The learning-consciousness connection

Jonathan Birch[1] · Simona Ginsburg[2] · Eva Jablonka[1,3]

Abstract

This is a response to the nine commentaries on our target article "Unlimited Associative Learning: A primer and some predictions". Our responses are organized by theme rather than by author. We present a minimal functional architecture for Unlimited Associative Learning (UAL) that aims to tie to together the list of capacities presented in the target article. We explain why we discount higher-order thought (HOT) theories of consciousness. We respond to the criticism that we have overplayed the importance of learning and underplayed the importance of spatial modelling. We decline the invitation to add a negative marker to our proposed positive marker so as to rule out consciousness in plants, but we nonetheless maintain that there is no positive evidence of consciousness in plants. We close by discussing how UAL relates to development and to its material substrate.

Replies to commentaries on "Unlimited Associative Learning: A primer and some predictions".

✉ Jonathan Birch
  j.birch2@lse.ac.uk

[1] Centre for Philosophy of Natural and Social Science, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, UK

[2] Natural Science Department, The Open University of Israel, 1 University Road, 4353701 Raanana, Israel

[3] The Cohn Institute for the History and Philosophy of Science and Ideas, Tel Aviv University, 6934525 Ramat Aviv, Israel

This is a response to the nine commentaries on our target article "Unlimited Associative Learning: A primer and some predictions" (Birch et al. 2020). We are grateful to the commentators for the critical dialogue they have opened up, for giving us an opportunity to clarify controversial and ambiguous issues, and for pointing out the limitations and potential extensions of the UAL framework.

Our responses will be organized by theme rather than by author. At the heart of the target article and the commentators' challenges is the relationship between UAL and phenomenal consciousness: the elusive property of there being "something it's like" to undergo a cognitive or perceptual process. Several commentators believe that UAL is not a good marker of phenomenal consciousness, suspecting instead that UAL may be a marker of something else, such as perceptual awareness understood in non-phenomenal terms (Masciari and Carruthers 2021), a new cognitive capacity rather than sentience (Irvine 2021; Godfrey-Smith 2021), or a set of attentional skills (Montemayor 2021). According to one commentary (Brown et al. 2021), UAL has no relevance to consciousness because a necessary higher-order ingredient is lacking. At the opposite extreme, Linson et al. (2021) believe that we are too restrictive in refusing to attribute consciousness to plants, which display remarkable adaptive plasticity. Several commentators (Rudrauf and Williford 2021; Levin 2021; Mallatt 2021) point to additional research directions with the potential to enrich and extend the UAL approach.

A note on "we". The three authors have their own views and don't agree on everything, but we have nonetheless written our responses in the first-person plural. Readers can consult our other recent work (Ginsburg and Jablonka 2019; Birch 2020a,b) to infer where the differences may lie.

From a list of capacities to a minimal functional architecture

Before responding to commentators, we want to introduce one important new idea to supplement the ideas in the target article. This will help us respond to several of the commentaries.

The target article introduced a list of capacities: global accessibility and broadcast; binding/unification and differentiation; selective attention and exclusion; intentionality; integration of information over time; an evaluative system; agency and embodiment; and registration of a self/other distinction. We suggested that possessing all of these capacities was plausibly sufficient for having a capacity for phenomenal consciousness. What we did not do in the target article was offer a model of how these capacities relate to each other. This may have made the list seem a little miscellaneous or disunified. For example, Brown et al. (2021) protest that "the features have almost nothing to do with one another".

