Conflict of interest. None of the authors declares any conflict of interest.

Notes

[1] Addressing all misrepresentations in the target article would require a response just as long.

[2] The fundamental identity of IIT is explicitly between an experience, with all its specific properties, and a cause–effect structure. It should be obvious that an experience cannot be identical to a single number, Φ (>0).

[3] The reference is to Wagner (2006), discussed and quoted in Haun and Tononi (2019).

[4] In accordance with the principle of maximal existence (Tononi, 2015). The border separates units inside and outside the PSC even though they may be connected.

[5] Findlay et al. (2019); Oizumi et al. (2014).

[6] In short, I only have immediate evidence for my own experience. Everything else is an inference, including the presence of consciousness in other human beings (in philosophy, this is known as the problem of other minds).

[7] If anything, rather than "panpsychist," IIT is much more restrictive than many functionalist notions about consciousness that would ascribe consciousness to intelligent machines that can broadcast information globally or monitor the trustworthiness of sensory information (Dehaene, Lau, & Kouider, 2017).

[8] This is a typical initial reaction to IIT even when all five postulates are duly considered (unlike Merker et al.). One might concede that certain brain regions may be uniquely suited to satisfying those properties and provisionally accept the empirical evidence indicating that those regions may constitute the PSC. But one still cannot imagine how a physical system – even one that is a maximum of intrinsic, structured, specific, unitary cause–effect power – would then "give rise" to experience. This attitude is understandable because as neuroscientists (and generally as scientists), we take the extrinsic perspective: We take a physical substrate – the brain – as our starting point and try to figure out how it works and performs its various functions. And, from this extrinsic perspective, there seems to be no way to derive the very fact that consciousness exists – the existence of an intrinsic, phenomenal perspective. This is why, when it comes to consciousness, the starting point must be consciousness and its properties, not a physical substrate. What we should expect of physics is to account for the properties of experience, not to "give rise" to them. Phenomenology is first and physics second: Physics is our way to characterize operationally – from within our own consciousness – what we infer to exist independent of it (Ellia et al., 2021).

[9] This point was made, for example, in Edelman and Tononi (2000).

[10] Just to give a sense, the cause–effect structure unfolded from a grid of eight binary units explored in Haun and Tononi (2019), to account for spatial experience is composed of 133 distinctions and $>10^{40}$ relations. As explicitly stated by the explanatory identity of IIT, an experience corresponds to an (immensely rich) cause–effect structure, not to a single number Φ, which only quantifies its irreducibility.

[11] Merker et al. mention early work that introduced a measure of complexity C_N (Tononi, Sporns, & Edelman, 1994) which is sensitive to the dual requirements of information and integration (based on the compositional sum of average mutual information between subsets and the rest of a system) and has widespread applicability. They point out approvingly that C_N (unlike Φ) was not claimed to be a measure of consciousness. This is correct, because it was already clear then that all essential properties of consciousness need to be satisfied in physical terms, among them intrinsicality and exclusion. In fact, other early work attempted to capture these other properties by introducing a measure of functional clustering to identify the borders of a maximally integrated entity (Tononi, McIntosh, Russell, & Edelman, 1998). But it was also clear that a proper measure of consciousness should eventually account for all its properties at once, including composition and information, and should be state-dependent rather than an average. Only in this way can we account for the fact that every experience exists in full, immensely structured and fully specific, "here and now" or, as elegantly put by Merker et al., "in a perpetual pristine present for which no content outside the state it is in exists" (sect. 3, para. 13).

References

Aaronson, S. (2014) Why I Am Not An Integrated Information Theorist (or, the Unconscious Expander).

Barbosa, L. S., Marshall, W., Albantakis, L., & Tononi, G. (2021). Mechanism integrated information. *Entropy 23*(3), 362.

