

# Computational explanation in neuroscience

Gualtiero Piccinini

Received: 6 July 2006 / Accepted: 8 August 2006 /  
Published online: 20 October 2006  
© Springer Science+Business Media B.V. 2006

**Abstract** According to some philosophers, computational explanation is proprietary to psychology—it does not belong in neuroscience. But neuroscientists routinely offer computational explanations of cognitive phenomena. In fact, computational explanation was initially imported from computability theory into the science of mind by neuroscientists, who justified this move on neurophysiological grounds. Establishing the legitimacy and importance of computational explanation in neuroscience is one thing; shedding light on it is another. I raise some philosophical questions pertaining to computational explanation and outline some promising answers that are being developed by a number of authors.

**Keywords** Computational explanation · Mechanistic explanation · Computational neuroscience · Cognitive neuroscience · Theoretical neuroscience · Computationalism · Pancomputationalism · Computational theory of mind · Models · Representation · Introspection

Some philosophers would think my title is an oxymoron: “in the language of neurology ..., presumably, notions like *computational state* and *representation* aren’t accessible” (Fodor, 1998, p. 96). For ease of reference, I will call this view ‘computational chauvinism’: when it comes to explaining cognitive capacities, computational explanation is proprietary to psychology—it does not belong in neuroscience.

Many neuroscientists beg to differ. They routinely appeal to computations (and representations, for that matter) in interpreting data, forming hypotheses, and building models. They publish in journals named *Neural Computation*, *Journal of Computational Neuroscience*, and *Network: Computation in Neural Systems*. Before accusing neuroscientists of overstepping their boundaries, philosophers should take a closer look at what neuroscientists are trying to say.

---

G. Piccinini (✉)  
Department of Philosophy, University of Missouri—St. Louis,  
599 Lucas Hall (MC 73), 1 University Blvd.,  
St. Louis, MO 63121-4400, USA  
e-mail: piccininig@umsl.edu

Computational explanation is explanation of a phenomenon in terms of computational processes. It is routinely employed to explain the behavior of calculators and computers. For instance, your PC lets you browse the web by, among other things, executing an appropriate program, and executing that kind of program constitutes a computation. By contrast, the phenomenon of radioactivity is usually not explained in terms of computation; rather, it is explained in terms of the microphysical process of atomic decay. Computational explanation is sometimes employed outside the domain of the familiar artifacts we use every day. I will here focus on what is sometimes called computationalism: the view that cognitive capacities are explained by computational processes.<sup>1</sup>

Computationalism is contentious. Some scientists and philosophers reject it, and some even find it offensive. But computationalism, in its many variants, has been at the center of contemporary sciences of mind over the last 50 years. For better or worse, computationalism has taken hold. Plenty of philosophers have noticed and offered ways to understand it. They have reached no consensus, however, on whether and how computational explanation fits within neuroscience.

## 1 Computational chauvinism

Computational chauvinism is correct in the following sociological respect. Many mainstream cognitive scientists put some form of computational explanation at the core of psychology, while paying little attention to the brain (e.g., Fodor, 1975; Gallistel, 1990; Johnson-Laird, 1983; Miller, Galanter, & Pribram, 1960; Newell & Simon, 1972; Pylyshyn, 1984). They assume psychologists can discover the correct computational explanation of cognitive capacities in relative autonomy from neuroscience. According to them, the role of neuroscientists is simply to discover the neural mechanisms that implement the computational process.

For instance, Allen Newell is well known for his theory (developed with Herbert Simon) that the mind is a physical symbol system (Newell & Simon, 1976; Newell, 1990). To a first approximation, a physical symbol system is a general purpose, stored-program computer. Over 40 years ago, Newell invited an audience of neuroscientists to discover “where in the nervous system I can find symbols” (Gerard & Duyff, 1962, p. 343). The underlying division of labor is clear: the psychologist finds the computational explanation; the neuroscientist’s task is to find its implementation.