Gánti's hallmarks of life can also appear miscellaneous—until one sees the chemoton model that ties them together. Ginsburg and Jablonka's (2019) book, *The Evolution of the Sensitive Soul*, argues that the hallmarks of consciousness are similarly interrelated and attempts to construct a minimal model (inspired by Gánti's chemoton) of how they interact to enable UAL (see Ginsburg and Jablonka 2019, Ch. 8). A simplified version of the architecture of UAL is given in Fig. 1. The basic picture
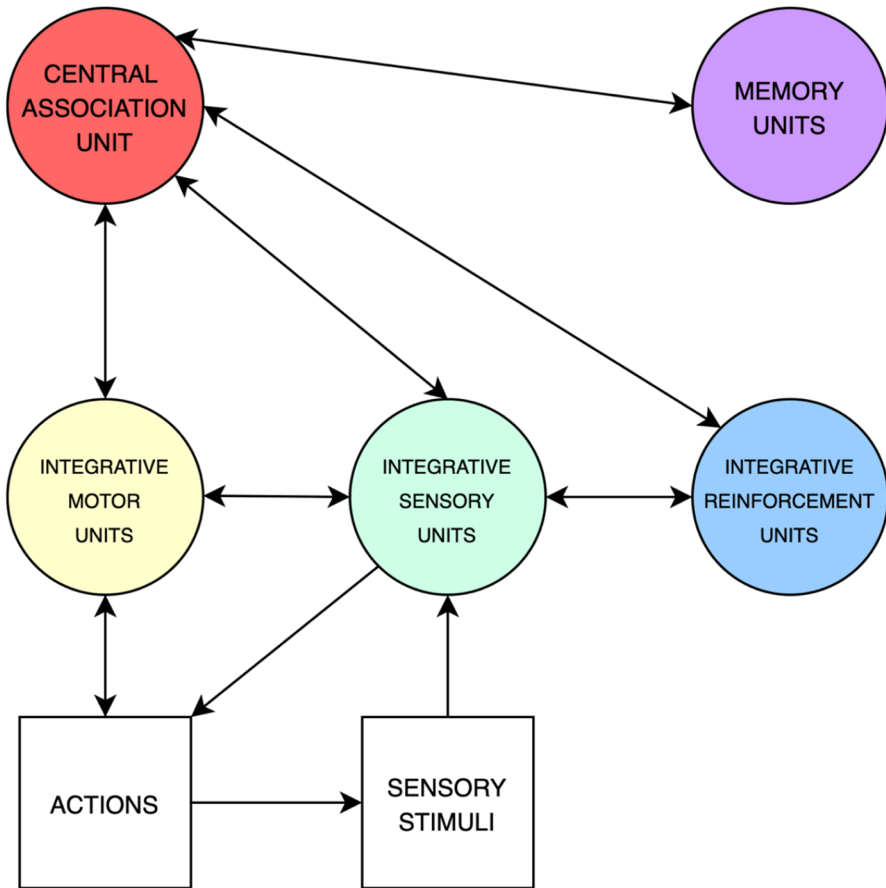
Fig. 1 A minimal UAL architecture. Unlimited Associative Learning is hypothesized to depend on recip-rocal re-entrant connections between sensory, motor, reinforcement and memory processing units, with a central association unit akin to a minimal global workspace at the core of the network. This architec-ture is proposed to be sufficient for conscious experience. (In Ginsburg and Jablonka 2019, Fig. 8.2, this architecture underlies learning, and therefore includes a specification of the temporal relations between inputs and outputs.)

is that a central association unit is linked to more specialized integrative processors that handle sensory information, motor information, memory storage, and evalu-ation. Within each unit there are various levels of hierarchical processing, which employ local synaptic memory and reinforcement mechanisms (not shown). The double-headed arrows represent two-way re-entrant connections between the units. We will call this the *minimal UAL architecture*.

The minimal UAL architecture helps clarify the relationship between UAL and the global workspace theory (GWT), the theory of consciousness with which it is most obviously compatible. A minimal global workspace, as depicted in Dehaene et al. (1998) and reproduced here as Fig. 2, is a neuronal system that receives
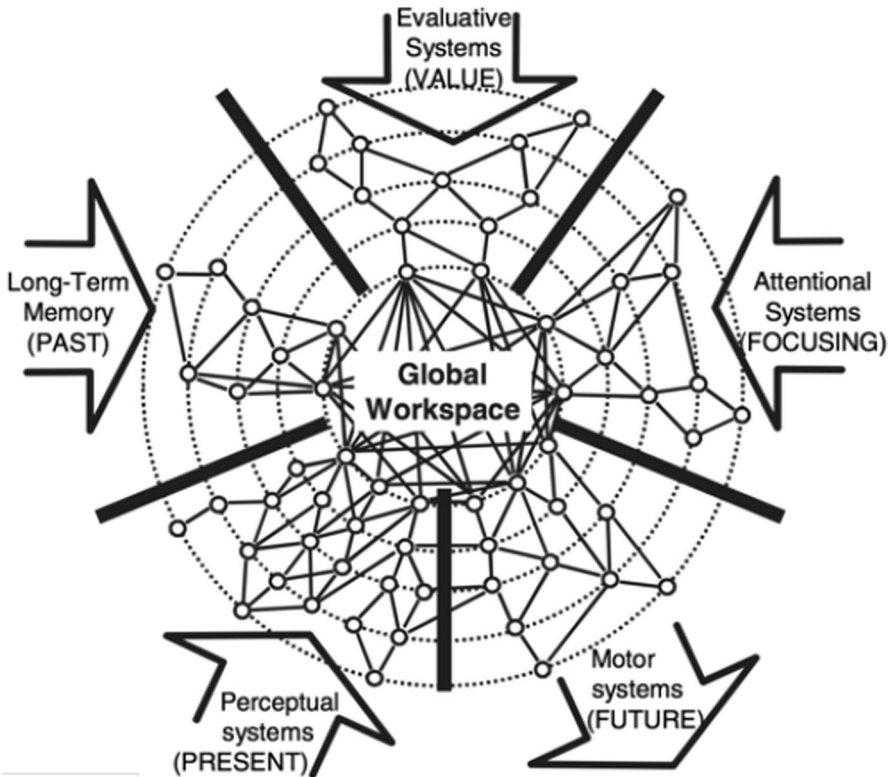
Fig. 2 The global workspace theory. The figure used by Dehaene et al. (1998) to illustrate the core elements of a global workspace. Note that broadcast to a wide range of consumer systems such as planning, reasoning and verbal report does not feature in the figure. © (1998) National Academy of Sciences. The copyright holder permits non-commercial reuse

