

Consciousness Generates Agent Action

Jonathan Delafield-Butt¹ & Colwyn Trevarthen²

1. Laboratory for Innovation in Autism, University of Strathclyde, 50 George Street, Glasgow G1 1QE, Scotland, United Kingdom. jonathan.delafield-butt@strath.ac.uk. <https://www.strath.ac.uk/staff/delafieldbuttjonathandr/>
2. School of Psychology, Philosophy, and Language Sciences, University of Edinburgh, Dugald Stewart Building, 3 Charles Street, Edinburgh EH8 9AD, Scotland, United Kingdom. c.trevarthen@ed.ac.uk.

Invited commentary in response to Merker, Williford, and Rudrauf's

“The Integrated Information Theory of consciousness: A case of mistaken identity”

for *Behavioural and Brain Sciences*

ACCEPTED AUTHOR MANUSCRIPT

29th September 2021

For the full copy-edited version and important target article, please see

<https://www.cambridge.org/core/journals/behavioral-and-brain-sciences>

Delafield-Butt, J. and Trevarthen, C. (2021). Consciousness Generates Agent Action. *Behavioral and Brain Sciences*. In press.

Consciousness Generates Agent Action

Jonathan Delafield-Butt & Colwyn Trevarthen

Abstract

Consciousness directs the actions of the agent for its own purposive gains. It re-organises a stimulus-response linear causality to deliver generative, creative agent action that evaluates the subsequent experience prospectively. This inversion of causality affords special properties of control that are not accounted for in IIT, which is predicated on a linear, deterministic cause-effect model. IIT remains an incomplete, abstract, and disembodied theory without explanation of the psychobiology of consciousness that serves the vital agency the organism.

Keywords. agency, affects, action, consciousness, panpsychism, embodiment

Consciousness Generates Agent Action

Integrated Information Theory is a clearly defined physical model that addresses key aspects of consciousness (Oizumi, Albantakis, & Tononi, 2014; Tononi, 2004). It gives attention to an interdependency of operations and differences in spatiotemporal scales, and illustrates basic functions of integration and differentiation of experience. It is also sufficiently generalised to accept the possibility for conscious processes in non-traditional systems, allowing for consideration of underpinnings that are panpsychological (Basile, 2010; Goff, 2019). But Merker, Williford, and Rudrauf (2021) do well to point out that IIT suffers errors of logic that render it incoherent as a primary theory of consciousness, and we agree it cannot be used to identify consciousness, by the rules of its theory. Beyond this concern of its logic, IIT suffers another structural shortcoming for a theory of consciousness. It omits the most important psychobiological function of consciousness as *generative*, an affective animation that guides physical movement for vital purpose (Reed, 1996; Trevarthen & Delafield-Butt, 2017a; Clark, 2019). This fundamental property of consciousness, with its aspects of affective evaluation made with knowledge, are neglected.

In IIT there is no perspective or purpose of a conscious agent. The ‘information’ it attends to in its measures of coherence are not anchored in the vital necessities of life, the essential evolved function of consciousness to serve the purpose of the organism as an agent with choice, aware of options for action (or inaction), and able to consolidate information for vital gains. This is the psychobiological essence of consciousness, which IIT has missed. Without adaptation of purpose to serve and protect the vitality of the organism, what is the point of consciousness? We perceive IIT as disembodied, abstract, and lacking the affective vital meaning of feelings, a cornerstone of all living consciousness (Langer, 1967; Panksepp & Biven, 2012; Solms, 2021; Vandekerckhove & Panksepp, 2011).

We can learn from detailed attention to the beginning of conscious human life in body movement. Infants reach to engage the future with innate and learned knowledge of its consequences – their neuro-motor systems are structured with an inherent prospective awareness, an intrinsic knowledge of self-generated futures with their pleasurable benefits and fearful dangers, which are experienced and remembered (Delafield-Butt et al., 2018; Trevarthen & Delafield-Butt, 2017a, 2017b). This is the essence of the primitive brainstem-mediated consciousness, evident in the intentional movements of the human foetus from the second trimester onwards (Delafield-Butt & Gangopadhyay, 2013). An active and sensible agency shapes that informs experience in the development of a human consciousness (Trevarthen & Delafield-Butt, 2017a). It does not require cortical processing.

But the experiments to validate Tononi’s IIT theory focus on interpretations of activity recorded in the neocortex – an area of the brain not required for basic human consciousness (Damásio, 2010; Merker, 2007). The theory lacks reference to psychobiological foundations in lived, embodied, affective experience of agile agency in the material world, perceptual evidence of which structures, informs, and unifies conscious experience in ‘The Self As Agent’ (Macmurray, 1959). We are surprised Merker has not recovered his own empirical and theoretical ideas into the debate. All biological evidence suggests consciousness must be purposive, needing evaluation of choices as affordances for future action. These are the embodied, ecological ingredients of the experience of life that make sense of conscious

awareness and subjective feelings (Langer, 1967). Merker set these out clearly in his case for the ‘liability of mobility’ in the evolution of consciousness (Merker, 2005). In all vertebrates, the primary integrator of exteroceptive, proprioceptive, and interoceptive information is in the brainstem, where affective evaluation, learning and memory motivate choice in the conscious agent for selection in volitional body movement (Merker, 2007).

