Finally, we remind our readers that all transcendent ideas which have led to wiser knowledge and purposiveness of life on earth led by human technology and reason have been dependent on sensitive attention to and measurement of the qualities, forms, and expressions of nature. This has been made clear by the works of Darwin (1872) and the contemporary science of cycles of prosperity and extinction in life on earth surveyed by Attenborough (2020). The latter, especially, teaches our collective human understanding about risks of over-extending our inventive technical industry, in all our communities and nations, and in our science.

Conflict of interest. There are no conflicting interests.

References

Attenborough, D. (2020). *A life on our planet: My witness statement and a vision for the future*. Ebury Press.

Basile, P. (2010) It must be true – But how can it be? Some remarks on panpsychism and mental composition. In: *The metaphysics of consciousness*, eds. P. Basile, J. Kiverstein, & P. Phemister, pp. 93–112. Cambridge University Press.

Clark, A. (2019). Consciousness as generative entanglement. *Journal of Philosophy 116*(12), 645–662.

Damasio, A. (2010). *Self comes to mind: Constructing the conscious brain*. Pantheon/Random House.

Darwin, C. (1872). *The expressions of emotion in man and animals*. London: Methuen.

Delafield-Butt, J. T. (2014) Process and action: Whitehead's ontological units and perceptuomotor control units. In: *Life and process*, ed. S. Koutroufinis, pp. 133–156. De Gruyter Ontos.

Delafield-Butt, J. T., Freer, Y., Perkins, J., Skulina, D., Schögler, B., & Lee, D. N. (2018). Prospective organization of neonatal arm movements: A motor foundation of embodied agency, disrupted in premature birth. *Developmental Science 21*(6), e12693.

Delafield-Butt, J. T., & Gangopadhyay, N. (2013). Sensorimotor intentionality: The origins of intentionality in prospective agent action. *Developmental Review 33*(4), 399–425.

Goff, P. (2019). *Galileo's error: Foundations for a new science of consciousness*. Pantheon Books.

Grossberg, S. (2021). *Conscious mind resonant brain*. Oxford University Press, quoted in https://blog.oup.com/2021/07/why-did-evolution-create-conscious-states-of-mind/.

Langer, S. (1967). *Mind: An essay on human feeling*. John Hopkins Press.

Leibniz, G. (1716) *La Monadologie*.

Macmurray, J. (1959). *The self as agent* (volume I of *The form of the personal*). Faber and Faber. (Paperback, 1969).

Merker, B. (2005). The liabilities of mobility: A selection pressure for the transition to consciousness in animal evolution. *Consciousness and Cognition 14*, 89–114.

Merker, B. (2007). Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral and Brain Sciences 30*, 63–134.

Oizumi, M., Albantakis, L., & Tononi, G. (2014). From the phenomenology to the mechanisms of consciousness: Integrated information theory 3.0. *PLoS Computational Biology 10*(5), e1003588.

Panksepp, J., & Biven, L. (2012). *The archaeology of mind: Neuroevolutionary origins of human emotions*. Norton.

Reed, E. S. (1996). *Encountering the world: Toward an ecological psychology*. Oxford University Press.

Solms, M. (2021). *The hidden spring: A journey to the source of consciousness*. Profile Books.

Strawson, G. (2006) Realistic monism: Why physicalism entails panpsychism. In: *Consciousness and Its place in nature*, ed. A. Freeman, pp. 3–31. Imprint Academic.

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience 5*(1), 42.

Trevarthen, C., & Delafield-Butt, J. (2017b) Intersubjectivity in the imagination and feelings of the infant: Implications for education in the early years. In: *Under-three year olds in policy and practice (policy and pedagogy with under-three year olds: Cross-disciplinary insights and innovations)*, eds. E. J. White and C. Dalli, pp. 17–39. Springer.

Trevarthen, C., & Delafield-Butt, J. T. (2015) The infant's creative vitality, In projects of self-discovery and shared meaning: How they anticipate school, and make It fruitful. In: *International handbook of young children's thinking and understanding*, eds. S. Robson & S. F. Quinn, pp. 3–18. Routledge.

Trevarthen, C., & Delafield-Butt, J. T. (2017a) Development of consciousness. In: *Cambridge encyclopedia of child development*, eds. B. Hopkins, E. Geangu & S. Linkenauger, pp. 821–835. Cambridge University Press.

Vandekerckhove, M., & Panksepp, J. (2011). A neurocognitive theory of higher mental emergence: From anoetic affective experiences to noetic knowledge and autonoetic awareness. *Neuroscience & Biobehavioral Reviews 35*(9), 2017–2025.

Whitehead, A. N. (1929). *Process and reality*. Macmillan.

