Consciousness cannot be separated from function

Michael A. Cohen and Daniel C. Dennett

1 Vision Sciences Laboratory, Department of Psychology, William James Hall, Harvard University, Cambridge, MA, 02138, USA
2 Center for Cognitive Studies, Department of Philosophy, Tufts University, Medford, MA, 02155, USA

Numerous theories of consciousness hold that there are separate neural correlates of conscious experience and cognitive function, aligning with the assumption that there are ‘hard’ and ‘easy’ problems of consciousness. Here, we argue that any neurobiological theory based on an experience/function division cannot be empirically confirmed or falsified and is thus outside the scope of science. A ‘perfect experiment’ illustrates this point, highlighting the unbreachable boundaries of the scientific study of consciousness. We describe a more nuanced notion of cognitive access that captures personal experience without positing the existence of inaccessible conscious states. Finally, we discuss the criteria necessary for forming and testing a falsifiable theory of consciousness.

The hard problem of consciousness is an impossible problem

A goal of neuroscience is to locate the neural correlates of consciousness: the minimal set of neuronal events leading to subjective awareness (see Glossary) [1–3]. Numerous influential theories hold that conscious experience has its own neural underpinnings that can be separated from all cognitive functions (i.e. attention, working memory, language, decision making, motivation etc.). Different theories equate consciousness with different correlates: recurrent activation between cortical areas [4–7], coalitions of ‘winning’ neurons [8–10], special microactivations distributed throughout the brain [11–13] or activity in the ventral stream [14]. Although the details of these theories vary, they all assert that conscious experience and cognitive functions have distinct neural correlates (Box 1).

This alleged division between experience and function is often mapped onto the distinction between the ‘hard’ and ‘easy’ problems of consciousness [3]. Under this view, the hard problem is answering the question of how phenomenal experience arises from physical events in the brain, whereas the easy problems are characterizing the mechanisms supporting cognitive functions. In this article we argue that, from an empirical perspective, the ‘hard problem’ is actually an impossible problem that inherently isolates consciousness from all current and future avenues of scientific investigation. All theories of consciousness based on the assumption that there are hard and easy problems can never be verified or falsified because it is the products of cognitive functions (i.e. verbal report, button pressing etc.) that allow consciousness to be empirically studied at all. A proper neurobiological theory of consciousness must utilize these functions in order to accurately identify which particular neural activations correlate with conscious awareness.

A motivation behind dissociative theories is the belief that theories associating awareness with access [15–17] cannot explain the richness of phenomenology. In other words, it is claimed that ‘phenomenology overflows access’ ([17], p. 487): we experience more than can possibly be captured by cognitive mechanisms that are known to have strict limits. Visual attention [18,19], working memory [20,21], dynamic tracking [22,23] and many other such processes have well-established capacity limits. Phenomenology, however, is claimed to have no such limitations. It is thought that when we look out onto the world we do not only see a few attended items; we see the whole world. Thus it is argued that although we are conscious of a variety of inputs we have access to only a small subset of these experiences [4,7,10,13].

Here, we analyze the data used to support the claim that phenomenology overflows access and show how these results can be accounted for under a pure access/functional view of consciousness. We then argue that dissociative theories are inherently unfalsifiable and beyond the scope of science, because inaccessible conscious states are intrinsically off-limits to investigation. With this in mind we end by describing the necessary components of a proper scientific theory of consciousness.

Evidence supporting the dissociation

What data support the view that consciousness occurs independently of, and can be experimentally dissociated

Glossary

Access consciousness: conscious states that can be reported by virtue of high-level cognitive functions such as memory, attention and decision making.
Awareness: the state of perceiving, feeling or experiencing sensations.
Easy problem of consciousness: understanding the mechanisms that support relevant functions such as language and attention.
Hard problem of consciousness: explaining phenomenal consciousness (e.g. the feeling of ‘what it is like’ [80]).
High versus low-level brain regions: in this context, the distinction between high and low-level brain regions roughly correspond to sensory and non-sensory (functional) regions. More specifically, ‘low-level’ regions are involved in the processing and discrimination of visual stimuli, whereas ‘high-level’ regions are involved in attention, language, and decision-making.
Phenomenal consciousness: the subjective aspect of experiencing the world (e.g. the experience of seeing the color red).
Box 1. Example dissociative theories

The partitioning of conscious experience from cognitive function is common in neurobiological theories of consciousness. Three representative theories are described below.