Nature has been uncooperative with this approach. Since Newell’s invitation, neuroscientists have made enormous progress: plenty of new techniques have emerged, yielding many results. Yet there doesn’t seem to be evidence of symbol systems (in Newell’s sense) in the nervous system. This undermines the strong form of autonomy of psychology from neuroscience assumed by computational chauvinists. If psychological theorizing is to make contact with actual neural mechanisms, it needs to be more responsive to neuroscientific evidence than computational chauvinists suppose

---

<sup>1</sup> Another use of ‘computation’ pertains to overt behavior alone. In this sense, a behavior of a system may or may not be described as computational. For instance, when people perform additions or multiplications they are said to perform computations, whereas when they run or climb they usually are not. The question remains as to what explains their behavior. If the explanation involves inner computational processes, then the behavior in question is subject to computational explanation.

(Churchland, 1986; Egan & Matthews, this issue; Feest, 2003; Keeley, 2000; Revonsuo, 2001).

To be sure, some psychologists do attempt to formulate computational explanations that are “neurally inspired” (Rumelhart & McClelland, 1986, p. 131). Their effort was behind the neo-connectionist movement that became popular in the 1980s. Their models take the form of processing units connected in networks, somewhat analogous to networks of neurons. But the data many of them use to constrain their models are purely behavioral rather than neural. The resulting models are often accompanied by little if any evidence that the postulated mechanisms are implemented in the brain. As a consequence, the relationship between many connectionist explanations and neural mechanisms remains obscure, and much connectionist work retains a flavor of computational chauvinism.<sup>2</sup>

Computational chauvinism faces another problem: neuroscientists have developed their own way of formulating explanations, including computational ones, and they have done so since the very origin of computationalism. Computational explanation was initially imported from computability theory into the science of mind by neuroscientists, who justified this move on neurophysiological grounds. The man most responsible for this move was Warren McCulloch, who co-wrote with Walter Pitts the paper that launched computationalism (McCulloch & Pitts, 1943; cf. Boden, 1991). McCulloch’s main empirical motivation was Edgar Adrian’s discovery that neural impulses are all-or-none affairs (Adrian, 1928; for some of the history see Frank, 1994; Piccinini, 2004). McCulloch thought the all-or-none law of nervous activity justified describing neural processes as logical operations over digital inputs. Roughly speaking, he saw neural nets as logic circuits and the brain as a digital computer (McCulloch, 1949).

At first, psychologists were more receptive than neuroscientists to McCulloch and Pitts’s idea. Many experimental neuroscientists were skeptical of McCulloch and Pitts’s theory, worrying that it didn’t take sufficiently into account known anatomical and physiological phenomena (e.g., see the discussions of von Neumann and McCulloch’s papers in Jeffress, 1951). Those neuroscientists who did accept computationalism eventually rejected McCulloch and Pitts’s version of it. They developed different formalisms, methodologies, and justifications. They made their formalisms more suited to modeling neurophysiological data. Eventually, they won converts. By now, the appeal to neural computation to explain cognitive phenomena is widespread in neuroscience (Anderson, Pellionisz, & Rosenfeld, 1990; Anderson & Rosenfeld, 1988, 1998).

Establishing the legitimacy and importance of computational explanation in neuroscience is one thing; shedding light on it is quite another. I will quickly introduce a few relevant issues. But first, a bit of terminology.

---

<sup>2</sup> For example, consider the standoff between connectionist and classicist models of English past tense acquisition (McClelland & Patterson, 2002; Pinker & Ullman, 2002). In the face of objections by classicists that their model was not powerful enough, connectionists switched from a model trained using the perceptron convergence procedure (Rumelhart & McClelland, and the PDP Research Group, 1986, p. 225) to a model trained with backpropagation. The latter is generally considered *less* neurologically plausible than the former. Furthermore, the main piece of quasi-neurological evidence discussed in this debate is the putative double dissociation between regular and irregular past tense verb processing. Such evidence was introduced in the debate *by* classicists to argue *against* connectionists.