Stephen Grossberg (2021) is right when he says, “*we consciously see in order to be able to reach*”. Sight is imaginative, about what to do, where and how. But, IIT adapts a linear causality, a ‘cause-effect’ logic in its account of how information and its effects at different locations in the brain are integrated and evaluated. Consciousness affords a further logic, an ‘action-effect’ causality of experienced *self-generated* agent action, animated with feeling and vital purpose. All conscious agents create their own causal flow, and this structures how their sensory information across all modalities becomes structured. We view Merker, Williford, and Rudrauf’s (2021) recalling of the perspective view of consciousness as an attempt to re-embodiment the concept of consciousness, to put the conscious agent back into the picture as a causal structure. This is a needed step.

We also draw caution to their conflation of panpsychism and IIT. Panpsychism is not IIT, but IIT has come to adopt elements of the metaphysics into its system, *ad hoc*. In our view this is not a justified integration of this complex metaphysics into a technical theory, but ‘cherry-picks’ a crude idea of ‘mind everywhere’ or accepts an equation of ‘physicalism as panpsychism’ (Strawson, 2006) without taking due care to understand the depths and structure of the philosophy assumed. Panpsychism and IIT require distinction.

The metaphysical frameworks of the two major panpsychist theorists, Leibniz (1716) and Whitehead (1929), present a highly structured account of mind as coincident with occasions, or events. These, like Merker’s original view of the “liabilities of mobility” (Merker, 2005), require an integration of experience and felt evaluation of affordances for choice of action – enabling volition and the intentional action of a conscious agent (Delafield-Butt, 2014). These complex metaphysical frameworks require not the empty, abstract, disembodied consciousness of IIT, but stand in agreement with more established notions of an ecological psychology of agents nested in environmental affordances (Reed, 1996). The benefit and reach of panpsychism means that conscious agency is not restricted to the level of whole organisms, be they humans, cats, mice, or worms, but to their organic subsystems, too. And also to simpler component organisms such as cells and organelles. Generalised principles of IIT that extend beyond the animal kingdom are potentially important, but as described they remain devoid of purpose, which leaves them unfaithful to panpsychism, and phenomenology as well. That IIT invites scientific discourse on this topic is welcome, but the outcome is somewhat incoherent, and incomplete.

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 Charles Darwin (1872) and the contemporary science of cycles of prosperity and extinction in life on earth surveyed by David 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.

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 P. Basile, J. Kiverstein, & P. Phemister (Eds.), *The Metaphysics of Consciousness* (pp. 93–112). Cambridge University Press.
- 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.
- Clark, A. (2019). Consciousness as Generative Entanglement. *Journal of Philosophy*, 116(12), 645-662.
- Delafield-Butt, J.T. (2014). Process and Action: Whitehead's Ontological Units and Perceptuomotor Control Units. In S. Koutroufinis (Ed.), *Life and Process* (pp. 133-156). Berlin/Boston: 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., and 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, 2021 *Conscious Mind Resonant Brain*. Oxford: 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*. Baltimore: John Hopkins Press.
- Leibniz, G. (1716). *La Monadologie*
- Macmurray, J. (1959) *The Self as Agent* (Volume I of *The Form of the Personal*) London: 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
- Merker, Williford, and Rudrauf (2021). *Behavioral and Brain Sciences* (forthcoming). The Integrated Information Theory of consciousness: A case of mistaken identity (Target article).
- 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*. New York: Norton.
- Reed, E. S. (1996). *Encountering the World: Toward an ecological Psychology*. Oxford: Oxford University Press.
- Solms, M. (2021). *The Hidden Spring: A journey to the source of consciousness*. London: Profile Books.

- Strawson, G. (2006). Realistic monism: Why physicalism entails panpsychism. In A. Freeman (Ed.), *Consciousness and Its Place in Nature*. (pp. 3-31). Charlottesville, VA: Imprint Academic.
- Tononi, G. (2004). An information integration theory of consciousness. *BMC Neuroscience*, 5(1), 42
- Trevarthen, C., & Delafield-Butt, J. T. (2017a). Development of Consciousness. In B. Hopkins, E. Geangu & S. Linkenauer (Eds.), *Cambridge Encyclopedia of Child Development* (pp. 821-835). Cambridge: Cambridge University Press.
- Trevarthen, C. and Delafield-Butt, J. (2017b). Intersubjectivity in the imagination and feelings of the infant: Implications for education in the early years. In, E. Jayne White and Carmen Dalli (Eds.), *Under-three Year Olds in Policy and Practice* (Policy and Pedagogy with Under-three Year Olds: Cross-disciplinary Insights and Innovations). Heidelberg: Springer, pp. 17-39.
- 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 S. Robson & S. F. Quinn (Eds.), *International Handbook of Young Children's Thinking and Understanding* (pp. 3-18). Abingdon, Oxfordshire & New York: Routledge.
- Vandekerckhove, M., & Panksepp, J. (2011). A neurocognitive theory of higher mental emergence: From anoetic affective experiences to noetic knowledge and auto-noetic awareness. *Neuroscience & Biobehavioral Reviews*, 35 (9), 2017-2025.
- Whitehead, A. N. (1929). *Process and Reality*. New York: Macmillan.

Conflicting Interests

There are no conflicting interests.

Funding Statement

None