**Local recurrency.** The best-known theory that embraces the separation between experience and function is the local recurrency theory put forth by Lamme [4,5] and Block [6,7]. According to this theory, visual information is processed in the cortex by an initial feedforward sweep in which representations of motion, color and shape are formed [61,62]. Although representations at this stage can be rather detailed, no conscious experience accompanies this processing. Such experiences only arise as a result of sustained RP between visual areas. However, the experiences that accompany RP are independent of all cognitive functions, especially attention [4]. Indeed, this theory explicitly maintains that local recurrency is the neural correlate of one and only one form of consciousness: phenomenal consciousness [28]. Access consciousness, which comprises functions such as working memory, language production and so on, is achieved when RP extends into the frontal cortex and engages higher-level functions.

**Microconsciousness.** Zeki’s theory of microconsciousness states that consciousness is not a unified state but is instead distributed in space and time [11-13]. Rather than emphasizing the flow of information between regions, like the local recurrency theory, this theory focuses on the activation of ‘essential nodes’ throughout the cortex. Each node represents different bits of information (e.g. color or motion) and the activation of each node generates its own microconsciousness. We have the impression of a unified consciousness because each of these individual representations is bound to others, post-experimentally, to form an accessed macroconsciousness [13]. It is at this macrolevel that functions such as language and decision making operate on the distributed experiences and lead to subjective reports. Thus, the micro/macro distinction again dissociates conscious experience and cognitive function [29].

**Coalitions of neurons.** Crick and Koch proposed that consciousness stems from ‘winning’ coalitions of neurons (sustained activation of a collection of neurons that are dedicated to the processing and representation of a particular stimulus or event) [8]. Under this view, coalitions supporting one representation compete with coalitions supporting other representations [9]. Only after a winning coalition becomes conscious can attention be diverted to it. Occasionally, only one coalition ‘wins’ at a time, leading to a relatively tight correlation between consciousness and attention. However, this correlation is not perfect. Koch and colleagues have written extensively about the existence of consciousness without attention [63,64,69,70], recently claiming that consciousness without attention is a form of phenomenal consciousness as described by Block and Lamme [10].

All of these theories have distinctive strengths, and some plausibility, but they also share a fundamental flaw: they posit the existence of conscious states that even the individual him or herself does not realize he or she is having. Highlighting this flaw might provide impetus for revision and improvement of these theories: rejecting the one shared feature of them all and leaving the other features to be sorted out empirically.

---

Figure 1. How much is consciously perceived at a given psychological moment? Those who argue for dissociating consciousness and function do so because they claim that awareness overflows conscious access. (a) In this case, the identities of all the letters on the screen are perceived. (b) However, McConkey and Rayner have shown that when uniform xs replace the nonfixated words of text, participants do not realize this has happened [65,66]. (c) If there is no other text on the screen besides the fixated word, then participants will notice this instantly. This elegantly demonstrates that although people are aware of the ‘presence’ of nonattended items in this case, they are actually not aware of the ‘identities’ of those items.

---

(a) Normal text  
(b) Multi-access  
(c) Strict access

<table>
<thead>
<tr>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>While reading this, you’re not conscious of most of the letters.</td>
<td>While xxxxxx xxxxxx xxxxxx, you’re not conscious of most of the letters.</td>
<td>While xxxxxx xxxxxx, you’re not conscious of most of the letters.</td>
</tr>
<tr>
<td>While reading this, you’re not conscious of most of the letters.</td>
<td>While xxxxxx xxxxxx xxxxxx, you’re not conscious of most of the letters.</td>
<td>While xxxxxx xxxxxx, you’re not conscious of most of the letters.</td>
</tr>
</tbody>
</table>

= fixation location

---

TRENDS in Cognitive Sciences
a more vivid, detailed world than can be reported. The world beyond focused attention is not in total ‘darkness’: when staring intently at a single item, one is still aware of some aspects of the scene around it [30,31]. Such a claim is obviously true. Dissociative theorists cite this fact as the primary example of phenomenology overflowing access. However, this is not a problem for theories that identify consciousness with function.