Barbosa, L. S., Marshall, W., Streipert, S., Albantakis, L., & Tononi, G. (2020). A measure for intrinsic information. *Scientific Reports 10*(1), 18803. doi: 10.1038/s41598-020-75943-4

Dehaene, S., Lau, H., & Kouider, S. (2017). What is consciousness, and could machines have it? *Science (New York, N.Y.) 358*(6362), 486–492.

Edelman, G. M., & Tononi, G. (2000). *A universe of consciousness how matter becomes imagination*. Basic Books.

Ellia, F., Hendren, J., Grasso, M., Kozma, C., Mindt, G., Lang, J., … Tononi, G. (2021) Consciousness and the fallacy of misplaced objectivity. *Neuroscience of Consciousness, 2021*(2), niab032.

Findlay, G., Marshall, W., Albantakis, L., Mayner, W., Koch, C., & Tononi, G. (2019) Dissociating Intelligence from Consciousness in Artificial Systems – Implications of Integrated Information Theory. Paper presented at the TOCAIS 2019 (Towards Conscious AI Systems), Stanford, CA.

Haun, A., & Tononi, G. (2019). Why does space feel the way it does? Towards a principled account of spatial experience. *Entropy 21*(12), 1160.

Hoel, E. P., Albantakis, L., Marshall, W., & Tononi, G. (2016). Can the macro beat the micro? Integrated information across spatiotemporal scales. *Neurosci Conscious 2016*(1), niw012. doi: 10.1093/nc/niw012

Hoel, E. P., Albantakis, L., & Tononi, G. (2013). Quantifying causal emergence shows that macro can beat micro. *Proceedings of the National Academy of Sciences of the United States of America 110*(49), 19790–19795. doi: 10.1073/pnas.1314922110

Marshall, W., Albantakis, L., & Tononi, G. (2018). Black-boxing and cause–effect power. *PLoS Computational Biology 14*(4), e1006114. doi: 10.1371/journal.pcbi.1006114

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. doi: 10.1371/journal.pcbi.1003588

Song, C., Haun, A. M., & Tononi, G. (2017). Plasticity in the structure of visual space. *eNeuro 4*(3), 1–7. doi: 10.1523/ENEURO.0080-17.2017

Tononi, G. (2012). The integrated information theory of consciousness: An updated account. *Archives Italiennes de Biologie 150*(2/3), 56–90.

Tononi, G. (2015) Integrated Information Theory Scholarpedia.

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. doi: 10.1038/nrn.2016.44

Tononi, G., McIntosh, A. R., Russell, D. P., & Edelman, G. M. (1998). Functional clustering: Identifying strongly interactive brain regions in neuroimaging data. *NeuroImage 7*(2), 133–149. doi: 10.1006/nimg.1997.0313

Tononi, G., Sporns, O., & Edelman, G. M. (1994). A measure for brain complexity: Relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Sciences of the United States of America 91*(11), 5033–5037. doi: 10.1073/pnas.91.11.5033

Wagner, M. (2006). *The geometries of visual space*. Psychology Press.

Consciousness, complexity, and evolution

Walter Veit

School of History and Philosophy, The University of Sydney, Sydney, NSW 2006, Australia.
wrwveit@gmail.com
https://walterveit.com/

doi:10.1017/S0140525X21001825, e61

Abstract

The idea that consciousness and complexity are closely related has been a major driver of the popularity of integrated information theory (IIT) of consciousness, despite its major formal, phenomenological, and neuroscientific shortcomings. Here, I argue that we can recover this intuition by replacing its biologically neutral notion of complexity with an evolutionary one that I shall dub “pathological complexity.”

