

Richard Brown, Joseph LeDoux, and David Rosenthal
“The Extra Ingredient”
Forthcoming in *Biology and Philosophy*

Birch et al (2020) propose an Unlimited Associative Model (UAL) as a transition marker for the appearance of consciousness in the animal kingdom. Their approach is explicitly modeled on the way in which origin-of-life science proceeds. Their strategy is to find a list of features that are jointly sufficient for the presence of conscious experience, and then to look for some conditions that would be required in order for the features on that list to occur. They propose that the following features are jointly sufficient for conscious experience:

First, there is *global accessibility and broadcast*: a minimal global workspace (in the sense of Dehaene [2014](#)) where information from perception, memory and evaluative systems is integrated and broadcast back to these and other systems. Second, there is *binding/unification and differentiation*: objects (e.g. a blue box) are perceived, not just fragmented features (boxness and blueness). Third, there is *selective attention and exclusion*: there are mechanisms for making some stimuli more salient than others. Fourth, there is *intentionality*: the capacity to represent the world and one’s own body. Fifth, there is *integration of information over time*, not just at a single time. Sixth, there is an *evaluative system*. Seventh, there is *agency and embodiment*. Eighth, and finally, there is *registration of a self/other distinction*.

Birch et al argue that UAL can serve as a marker indicating the presence of these features, so that questions about the origin of consciousness can be translated into questions about the origin of UAL. But there are other accounts of the origin of consciousness, such as higher-order theory (e.g., LeDoux 2019; Rosenthal 2005, chs. 7, 10; and others), which Birch et al simply ignore.

And unfortunately, Birch et al also offer no argument that their proposed features do jointly suffice for the presence of conscious experience, beyond saying that many theorists will accept it. That’s possible. The list is in effect a compilation of features advanced by various theories as being essential to consciousness; so many theorists will find something to like. But without evaluating the claims those different theories make about their favorite feature, the list remains unconvincing.

Birch et al are clear that by sufficiency they mean to rule out the relevance of philosophical zombies -i.e., physical duplicates that lack consciousness. But they give no reason to think that the joint realization of all of their features could not occur without any conscious experience at all. And there are reasons, considered below, to think that the features they list either have no connection with consciousness or at most occur only in connection with somewhat special types of conscious experience.

Birch et. al. see their model as incompatible with higher-order-thought (HOT) theories of consciousness, on which a state is conscious if one is in some suitable way aware of

that state. They see higher-order (HO) awareness as an “extra ingredient”. But since Birch et al go on to say that “[t]his is not the place for a detailed discussion of HOT theories,” they don’t address why they take HO awareness to be an extra ingredient or why HOT theorists are convinced that it’s needed.

So a few words about that. It’s central to any understanding of what it is for a psychological state to be conscious that we know how conscious mental states differ from subliminal psychological states that occur without being conscious. When one perceives something subliminally, one is wholly unaware of perceiving that thing, and will sincerely deny doing so. It follows that if a state does occur consciously, one is in some suitable way aware of that state.

This HO awareness is central to our commonsense understanding of what it is for a psychological state to be conscious. And it is central also to all experimental work on consciousness; researchers rely on distinguishing conscious states from subliminal and other unconscious states by whether subjects report being in the state in question. And a verbal report is an indication that the subject is aware of the state; without such awareness, the subject would be unable to report. Indeed, Birch et al in effect recognize this when they write that “it is only in humans that we can independently verify, through verbal report, that a stimulus was consciously experienced.”

So the HO awareness that HOT theory posits as accompanying every conscious state is by no means an extra ingredient. And HO theorists have no difficulty describing why that posit is crucial. The HO awareness is needed to explain how conscious states differ from psychological states that aren’t conscious (Brown, Lau, and LeDoux 2019; Lau and Rosenthal 2011; Rosenthal 2000).

Birch et al write “that the neural signatures of [conscious] experience, whatever they turn out to be, will also be correlated with UAL.” Perhaps. But it’s not enough to look for the neural signatures of conscious experience if we can’t independently distinguish conscious from nonconscious psychological states. Moreover, we can’t simply assume that we know which states are likely to be conscious in nonhuman animals, or indeed whether any are. We need a reliable way to distinguish conscious from unconscious psychological states in cases we understand. At present, those cases are all human. Once we get a reliable way to differentiate for humans conscious from unconscious psychological states, we can propose various ways that other creatures ‘might’ be conscious. But we will not have the same degree of certainty that we have from human verbal report. For this reason it is currently altogether premature to be talking about transition markers.

