

Book Review: Minds and Computers

Phillippe Verdoux
Institute for Ethics and Emerging Technologies

Review of *Minds and Computers: An Introduction to the Philosophy of Artificial Intelligence* by Matt Carter (Edinburgh University Press, 2007). ISBN 978-0748620999. \$32, Paperback.

There are different reasons a given problem may be considered abstruse. One possibility is that the problem may be simple in terms of its component parts but include concepts that the human mind finds difficult to grasp. Or, alternatively, the problem may be easy to grasp conceptually but involve too many parts for the mind to keep in order. The former concerns the *qualitative* status, or “type,” of the problem, while the latter its *quantitative* character, or “size” (McGinn 2005, p. 331). The point is that, in a significant way, the field of artificial intelligence (AI) includes *both* elements of abstruseness: on the one hand, the project of creating machine (super)intelligence is predicated on a number of sophisticated *philosophical* theses concerning the nature of mind. On the other hand, given that AI research is situated within the highly composite super-discipline of cognitive science, a deep understanding of AI requires that one know at least *something* about the many fields of empirical research surrounding AI, such as computer science, psychology, neuroscience and linguistics.

Thus, any book attempting to provide an informative introduction to AI immediately confronts the problem of making accessible to students a constellation of issues that are both quantitatively and qualitatively abstruse. Making matters worse, the author must navigate the *breadth-depth tradeoff*: greater detail about any single topic means less total topics covered, and more total topics covered means less said about any single issue. Such is our “finitary predicament,” as Christopher Cherniak (1990) has called it.

In his first book *Minds and Computers*, Matt Carter provides a comprehensive introduction to the field of AI research. The book begins with a predictable walk through dualism, behaviorism, psychophysical identity and functionalism (culminating in a chapter on computationalism), offering a crisp and insightful explication of each of these issues. But Carter nicely distinguishes himself from other authors by including a number of information-packed chapters on extra-philosophical topics, as the titles “Formal Systems,” “Games” and “Human Language” suggest. This adds to the philosophical foundation established early on by providing a rich interdisciplinary context in which to situate AI. As Carter points out, if one’s philosophical aim is to spell out the functionalist theory – a theory that leaves as unspecified “black boxes” what exactly mental states and the mechanisms causally connecting them *are* – in computational terms, then one ought to have an understanding of what exactly constitutes a *computation*. With notable facility, Carter provides a rigorous account of this concept, as well as related concepts like “effectivity,” “algorithm” and “register machine”; Carter then shows how to integrate the notion of computation into the functionalist framework. The result is a substantive, non-boxological theory of mentality.

Despite many positive features, Carter’s book is vulnerable to a number of criticisms. For example, one finds several vexatious errata – in some cases more accurately describable as “thinkos”¹ – throughout the book. On page 32, for example, Carter asserts without any citation that there are “ten billion neurons in the brain.” But in fact there are an estimated *one hundred*

billion neurons crowding each one of our crania. And the fact that the *word* instead of the *numeral* was used here suggests a thinkological, rather than a mere typographical, error. But enough caviling.

More significantly, it is not clear what the fourth chapter entitled “Neuroanatomy” adds to the book.² First, this chapter is largely a *list* of technical neuroanatomical terms. It is therefore doubtful that even the most ambitious reader not already familiar with neuroscience would take much away from it,³ or for that matter use it as a resource for future research (that is, in place of a textbook). Furthermore, Carter does not once mention the phenomenon central to contemporary neuroscience, namely the *action potential*.⁴ In any introductory chapter on the subject, it seems to me imperative that this phenomenon be mentioned, if not explicated in some detail. Finally, the information provided in “Neuroanatomy” is hardly referred back to in subsequent chapters. Thus, while it is unarguably *a good thing* for students interested in the philosophy of mind and AI to know the rudiments of neuroscience, as far as I can tell Chapter 4 is a largely superfluous appendage to the already-bulky body of *Minds and Computers*.