The integrated information theory of agency

Hugh Desmond and Philippe Huneman

Institute of History and Philosophy of Science and Technology, Université Paris 1 Panthéon-Sorbonne – CNRS, 75006 Paris, France.
hugh.desmond@gmail.com
www.hughdesmond.net
philippe.huneman@gmail.com
https://philippehuneman.wordpress.com

doi:10.1017/S0140525X21002004, e0

Abstract

We propose that measures of information integration can be more straightforwardly interpreted as measures of agency rather than of consciousness. This may be useful to the goals of consciousness research, given how agency and consciousness are "duals" in many (although not all) respects.

Once consciousness is analyzed as "efficient network activity," it is manifested across a broad range of systems, and its meaningfulness as a concept becomes diluted. That is, as Merker *et al.* (forthcoming) successfully show, a fundamental challenge for the integrated information theory (IIT) of consciousness (Tononi, 2008; Tononi, Boly, Massimini, & Koch, 2016).

As a way forward, we propose that measures of information integration can more straightforwardly be interpreted as measures of agency rather than of consciousness. Because agency can be defined in terms of behavioral patterns, it avoids the problems arising from quantifying "first-person perspective" properties by means of "third-person perspective" measures. As the conceptual dual of consciousness, agency may, thus, deserve a more prominent place in consciousness research.

Agency and integration. Agency is increasingly of interest to biologists, as many developmental patterns and behaviors (including those of plants) are characterized in agential terms. It is part of a trend to assign a greater theoretical role to organisms as such in our understanding of evolution: Organisms do not simply passively undergo evolutionary processes but actively shape their selective environment (Laland, Matthews, & Feldman, 2016) and respond in a goal-directed manner to opportunities or "affordances" in their environment (Walsh, 2015). At a very general level, "agency" refers to how organisms exhibit goal-directed behaviors in response to environmental change.

Talking about goals in this way activates old worries about teleology and anthropomorphism (i.e., agency as human-like intentionality). However, in practice, the pay-off for explaining behaviors as goal-directed lies in accounting for patterns of behavioral robustness: An organism's "goal" is simply what it attempts to achieve through various means, even when it is perturbed or challenged by an environmental change. In other words, agency refers to how a (1) small number of goals can

Fig. 1 - B/W online

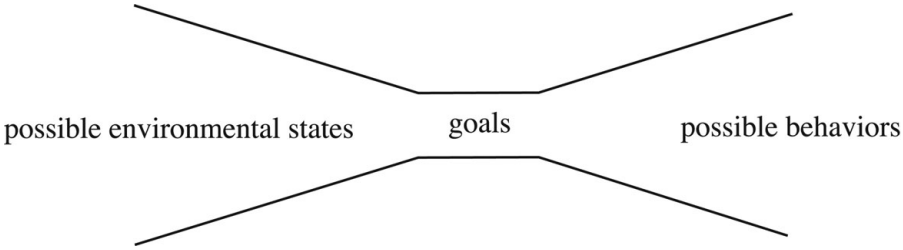

Fig. 1. (Desmond and Huneman). The bow-tie architecture of agency.

account for patterns of connectivity between (2) a large number of possible environmental states, and (3) a large number of possible behaviors.

This explanatory structure describes a bow-tie architecture (see Fig. 1) where environmental states and behaviors are integrated in virtue of the "goals" present. Note that the figure does not illustrate any fine-grained connections between environmental states and behaviors. What it does illustrate is the explanatory general structure of agency, where the "goals" are used to explain how environmental states and behaviors are informationally integrated (for an information-theoretic treatment of scientific explanation, see e.g., Desmond, 2019).

When the issue is put in this way, suggestive parallels with theories of consciousness emerge. The IIT posits consciousness as the integration of various experiential properties (Tononi, 2008), but also global workspace theory (Dehaene & Changeux, 2011, p. 11) posits a similar structure, where consciousness is a global broadcast mechanism, integrating input and output systems.

However, unlike theories of consciousness, an integrated information theory of agency would not need further justification of why such bow-tie architectures should be identified with agency. Goal-directedness is a third-person concept and in this way it can be unproblematically fleshed out in terms of input–output patterns. Agency as a concept just *is* a type of pattern of connectivity between environmental states and behaviors. There is no need to posit a counterpart to "qualia" or some ineffable subjective quality.

The underlying reason for this is that agency is an explanatory concept rather than one that refers to an empirical state-of-affairs. Informally, agency could be said to be more similar to "Newton's law of inertia" rather than to "white snow." If one is habituated to thinking of agency in terms of intentionality (or the presence of some form of mentality), this view of agency may require a gestalt-switch. The concept of agency imposes a structure on observed behavior, and if the observed behavioral patterns do not exhibit this general structure, there is simply no need to describe them as "agential" (see discussion in Desmond & Huneman, 2020).