The evolution of consciousness and complexity have for a long time been seen as inherently linked. In thinking about the possibility of sentience in machines and nonhuman organisms, our collective willingness to attribute phenomenological experience to them seems to be primarily driven by a measure of their complexity. For organisms such as jellyfish that are too simple in terms of their behavioral repertoire and nervous system organization, it seems all but impossible to attribute them the rich kind of mindedness that we associate with human consciousness. Furthermore, it’s precisely the biological complexity of cephalopods that has led to calls for a recognition of their sentience and hence for them to be included in animal welfare science and legislation (Browning, 2019; New England Anti-Vivisection Society et al., 2020). It’s in this context that Tononi’s (2004, 2005, 2008, 2012) integrated information theory (IIT) has become extremely popular in the public, despite its comparative unpopularity within the larger scientific community.

The identification of consciousness with *integrated information* itself, once a certain complexity of measured by “phi” Φ is reached, has long been viewed as problematic – indeed, untestable. Merker, Williford, and Rudrauf provide us with a barrage of well-worked out arguments against IIT, showing that it fails for a variety of formal, phenomenological, and neuroscientific reasons. Nevertheless, the popularity of IIT is not entirely ill-motivated and my goal here is to recover some of its merits while offering an additional criticism of the project. IIT has obvious virtues such as the *in-principle* applicability to systems very different from ourselves, allowing them to be placed along a continuum from more to less conscious (Tononi & Koch, 2015). Furthermore, the idea of beginning with minimal theoretical commitments and a very simple model is familiar from other sciences studying complex phenomena (Veit, 2019a, 2019b). Aided by the perceived link between complexity and consciousness, one may be forgiven for thinking that IIT provides us with a good starting point. Yet, instead of providing us with a simple and general framework that can be tested and improved, Merker, Williford, and Rudrauf convincingly argue that the entire framework is resistant to change and empirical progress, thus holding us back, rather than enabling us to move forward.

An additional problem of the theory is its failure to take evolutionary considerations on board; and this may be its greatest problem yet. Now, there is admittedly something deeply right about IIT. Of course, complexity matters! No one can deny that and the explicit acceptance of a gradualist model of consciousness seems to lend itself very well to evolutionary considerations (Veit & Huebner, 2020). But it cannot *just* be a mere one-dimensional scale of information integration that matters for subjective experience. The biological world not only contains gradations, but also varieties, and this should be part and parcel of a biological account of consciousness. The IIT simply asserts that it’s their chosen measure of complexity that matters – more so that it’s all that matters – but doesn’t offer any compelling reasons why this should be so (Browning & Veit, 2021). As Dan Dennett once said at the 2017 NYU Animal Consciousness conference: “Complexity matters, but which complexity?” In order to determine which complexity *matters* for consciousness, we cannot avoid the teleonomic question of what consciousness is *for*. But this question has deliberately received very little attention within both the IIT framework and the science of (human) consciousness, despite the Darwinian insight that it’s only once we address the function of a complex biological phenomenon that we can truly begin to understand it.

In an effort to begin with a minimal and theoretically neutral model, IIT deliberately avoids commitments to the evolutionary rationale of consciousness. But this silence on biological matters is ultimately the primary reason why the theory must be abandoned. From an evolutionary perspective, we must focus on the complexity of a new biological mode of being (Browning & Veit, 2021). This is why an investigation of consciousness as a *response to*, rather than a mere *product of* complexity ought to be how we can begin a true biological science of consciousness; one that emphasizes varieties and gradations instead of an all-or-nothing quality, as was advocated early on by Griffin (1976) and more recently by Godfrey-Smith (2019) in their emphasis on the different lifestyles of animals. Instead of asking for a neutral measure of complexity that constitutes consciousness, we deliberately ask for an evolutionary loaded sense of living complexity that makes consciousness *worth* having – this I shall refer to as “pathological complexity”: the complexity of an organism’s life history (Veit, 2021). Does that just replace one theoretically intractable notion of complexity with another? Not quite, because there already exists a science that has done the work for us. Pathological complexity can be operationalized as the computational complexity of the optimization problem studied by state-dependent or state-based behavioral and life history theory. It is a biological measure of complexity that scales up alongside organism’s degrees of freedom and that exploded during the Cambrian (Veit, 2022). Based on this notion of evolutionary complexity, we can build an alternative framework for the study of consciousness on what I call the pathological complexity thesis:

Pathological complexity thesis: The function of consciousness is to enable the agent to respond to pathological complexity.