In addition, there are a number of claims in Carter’s book that are likely to engender confusion in the careful reader. In Chapter 3, Carter distinguishes between “psychological” and “philosophical” (also “analytic” or “logical”) behaviorism. B.F. Skinner, Carter states, is “by far the most influential psychological behaviourist after Watson” (Carter 2007, p. 21). So far so good. But Carter then writes that “psychological behaviourism [...] is a *methodological* view – a doctrine concerning the way in which one should go about doing psychology” (Carter 2007, p. 23). The problem is that psychological behaviorism and methodological behaviorism are standardly construed as *logically distinct theses*: holding the normative view that psychology is the science of behavior and not the mind does *not* entail (or *vice versa*) the explanatory view that one can account for organismal behavior without reference to any event, state or phenomenon “inside” the individual – that is, “the sources of behavior are external (in the environment), not internal (in the mind)” (Graham 2007).⁵ Students who pursue the philosophy of mind beyond this book will no doubt be surprised to discover that psychological and methodological behaviorism are quite distinct – although by no means mutually exclusive – positions.

Furthermore, after initially characterizing analytic behaviorism as “a reductive semantic thesis” (Carter 2007, p. 23), Carter claims that this version of behavioristic theory⁶ conflates the intensional issue of *what constitutes* a mental state with the extensional issue of *which things* in the world actually count as mental states. Carter then explains that:

This is because [analytic] behaviourists are *eliminativists* about mental states and, hence, do not believe there is a substantive identification to be made. They hold that talk of mental states is, in fact, just talk about dispositions to behave – behaviourism is an ontologically eliminative and semantically reductive theory of mind. (Carter 2007, p. 35)⁷

But on the standard reading, this is not quite accurate: while the radical behaviorism of B.F. Skinner and company was motivated by eliminativist aspirations, the analytic behaviorism of Ryle and others is best understood as a purely reductionistic thesis (Rey 1997, p. 96).⁸ And *reduction does not entail elimination*.⁹ Consider the case of knowledge: by reducing knowledge through conceptual analysis *to* justified true belief, epistemologists were in no way suggesting that knowledge does not exist. The same holds true in the case of phenomena like heat and mean molecular motion, or water and H₂O (i.e., heat and water still *exist* despite being reduced). With respect to analytic behaviorism, then, exponents of this view aimed to provide a semantic analysis

of mentalistic expressions – 'Phil *loves* Whitney,' or 'Stephanie *believes* in supernatural deities' – an analysis given in terms of observable behavioral dispositions. (This move is “reductive” because it attempts to explicate mentalistic language in entirely non-mentalistic language.) Analytic behaviorists are still happy to talk about mental states, as Ryle himself affirms,¹⁰ but when asked what is *meant by* mental state terms like 'loves' and 'believes', the exponent of this position will promptly cite some repertoire of dispositions to behave in particular ways, given particular antecedent conditions.¹¹

Another way of articulating this criticism is by relating analytic behaviorism to the semantic thesis of *irreferentialism* (Rey 1997, p. 140-151). According to this view, mentalistic terms “don't even *purport* to refer to” the mental phenomena postulated by Cartesian dualists.¹² Instead, such terms are semantically similar to 'John Doe' (or 'the average American') in the following way: if someone thinks that 'John Doe' refers to any *actual individual*, he or she would be seriously misunderstanding the term's ordinary use in the language. Thus, one is not so much making the *eliminativist claim* that John Doe fails to be a member of the ontology as the *semantic claim* that 'John Doe' is not the sort of term that even *attempts to* pick out any particular person.¹³ Compare this to the term 'phlogiston,' which does indeed purport to refer (i.e., to a particular chemical released during combustion). But it turns out that, after some empirical investigation, this referent is non-existent. Modern science is therefore ontologically eliminativist with respect to phlogiston while analytic behaviorists, in contrast, are mere irreferentialists with respect to mental state terms; and the semantic analyses given are an attempt to reveal this irreferentiality.¹⁴