Agency and consciousness as duals. Whether or not the apparent isomorphism between theories of consciousness and the structure of agency is more than skin-deep is not a question we systematically address here. Instead, we offer a general rationale why the appearance of such isomorphisms should not be surprising. If agency refers to the "activity" of the organism in relation to the environment, consciousness in its broadest sense denotes the "passivity" of the organism. A synonym for consciousness – sentience – makes this passivity more apparent: The capacity of "feeling" refers to how an organism "undergoes" its environment (think also of "e-motion": being moved). Agency and consciousness are different sides of the same fundamental coin of organism-environment relationality. One cannot have activity without passivity, and vice versa.

In mathematics, dual concepts are used to integrate two different ways of looking at a same object (Atiyah, 2007). Similarly, agency and consciousness can, at a fundamental level, be viewed as "duals." And just as invoking the dual operator in mathematics may help solve otherwise intractable problems, perhaps some of the challenges facing our understanding of consciousness can be addressed by invoking agency. For instance, it is likely that the evolution of consciousness can only be understood by simultaneously understanding how agency evolved. This is reflected in how greater sensorimotor control has evolved in tandem with various proxies of consciousness such as cognitive systems (Godfrey-Smith, 2020; van Duijn, Keijzer, & Franken, 2006).

Could an IIT of agency avoid the equivalent of panpsychism, which seems unavoidable once consciousness is naturalized and/or de-anthropomorphized? Panpsychism's dual is "panagentialism": seeing agency everywhere. In other contexts this has been called hyper-agency detection (cf. Atran, 2002). We believe panagentialism can be more easily defused, because of a subtle asymmetry between agency and consciousness (at least as the latter is typically understood). Attributing agency is an explanatory strategy to make sense of behavioral complexity – not a statement about the ontological makeup of the world. Panagentialism is thus simply a (poor) explanatory practise. Note that there may also be "no facts of the matter" regarding consciousness (Carruthers, 2020). In that case, panpsychism would also be a poor explanatory practise, and agency and consciousness would be true duals.

Financial support. This research was funded by the ANR-DFG funded project "The Explanatory Scope of Generalized Darwinism: Towards Criteria for Evolutionary Explanations Outside Biology."

Conflict of interest. The authors declare that there is no conflict of interest.

References

Atiyah, M. (2007). Duality in Mathematics and Physics. Transcription of a talk delivered at the Institut de Matemàtica de la Universitat de Barcelona.

Atran, S. (2002). *In gods we trust: The evolutionary landscape of religion*. Oxford University Press.

Carruthers, P. (2020). Stop caring about consciousness. *Philosophical Topics 48*(1), 1–20

Dehaene, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron 70*(2), 200–227. https://doi.org/10.1016/j.neuron.2011.03.018.

Desmond, H. (2019). Shades of grey: Granularity, pragmatics, and non-causal explanation. *Perspectives on Science 27*(1), 68–87. https://doi.org/10.1162/posc_a_00300

Desmond, H., & Huneman, P. (2020). The ontology of organismic agency: A Kantian approach. In: *Natural born monads: On the metaphysics of organisms and human individuals* (pp. 33–64), eds. A. Altobrando & P. Biasetti, De Gruyter.

Godfrey-Smith, P. (2020). Varieties of subjectivity. *Philosophy of Science 87*(5), 1150–1159. https://doi.org/10.1086/710541

Laland, K., Matthews, B., & Feldman, M. W. (2016). An introduction to niche construction theory. *Evolutionary Ecology 30*, 191–202. https://doi.org/10.1007/s10682-016-9821-z

Tononi, G. (2008). Consciousness as integrated information: A provisional manifesto. *The Biological Bulletin 215*(3), 216–242. https://doi.org/10.2307/25470707

Tononi, G., Boly, M., Massimini, M., & Koch, C. (2016). Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience 17*(7), 450–461. https://doi.org/10.1038/nrn.2016.44

van Duijn, M., Keijzer, F., & Franken, D. (2006). Principles of minimal cognition: Casting cognition as sensorimotor coordination. *Adaptive Behavior 14*(2), 157–170. https://doi.org/10.1177/105971230601400207

Walsh, D. (2015). *Organisms, agency, and evolution*. Cambridge University Press. https://doi.org/10.1017/CBO9781316402719

Searching in the wrong place: Might consciousness reside in the brainstem?