Similar to the IIT, the pathological complexity thesis offers us a general theoretical framework and model for thinking about consciousness. As such, it will inevitably share some of the problems all theories at this level of generality are faced with. But it overcomes one central problem of the IIT: This is a sense of complexity that makes evolutionary *sense*. And it’s this important feature that can serve as a scaffold for future bottom-up approaches that take Darwinian thinking seriously, emphasizing both gradations and varieties of subjective experience in animal life, and allowing us to make predictions about the phenomenological complexity of other animals – which can then feed back into our understanding of the pathological complexity they evolved to deal with.

Financial support. This research was supported under Australian Research Council’s Discovery Projects funding scheme (project number FL170100160).

Conflict of interest. None.

References

Browning, H. (2019). What is good for an octopus? *Animal Sentience*, *26*(7). https://doi.org/10.51291/2377-7478.1476

Browning, H., & Veit, W. (2020). The measurement problem of consciousness. *Philosophical Topics*, *48*(1), 85–108. https://doi.org/10.5840/philtopics20204815

Browning, H., & Veit, W. (2021). Evolutionary biology meets consciousness: Essay review of Simona Ginsburg and Eva Jablonka's *The Evolution of the Sensitive Soul. Biology & Philosophy*, *36*(5), 1–11. https://doi.org/10.1007/s10539-021-09781-7

New England Anti-Vivisection Society, *et al.* (2020). Petition to include cephalopods as "animals" deserving of humane treatment under the public health service policy on humane care and use of laboratory animals. Harvard Law School Animal Law & Policy Clinic, 1–30. https://doi.org/10.13140/RG.2.2.27522.30401

Godfrey-Smith, P. (2019). Evolving across the explanatory gap. *Philosophy, Theory, and Practice in Biology*, *11*, 1–24. https://doi.org/10.3998/ptpbio.16039257.0011.001

Griffin, D. R. (1976). *The question of animal awareness: Evolutionary continuity of mental experience*. Rockefeller University Press.

Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, *5*(42), 1–22.

Tononi, G. (2005). Consciousness, information integration, and the brain. *Progress in Brain Research*, *150*, 109–126.

Tononi, G. (2008). Consciousness as integrated information: A provisional manifesto. *Biological Bulletin*, *215*, 216–242.

Tononi, G. (2012). Integrated information theory of consciousness: An updated account. *Archives Italiennes de Biologie*, *150*, 290–326.

Tononi, G., & Koch C. (2015). Consciousness: Here, there and everywhere? *Philosophical Transactions of the Royal Society, B*, *370*, 20140167.

Veit, W. (2021). The evolution of knowledge during the Cambrian explosion. *Behavioral and Brain Sciences*, *44*, e47. https://www.doi.org/10.1017/S0140525X20000771

Veit, W. (2019a). Model pluralism. *Philosophy of the Social Sciences*, *50*(2), 91–114. https://doi.org/10.1177/0048393119894897

Veit, W. (2019b). Modeling morality. In L. Magnani, A. Nepomuceno, F. Salguero, C. Barés & M. Fontane (eds), *Model-Based reasoning in science and technology* (pp. 83–102). Springer. https://doi.org/10.1007/978-3-030-32722-4_6

Veit, W. (2022). *Health, Agency, and the Evolution of Consciousness*. Ph.D. thesis, University of Sydney. Manuscript in preparation.