## 2 Computational, cognitive, and theoretical neuroscience

Computers are useful in neuroscience in many ways. For instance, they may be used to construct detailed neural maps.<sup>3</sup> This is not what is usually meant by computational neuroscience.

Computational neuroscience, in the most common sense of the term, is the construction and use of computational models of neural processes, analogously to the construction and use of computational models of other processes within other sciences (Cruse, 2001). Although many computational neuroscientists believe that neural systems perform computations (e.g., Churchland, Koch, & Sejnowski, 1990), some are more cautious about computationalism (e.g., Perkel, 1990). By itself, developing computational models of neural phenomena does not commit neuroscientists to the conclusion that brains compute any more than developing computational models of, say, the Big Bang commits cosmologists to the view that the universe is a computer. If neural systems compute, this needs to be established by more than the existence of computational neuroscience.

Cognitive neuroscience is that part of neuroscience that studies how the brain performs cognitive functions (Bechtel, 2001; Gazzaniga, 2000). Cognitive neuroscience includes a strong experimental tradition as well as the use of computational and theoretical tools. Hence, it overlaps with computational neuroscience. A cognitive neuroscientist may be more of an experimentalist or more of a theorist, may or may not build computational models, and may or may not endorse computationalism.

Theoretical neuroscience is the use of theoretical constructs and mathematical techniques to understand the brain (Dayan & Abbott, 2001). The relevant techniques and constructs prominently include computational models, so the term ‘computational neuroscience’ is often used as a synonym of ‘theoretical neuroscience’ (Grush, this issue). As I pointed out, this by itself does not commit a theoretical neuroscientist to computationalism. Theoretical neuroscientists also use theoretical constructs and mathematical techniques that have no direct connection with computability theory. Examples include information theory, statistics, various branches of dynamical systems theory, and even quantum field theory. Hence, strictly speaking, computational neuroscience is only one component of theoretical neuroscience among others. Furthermore, although there are many theoretical neuroscientists who are computationalists, there are others who reject computationalism or are noncommittal about it (Edelman, 1992; Freeman, 2001; Globus, 1992; Perkel, 1990).

In sum, one should not be misled by terminological choices such as the use of ‘computational neuroscience’ as a synonym for ‘computationalism’ or ‘theoretical neuroscience’ into thinking that if one develops a theory of the brain, then one must build computational models of the brain, or that one understands cognitive neural functions only if one obtains their computational explanations, or that one builds computational models of neural processes only if one believes that the brain computes. Neuroscience is more nuanced than that. Contrary to what is sometimes implied, theoretical neuroscience is broader than computational neuroscience, and neither of them commits its practitioners to computationalism.

<sup>3</sup> David van Essen of Washington University in St. Louis has led the development of sophisticated digital brain maps, which summarize anatomical and functional data. The resulting database is accessible at <http://sumsdb.wustl.edu:8081/sums/index.jsp>

So, does the brain really compute? Answering this question requires getting clear on several background issues, all of which are independently interesting.

### 3 Computation and observer-dependence

Is being computational a genuine property or is it just a projection of the mind? John Searle, for one, maintains that computation is observer-dependent in the strong sense that we are free to ascribe any computation to any process as we please and there is no fact of the matter as to whether our ascriptions are correct (Searle, 1992, Chap. 9). If computation is observer-dependent, then putative computational explanations may be just ways of labeling things that do not get at their real properties—mere pseudo-explanations. Under this view, computationalism is not so much false as vacuous.

One way to address Searle's challenge is to grant that computation is observer-dependent but maintain that computation is explanatory nonetheless. This is the position of Churchland and Sejnowski, 1992. In his essay, Oron Shagrir fleshes out this position by providing suitable notions of observer-dependence and computational explanation.

Another way out is to reject Searle's contention and maintain that computation is in the world. Many authors have argued that computation may be seen as a way of capturing some aspects of the causal structure of the world (Chalmers, 1996a; Copeland, 1996; Smith, 2002). This raises a further question: which physical processes deserve to be described as computational? Perhaps all?