content from perceptual systems, memory systems, evaluative systems and attentional systems, integrates the content into a unified, coherent representation of the world, and broadcasts this content back to the input systems and onwards to motor systems. In humans, a global workspace ensures that local processors across the cortex are operating "on the same page" in respects that matter.

Proponents of GWT often add that included under "motor" systems are mechanisms of verbal report and rational planning. They add this to emphasize that the global workspace theory can explain the reportability of conscious experience in humans. But we don't interpret this link to verbal report and rational planning as part of what it is to have a minimal global workspace. There is a reasonable interpretation of the theory on which this is an optional add-on, present in many adult

humans but absent in infants, minimally conscious adults, and most animals with global workspaces. Defenders of the global workspace theory should endorse this interpretation unless they want to maintain that verbal report and rational planning are necessary for conscious experience.[1]

Possessing the minimal UAL architecture is a lot like possessing a minimal global workspace. There is one important difference we want to emphasize, namely that we see no need for dedicated attentional networks that are *distinct* from local perception, memory and evaluation networks. We find it plausible that local attentional networks exist *within* perceptual, memory and evaluative systems—that processes of selective exclusion and amplification are *diffused throughout* the minimal UAL architecture.

We see this minimal architecture as compatible with, but not committed to, the predictive processing movement in cognitive neuroscience (e.g., Friston 2010; Clark 2016). It is well aligned with a picture on which the central association unit is engaged in modelling the agent and the surrounding world, and on which its outputs are predictions of sensory and motor information. The local integrative processors may then be interpreted as sending prediction error signals back to the association unit, allowing the models to be updated. GWT was developed before the advent of predictive processing, so top-down signalling of predictions *from* the workspace *to* the sensory input systems is not usually emphasized in that literature. In light of predictive processing, we think it is a plausible hypothesis that these top-down signals in fact play an important role.

Less consensus than we thought?

We said in the target article that our list of capacities was plausibly sufficient for a capacity for conscious experience. The minimal UAL architecture implies the presence of at least a minimal form of all the capacities in the list. So, if the list of capacities (understood minimally) is sufficient for experience, then so is the minimal UAL architecture.

Have we made a controversial claim here? We thought the sufficiency of our list could be a point of near-consensus, due to (a) the popularity of GWT, and (b) the fact that some alternatives to GWT, such as Lamme's (2006) recurrent processing theory (RPT) and Block's (2011a) fragile short-term memory theory, take integrative perceptual processing *alone* to be sufficient for conscious experience, and so should find the minimal UAL architecture to be more than sufficient. We realized we would not be bringing along defenders of the higher-order thought (HOT) view, such as Brown et al. (2021), but other commentaries give other reasons for dissent (Masciari and Carruthers 2021; Montemayor 2021).

---

[1] Dehaene (2014, p. 246) appears sympathetic to the "optional add-on" interpretation, writing that "I would not be surprised if we discovered that all mammals, and probably many species of birds and fish, show evidence of a convergent evolution to the same sort of conscious workspace.".

Let's start with HOT. We see "HOT" as an umbrella term for a diverse family of positions. We will focus here on "actualist" versions, which posit that mental states become phenomenally conscious when they are the target of an *actual* higher-order thought (such as the thought that I am seeing a blue sky).[2] Any such view runs into trouble when we ask: What determines the content of a conscious experience—is it the first-order mental state or the higher-order thought? Suppose, for example, I see a blue sky and think that I am seeing a red sky. What is my conscious experience like: red or blue? If we say "blue", the content of the HOT is now irrelevant to the content of the experience, raising the question of why it is needed at all. If we say "red", the content of the HOT is now all-important, clashing with evidence that our experiences have fine-grained content that outruns our concepts (Neander 1998; Block 2011b; Block 2019; Carruthers 2019). As Block (2019, p. 202) puts it, "the cognitive system that according to [HOT] generates conscious experience is simply too coarse grained to explain normal human perceivers consciously seeing a million colors even though they have concepts of only a tiny fraction of those colors."

