

Timothy O'Connor

Consciousness, Free Will, and the Sciences of the Mind

Abstract In his review of the trio of philosopher-scientist dialogues on the nature and capacities of the human mind, Paul Thagard (2018) advocates clearly and forcefully for a fairly extreme position, which he advances as preferable to an equally extreme alternative. I will suggest a middle path that becomes attractive when one attends not just to the range of data now pouring forth from the sciences of mind but also to our own experience as minded individuals.

Keywords consciousness, free will, emergence, qualia, dualism, cognitive science, first-person perspective

In his critical review of the essays in this issue's special symposium, Paul Thagard (2018) begins on a note with which I heartily agree: Philosophy and science have distinct and important contributions to make to a proper understanding of ourselves as biologically-rooted, thinking, feeling, and intentionally acting beings. But what does philosophy bring to the table in this discussion? According to Thagard, philosophy's distinctive contributions concern method, and not content. Philosophy is more general and in places more normative than science; it also can provide critical perspective on the methodology of the sciences, which is particularly helpful at the inter-disciplinary crossroads of the cognitive sciences. All of this seems correct to me. But it leaves out philosophy's most important contribution to an understanding of the human mind.

Human beings are both objects and subjects; accordingly, adequately understanding ourselves requires study and theorizing that integrates the third- and first-personal points of view. An important contribution of philosophy to its dialogue with the human sciences, and one that Thagard appears to dismiss, is to systematize the first-person perspective on the nature and powers of the mind. This longstanding philosophical project is famously contentious, and it is understandable that *scientists* should prefer to confine their studies to data that can be verified and measured inter-subjectively in accordance with familiar methods. But for those who have the goal to integrate all sources of information into a

Timothy O'Connor (✉)

Department of Philosophy, Indiana University, Bloomington, IN 47405, USA

E-mail: toconnor@indiana.edu

comprehensive understanding of our mental lives, it would be more than short-sighted to dismiss the reality of first-personal phenomena. For first-personal sensory-cognitive experience of scientific inquirers is the portal through which all scientific data comes—a point to which I will return.

If we adopt this methodological stance of seeking to weave together first- and third-personal perspectives, what should we say about the three core questions addressed by Thagard and our other symposiasts?

1 What Is Consciousness?

Thagard suggests that we analyze the *concept* of consciousness by identifying standard examples of conscious states, their typical features, and the sorts of explanations in which the concept of consciousness is commonly used. This multi-pronged procedure seems sensible enough. However, his unargued assumption that consciousness must be a wholly mechanistic and materialist phenomenon informs his list of its typical features: awareness, shifts in attention, beginnings and ends of states, and a degree of unity (see Thagard 2018). These are structural-relational, rather than intrinsic features, the sort one would cite for something one grasps somewhat indirectly, and lacking an understanding of its deeper, fundamental character. But from a first-person perspective, what seems most characteristic of consciousness is that it is constituted by qualities that are inextricable from that very perspective: they are qualities concerning how things *seem to* the conscious subject—how objects around one seem to be in terms of shape, color, and sound, and how oneself seems to be in terms of beliefs, desires, emotions, and pain/pleasure. And the awareness of those qualities is distinctive, too: it is (or presents itself as) maximally direct, not mediated by representations in the way that our awareness of the objects of our environment is mediated by our experiential representations of them.

This first-person perspective and its implications for an uncompromising materialism have been well articulated by many authors, including Frank Jackson, in his thought experiment regarding Mary, an imagined future neuroscientist of color vision, and Thomas Nagel, in his reflections on the limits of our ability to grasp how things seem to bats and other creatures who cognitively and perceptually differ from us to a great extent (Jackson 1982; Nagel 1974). These discussions bring out a seemingly insuperable difficulty in identifying the qualities of subjective experience with plausible neural candidates.

Consider the experience of hearing a short musical note from a trumpet. As one attends to that portion of one's auditory field, one has a substantive and determinate grasp of the phenomenal sound quality, enabling one to infer straightaway that it is intrinsically *simple* in the following sense: It is a smooth,

uniform, homogeneous quality, one that is not further dissectible into structural components or patterns. This simple, non-structural quality differs primitively from other, equally simple, determinate phenomenal sounds. And this feature of the intrinsic simplicity of the consciously-perceived elements of conscious experiences thoroughly pervades the phenomenology of perceptual experience in all of the basic sense modalities. (Though I can't argue it here, the range of such subjective qualities is not restricted to "raw feels," but plausibly includes intentional properties of conscious intentional states.) Such structurally simple conscious qualities cannot plausibly be identified with any physical state. A musical note from an instrument generates a complex waveform that is processed by the ear's cochlea and the auditory pathways and structures of the nervous system. This is an event involving a highly structured concatenation of electro-chemical impulses and exchanges taking place among billions of neurons in the auditory cortex, along with highly complex chemical activity taking place across cell membranes and synaptic connections. The felt sound quality in the auditory experience lacks any isomorphic structure to correspond to the neurobiological (let alone microphysical) complexity—it lacks any of the grainy, particulate, discontinuous spatiotemporal structure or hierarchical composition of its neurobiological and ultimately microphysical correlates.