### 4 Pancomputationalism?

Some people think everything is computational. They do so for different reasons. Some commit the fallacy of believing that pancomputationalism is a consequence of the Church-Turing thesis (cf. Copeland, 2000; Piccinini, forthcoming a). Others think that everything is computational because computation is observer-dependent (Searle, 1992). Yet others believe computation is the most fundamental physical process and everything is made out of it (Wolfram, 2002).

From pancomputationalism, computationalism follows trivially: if everything is computational, the brain certainly is. But this argument provides only cold comfort to most computationalists. McCulloch and his followers aimed at explaining peculiarly mental characteristics, such as rationality and intentionality, by exploiting the analogy between brains and computing devices. If the same analogy holds between computing devices and everything, it is doubtful that computationalism retains its distinctive explanatory power (cf. Grush, 2001).

Shagrir (this issue) responds that computationalism remains explanatory in the face of pancomputationalism. Elsewhere, I have offered a different diagnosis: pancomputationalism may be true or close to true only in the trivial sense that everything can be described to some degree of approximation by a computational model; in the sense that matters to computationalism—i.e., in the sense in which computation may be relevant to explaining mental capacities—pancomputationalism is far from true (Piccinini, forthcoming b).

Be that as it may, it should be clear by now that there are different ways of describing a process computationally. At the very least, computations may be used to model the behavior of a system, or they may be used to explain it.

## 5 Modeling versus explaining

Is computation just a tool for building models? Is its role in neuroscience analogous to the role it plays in other disciplines, or is there something more to it? If we wish to reject pancomputationalism, we need a way to draw a line between systems that compute and systems that don't, so that we can discern on which side of the line brains fall. Even if we think everything is computational, we may still wish to distinguish explanatory computational models from non-explanatory ones. Hence, the distinction between explanatory and non-explanatory models is orthogonal to the question of which systems compute.

Carl Craver (this issue) draws a general distinction between explanatory and non-explanatory models. To be sure, Craver points out the many ways in which good models—whether explanatory or not—are useful: they summarize large bodies of data and predict the behavior of a system under a variety of conditions. Explanatory models do more than that. According to Craver, they also provide mechanisms: they describe how the system's components and their activities are organized together so that they exhibit the phenomena.

Mechanistic explanations come at many levels. Many philosophers of mind are used to speaking of two: the cognitive (or mental) level and the neurological level. This two-levelism, as William Lycan (1987) calls it, fits well with computational chauvinism. For if there are two levels, they can be evenly divided between psychologists and neuroscientists: psychologists remain in charge of the cognitive level, where computational explanations are given, while neuroscientists can take care of the neural level, where implementations are found. Could anything be fairer?

Philosophers of neuroscience know better. There are many more than two levels. Going from the top down, the behaving animal contains the nervous system, which contains neural systems, which contain brain areas and nuclei, which contain networks of neurons, which are made out of various cellular components, which themselves span multiple levels. And this is just a glimpse—a lot more should be said in an adequate discussion of levels.

Much is known about neurons and small networks; that is where most of the action in computational neuroscience is. When we move from small networks to brain areas and nuclei, or all the way to neural systems, it becomes harder to construct illuminating explanations. Part of the problem is lack of data. While neuroscientists have devised exquisite techniques for recording the activities of individual neurons, it is hard to detect what a whole area (containing millions of neurons) does.

In recent decades, however, brain-imaging techniques<sup>4</sup> have emerged, yielding insights into the activities of neural areas and the systems they compose. As data about systems become available, theorists are looking for ways to make sense of them. Frances Egan and Robert Matthews (this issue) point out that there are now enough data and enough techniques for neuroscientists to model neural systems and

<sup>4</sup> The most prominent imaging techniques are positron emission tomography (PET), functional magnetic resonance imaging (fMRI), and event related potentials (ERPs).

formulate explanations of cognitive capacities at that level. Furthermore, they argue that the project of modeling neural systems may be pursued with a large degree of autonomy from both cognitive psychology and the study of neurons and networks. They propose the autonomous study of neural systems as a “third way” between purely top–down approaches (i.e., computational chauvinism) and purely bottom–up approaches.

