Unlimited associative learning and the origins of consciousness: the missing point of view

David Rudrauf¹ · Kenneth Williford²

Received: 11 December 2020 / Accepted: 30 August 2021 / Published online: 8 September 2021
© The Author(s) 2021

Introduction

In their article, Birch, Ginsburg, and Jablonka (see this issue) argue that consciousness is a “mode of being” that can be sufficiently characterised, for practical purposes, by a single summary concept or property, that of unlimited associative learning (UAL), as an open-ended form of learning a system uses to learn about the world and itself. UAL would subsume a list of jointly sufficient capacities on the importance of which for consciousness there is some consensus: global accessibility and broadcasting, binding/unification and differentiation, selective attention and exclusion, intentionality, integration of information over time, the presence of an evaluative system, agency and embodiment, and the registration of a self/other distinction. UAL could further be characterized by the following “natural cluster” of features related to different aspects of processing: *compound stimuli*, *novel stimuli*, *second-order conditioning*, *trace conditioning*, *flexible*, *easily rewritable associations with value*. Thus according to the authors, UAL “plausibly requires the existence of functionally coupled systems with all of the hallmarks of consciousness”.

UAL, the authors argue, could in turn serve as a single *evolutionary transition marker*, offering testable predictions on a “firm theoretical and methodological footing”, to probe the phylogenetic tree of animals, and draw a line between non-conscious and conscious animals. As a result, the marker could be used to identify when in evolution, and across which branches of the evolutionary tree, consciousness is likely to have appeared. The authors, and we agree with them on this point, reject the option of taking language as part of the definition of such a transition marker. This would automatically make *homo sapiens* the only holder of consciousness. But consciousness is very likely to have appeared earlier in evolution, setting the ground for that consciousness *homo sapiens* is able to report on to its conspecifics verbally.

✉ David Rudrauf
david.rudrauf@unige.ch

¹ Faculty of Psychology and Education Science, Swiss Center for Affective Science, University Center of Computer Science, University of Geneva, Geneva, Switzerland

² Department of Philosophy and Humanities, University of Texas at Arlington, Arlington, TX, USA

One question that may come to mind in appraising this theory is whether learning, in and of itself, whether UAL or other forms of learning, could be a sufficient basis for characterizing if not understanding consciousness, even in the pragmatic perspective of finding a single evolutionary transition marker. However, there is perhaps a more puzzling question that arises from the very target definition of consciousness that the authors put forth to set the stage for their theory. According to them, a conscious system is an “experiencing subject”, that “has a subjective point of view on the world and its own body”, which is at the core of the generation of an “elusive phenomenal consciousness”, so that “it feels like something to be that system”.

One may then wonder why such a “subjective point of view”, so central in their definition, does not somehow make it into their list of properties, as a necessary, though not sufficient condition, for considering a system to be conscious. It would seem that their transition marker would at least also require an operational theory of such a point of view, which would be more directly and explicitly related to their definition of consciousness, and which could then be used to probe living systems in an evolutionary perspective by demonstrating the presence of the relevant type of processing.

Following the authors’ own premises, we should conclude that a valid characterization of consciousness in nature would thus require us to establish a theory of such viewpoints, or “viewpoint theory”, in a manner that is non-trivial and sufficiently operational to be used in a pragmatic way. The importance of a “subjective point of view” to first-person phenomenology certainly makes the (near) consensus, and this seems to hold for the authors of the target article, at least in their working definitions of consciousness. The philosopher Nagel (1974), in his highly influential article, “What is it like to be a bat?”, made such a subjective point of view a key ingredient of his somewhat vague notion, endorsed by the authors, that an integral aspect of consciousness is that “it feels like something to be that system”. Pioneering contributions to viewpoint theory can be traced to Trehub (1977, 2007) and Lehar (2003). The viewpoint has been further characterized theoretically and experimentally, sometimes under the notion of a first-person “egocenter”, as operating, in a pivotal manner, at the core of the inferential processing allowing organisms to adaptively relate multimodal sensory information and intentional action in the world (Merker 2013, p. 13; see also Merker 2012, p. 54).

Viewpoint-dependent computation also plays a pivotal role in the ability of at least human consciousness for imaginary perspective taking (Rudrauf et al. 2017). Perspective taking relies on the spatial framing of multimodal sensory and affective information, and conditions the way embodied agents represent and interact with objects and others. Indeed, humans, and probably other animals (see, e.g., Premack and Woodruff 1978), are social beings that are capable of implementing Theory of Mind (ToM), which itself relies on perspective taking, and the disruption of which can lead to profound alterations of the normal development of consciousness, as in Autism Spectrum Disorders (ASD) (Baron-Cohen 1989; see also Rudrauf and Debané 2018).

But how can we approach the challenge of establishing an operational mechanism capable of accounting for the “subjective point of view” so central to consciousness?