The world beyond focal attention is not in darkness because when attention is not entirely engaged by a primary task, and it is unclear if attention can ever be entirely engaged using psychophysical techniques, excess attentional resources are automatically deployed elsewhere [32–35]. Thus, certain items are processed through focal attention, whereas others are processed via distributed attention [36,37]. Focal attention often leads to high resolution percepts whereas the percepts from distributed attention are at a lower resolution but with certain basic elements preserved [36–48] (Figure 2). It is inaccurate to say that information outside the focus of attention receives zero attention. Information not processed by focal attention can nevertheless be the target of other types of attention: distributed, featural, spatial, internal and so on [49].

Is this degraded visual information an example of phenomenology overflowing access? Indeed, the degraded information is consciously perceived. However, the function supporting this perception is simply distributed, rather than focal, attention. In fact, when attention is engaged in a sufficiently difficult task, observers can fail to perceive even coarse and degraded information, such as the gist of a scene, because of inattentive blindness [50]. This information is undoubtedly accessed because observers explicitly report seeing more than what is focally attended (the idea that such information is indeed accessed has been recognized by Block, see [7], p. 487).

Once it is recognized that distributed attention leads to degraded but accessed percepts, the motivation for claiming that this degraded information is an example of inaccessible conscious states disappears. The world beyond focused attention is not in darkness because there are functional resources (in this case, multiple forms of attention) dedicated to processing that information (Figure 1c).

Is there more to phenomenology?

Dissociative theories claim that there is phenomenology over and above the accessed information previously described. However, various empirical results cast doubt upon this claim. In a modified version of the Sperling paradigm, where letters are sometimes unexpectedly replaced with pseudo-letters, participants still claim to see only letters [51]. Another example of this phenomenon can be seen in Figure 2. When participants are instructed to fixate at the center of a screen, two images can be successively presented in the same location, with a blank image briefly separating the two, and the drastic changes between the images go unnoticed (a phenomenon known as change blindness). If participants are conscious of the identities of all elements in the scene, as has been repeatedly claimed by dissociative theorists, then participants should instantly notice the pseudo-letters or the scrambled image. The fact that they do not suggests that participants are overestimating the contents of their own experience.

Even though people do not notice these changes, the illusion of seeing more still needs to be explained. Why is it that people overestimate the richness of their conscious perceptions [52]? The nature of this illusory experience still needs to be explained and should be the focus of future empirical work. Functionalist accounts can study this by varying the prior expectations and confidence levels of participants in a variety of paradigms [27,51]. Dissociative theories, meanwhile, ‘explain’ this illusion by relying on
inaccessible conscious states that, as the next section describes, inherently prevent the possibility of confirmation or falsification.

**The perfect experiment**

Currently, no experimental results uniquely support the existence of consciousness independent of function and access. Could future experiments accomplish this? We argue that all theories of consciousness that are not based on functions and access [4–14] are not scientific theories. Consider perhaps the most drastic experiment possible, the ‘perfect’ experiment: imagine that, in the future, surgeons are able to isolate the parts of the visual cortex that represent color while wholly preserving their activation patterns. After this surgery, the areas involved in color perception (visual area V4, inferotemporal cortex etc.) behave normally but are simply unable to project to higher brain areas [53–57]: perfect isolation. Although the color areas are isolated, all other visual areas (e.g. motion, luminance, object recognition etc.) are untouched and project to higher-level regions in a normal manner (Figure 3). Such a clean separation of one aspect, color, of visual perception is profoundly unrealistic but this idealization provides a simplification that is revealing of the key flaw in theories that dissociate function from consciousness.

According to all the theories discussed above, or possible theories based on the experience/function divide, whatever is necessary for color consciousness will be preserved in these color areas. If these theories are mutually exclusive, then we can imagine a different participant for each particular theory. All that matters is that we do not allow these isolated areas of a supposed type of phenomenal consciousness to interact with other cognitive functions.