One way out of this bind is suggested by Lau's (2019) "perceptual reality monitoring" version of HOT, on which the content of the experience is determined by the first-order state, and the higher-order state merely "indexes which of a number of possible first-order states is a reliable reflection of the current world". On this version, the higher-order state is a tag that marks some first-order states as reliable, and this tag is said to transform a perceptual representation from unconscious to conscious without changing its content. This account clearly has some work to do in relation to affective experiences, imagination, conscious memory and afterimages (though see Lau 2019 for some tentative discussion of these topics). Moreover, we are not sure this view properly belongs in the HOT family at all, since there is no representation of first-order representations.[3] The unification of sensory and motor information to construct a perceptual representation of the world is part of our set of sufficient conditions, and we are open to the idea that the marking of some sensory inputs as more reliable than others might be part of the story about how that unification happens.

Carruthers, once a higher-order theorist, recently rejected the view in favour of GWT (Carruthers 2017). Masciari and Carruthers' (2021) critique should be understood in that context. They frame their concern as a dilemma: either the UAL project claims no relevance to phenomenal consciousness, in which case it is "worthy but dull", or it does claim to be relevant to phenomenal consciousness, in which case it begs the question. Our target article is clearly about phenomenal consciousness, though it tries to stay out of the swamp of unnecessary philosophical jargon.

---

[2] This contrasts with "dispositionalist" HOT, which posits that mental states become phenomenally conscious when they are *available* to become the target of a higher-order thought. Such a view was once defended by Carruthers (2000), but he has since abandoned it (Carruthers 2017).

[3] Brown et al. (2019) distinguish HOT (higher-order theory) from HOTT (higher-order thought theory) in order to make room for views such as Lau's, which are examples of HOT without HOTT. In this confusing terminology, "higher-order" seems to mean little more than "*something* more than first-order representation is required, but not necessarily a higher-order representation", and the UAL framework is consistent with this.

The first paragraph says "a conscious system … is capable of generating that elusive property that philosophers call phenomenal consciousness" (2020, p. 1). So the accusation of "begging the question" is the one to discuss.

Masciari and Carruthers write:

> All theorists can accept that the list of properties proposed by Birch et al. *accompanies* phenomenal consciousness in humans (because humans are phenomenally conscious and all humans possess those properties). But there is no agreement that all those properties play any role in the explanation of consciousness. For example, although humans do possess a global workspace, of course (working memory), many deny that entry into that workspace is what constitutes a state as phenomenally conscious (Block, 2007, 2011; Boly et al., 2017; Haun et al., 2017). Moreover, even if there were agreement on the set of properties sufficient for phenomenal consciousness in the human case, there would be no agreement about how "pared down" each of those properties can become while still being collectively sufficient for first-person phenomenal consciousness (as opposed to some lesser degree of access-consciousness or some limited form of perceptual awareness).

There are two criticisms being made here. One is that our list of capacities (even if not pared down at all) would not be widely seen as nomologically sufficient for a capacity for phenomenal consciousness. Block is cited, but we think our list of capacities implies a capacity for fragile short-term memory of the type Block takes to be linked to phenomenal consciousness. This is needed to integrate sensory information across short temporal gaps. We also take our sufficiency claim to be compatible with Merker's (2005, 2007) midbrain-based theory, on which a set of conditions even more minimal than ours is posited to be sufficient. What about Integrated Information Theorists? This is not the place for a long digression on the Integrated Information Theory (IIT), but we think they should get on board, in so far as they agree that an *evolved* system with our list of capacities would have non-zero integrated information (or "phi"). There are theoretically possible feedforward ("zero phi") architectures for realizing UAL (Doerig et al. 2019), but no reason to think they could realistically have evolved (cf. our discussion of AI in the target article).

The second claim is that, granting that the capacities are nomologically sufficient for consciousness in a full-blown human form, there would still be no consensus regarding how pared down each capacity could be while still generating a sufficient condition. Our proposal is that, if you strip away the elements of human conscious processing that are fairly clearly dispensable to conscious experience, such as language, rational planning, inner speech, imaginative mental imagery, social cognition, normative cognition and mind-reading (theory of mind), you are left with some version of the minimal UAL architecture, and this is still sufficient.

Masciari and Carruthers would probably object to "fairly clearly". The dispensability of language for consciousness is supported by the retrospective reports of adults who have experienced aphasia (Koch 2019). The dispensability of inner speech and imaginative mental imagery is supported by cases of aphantasia (Whiteley 2020). The dispensability of other sophisticated capacities such as mind-reading is less robustly established, but there is no theoretical reason to see them as more

central to conscious experience than language unless one endorses a HOT theory of consciousness, and we do not.

Montemayor's (2021) concern is that UAL is fully explained by a suite of attentional abilities that can be exercised unconsciously. We disagree that UAL is fully explained by attention alone. Consider the element that involves forming discriminating novel, compound stimuli. Attention *alone* does not confer that ability. It is a perceptual ability, one that may well depend on exactly the kind of integrative perceptual processing emphasized by the recurrent processing theory. The minimal UAL architecture includes perceptual, cognitive and motor elements.