The problem has recently been addressed by the Projective Consciousness Model (PCM, Rudrauf et al. 2017, 2020; Williford et al. 2018). The PCM explicitly integrates a formal theory of the subjective point of view and of its causal contribution to biological cybernetics, or more specifically to active inference. Active inference is a method through which autonomous embodied agents optimize their actions based on the prediction of the sensory impact of their expected outcome, and based for instance on the minimization of a quantity such as Free Energy, which reflects the divergence between expectations and outcomes (Friston 2010; Friston et al. 2017). The PCM places consciousness, at least in humans, at the core of the process of active inference, through the concept of a Field of Consciousness (FoC). The concept straightforwardly derives from considering the manner through which the phenomenology of human consciousness integrates multimodal information into a 3-dimensional supramodal workspace, which is conspicuously organized in a perspectival manner.

Without entering into the details here, the PCM forefronts the perspectival structuration of conscious experience around an elusive point of view within an internal, virtualized space, the FoC. The FoC allows for the access and representation of information in a projective geometrical manner, in such a way that the “workspace” of consciousness relies on a 3-dimensional projective space, subject to projective transformations for calibration and perspective taking (see Rudrauf et al. 2017, 2020; Williford et al. 2018).

We have shown how the principles of the PCM could explain perceptual illusions such as the Moon Illusion and Ames Room, as a result of the calibration of a 3-dimensional projective frame under free energy minimization (Rudrauf et al. 2020). We previously offered preliminary descriptions of how the PCM principles could integrate affective processing and imaginary projections to support active inference (Rudrauf et al. 2017; Rudrauf and Debbané 2018), and we are currently working on further developing models of appraisal and ToM based on the same mechanisms in order to govern artificial (virtual or robotics) agents.

Through the FoC, an agent can causally relate perception, i.e., representations directly informed by sensory evidence, and the imagination, i.e., representations informed by prior beliefs and preferences only, in order to perform active inference. The process can support the situated appraisal and reappraisal of multimodal sensorimotor, affective and social information, within a subjective referential in which orientation, directions, and relations of incidence are pivotal, in order to motivate and program action. Internal models of the beliefs and preferences attributed to other agents (including of other agents’ beliefs about the beliefs and preferences of others), can be integrated into the model so that the agent can employ ToM as part of active inference, through simulations within its own FoC.

Thus according to the PCM, 3-dimensional projective geometrical processing is a necessary component of a theory of consciousness, encoding its inherent subjective point of view and articulating its role in information integration in a manner that is consistent with the phenomenology of subjective experience and simultaneously makes sense of its role in biological cybernetics. Accordingly, 3-dimensional projective geometrical processing of sensorimotor information should at least be on the list proposed by the authors to characterize valid evolutionary transition markers for

consciousness. Without such a projective mechanism, the list of capacities essential to characterize consciousness would not be sufficient, we maintain. UAL as such, although perhaps important for understanding important aspects of consciousness, does not entail such essential property, and thus cannot be a fully satisfactory criterion to probe the evolution of consciousness in nature.

Funding Open Access funding provided by Université de Genève.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Baron-Cohen S (1989) Joint-attention deficits in autism: towards a cognitive analysis. *Dev Psychopathol* 1(3):185–189
- Friston K (2010) The free-energy principle: A unified brain theory? *Nat Rev Neurosci* 11(2):127
- Friston K, FitzGerald T, Rigoli F, Schwartenbeck P, Pezzulo G (2017) Active inference: a process theory. *Neural Comput* 29(1):1–49
- Lehar S (2003) *The world in your head: a gestalt view of the mechanism of conscious experience*. Erlbaum, Mahwah, NJ
- Merker B (2013) Body and world as phenomenal contents of the brain's reality model. In: Pereira A, Lehmann D (eds) *The unity of mind, brain and world: current perspectives on a science of consciousness*. Cambridge University Press, Cambridge, pp 7–42
- Merker B (2012) From probabilities to percepts: a subcortical “global best estimate buffer” as locus of phenomenal experience. In: Edelman S, Fekete T, Zack N (eds) *Being in time dynamical models of phenomenal experience*. John Benjamins, Amsterdam, pp 37–79
- Nagel T (1974) What is it like to be a bat? *Philos Rev* 83:435–450
- Premack D, Woodruff G (1978) Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences* 1(4):515–526
- Rudrauf D, Debbané M (2018) Building a cybernetic model of psychopathology: beyond the metaphor. *Psychol Inq* 29(3):156–164
- Rudrauf D, Bennequin D, Granic I, Landini G, Friston K, Williford K (2017) A mathematical model of embodied consciousness. *J Theor Biol* 428:106–131
- Rudrauf D, Bennequin D, Williford K (2020) The Moon illusion explained by the projective consciousness model. *J Theor Biol* 507:110455
- Trehub A (1977) Neuronal models for cognitive processes: networks for learning, perception and imagination. *J Theor Biol* 65:141–169
- Trehub A (2007) Space, self, and the theater of consciousness. *Conscious Cogn* 16:310–330
- Williford K, Bennequin D, Friston K, Rudrauf D (2018) The projective consciousness model and phenomenal selfhood. *Frontiers Psychol*. <https://doi.org/10.3389/fpsyg.2018.02571>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.