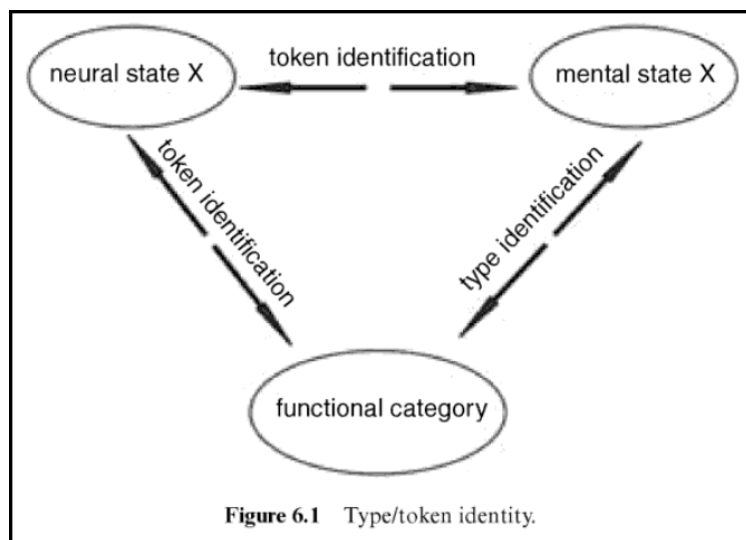
Yet another point of confusion stems from Carter's discussion of “embodied experience.” After describing Searle's Chinese room objection to strong AI, Carter argues that “the lesson to draw from the Chinese room thought experiment is that *embodied experience* is necessary for the development of semantics” (Carter 2007, p. 179). Carter elaborates:

In order for our mental states to have meaning, we must have antecedent experience in the world, mediated by our sensory apparatus. In other words, semantics do not develop in isolation but, rather, this development is conditional on experience in relation to the empirical world. (Carter 2007, p. 179)

But this does not, Carter contends, problematize computationalism. Rather, “it merely shapes the explanatory burden on the computationalist, requiring them to provide a computational account of the meaning conferring mechanisms” (Carter 2007, p. 180). Now, it may well be a shortcoming of mine that I find Carter's claims opaque. There are of course numerous theories of intentionality (said to be) compatible with the computational-representational theory of thought, including the teleosemantic account of Millikan (1984) and Fodor's (1990) asymmetric dependence theory of content. But talk of *embodiment* suggests something rather different: as readers are no doubt aware, a defining feature of functionalist theories is that they posit mental states as *disembodied* entities. This follows from the thesis of – as Carter nicely puts it elsewhere – “ontological neutrality,” also called “substrate independence” and “multiple realizability.” Thus, even if Carter's contention turns out to be coherent (I withhold any judgment due to insufficient argumentation), it certainly deserves, at least in my opinion, more discussion than it gets (see Carter 2007, p. 179-180, 206). Then again, given the breadth-depth tradeoff mentioned above, such discussion would not be free.

These are just a few *of only a few* problems that I could discern in *Minds and Computers*. It is worth emphasizing again that far more could be said about what Carter gets right – and often in a terse but informative way – than what he gets wrong or muddles. Take as a quick example

Carter's Figure 6.1 (below), found on page 46. No doubt many students new to the philosophy of mind become frustratingly puzzled by the relation between functionalism and reductionism, especially given the phenomenon of multiple realizability. Thus, one is told that functionalism is *anti-reductionistic* – but in what sense exactly? In a token or a type sense? With respect to neural or functional states? Is it therefore not compatible with physicalism? Using Figure 6.1, Carter makes explicit what is usually only *tacit* in discussions of functionalism, namely that functionalist theories identify *types* of mental states with *types* of functional states (a type-type identity), but not *types* of mental – or functional – states with *types* of neural states (a token-token identity). Carter also explains how the token-token identity of functionalism is not empirically vacuous, since “psychological inquiry, on the functionalist account, is a matter of determining and investigating the characteristic functions of particular types of mental states” (Carter 2007, p. 47). I can imagine clarifications like this saving the neophyte much time in trying to figure out *in exactly which ways* functionalism is and is not reductionistic.



In conclusion, it was a pleasure reading *Minds and Computers*, and no doubt beginning students in philosophy will find Carter's debut book intellectually rewarding. This goes of course for that growing demographic of students interested in the philosophy of technology, since AI research combines some of the more sophisticated work in contemporary philosophy with a multiplicity of fascinating issues relating to cutting-edge technology. But, as already mentioned, the claim of strong AI is fundamentally *philosophical* in nature, and thus its status ultimately depends not on the state of advanced technology but on its status as a philosophical thesis.¹⁵