But what about the claim that "the empirical evidence robustly and abundantly shows that most forms of attention, and certainly those relevant to UAL, can occur unconsciously"? We have not seen any evidence that the whole UAL package (i.e. the learning of chains of associations between novel, compound stimuli that bridge temporal gaps and can be flexibly revalued) can be performed unconsciously. In fact, there is some evidence that learning capacities which involve only some elements of UAL, such as first-order trace conditioning (Clark et al. 2002), first-order instrumental conditioning (Skora et al. 2021), and first-order reversal learning (Travers et al. 2018) already require conscious experience of the stimuli, and this evidence is part of the empirical motivation for the UAL view. The evidence is inconclusive (as Birch 2020a emphasizes) but we take the evidence that exists to be mostly supportive. We seem to be reading the empirical literature differently.

Overemphasizing learning?

Several commentators (Irvine 2021; Rudrauf and Williford 2021; Godfrey-Smith 2021) suggest that the UAL framework overemphasizes the evolutionary significance of learning, while underplaying the importance of egocentric spatial representation. Irvine (2021) suggests an additional transition marker: generating "egocentric representations of itself acting in space, where actions are goal-directed and selected in a top-down manner". As Irvine notes, Merker (2005) and Barron and Klein (2016) can be seen as positing such a transition marker, albeit without calling it a transition marker. In a similar vein, Rudrauf and Williford (2021) propose "3D projective geometrical processing" as a capacity that belongs in our list, and one that points to a deficiency of UAL as a transition marker, in so far as UAL does not require this sort of processing.

The emphasis on 3D projective geometry strikes us a somewhat vision-centric way of thinking about consciousness. An animal without 3D representation of itself in extended space could intuitively still be conscious. It might have conscious experiences of olfactory and tactile stimuli, for example. It is a virtue of the UAL approach that it offers a transition marker we could look for even in animals without sophisticated distance vision. The ideas of an egocentric model, stable perspective or point of view are somewhat broader, and do not have to be understood in terms of 3D spatial representation. They are closer to the capacity to register a difference between self-caused and other-caused stimuli and to process the two types of stimulus differently, which we take to be a requirement of UAL.

Godfrey-Smith (2021) calls for a focus on "an active animal's real-time response to events in the absence of learning" and "the discrimination of self-caused from other-caused sensory events, in a context where one's rapid motions cause cascades of sensory stimuli mixed in with those derived exogenously". This is a similar point to those of Irvine and Rudrauf and Williford but made in a less vision-centric way, stressing the importance of self-world registration rather than 3D spatial modelling. We are not sure what to make of the suggestion that "attention, binding, integration and so on have roles that outrun UAL". Outrun how? This claim can be read in a way which is compatible with the target article. An implication of Claim 4 in the target article is that selection for UAL at least partly drove the evolution of the functional architecture that supports it. However, once a lineage has these capacities, it can do things other than UAL with them (e.g., online, "real-time" discrimination of stimuli). Moreover, some of these capacities, such as perceptual binding and selective attention, may develop earlier in ontogeny than others, and may be put to use for other functions before they are used for UAL, since UAL also involves long-term memory. Ontogeny does not recapitulate phylogeny in this case.

Godfrey-Smith suggests that UAL may be a transition marker but not an evolutionary "landmark", the idea being that the "real-time" consequences of self-other registration for action guidance may have been more important in driving the Cambrian explosion than its consequences for learning. That could be right: we conceded in the target article that "the assumption that associative learning drove the Cambrian explosion is not itself experimentally testable" (Birch et al. 2020, p. 13).

There is much more on the role of limited associative learning and UAL in the Cambrian explosion in *The Evolution of the Sensitive Soul* (Ginsburg and Jablonka 2019, Ch. 9). Part of the story is that limited associative learning had already begun to evolve in animals with very simple brains during the Ediacaran era, which preceded the Cambrian. The evolution of associative learning during this era began to drive change in the behaviour, physiology and morphology of the learning species and the organisms that interacted with them. As the scope and complexity of associative learning increased, so did the pace of ratcheting interactions (within and between species) that contributed to the Cambrian explosion. The evolution of UAL had many evolutionary effects, one of which, we suggested, was the evolution of forgetting and the sophistication of the stress response. Mallatt (2021) expresses scepticism about the possibility of ever testing these predictions. Although the study of these co-evolutionary relations is demanding, we are more optimistic than Mallatt, believing that with current bioinformatic techniques these relations are amenable to analysis.

Positive and negative markers

There are several limitations to the UAL framework, only some of which were discussed in the target article, and which rightly worry some of the commentators. Before we start looking at these, however, we would like to respond to a point made by Linson et al. (2021), which, they maintain, challenges the very notion of unlimited heredity (UH) and UAL as transition markers. They argue that the number of

realized variations in a system of UH or UAL is shaped by the interactions of the organism with its environment. This is true, but it is not clear why it is relevant to the proposition that there is a (practically) unlimited number of possible hereditary and behavioural variants, only a very small number of which are realized during the organism's life or during the evolutionary history of its lineage. Moreover, it is not just the number of theoretical structural-combinatorial possibilities, but also the number of phenotypically realizable possibilities that is vast (practically unlimited) given the richness of the potentially changing reciprocal interactions of the organism with its world over time (see Jablonka and Lamb 2020 for a discussion of these reciprocal interactions).