The systems level is one of the links between the properties of networks of neurons and the capacities of organisms. As Egan and Matthews’ proposal makes clear, much is still unknown about the properties of levels intermediate between neurons and behaving organisms. The big challenge facing neuroscience is to work out the relevant properties of each level and integrate them into one mechanistic picture (cf. Bechtel, 2001; Craver, forthcoming; Revonsuo, 2001).

Suppose we find a general notion of explanation—mechanistic explanation, perhaps—that is adequate to neuroscience. Are any neuroscientific explanations computational? If computational explanation is defined to be the same as mechanistic explanation (or whatever kind of explanation neuroscientists give), then the answer is trivially, yes. But if we simply identify computational and mechanistic explanation, we lose that distinction—between mechanisms that compute and mechanisms that don’t—that motivated computationalism in the first place.

If, instead, we reject a crude identification between mechanistic and computational explanation, then we need an account of their relationship. Are any neuroscientific explanations computational? What is distinctive about computational explanations?

## 6 Computation with or without representation

Computation is often thought to be a process of manipulating representations. If representation is taken as a necessary condition for computation, this becomes the semantic view of computation: “no computation without representation” (Fodor, 1981). According to this view, in order to define the notion of computation that is relevant to explaining cognitive capacities, we first need to individuate special entities that possess semantic properties—representations. When those are found, computations may be defined as processes that preserve the semantic properties of representations. Even more strongly, some authors argue that computational states are individuated at least in part by their semantic properties (Crane, 1990; Peacocke, 1999; Shagrir, 2001, this issue).

The semantic view of computation promises to fulfill the original aim of computationalism: contributing to an explanation of peculiarly mental capacities, such as rationality and intentionality. For if mental capacities are explained by the manipulation of internal representations and computations are just the relevant manipulations, then the semantic view seems to finally provide a robust notion of computational explanation that fits the needs of computationalism: computational explanation as explanation involving semantic content. Furthermore, it may seem that the semantic view of computation exorcises the specter of pancomputationalism at the same time. For it seems that very few things—minds included—possess genuine semantic properties. If so, very few things are subject to genuine computational explanation, and computationalism, based on a semantic view of computation, appears to be a robustly explanatory theory of cognition.

But the story doesn't end here. As Shagrir (this issue) points out, it is not entirely clear which notion of representation, and which notion of representational manipulation, is involved in computational explanation. If we restrict the notion of representation to that of *mental* representation, we seem to obtain a robust notion of computationalism, but we also lose the analogy between minds and computing artifacts that originally motivated computationalism. Most computers and calculators, by most accounts, do not possess mental representations. And without the analogy with computing devices, we are left with no independent grip on the notion of computational explanation.

If, however, we broaden our notion of representation to include the states of computers and calculators, we drift back towards pancomputationalism. For if our notion of representation is loose enough, everything can be interpreted as possessing semantic properties. Shagrir takes this route but adds that nevertheless, computational explanation remains helpful and nontrivial—it explains how a system performs an information processing task.

Faced with the difficulties of the semantic view, others prefer to abandon it altogether, and opt for a non-semantic view of computation (Egan, 1995; Stich, 1983). After all, computations can be defined independently of any semantic interpretation of the computational states, inputs, and outputs (cf. Machtey & Young, 1978). I subscribe to a non-semantic view of computation too. In my view, computational explanation is a specific kind of mechanistic explanation—roughly, a mechanistic explanation that characterizes the inputs, outputs, and sometimes internal states of a mechanism as strings of symbols, and provides a rule, defined over the inputs (and possibly the internal states), for generating the outputs (Piccinini, forthcoming c, d).