Veit, W., & Huebner, B. (2020). Drawing the boundaries of animal sentience. *Animal Sentience*, *29*(13). http://doi.org/10.51291/2377-7478.1595

Functional theories can describe many features of conscious phenomenology but cannot account for its existence

Max Velmans

Department of Psychology, Goldsmiths, University of London, London SE14 6NW, UK.
m.velmans@gold.ac.uk
https://www.gold.ac.uk/psychology/staff/velmans/

doi:10.1017/S0140525X21001813, e62

Abstract

Merker, Williford, and Rudrauf argue persuasively that integrated information is not identical to or sufficient for consciousness, and that projective geometries more closely formalize the spatial features of conscious phenomenology. However, these too are not identical to or sufficient for consciousness. Although such third-person specifiable functional theories can describe the many *forms* of consciousness, they cannot account for its *existence*.

Merker, Williford, and Rudrauf have provided a thoughtful, and, in my view, decisive critique of the integrated information theory (IIT) claiming that "consciousness is one and the same as integrated information" (Oizumi, Albantakis, & Tononi, 2014; Tononi, 2008). Rather, Φ (the formal measure of integrated information within IIT) is one measure of network efficiency that can be applied to network information processing in general. For this reason, information integration efficiency can be doubly dissociated from consciousness. For example, there can be efficient information flows in complex economic, social, and transportation systems that are far removed from those usually thought to have a unified, integrated consciousness, and there is extensive evidence for efficient *unconscious* integrated information processing in systems that do have consciousness, namely human minds (see e.g., Kihlstrom, 1996; Velmans, 1991). If so, integrated information processing is not a sufficient condition for consciousness.

Given the inability of IIT to clearly demarcate systems (or subsystems) that don't have consciousness from those that do, Merker et al. suggest an alternative measure of what it is like to have a conscious "point of view" employing a form of projective geometry that specifies the "point-horizon" structure of human consciousness (Rudrauf et al., 2017). As with the other theories that focus on the projected, three-dimensional nature of conscious phenomenology (e.g., Lehar, 2003; Pereira, 2018; Revonsuo, 2006; Trehub, 2007; Velmans, 1990, 2008, 2009), this specification of self-location within a three-dimensional phenomenal world captures a central, but often ignored feature of human conscious phenomenology that has many important consequences for understanding consciousness (cf. Velmans, 2009). However, it does not follow that the ability to generate and implement such a geometry demarcates conscious entities from nonconscious ones. As with integrated information, double dissociations between systems that instantiate projective geometries and consciousness are easy to envisage. For example, one can devise complex *nonconscious*-guided missile systems that have sophisticated ways of computing and navigating their own location within a surrounding terrain. Conversely, although projective geometries can be used to formalize self-location and navigational aspects of human consciousness, there are many "qualia" of consciousness that cannot be formalized in this way, such as the taste of coffee or the smell of a rose.

Which leads to a deeper question: Is it possible, even in principle, to specify a third-person observable, functional principle that demarcates conscious entities from nonconscious ones? In Velmans (2009, 2012), I have surveyed many suggested criteria for distinguishing entities that are conscious from those that are not. Broadly speaking, theories about the distribution of consciousness divide into *continuity* and *discontinuity* theories. Discontinuity theories all claim that consciousness emerged at a particular point in the evolution of the universe. They merely disagree about which point. Most try to define the point of transition in functional terms, although they disagree about the nature of the critical function. Some think consciousness "switched on" only in humans, for example, once they acquired language or a theory of mind. Some believe that consciousness emerged once brains reached a critical size or complexity. Others believe it co-emerged with the ability to learn, or to respond in an adaptive way to the environment. However, it is well recognized in modern philosophy of mind that all such theories face the problem of "brute emergence." If consciousness was entirely absent before the emergence of a critical function, what is it about that function that suddenly creates consciousness?

In Velmans (2009, 2012), I have argued that such theories confuse the conditions for the *existence* of consciousness with the added conditions that determine the many *forms* that it can take. Who can doubt that verbal thoughts require language, or that full human self-consciousness requires a theory of mind? Without internal representations of the world, how could