A transition marker, be it UH or UAL, is a *positive* marker: it purports to tell us when an entity is alive/conscious, but does not purport to tell us which entities are not alive/conscious. Mallatt (2021) finds this a significant limitation, arguing that we should be bolder: we should say that, if an animal does not even have integrative internal maps of the external world and its own body, it is not a serious candidate for consciousness. We can see the attractions of this idea. Representation is undoubtedly a key part of the UAL architecture, and representations with a map-like format, in which there is an isomorphism between the structure of the representation and the structure of some part of the body or external world, are extremely important. We just don't want to get drawn into inferring the *absence* of conscious experience from the *absence* of any component of UAL, no matter how central it is. By restricting ourselves to positive markers, we can remain neutral on negative markers—and the more we can remain neutral about, the more chance there is that we can forge a consensus around what we do assert.

A general remark: when considering evolutionary transitions, there are bound to be grey areas that defy clear definition. For example, the point at which a complex chemical system can be defined as living is moot and depends on one's theory of life. Is a simple version of Gánti's chemoton that lacks a replicating polymer alive? Or is a replicating polymer lacking a membrane separating it from the world (on the surface of a rock) alive? The idea that there is a sharp transition, suggested by the metaphor of a light bulb that is either on or off, is not compelling in the case of life. One can instead think of the transition using the metaphor of a dimmer switch. We suspect that something similar may be true of the transition from conscious to nonconscious life.

So... what about plants?

Linson et al. (2021) argue that our sufficient conditions are animal-centric. They also suggest that plants manifest many of the key features of UAL. Ginsburg and Jablonka (2021) have criticised elsewhere the idea that adaptive plasticity alone suffices for sentience, so here we focus on what we know plants can and cannot learn. We use the term learning in a literal sense, which goes beyond plasticity and priming and includes encoding, storage and recall.

The fact that animals co-evolved with plants has no bearing on the question whether plants are or are not conscious, even though some interesting traits in plants

may have evolved because of selection by conscious animals (Mallatt et al. 2020; Jablonka 2021). Plants display remarkable adaptive plasticity including many forms of priming, but they do not show UAL, and the evidence for any associative learning in plants, is, at best, scant. There is evidence that simple non-associative learning, such as learning by habituation, is found in unicellular organisms such as ciliates, and it is easy to model epigenetic memory and simple forms of learning in single cells (Ginsburg and Jablonka 2009). In multicellular organisms, learning requires complex communication among different cells and tissues, and investigations of non-associative learning in plants have generated conflicting results (Abramson and Chicas-Mosier 2016; Adelman 2018).

Crucially, though, there is neither evidence for UAL in plants nor any evidence of the processes that enable it: no evidence of integrative representations of world, body and prospective action, no evidence of an integrative memory system for storing such representations, no evidence of a flexible valuation system, and no evidence for learning across temporal gaps (see Taiz et al. 2019 for a recent critique). Nor is there evidence for neural-like processes and structures that can implement UAL. Such processes and structures have only been found in (some) animals with a central nervous system (Ginsburg and Jablonka 2019; Feinberg and Mallatt 2016; Mallatt, 2021). There is a single study of limited associative learning (conditional sensitization) in pea plants (Gagliano et al. 2016), but it did not replicate (Markel 2020) and in any case did not provide significant evidence against a plausible null hypothesis (Taiz et al. 2019). This limited form of associative learning, even if shown convincingly, would still lack the features of UAL.

In short, we do not rule out conscious experience in plants, because we are not in the business of providing negative markers of conscious experience. But there is no evidence that they satisfy our positive transition marker.

Development and material substrates

Levin (2021) takes the UAL project into a new realm of research, raising three important issues. First: at what point in ontogeny does the minimal UAL architecture come online? We believe that the development of the nervous system can give us a good idea as to when the basic structures that implement UAL are in place, and therefore, with all the inevitable qualifications, when the developing being is capable of consciousness. Whether it is actually *manifesting* that capacity is another question, because it may be in a state of deep, dreamless sleep. The UAL architecture may develop some time before its manifestation.

