J. Neurosci.* 34, 14233–14242.
15. Binder, J.R., and Desai, R.H. (2011). The neurobiology of semantic memory. *Trends Cogn. Sci.* 15, 527–536.
16. Tulving, E. (1983). *Elements of episodic memory*, (New York: Oxford).
17. Schacter, D.L., Benoit, R.G., and Szpunar, K.K. (2017) Episodic future thinking: mechanisms and functions. *Curr. Opin. Behav. Sci.* 17, 41–50.
18. Neubert, F.X., Mars, R.B., Thomas, A.G., Sallet, J., and Rushworth, M.F. (2014). Comparison of human ventral frontal cortex areas for cognitive control and language with areas in monkey frontal cortex. *Neuron* 81, 700–713.
19. Semendeferi, K., Teffer, K., Buxhoeveden, D.P., Park, M.S., Bludau, S., Amunts, K., Travis, K., and Buckwalter, J. (2011). Spatial organization of neurons in the frontal pole sets humans apart from great apes. *Cereb. Cortex* 21, 1485–1497.
20. Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends Cogn Sci.* 4, 417–423.
21. Miller, E.K. (2013). The “working” of working memory. *Dialogues. Clin. Neurosci.* 15, 411–418.
22. Carruthers, P. (2015). *The Centered Mind: What the Science of Working Memory Shows Us About the Nature of Human Thought*. Oxford: Oxford University Press.
23. Persuh, M., LaRock, E., and Berger, J. (2018). Working memory and consciousness: The current state of play. *Front. Hum. Neurosci.* 12 <https://doi.org/10.3389/fnhum.2018.00078>.
24. Bergström, F., and Eriksson, J. (2018). Neural evidence for non-conscious working memory. *Cereb Cortex* 28, 3217–3228.
25. Postle, B.R. (2016). How does the brain keep information “in mind”? *Curr Dir Psychol Sci.* 25, 151–156.
26. Schacter, D.L., and Addis, D.R. (2007). On the constructive episodic simulation of past and future events. *Behav. Brain Sci.* 30, 331–332.
27. Fleming, S.M., Weil, R.S., Nagy, Z., Dolan, R.J., and Rees, G. (2010). Relating introspective accuracy to individual differences in brain structure. *Science* 329, 1541–1543.
28. Gilboa A., and Marlatte G. (2017) Neurobiology of schemas and schema-mediated memory. *Trends Cogn. Sci.* 21, 618-631..

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