Marshall Devor, Mary Koukoui and Mark Baron

Department of Cell and Developmental Biology, Silberman Institute of Life Sciences, and the Center for Research on Pain, The Hebrew University of Jerusalem, Jerusalem 91904, Israel.
marshlu@mail.huji.ac.il; mary.koukoui@mail.huji.ac.il; mark.baron@mail.huji.ac.il

doi:10.1017/S0140525X21001928, e0

Abstract

Doubtless, the conscious brain integrates masses of information. But declaring that consciousness imply "emerges" when enough has accumulated, doesn't really explain how first person experience is implemented by neurons. Moreover, empirical observations challenge integrated information theory's (IIT) reliance on thalamo–cortical interactions as the information integrator. More likely, the cortex streams processed information to a still-enigmatic consciousness generator, one perhaps located in the brainstem.

General anesthetics are remarkable small molecules that transiently erase consciousness. They eliminate the sense of "me" together with sensory perception, movement, and memory, while leaving complex autonomic and housekeeping functions, and even early-stage sensory processing, mostly unchanged. Anesthetics also slash information content by synchronizing burst-firing in large populations of cortical neurons (slow-wave electroencephalogram [EEG]), and lessening functional connectivity (Pal et al., 2020; Volgushev, Chauvette, & Timofeev, 2011). Both changes should sharply reduce Φ, the integrated information theory's (IIT) metric of consciousness. Two contrasting models attempt to explain how anesthetics work: direct suppression of spinal and cortical activity, and interference with dedicated circuitry-designed evolutionarily to switch between consciousness and unconsciousness (Minert, Yatziv, & Devor, 2017). Both models meet IIT's requirement that reduced Φ should reduce consciousness. Sometimes, however, when cortical Φ is expected to collapse, no equivalent loss-of-consciousness is observed. Some examples:

- Unilateral intracarotid delivery of anesthetics (Wada test) induces slow-wave cortical EEG bilaterally. This should drastically reduce Φ, but subjects retain consciousness and memory formation (Halder, Juel, Nilsen, Raghavan, & Storm, 2021).
- Cortical slow-wave EEG also dominates non-rapid eye movement (NREM) sleep, presumably with low Φ. Yet NREM sleep supports dreaming, including the subjective sense of "being there," just like in REM sleep with its wake-like EEG (McNamara et al., 2010).
- IIT sees the thalamo–cortical system as the maximally differentiated/integrated "main-complex" that generates high Φ (Tononi & Sporns, 2003). However, large thalamic lesions in animals and humans can be compatible with retained or recovered consciousness (Fuller, Sherman, Pedersen, Saper, & Lu, 2011; Jimenez Caballero, 2010; Sakamoto, Okubo, Kanamaru, Suzuki, & Kimura, 2015).
- Large occipital cortex lesions surely reduce Φ. But the result is perceptual blindness, not impaired consciousness. It's like turning off the lights. The unawareness of visual stimuli is like the unawareness we all have of magnetism and ultraviolet stimuli. A skill is lost, but "I" am still present. Similarly, surgical hemispherectomy, prefrontal lobotomy, and congenital cortical thinning may cause disability, but not necessarily disordered consciousness (Feuillet, Dufour, & Pelletier, 2007).
- Auras in epilepsy, and direct cortical stimulation, are almost never painful, and cortical lesions rarely if ever cause analgesia (Penfield, 1975). However, broad consensus sees consciousness as essential for experiencing pain. It follows that cortex is not necessary for experiencing pain or being conscious.
- Conversely, stimulating certain subcortical sites does evoke pain reliably, and small brainstem lesions frequently induce unconsciousness, including loss of pain. Simple logic demands that the brainstem, perhaps its super-integrative isodendritic core, should be a lead candidate as "main-complex" and generator of high Φ and consciousness. But, in practice, proposals that the "primitive" brainstem might be primary are rarely taken seriously and are typically dismissed with metaphors like "the brainstem is a switch…" or "just a power-supply…." Ignoring the fact that mitochondria supply each neuron's power, should feelings of indignity drive scientific theory? Consider the following realignment of the roles of cortex versus brainstem in the generation of consciousness.

(1) The brainstem did appear early in evolution. But, with so little modern research dedicated to it, are we certain it stagnated ever since? In all vertebrates including man, the brainstem exerts broad control over the entire neuraxis including active control of state-of-consciousness. And its unique architecture, particularly the isodendritic structure of neurons in its reticular core, appears specialized for multimodal integration (Hobson & Scheibel, 1980; Ramon-Moliner & Nauta, 1966). Can one confidently reject the possibility that the raw feel of being conscious resides here, not in the cortex? Note also the consequences of this question for vegetative patients with an intact brainstem who appear to be awake but are currently presumed to be unconscious and incapable of experiencing pain.

(2) Over time, species evolved ever improving faculties to optimize niche-adapted engineering skills in the realms of sensory processing, motor performance, memory, and cognitive/predictive abilities. This process accelerated in early mammals with the appearance of cortical brain architecture. The "cortical" arrangement of neurons, in contrast to