Thus, simple reflection on the qualitative character of conscious experiences and other conscious states suggests that they instantiate qualities that are not identical to or "realized by" physical states of the brain, though they are doubtless caused and sustained by some such physical states. What are the implications of this result for our understanding of the human mind?

2 What Is Mind?

Thagard is well aware of this common, "naïve" approach to understanding consciousness. But he dismisses it on the grounds that it leads straight to a bankrupt general philosophy of mind, mind-body dualism, on which the mind is a wholly immaterial substance interacting with the brain (or alternatively, to dualism's radical cousin, idealism, that reduces the physical side of this partition to the mental). But why this forced choice between two stark alternatives?

Students of the Western philosophical tradition trace Thagard's materialism and the "mind-body"/ "substance" dualism that he rejects to the ancient Greeks Democritus and Plato, respectively. But there was also the mediating position of Aristotle. The details of Aristotle's understanding of human beings are no more important for my purposes than those of Democritus's account are for Thagard's. What I wish to draw from Aristotle is simply his broad vision of human nature: wholly rooted in the physical world, yet not reducible to the body's constituent

parts and their interactions. How might we develop this vision in the 21st century?

A human mind is not (as Plato held) a self-contained substance that is a dual to the body and its brain—well and good. It is wholly composed of physical parts (as we now know, at a certain level of decomposition, neurons). But the interactions of those parts result in part in irreducible, conscious states of the whole human subject. To put it in a contemporary idiom, these states *emerge* from those interactions, where “emerge” signals a combination of dependency and irreducibility.

Many materialist philosophers judge emergentism to be no less problematic on broadly empirical grounds than mind-body dualism. But there has been an antireductionist wave in some of the sciences, such that it is widely thought that complex, organized systems involve an interplay of “bottom up” and “top down” properties and explanatory principles (Laughlin 2005; Noble 2006; Tse 2013; Ellis 2016). It is a difficult question how precisely to characterize emergent properties and processes.¹ Perhaps the way “higher level” phenomena emerge from lower level phenomena differs from case to case. What seems to be distinctive about emergent mental phenomena is that they are irreducible in a very strong sense: since conscious states appear not be even token identical to physical states, they are not physically “realized.” This strong form of irreducibility implies, in turn, that the *dependency* of conscious mental states upon physical states must be (merely) causal, and not constitutive. A human mind will then be an evolving interplay of complex information-bearing physical states and conscious mental states that confer a strong perceived unity on the whole person. To say anything more about how to understand this unity, we would need to take a stand on contentious ontological issues concerning the general categories of property and individual/substance.²

3 What Is Free Will?

Thagard’s materialist assumption and false choice regarding the alternative also mar his discussion of whether human beings have free will “in the traditional sense”—whether it is sometimes true that we are able to choose any of a plurality of actions. He assumes that all our choices, including deliberate, conscious choices, are entirely the resultants of neural mechanisms we have not yet identified. He appears further to assume that these mechanisms always evolve

¹ For discussion and proposals, see O’Connor and Wong (2015), Gillett (2016), and Humphreys (2016).

² I develop my own position on this question in O’Connor (2018).

deterministically (or near-enough-to-determinism-as-makes-no-difference). The (sole) alternative to these assumptions, he assures us, is the assumption that human beings are “supernatural agents,” by which I take it he means non-physical minds as understood by mind-body dualists.

But the alternative, emergentist view noted above easily squares with the reality of human freedom: the capacity to consciously choose for reasons can emerge from lower-level neural mechanisms. Supposing ourselves to have such a power of conscious choice does not make us out to be supernatural entities: “Natural” is not synonymous with “reductionistic materialism.” It can also be causally influenced by myriad such factors without those factors uniquely determining the choices I make: Causation is not causal determination. For an event *e*'s to be caused is, roughly, for there to be factors that contributed to *e*'s occurring and without which *e* would not have occurred, or would have been less likely. It is consistent with an event's having a cause, or many causes, in this sense, that it might not have occurred, even given the total situation (including those causes)—consistent, that is to say, with its not being the uniquely possible outcome—with its not being determined to occur. I can have an evolving propensity less than 1 to choose in particular ways up to the moment (short interval) of choice itself. So, to be more concrete, my wanting to drink beer and make merriment can be a significant cause of my going to the pub, a factor that made it more or less likely that I would do so, even if it remained possible that I stayed home and worked on an overdue article (in which case a very different desire would have caused my action). Whichever I do will have causes, without their precluding my acting differently. Given this, Thagard's invitation to us (see Thagard 2018) to equate “not determined” with “random” appears groundless.

