**teorema**
Vol. XXXVII/3, 2018, pp. 167-69
ISNN 0210-1602
[BIBLID 0210-1602 (2018) 37:3; pp. 167-69

# Comment on Gessell and de Brigard

Daniel Dennett

This essay is a fine example of the virtue, indeed the necessity, of combining philosophical and neuroscientific expertise if we want to make progress in solving "the mind-body problem." It's complicated, as one says, and when the complications push otherwise reasonable and even compelling sketches of solutions into fantasyland, we need to reassess. Gessell and de Brigard (G&B) point to a "discontinuity" looming now that the dreams of GOFAI are being replaced with more biologically realistic visions of the not-so-bureaucratic architecture of the brain. Neurons, it now appears, have agendas, and are themselves nontrivial intentional systems. The good news, as G&B clearly articulate, is that we can still have the Turing-inspired, McCulloch-Pitts-inspired vision of an ultimately non-intentional, mechanistic account down in the engine room, but it will be down in the engine room of neurons, not memory systems, language modules, or world-knowledge storehouses. The bad news is that the sort of intelligently designed "politburo" architectures (Eric Baum's evocative term) of GOFAI, which have inspired more than a generation of cognitive scientists, now look seriously over-idealized. Or rather: *mis*-idealized. Science cannot proceed without "oversimplified" models, but some modeIs lead us astray instead of providing insights.

I would put it this way: the top-down-designed models of GOFAI presuppose too much comprehension in the processes that designed the hierarchies in the first place and that redesign them on the fly. (The extended example of designing the elevator-controller in FBBB provides a clear view of the means-ends analysis and ontological inventory that go into such efficient and provably reliable systems; no such design process laid out the blueprint for the human brain.) There have been murmurs of this for decades, with doubts expressed about "boxology" and somewhat quixotic campaigns for adopting, for instance, a "dynamical systems" approach to replace "computationalism," but now the skepticism is setting in from other sides as well.

All this is highlighted by G&B's crisp account of the background history of functionalism from Lewis and Putnam on to homuncular functionalism, and they are exactly right about the points I made against any straightforward identification of particular (folk-)psychological states, such as *the belief that it is raining* or *the desire for ice cream now,* with particular logical states in the machine table of any Turing machine. But G&B note that I have neglected a question this raises: "What if the things that matter to neural networks, or even neurons, are not the things that matter to us?" If, as G&B put it, "neurons are computing predictions about the presence or absence of things that matter *to them,"* how will we ever get to an account of their activities that explains what is happening in terms of the presence or absence of things that matter to *us*? The discontinuity as I see it is between the use of the intentional stance at the personal level and its use at sub-personal levels. In a way, I have been stressing this for years, insisting that your brain doesn't understand English; you do. Your brain isn't in pain; you are. No privileged area of your brain enjoys the rosy hues of the sunset; you do. But how do we get *up* to the personal level from the sub-personal level?

When we analyze a whole person into a smallish crew (or gaggle, since there is no captain) of highest-level sub-personal intentional systems, what will be the content of *their* beliefs and desires (the information they are in charge of, the tasks that fall to them given their talents and goals)? First, we must remind ourselves that they would be competent without comprehension of the whole-person sort; they could have, perhaps, the sort of behavioral comprehension that non-speaking, non-reflective animals have [see FBBB, pp. 94-101]. This minimal and relatively myopic comprehension would not prevent them from achieving coordination, joining forces on occasion, becoming dependent on each other when not in competition, etc. But the "reporting relations" would not be much like the bureaucratic hierarchies of GOFAI—a crutch for the imagination going way back to the TOTE units (Test Operate Test Exit) of Miller Galanter and Pribram (1960). A key feature of any defensible account of the transmission of personal level semantic information by neural events will have to treat the senders and receivers as largely unwitting (uncomprehending) transmitters. Somehow the larger systems of which they are parts get to use (and hence appreciate, in a diminished sense) the information. For more on the vicissitudes and options for making sense of neuronal signaling, see Cao (2012) and Huebner (2013).

The problem, as I now see it thanks in part to G&B's probing, is that in order to handle the personal level, we need to tackle *at the same level of detail* the decomposition of the person-who-enjoys, the person-who-understands that we have achieved in our (still partial) understanding of the inbound path of perception, discrimination, etc. We must ask and answer the Hard Question, "And then what happens?" [(1991), p. 255]. The neural activity that underlies the (personal level) reactions, reflections, reorganizations, and self-attributions that are the normal sequelae of becoming conscious of something is just as inaccessible to introspection as the perceptual processing that precedes consciousness. So instead of postponing analysis of the Inner Witness, we need to figure out how the illusion of an Inner Witness is composed of neural processes that are themselves as unconscious as reflex arcs.