When shown a colored apple what will our hypothetical participants say? They will surely not say that they see any colors because the areas responsible for processing color have been isolated from higher-level areas, including language production. They will be able to identify the object as an apple because visual areas responsible for all other aspects of visual cognition are intact and connected to these higher-level regions. Thus, they are simply color-blind. We can imagine them saying, ‘I know you say my color areas are activated in a unique way, and I know you believe this means I am consciously experiencing color but I’m looking at the apple, I’m focused on it, and I’m just not having any experience of color whatsoever’ (Box 2).

Moreover, imagine that, before the surgery, that particular shade of red would reliably agitate or excite the patient. Would the patient have such feelings now and say something like, ‘I don’t see red but I notice that I’ve gotten a little tense’? As described here, the patient would not because such affective, emotional or ‘limbic’ reactions are themselves the types of functions that we are isolating from the color area. To be excited or calmed or distracted by a perceptual state of red discrimination is already to have functional access to that state, however coarse-grained or

---

**Box 2. What if we gave the isolated color area the ability to communicate?**

Is it possible that even though the subject is not conscious of red, the isolated color area itself is experiencing color (similar to the way the right hemisphere of split-brain patients is often described) [11]? What would happen if we supplied a reporting mechanism for the isolated color area?

Imagine the reporting mechanism is nothing more than the simple hardware needed to actually transmit a message (e.g. a speaker). If this device were connected to the color area, then it seems clear that there would be no reports of color consciousness. The cognitive functions needed to select a particular thought, decide how best to describe it, and to execute that action are still absent, preventing any type of response from being formed or conveyed.

Whereas if the color area were connected to a more sophisticated reporting mechanism that was endowed with these functions there would probably be reports of color consciousness. However, this is not because the color area is experiencing its own isolated consciousness; rather, it is because the color area is now connected to the functions that are crucial for consciousness. By connecting the color area to a mechanism endowed with the relevant functions, the previously unconscious color information can now be accessed by a broader cognitive system.
incomplete, because such a reaction can obviously affect decision making or motivation (Box 3).

In spite of this frank denial by subjects, theories that posit dissociation between consciousness and function would necessarily assume that participants of the ‘perfect experiment’ are conscious of the apple’s color but simply cannot access that experience. After all, the conditions these theories stipulate for phenomenal consciousness of color are all met, so this experiment does not disprove the existence of isolated consciousness; it merely provides another particularly crisp example of consciousness without access.

However, there is a crucial problem with this logic. If this ‘perfect experiment’ could not definitively disprove dissociative theories, then what could? The subject manifests all the functional criteria for not being conscious of color so what would ground the claim that the subject nevertheless enjoys a special kind of consciousness: phenomenal consciousness without access consciousness (Box 2)?

The domain of a science of consciousness
What the perfect experiment demonstrates is that science necessarily relies on cognitive functions in order to investigate consciousness. Without input from subjects, input that is the product of such functions, theorists are left to define consciousness based on certain types of activation that are independent of a subject’s own experience. It has been claimed that separating consciousness from other cognitive functions is required because it ‘is a prerequisite for using the term [consciousness] at all’ ([15], p. 500).

What does it mean to study consciousness without function? Inevitably, theories motivated by this view will define consciousness in their own way (local recurrence, microconsciousness, coalitions of neurons, etc.) and say that whenever that criterion is met, consciousness must occur. But how do we set this criterion?

For example, what reason is there to think that local recurrence is conscious experience? Could local recurrence simply be a form of unconscious processing? It cannot be based on subjective reports because these reports are the direct result of cognitive functions. When an observer says, ‘But in the Sperling display I don’t just see a few letters on the screen, I see all the letters,’ there is no reason to believe that such an experience occurs independent of function.

The fact that the observer is reporting on this visual experience proves that the experience has been accessed by the broader cognitive system as a whole. Lamme writes, ‘You cannot know whether you have a conscious experience without resorting to cognitive functions such as attention, memory or inner speech’ ([15], p. 499). If this is true, then what reason is there to think this particular type of activation should be classified as correlating with conscious experience? What does it mean to have a conscious experience that you yourself do not realize you are having? In the face of such clear grounds for doubting such a conscious experience, dissociative theories need to provide a reason for claiming that these isolated types of activation involve any kind of consciousness.