Now, suppose that one way or another, we obtain a satisfactory notion of mechanistic explanation of mind, whether computational or not. How does mechanistic explanation relate to the 'subjective' side of the mind?

## 7 Neuroscience and introspection

Many neuroscientists and psychologists are skeptical of subjective experience. Scientific theories, they say, should only be based on hard, third-person data. Their attitude is captured by Daniel Dennett's heterophenomenology (1991, 2003). According to Dennett, scientists should not trust introspective reports. They should consider introspective reports a behavior like any other, to be explained by a complete scientific theory of mind.

Not everyone is satisfied with this skeptical approach. Subjects appear to have access to precious sources of information about their own mental conditions. Tapping those sources may be helpful or even indispensable to a complete science of mind. Recently, some scientists and philosophers have recommended adding to our scientific methodology first-person methods, that is, methods of investigation that rely on subjects' access to private, first-person evidence (Chalmers, 1996b; Goldman, 1997).

A prominent first-person method is phenomenology, in Edmund Husserl's sense. There is now a research program, articulated most explicitly by Francisco Varela, which aims at merging neuroscience and phenomenology (Thomson, Lutz, & Cosmelli, 2005; Varela, 1996). One active area within this tradition is the study of "the phenomenology of time consciousness"—roughly speaking, the aspects of conscious experience that have to do with time. For instance, a recent study by Dan Lloyd



analyzes fMRI data for clues about the temporal structure of experience (Lloyd, 2002). (In so far as Lloyd relies directly on imaging data, his approach is an example of the “third way” proposed by Egan and Matthews.)

Rick Grush (this issue) critically examines previous proposals, such as Lloyd’s, and then offers a novel way of bridging neuroscience and the phenomenology of time consciousness. Grush’s model is based on his view that neural representation is made possible by the neural emulation of the dynamic interaction between self and environment (Grush, 2004). Thus, Grush’s study is an example of neurophenomenology—one way that neuroscience may accommodate subjective experience.

There is another way. Introspective reports may be taken seriously as sources of evidence without needing to embrace first-person methods. Already many years ago, Allen Newell and Herbert Simon went beyond the traditional skeptical approach towards introspective reports. They developed protocol analysis, which was an important source of data for their theory of human problem solving (Newell & Simon, 1972). Protocol analysis is a methodology for extracting reliable information about mental states and processes from introspective reports, using public (third-person) procedures. The methodology was later refined by Simon in collaboration with K. Anders Ericsson (Ericsson, 2003; Ericsson & Simon, 1993). Protocol analysis may be divorced from computational chauvinism and generalized to other kinds of reports and tasks beyond those reviewed by Ericsson and Simon. The result is a general third-person methodology for using introspective reports as a source of evidence (Piccinini, 2003).

## 8 Conclusion

I have briefly discussed some of the issues that surround computational explanation in neuroscience and the way the present essays contribute to them. I did not discuss many other issues: Are explanations of mental capacities individualistic or anti-individualistic (in the sense of Burge, 1986)? Are brains computationally equivalent to Turing machines, less powerful, or more powerful? Are brains computers at all? How are higher levels related to lower levels? The list could go on. To answer these questions adequately, much more work remains to be done.

**Acknowledgements** The impetus for the papers collected here came from a workshop on Computational Explanation in Neuroscience that was held at Washington University in St. Louis on November 5–6, 2004. Among the presenters were Carl Craver, Frances Egan, and Oron Shagrir, though only Craver’s essay originates from his presentation. Many thanks to John Bickle, José Bermudez, the Philosophy–Neuroscience–Psychology Program, Arts and Sciences at Washington University, the Park Foundation, and all the participants in the workshop. Thanks to Carl Craver, John Gabriel, and Oron Shagrir for comments on this article. Preparation of this article was supported by the National Endowment for the Humanities and a Research Award from the University of Missouri—St. Louis. The views expressed here do not necessarily reflect those of these institutions.