The second issue is what the UAL framework has to say about the cognitive continuity between different stages of the life cycle. Understood broadly, memory is based not only on neural-synaptic mechanisms but also on non-neural memory systems including localized and distributed epigenetic memory and memory stored in bioelectrical fields (Manicka and Levin 2019). The non-neural memory systems can form a cognitive bridge between different phases of life, between the stored memories of the caterpillar that can be recalled by the butterfly in spite of the massive brain re-organization that occurred during metamorphosis. Such transferred memories

may be also the basis for the continuity of the sense of self, although the assumption that the sense of self of the caterpillar and that of the metamorphosed adult are continuous at the phenomenological level is by no means decided. Although memories can be transferred, it is not clear how much needs to be transferred to establish a continuous unity of self, and it is also not clear how having a separate body changes things (in a regenerating planarian that grows a new head, for example). The coherence of the sense of self and the unity and disunity of consciousness is an important question, not just for studies of regeneration and morphogenesis but for all studies of development. As Godfrey-Smith (2020) has emphasized, the question of unity and disunity is important for our understanding of human pathologies, as well as for understanding the varieties of consciousness of organisms with different forms of neural organization.

The third issue is that of derived UAL in human-designed hybrid artefacts made of cells and non-biological materials. UAL may be regarded as implying a basic form of domain-general intelligence, and any system that implements it requires a flexible system that can monitor and evaluate the many aspects of the entity and the myriad ways in which it is affected by its reciprocal interactions with the environment. The question of what the implementation of UAL in artificial entities requires (whether it requires something like a central nervous system, for example) opens the way for a better understanding of the relations between UAL and its material realization, and therefore for a better understanding of the relation between UAL, consciousness-as-we-know-it and consciousness-as-it-could-be.

References

Abramson C, Chicas-Mosier A (2016) Learning in plants: lessons from *Mimosa pudica*. Front Psychol 7:417. https://doi.org/10.3389/fpsyg.2016.00417

Adelman BE (2018) On the conditioning of plants: a review of experimental evidence. Perspect Behav Sci 41:431–446. https://doi.org/10.1007/S40614-018-0173-6

Barron AB, Klein C (2016) What insects can tell us about the origins of consciousness. Proc Natl Acad Sci USA 113:4900–4908. https://doi.org/10.1073/pnas.1520084113

Birch J (2020a) In search of the origins of consciousness. Acta Biotheor 68:287–294. https://doi.org/10.1007/s10441-019-09363-x

Birch J (2020b) The search for invertebrate consciousness. Noûs. Advance online publication, https://doi.org/10.1111/nous.12351

Birch J, Ginsburg G, Jablonka E (2020) Unlimited associative learning and the origins of consciousness: a primer and some prediction. Biol Phil. https://doi.org/10.1007/s10539-020-09772-0

Block N (2019) Empirical science meets higher-order views of consciousness: reply to Hakwan Lau and Richard Brown. In: Pautz E, Stoljar D (eds) Blockheads! MIT Press, Cambridge, MA, Essays on Ned Block's philosophy of mind and consciousness, pp 199–213

Block N (2011) The higher order approach to consciousness is defunct. Analysis 71:419–431. https://doi.org/10.1093/analys/anr037

Brown R, LeDoux J, Rosenthal D (2021) The extra ingredient. Biol Philos 36:16. https://doi.org/10.1007/s10539-021-09797-z

Carruthers P (2017). In defence of first-order representationalism. J Conscious Stud 24(5–6):74–87. http://www.jstor.org/stable/41237346

Carruthers P (2019) Human and animal minds: the consciousness questions laid to rest. Oxford University Press, Oxford

Carruthers P (2000) Phenomenal consciousness: a naturalistic theory. Cambridge University Press, Cambridge, UK

Clark RE, Manns JR, Squire LR (2002) Classical conditioning, awareness, and brain systems. Trends Cogn Sci 6:524–531. https://doi.org/10.1016/s1364-6613(02)02041-7

Clark A (2016) Surfing uncertainty: prediction, action, and the embodied mind. Oxford University Press, New York

Dehaene S (2014) Consciousness and the brain: deciphering how the brain encodes our thoughts. Viking Press, New York

Dehaene S, Kerszberg M, Changeux JP (1998) A neuronal model of a global workspace in effortful cognitive tasks. Proc Natl Acad Sci USA 95(24):14529–14534. https://doi.org/10.1073/pnas.95.24.14529

Doerig A, Schurger A, Herzog M (2019) The unfolding argument: why IIT and other causal structure theories cannot explain consciousness. Consc Cogn 72:49–59. https://doi.org/10.1016/j.concog.2019.04.002

Feinberg TE, Mallatt J (2016) The ancient origins of consciousness: how the brain created experience. MIT Press, Cambridge, MA

Friston K (2010) The free-energy principle: a unified brain theory? Nat Rev Neurosci 11:127–138. https://doi.org/10.1038/nrn2787

Gagliano M, Vyazovskiy VV, Borbely AA, Grimonprez M, Depczynski M (2016) Learning by association in plants. Sci Rep 6:38427. https://doi.org/10.1038/srep38427