The future of scientific theories of consciousness
It is clear, then, that proper scientific theories of consciousness are those that specify which functions are necessary for consciousness to arise. A true scientific theory will say how functions such as attention, working memory and decision making interact and come together to form a conscious experience. Any such theory will need to have clear and testable predictions that can in principle be verified or falsified. Most importantly, such theories will not claim that consciousness is a unique brain state that occurs independently of function; instead, the focus will be placed on the functions themselves and how they interact and come together to form consciousness.

There are several theorists who have already realized the need for functions in developing theories of consciousness. Dehaene and colleagues [16] have put forth a global neuronal workspace model that claims consciousness is defined by the orientation of top-down attention, long-distance feedback loops that extend into parietofrontal networks, and conscious reportability. Similarly, Kouider and colleagues [27] have discussed at great length how information that is in consciousness relies on a hierarchy of representational levels. Under this view, each level corresponds to different cognitive mechanisms responsible for different units of representation.

It is important to stress that both of these theories are merely the beginning, rather than the end, of the study of consciousness. There is still much work to be done in regards to how these functions and mechanisms interact. In Dehaene et al.’s theory, for example, a more thorough and specific understanding of the type of parietofrontal activation [16] and how it relates to the formation of memories and decisions is still necessary. The upshot of function-based theories is that they make claims about consciousness that can be tested and examined scientifically.

Although there are certainly those who disagree with the specifics of the theories put forth by Dehaene et al. and Kouider et al. [4–14], these are disagreements that can eventually be settled through more rigorous examination and testing. The same cannot be said of theories that maintain that consciousness occurs independent of
function. As the perfect experiment illustrates, such theories inherently prevent any future avenue for scientific research.

**Concluding remarks**

Understanding the necessary relation between function and experience reveals that the so-called hard problem of consciousness should be reclassified. Far from being a formidable obstacle to science, it achieves its apparent hardness by being systematically outside of science, not only today's science but any science of the future that insists on dissociating consciousness from the set of phenomena that alone could shed light on it. This is not to suggest that consciousness is a mystery that the human mind cannot comprehend [58]. It is simply that whatever mysteries and puzzles might continue to baffle us, we should not cripple our attempts at understanding by adopting a concept of consciousness that systematically blocks all avenues of further research.

The issues raised here generalize beyond the specific theories discussed [4–14]. Any theory wherein the neural correlates of conscious experience are separate from the neural correlates of cognitive function is ultimately doomed. No matter the specifics of the theory – C-fibers firing, grandmother cells, winning coalitions, microconsciousness, recurrent processing (RP) and so on – it is always possible in principle to isolate this activation. Such imagined isolation, however, actually removes the experience in question from further testing, scrutiny and verification. Although these theories might provide considerable insight into the formation of internal representations of the sensory and perceptual world, that is not enough to explain one’s personal awareness. A proper theory of consciousness cannot exclusively focus on how the brain forms and maintains representations. Such a theory must also explain in functional terms how those representations are experienced and accessed by the multiple functions constituting an observer [59]. Theories that do not acknowledge this are fundamentally incapable of explaining the full scope of consciousness.

**Acknowledgments**

This work was supported by a National Science Foundation Graduate Research Fellowship (M.A.C.). Special thanks to Justin Junge for extensive discussion and comments on the project. Thanks to Arash Afraz and Maryam Vaziri Pashkam for helpful discussions, and to Ray Jackendoff, Sid Kouider, Ken Nakayama, Jordan Suchow, and two anonymous reviewers for comments on the manuscript. Thanks to Jeremy Freeman for providing the images for Figure 2.

**References**

43 Halberda, J. et al. (2006) Multiple spatially overlapping sets can be enumerated in parallel. Psychol. Sci. 17, 572–576
56 Super, H. et al. (2001) Two distinct modes of sensory processing observed in monkey primary visual cortex (V1). Nat. Neurosci. 4, 304–310