References

- Carter, M. 2007. *Minds and Computers*. Edinburgh: Edinburgh University Press Ltd.
- Cherniak, C. 1990. *Minimal Rationality*. Cambridge, MA: MIT Press.
- Dennett, D. 2006. From Typo to Thinko: When Evolution Graduate to Semantic Norms. In *Evolution and Culture*. S.C. Levinson and P. Jaisson. Cambridge, MA: MIT Press. http://ase.tufts.edu/cogstud/papers/Dennett_From_Typo_to_Thinko.pdf
- Fodor, J. 1990. *A Theory of Content*. Cambridge, MA: MIT Press.
- Graham, G. 2009. Behaviorism. In *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta (ed.). <http://plato.stanford.edu/archives/win2009/entries/behaviorism/>
- McGinn, C. 2006. Can We Solve the Mind-Body Problem? In *The Philosophy of Mind: Classical Problems/Contemporary Issues*. B. Beakley and P. Ludlow (eds). Cambridge: MIT Press, 321-337.
- Millikan, R 1984. *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press/Bradford Books.

Rey, G. 1997. *Contemporary Philosophy of Mind: A Contentiously Classical Approach*. Cambridge, MA: Blackwell Publishers.

Ryle, G. 2006. From *The Concept of Mind*. In *The Philosophy of Mind: Classical Problems/Contemporary Issues*. B. Beakley and P. Lodlow (eds.). Cambridge, MA: MIT Press, 77-85.

Endnotes

- 1 I borrow this term from Dennett (2006), who uses “thinko” (originally a hacker’s term) on the model of “typo.” The term “thinkological” is my neologism.
- 2 Carter states at the beginning that he modestly hopes to convey a sense of the extraordinary complexity of the brain. While this is, of course, commendable in itself, I am again not sure that it warrants an entire chapter, especially when important details about other issues – such as that of “embodied experience” (discussed later in this review) – are given short shrift.
- 3 And of course if one already knows about neuroanatomy, he or she is likely to skip this chapter altogether.
- 4 Although Carter does mention in passing that “the soma will discharge an electrical impulse along its axon” (Carter 2007, p. 34).
- 5 In opposition to this claim of logical distinctiveness, Carter claims that analytic behaviorism “clearly entails” psychological behaviorism. (To be clear, the exact quote is “Clearly the former entails the latter,” where “the former” refers to analytic behaviorism and “the latter” to psychological behaviorism.)
- 6 If one is not extremely careful and failed to catch Carter’s statement, on page 23, that “henceforth, when I make reference simply to ‘behaviourism’, I will be referring to the philosophical variety,” the reader would no doubt lose track of the discussion.
- 7 A more ambiguous passage expressing a similar idea can be found on page 26.
- 8 Although philosophers like Julia Tanney have argued that the standard characterization of Ryle as an analytic behaviorist is simply wrongheaded. For the purposes of this paper, I will not deviate from the usual reading of Ryle.
- 9 See Rey 1997, p. 22-23 for helpful discussion.
- 10 As Ryle himself writes in *The Concept of Mind*: “I am not, for example, denying that there occur mental processes. But I am saying that the phrase ‘there occur mental processes’ does not mean the same sort of thing as ‘there occur physical processes’ (Ryle 2006, p. 84).
- 11 Furthermore, Carter writes in the quoted passage that analytic behaviorists do not believe there is a *substantive identification* to be made. But in Chapter 3, Carter describes this very position as “a substantive theory of what mental states *are*” (Carter 2007, p. 23). Indeed, the substantive identification *just is* the reductive connection analytic behaviorists make between mentalistic talk, on the one hand, and talk of behavioral dispositions, on the other.
- 12 If one likes, *here* is where Ryle’s famous “category mistake” makes its appearance.
- 13 By analogy, one might imagine a whole population of “Cartesians” who hold that the term “John Doe” *does* refer to some actual individual. The “analytic behaviorists” then come along and say “No, that is not how this term is ordinarily used, as I will now show through a careful semantic analysis of the term. John Doe, you will see, is a mere *philosopher’s myth*.” Such an analysis would then aim to reveal the term’s ordinary meaning – that is, to show that it doesn’t even *purport* to refer to anything.
- 14 See in particular Rey 1997, p. 143, for a careful discussion of the subtle but significant differences between eliminativism and irreferentialism.
- 15 Incidentally, a good interview with Matt Carter can be found at The Philosopher’s Zone, <http://www.abc.net.au/rn/philosopherszone/stories/2008/2121606.htm>. Also, I’d like to thank Matt Carter for discussing his book with me via email.