Ginsburg S, Jablonka E (2009) Epigenetic learning in non-neural organisms. J Biosci 34(4):633–646. https://doi.org/10.1007/s12038-009-0081-8

Ginsburg S, Jablonka E (2019) The evolution of the sensitive soul: learning and the origins of consciousness. MIT Press, Cambridge, MA

Ginsburg S, Jablonka E (2021) Sentience in plants: a green red herring? J Conscious Stud 28(1–2):17–33

Godfrey-Smith P (2020) Metazoa: animal life and the birth of the mind. Farrar, Straus and Giroux, New York

Godfrey-Smith P (2021) Learning and the biology of consciousness: a commentary on Birch, Ginsburg, and Jablonka. Biol Philos 36:44. https://doi.org/10.1007/s10539-021-09820-3

Irvine E (2021) Assessing unlimited associative learning as a transition marker. Biol Philos 36:21. https://doi.org/10.1007/s10539-021-09796-0

Jablonka E (2021) Signs of consciousness? Biosemiotics. https://doi.org/10.1007/s12304-021-09419-x

Jablonka E, Lamb M (2020) Inheritance systems and the extended synthesis. Cambridge University Press, Cambridge, UK

Koch C (2019) Consciousness doesn't depend on language. Nautilus, 16 September 2019. https://nautil.us/issue/76/language/consciousness-doesnt-depend-on-language

Lamme VAF (2006) Towards a true neural stance on consciousness. Trends Cogn Sci 10(11):494–501. https://doi.org/10.1016/j.tics.2006.09.001

Lau H (2019) Consciousness, metacognition, and perceptual reality monitoring. BioRxiv. https://doi.org/10.31234/osf.io/ckbyf

Levin M (2021) Unlimited plasticity of embodied, cognitive subjects: a new playground for the UAL framework. Biol Philos 36:17. https://doi.org/10.1007/s10539-021-09792-4

Linson A, Ponkshe A, Calvo P (2021) On plants and principles. Biol Philos 36:19. https://doi.org/10.1007/s10539-021-09793-3

Mallatt J (2021) Unlimited associative learning and consciousness: further support and some caveats about a link to stress. Biol Philos 36:22. https://doi.org/10.1007/s10539-021-09798-y

Mallatt J, Blat MR, Draguhn A, Robinson DG, Taiz L (2020) Debunking a myth: plant consciousness. Protoplasma Xx. https://doi.org/10.1007/s00709-020-01579-w

Manicka S, Levin M (2019) The cognitive lens: a primer on conceptual tools for analysing information processing in developmental and regenerative morphogenesis. Philos Trans Royal Soc B 374:20180369. https://doi.org/10.1098/rstb.2018.0369

Markel K (2020) Pavlov's pea plants? not so fast. An attempted replication of Gagliano et al. (2016). eLife 9:e57614. https://doi.org/10.1101/2020.04.05.026823

Masciari CF, Carruthers P (2021) Perceptual awareness or phenomenal consciousness?A dilemma. Biol Philos 36:18. https://doi.org/10.1007/s10539-021-09795-1

Merker B (2005) The liabilities of mobility: a selection pressure for the transition to consciousness in animal evolution. Conscious Cogn 14:89–114. https://doi.org/10.1016/S1053-8100(03)00002-3

Merker B (2007) Consciousness without a cerebral cortex: a challenge for neuroscience and medicine. Behav Brain Sci 30:63–134. https://doi.org/10.1017/S0140525X07000891

Montemayor C (2021) Attention explains the transition to unlimited associative learning Better than consciousness. Biol Philos 36:20. https://doi.org/10.1007/s10539-021-09794-2

Neander KN (1998) The division of phenomenal labour: a problem for representational theories of consciousness. Philos Perspect 12:411–434. https://doi.org/10.1111/0029-4624.32.s12.18

Rudrauf D, Williford K (2021) Unlimited associative learning and the origins of consciousness: the missing point of view. Biol Philos 36:43. https://doi.org/10.1007/s10539-021-09819-w

Skora LI, Yeomans MR, Crombag HS, Scott RB (2021) Evidence that instrumental conditioning requires conscious awareness in humans. Cognition 208:104546. https://doi.org/10.1016/j.cognition.2020.104546

Taiz L, Alkon D, Draguhn A, Murphy A, Blatt M, Hawes C, Thiel G, Robinson DG (2019) Plants neither possess nor require consciousness. Trends Plant Sci 24:677–687. https://doi.org/10.1016/j.tplants.2019.05.008

Travers E, Frith CD, Shea N (2018) Learning rapidly about the relevance of visual cues requires conscious awareness. Q J Exp Psychol 71:1698–1713. https://doi.org/10.1080/17470218.2017.1373834

Whiteley CMK (2020) Aphantasia, imagination and dreaming. Philos Stud, Adv Online Public. https://doi.org/10.1007/s11098-020-01526-8