Yes, but is there evidence that we have such a power of choice? Well, we can use the word “evidence” in more or less restrictive ways. But one thing we can say is that we all regularly have experiences as of exercising a power to choose, and the accompanying belief that we are able to exercise that power in a multiplicity of ways—that we can choose otherwise—also appears to be a natural belief. It is also a belief that is implicated in our interpretation of social reality, as we regularly praise and blame people for how they exercise their capacity to choose based on the assumption that it was open to them to have chosen otherwise.

Experiences of and natural beliefs regarding our capacity to make conscious choices could be illusory. But we generally, and rightly, take our most fundamental, universal beliefs or dispositions to belief to be reasonable, absent compelling evidence to the contrary. And I suggest we have nothing like such countervailing evidence when it comes to the reality of human free will. There has been a spate of interesting recent work in neuroscience and cognitive and social psychology that some take to be strong evidence against human free will.

(Daniel Wegner's *The Illusion of Conscious Will* is a slightly dated, but still useful and engaging summary of some of this work.) But when one carefully considers these studies, the evidence is not very impressive at all.³ In some cases, the studies merely show how malleable our beliefs are after the fact about what we did or did not consciously decide to do; in others, that there are likely external, *non-determining* causal influences on conscious choices that the agents are not, or are not fully, consciously aware of; and still others involve unusual cognitive/volitional disorders, such as the "anarchic hand" syndrome, where the agent is acting purposefully but in a way that is not consciously mediated and that is seemingly at odds with his conscious goals. (Some would not only mention but give pride of place to the neuroscientific studies of Benjamin Libet and various successors. But multiple philosophers and scientists have convincingly shown that Libet's studies do not have the anti-free-will implications that Libet thought they did.)⁴ These studies do lend evidence to the theses that human beings are highly fallible about their own behavior; that we are subject to causal influences, sometimes unwittingly; and that our control over our own behavior is a highly fragile thing, and can be diminished or destroyed as a result of subtle disturbances to our neural architecture. But why would anyone take evidence of limits and fragility to indicate that free will is an illusion? That conclusion could be drawn only by one persuaded that "the traditional idea of free will" is the idea of a magical, god-like capacity that transcends all nature—by one inclined to accept as exhaustive the stark binary of "pure"/no freedom. Anyone familiar with imaginative human writing stretching back to ancient wisdom literature of many cultures would be unlikely to accept such a false choice.

Thagard gives an interesting and original reason of his own for doubting that human beings have free will. As he notes, much of what we do is automatic, not consciously deliberated upon. It is plausible to think that, if we exercise free will, we do so when we are consciously aware of the options before us and consciously decide which option to take. However, contends Thagard,

People rarely choose to make their decisions consciously and deliberately rather than automatically. It is an open question why people, who normally carry out most of their behaviors automatically, are sometimes spurred to think more consciously and deliberately about what they do; but there is no reason to believe that people actually choose to operate in deliberative mode when they want to. (Thagard 2018, 389–90)

³ I discuss such cases in O'Connor (2009).

⁴ For a patient, thorough, and philosophically very careful analysis of studies in the Libet tradition, see Mele (2009). For a shorter and trenchant discussion by a fellow neuroscientist, see Clarke (2013). For my own assessment of these studies, see O'Connor (2009).

I take Thagard here to be arguing that, if we are to “have free will”—i.e., if it is ever true that we decide and act freely—then we must have a reliable ability to choose when we consciously deliberate about what to do. That is, we must have consciously decided that a necessary condition on our freely deciding is realized. It is not enough, in his view, that we just *find ourselves* wanting to consciously decide what to do on a given occasion. But why think that? Indeed, there are two good reasons to think that Thagard’s condition on having free will is implausible.