Happily, both theoretical and experimental work is already well underway on asking and answering these questions. Instead of providing an immense bibliography, I will just list, in alphabetical order, the names of a few of the best researchers in my opinion. Their work can be found easily on the Internet. Peter Carruthers, Axel Cleeremans, Michael Cohen, Stanislaus Dehaene, Michael Graziano, Yul H. R. Kang, Nancy Kanwisher, Hakwan Lau, Gustav Markkula, Alva Noë, David Rosenthal, Claire Sergent, and Michael Shadlen—with apologies to those whose names unaccountably escaped me while writing this under extreme time pressure. This is the topic that is now occupying most of my attention [see Dennett (2018)].

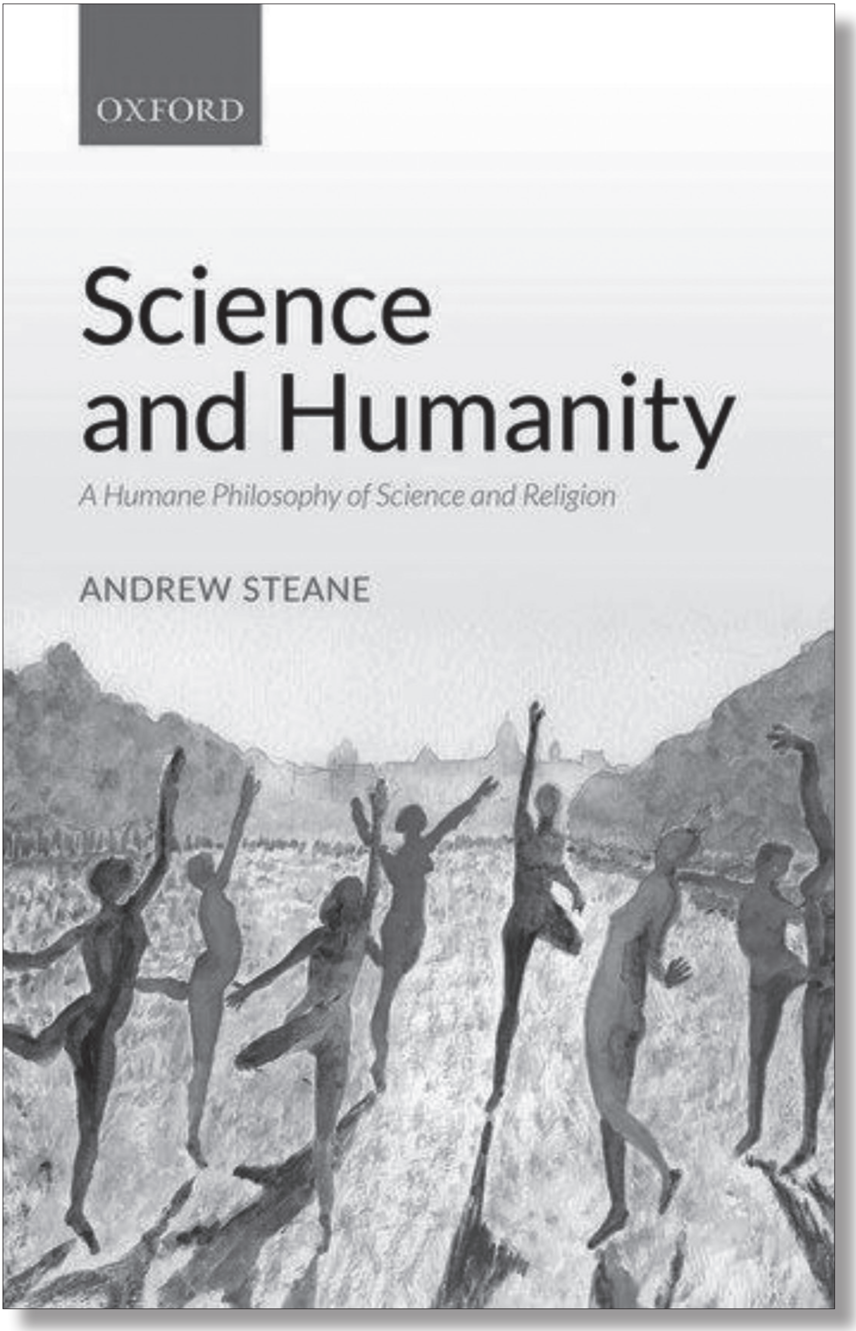
## REFERENCES

CAO, R. (2012), "A Teleosemantic Approach to Information in the Brain," *Biology and Philosophy* 27 pp. 49-71.

DENNETT, D. (2018), "Facing Up to the Hard Question of Consciousness," *Phil.Trans.R.Soc. B* 20170342.

HUEBNER, B. (2013), *Macrocognition: A Theory of Distributed Minds and Collective Intentionality,* Oxford University Press.

MILLER, G., GALANTER, E. and PRIBRAM, K. H. (1960), *Plans and the Structure of Behavior,* New York: Henry Holt.



OXFORD

# Science and Humanity

*A Humane Philosophy of Science and Religion*

ANDREW STEANE