Let’s turn then to the list of features Birch et al see as jointly sufficient for conscious experience. The first, global accessibility and broadcast (Dehaene 2014), is altogether independent of conscious experience. Peripheral vision is often conscious, but lacks global accessibility, and subliminal perceptions can have striking effects on downstream psychology processing. The second, binding, is of course typical of much conscious experience, but there is every reason to think it occurs with most unconscious

perceiving as well (e.g. binding of motion and contrast are part of the perception of a moving visual object but until those features are bound to object memory a conscious perception of what the moving object is will not occur). As for their third feature, there is now an extensive literature on the occurrence of selective object attention that occurs unconsciously (e.g., Norman et al 2013). Their fourth feature, representing nonmental reality, also plainly occurs unconsciously, as does agency (their seventh) and a distinction between self and other (which can occur as an immune response, for example) (eighth). And though human conscious experience typically integrates over time (fifth) and might often have evaluative implications (sixth), there is no reason to think either has any necessary connection with all conscious experience (the cerebellum unconsciously integrates motor signals over time, and valuation is present at the receptor levels for innately coded stimuli--tissue damage, for example).

Might each of the eight features occur without consciousness but still not jointly? That's not impossible, but Birch et al give no reason to think it's so. And because the features have almost nothing to do with one another, apart from apparently being pulled from various currently popular theories of consciousness, it's unlikely that they would work together or that their joint occurrence would be sufficient for conscious experience. Because of this we expect that the predictions discussed by Birch et al will go unconfirmed. And though the jury is still out there is some work that suggests that this may be so (Pessiglione M, et al 2008; Schultz 2013; Balderston et al 2014; Hopkins et al 2015)

Given the failure of the list of features to differentiate conscious from unconscious psychological states, the need for an appeal to the HO awareness posited by HO theories is all the more pressing. It is by no means an "extra ingredient."

WORK CITED

Balderston NL, Schultz DH, Baillet S, Helmstetter FJ. Rapid amygdala responses during trace fear conditioning without awareness. *PLoS One*. 2014 May 13;9(5):e96803. doi: 10.1371/journal.pone.0096803. PMID: 24823365; PMCID: PMC4019542.

Birch J, Ginsbur, S and Jablonka E. Unlimited Associative Learning and the origins of consciousness: a primer and some predictions. *Biol Philos* **35**, 56 (2020). <https://doi.org/10.1007/s10539-020-09772-0>

Brown R, Lau H, LeDoux JE (2019) Understanding the Higher-Order Approach to Consciousness Trends in Cognitive Sciences, Month 2019, Vol. xx

Dehaene, S and Naccache L (2001), "Towards a Cognitive Neuroscience of Consciousness: Basic Evidence and a Workspace Framework," *Cognition* **79**, 1-2 (April): 1-37.

Dehaene S (2014) *Consciousness and the brain: deciphering how the brain codes our thoughts*. Penguin Books, New York

Hopkins LS, Schultz DH, Hannula DE, Helmstetter FJ. Eye Movements Index Implicit Memory Expression in Fear Conditioning. *PLoS One*. 2015 Nov 12;10(11):e0141949. doi: 10.1371/journal.pone.0141949. PMID: 26562298; PMCID: PMC4642991.

Lau, Hakwan, and David Rosenthal (2011), "Empirical Support for Higher-Order Theories of Conscious Awareness," with Hakwan Lau, *Trends in Cognitive Sciences*, 15, 8 (August): 365-373.

LeDoux JE (2019) *The Deep History of Ourselves: The Four-Billion-Year Story of How We Got Conscious Brain*. New York, Viking.

Norman, Liam J., Charles A. Heywood, and Robert W. Kentridge (2013), "Object-based Attention without Awareness," *Psychological Science*, 24, 6 (June): 837– 843.

Pessiglione M, Petrovic P, Daunizeau J, Palminteri S, Dolan RJ, Frith CD. Subliminal instrumental conditioning demonstrated in the human brain. *Neuron*. 2008 Aug 28;59(4):561-7. doi: 10.1016/j.neuron.2008.07.005. PMID: 18760693; PMCID: PMC2572733.

Rosenthal, D (2005) *Consciousness and Mind*, Oxford University Press

Schultz DH, Balderston NL, Geiger JA, Helmstetter FJ. Dissociation between implicit and explicit responses in postconditioning UCS revaluation after fear conditioning in humans. *Behavioral Neuroscience*. 127: 357-68. PMID 23731073 DOI: 10.1037/a0032742