First, there are other necessary conditions on our exercising free will that we do not standardly suppose we must have decided to be met. When I act freely, I am aware of more than one alternative that is attractive to me, given my preferences and my beliefs about my circumstances and abilities. But I do not *decide* to make it the case that I am aware of more than one alternative—I just am. Similarly, it seems natural to suppose, I don’t decide *now will be an occasion where I consciously decide what I will do, rather than just acting automatically*. Instead, for whatever reason, I just am poised, consciously, to decide what to do. It might be a problem if I am so poised only very rarely and somewhat randomly, but that certainly doesn’t seem to be the case. I consciously decide what to do several times a day, and it seems pretty predictable when that is the case: when first of all I am reasonably alert, not overwhelmed by some experience or concern, etc, and when there is a matter of some importance to be decided and there is not just one “obvious” thing to do in the circumstances, given my preferences, but rather a multiplicity of options that have some attraction to me.

A second reason not to follow Thagard in supposing that freely and consciously deciding what to do requires that one have first consciously decided so to decide is that such a condition is unrealizable. For what shall we say of that prior conscious decision: did I freely decide to engage in it or did I just “find myself” consciously poised so to decide? If the latter, then it seems Thagard should conclude that the decision that it leads to is unfree. But if the former, then we will be led to require a regress of consciously deciding to consciously decide to consciously decide... This is absurd. Freedom of choice does not require that there be no preconditions on its exercise for which we are not consciously responsible. Nature *bestows* on us certain capacities not of our own making. For any of them to work properly, certain conditions also must be in place that, again, for which we are not responsible. Freedom of choice is simply a special case of that general truth about finite beings such as ourselves.

The results of science are the products of the activity of scientists. To embrace those results as rationally grounded, we must assume that intentional action at many levels was effectively engaged in and that the actors had *reliable* awareness of what they were up to and why, including at the final stage of reporting on methods and results. One certainly may argue unproblematically that human action and self-awareness are prone to error and ignorance in a variety of specific

forms. Our grasp of our own motivations is imperfect, we are sometimes self-deceived, and it is not always easy to come to a more accurate self-understanding even when we learn of the flaws in our cognitive design. But if the experiences and beliefs pervading consciously willed choices are systematically illusory, then our understanding of the nature and significance of scientific activity, too, is called sharply into question, including the activity that led us to that very conclusion of systematic illusion in conscious choice. Thus, to deny the efficacy of conscious will as a general matter is to saw off the branch on which one sits. While the human sciences undoubtedly will deepen and in places correct our self-understanding, they cannot, in the end, wholly divorce themselves from it.

References

- Clarke, P. 2013. "The Libet Experiment and Its Implications for Conscious Will," Faraday Paper 17, accessed on June 10th, 2018 from [https://www.faraday.st-edmunds.cam.ac.uk/resources/Faraday Papers/ Faraday Paper 17 Clarke_EN.pdf](https://www.faraday.st-edmunds.cam.ac.uk/resources/Faraday%20Papers/Faraday%20Paper%2017%20Clarke_EN.pdf)
- Ellis, G. 2016. *How Can Physics Underlie the Mind? Top Down Causation in the Human Context*. Berlin: Springer-Verlag.
- Gillett, C. 2016. *Reduction and Emergence in Science and Philosophy*. Cambridge: Cambridge University Press.
- Humphreys, P. 2016. *Emergence*. Oxford: Oxford University Press.
- Jackson, F. 1982. "Epiphenomenal Qualia." *Philosophical Quarterly* 32: 127–36.
- Laughlin, R. 2005. *A Different Universe: Reinventing Physics from the Bottom Down*. Cambridge, MA: Basic Books.
- Mele, A. 2009. *Effective Intentions: The Power of Conscious Will*. New York: Oxford University Press.
- Nagel, T. 1974. "What Is It Like to Be a Bat?" *Philosophical Review* 83: 435–56.
- Noble, D. 2006. *The Music of Life: Biology Beyond Genes*. Oxford: Oxford University Press.
- O'Connor, T. 2009. "Conscious Willed and the Emerging Science of Brain and Behavior," in *Downward Causation and The Neurobiology of Free Will*, edited by George F. R. Ellis, Nancy Murphy, and Timothy O'Connor, 173–86. New York: Springer.
- O'Connor, T. 2018. "Emergent Persons," in *The Blackwell Companion to Substance Dualism*, edited by Jonathan Loose, Angus Menuge and J.P. Moreland. Oxford: Blackwell, 369–76.
- O'Connor, T., and Wong, Hong Yu. 2015. "Emergent Properties," *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, accessed on June 10th, 2018 from <https://plato.stanford.edu/archives/sum2015/entries/properties-emergent/>
- Thagard, P. 2018. "Mind, Consciousness, and Free Will." *Frontiers of Philosophy in China* 13.3: 377–93.
- Tse, P. 2013. *The Neural Basis of Free Will: Criterial Causation*. Cambridge, MA: MIT Press.
- Wegner, D. 2002. *The Illusion of Conscious Will*. Cambridge, MA: MIT Press.