**How Computation Explains**

Abstract

This paper explores issues in the metaphyics of mind, and in our current attempts at an interdisciplinary approach to it.

The 20th century saw a monumental shift in our understanding of the mind, triggered by *computational cognitive science* — the use of tools, concepts, and formalisms from the computer sciences to explain the brain. As a result, there is a widespread belief that minds are computers, that *having a mind* is *having a computational structure of a certain type*. After all, computational cognitive science seems to assume or posit that the brain is a computer, and it does so precisely in order to explain the brain’s mental capacities. So identifying the mind with the brain’s computational structure seems to be a plausible interdisciplinary approach to the metaphysics of mind (e.g., see Chalmers, 2011; Peacocke, 1994; Piccinini, 2015). I argue that this is a mistake, caused by a failure to take the relevant scientific work seriously on its own terms.

 Taking that scientific work seriously means looking at the reason computational tools, concepts, and formalisms are explanatory of the brain, and a careful examination reveals that the reason is *not* that the brain falls into some metaphysical category, computer. Rather, computational cognitive science is an example of *domain transfer* — a transfer of tools and strategies from one domain to another, like when private companies borrow NASA’s testing procedures for rocket components to test advertising campaigns (Edsel, 2016), or when we use motivational strategies from video games to create more engaging university courses (Miller, 2014). But domain transfers typically make no metaphysical commitments: advertising campaigns are not rocket ships, and university courses are not games. I argue that the same is true of computational notions in cognitive science: they provide a useful set of tools and resources that bring with them no metaphysical commitments whatsoever. This is not to say that the idea of the brain as a computer is “just a metaphor” — it’s far more than that. It’s a sophisticated set of tools and strategies that are successfully applicable to the brain for good and interesting reasons. But the success of computational cognitive science gives us no more reason to think the brain *is* a computer than the success of NASA’s testing procedures in advertising give us reason to think ad campaigns *are* rocket ships.

 The motivation for the computational metaphysics of mind was that our mentality is explained, in cognitive science, by the brain *being* a computer of a certain sort, so it is plausible that the mind *is* that computational structure. But if cognitive science does not explain the brain by assuming that it *is* a computer of any sort, this motivation is undercut. I suggest two possible next steps for an interdisciplinary philosophy of mind. First, we could look more carefully at what cognitive science actually thinks the brain is: it is at least committed to the brain having a certain causal structure — maybe that’s where we could locate mentality. Second, and more radically, we could abandon the search for a metaphysics of mind altogether, accepting that mentality is not to be explained by appeal to any metaphysical category that minded beings belong to.

**Works Cited**

Chalmers, D. J. (2011). A Computational Foundation for the Study of Cognition. *Journal of Cognitive Science*, *12*, 323–357.

Edsel, A. (2016). *Breaking Failure*. New Jersey: FT Press.

Miller, M. (2014). *Minds Online: Teaching Effectively with Technology*. Cambridge MA: Harvard University Press.

Peacocke, C. (1994). Content, Computation and Externalism. *Mind & Language*, *9*(3), 303–335.

Piccinini, G. (2015). *Physical Computation: A Mechanistic Account*. Oxford University Press.