## References

- Adrian, E. D. (1928). *The basis of sensation: The action of the sense organs*. New York: Norton.
- Anderson, J. A., Pellionisz, A., & Rosenfeld, E. (1990). *Neurocomputing 2: Directions for research*. Cambridge, MA: MIT Press.
- Anderson, J. A., & Rosenfeld, E. (Eds.) (1988). *Neurocomputing: Foundations of research*. Cambridge, MA: MIT Press.

- Anderson, J. A., & Rosenfeld, E. (Eds.) (1998). *Talking nets: An oral history of neural networks*. Cambridge, MA: MIT Press.
- Bechtel, W. (2001). Cognitive neuroscience: Relating neural mechanisms and cognition. In P. Machamer, R. Grush, & P. McLaughlin (Eds.), *Theory and method in the neurosciences* (pp. 81–111). University of Pittsburgh Press.
- Boden, M. (1991). Horses of a different color? In W. Ramsey et al. (Eds.), *Philosophy and connectionist theory* (pp. 3–19). Hillsdale: LEA.
- Burge, T. (1986). Individualism and psychology. *Philosophical Review*, 95, 3–45.
- Chalmers, D. J. (1996a). Does a rock implement every finite-state automaton? *Synthese*, 108, 310–333.
- Chalmers, D. J. (1996b). *The conscious mind: In search of a fundamental theory*. Oxford: Oxford University Press.
- Churchland, P. S. (1986). *Neurophilosophy*. Cambridge, MA: MIT Press.
- Churchland, P. S., Koch, H., & Sejnowski, T. (1990). What is computational neuroscience? In E. L. Schwartz (Ed.), *Computational neuroscience* (pp. 46–55). Cambridge, MA: MIT Press.
- Churchland, P. S., & Sejnowski, T. J. (1992). *The computational brain*. Cambridge, MA: MIT Press.
- Copeland, B. J. (1996). What is computation? *Synthese*, 108, 224–359.
- Copeland, B. J. (2000). Narrow Versus Wide Mechanism: Including a Re-Examination of Turing's Views on the Mind-Machine Issue. *The Journal of Philosophy*, XCVI, 5–32.
- Crane, T. (1990). The language of thought: No syntax without semantics. *Mind and Language*, 5(3), 187–212.
- Craver, C. F. (forthcoming). *Explaining the brain*. Oxford: Oxford University Press.
- Cruse, H. (2001). The explanatory power and limits of simulation models in the neurosciences. In P. Machamer, R. Grush, & P. McLaughlin (Eds.), *Theory and method in the neurosciences* (pp. 138–154). University of Pittsburgh Press.
- Dayan, P., & Abbott, L. F. (2001). *Theoretical neuroscience: Computational and mathematical modeling of neural systems*. Cambridge, MA: MIT Press.
- Dennett, D. C. (1991). *Consciousness explained*. Boston: Little, Brown & Co.
- Dennett, D. C. (2003). Who's on first? Heterophenomenology explained. *Journal of Consciousness Studies*, 10(9–10), 19–30.
- Edelman, G. M. (1992). *Bright air, brilliant fire: On the matter of the mind*. New York: Basic Books.
- Egan, F. (1995). Computation and content. *Philosophical Review*, 104, 181–203.
- Ericsson, K. A. (2003). How to elicit verbal reports that provide valid unobtrusive externalization of concurrent thinking? *Journal of Consciousness Studies*, 10, 9–10.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis*. Cambridge, MA: MIT Press.
- Feest, U. (2003). Functional analysis and the autonomy of psychology. *Philosophy of Science*, 70, 937–948.
- Fodor, J. A. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.
- Fodor, J. A. (1981). The mind–body problem. *Scientific American*, 244. In J. Heil (Ed.), *Philosophy of mind: A guide and anthology* (pp. 168–182). Oxford: Oxford University Press (Reprinted 2004).
- Fodor, J. A. (1998). *Concepts*. Oxford: Clarendon Press.
- Frank, R. G. (1994). Instruments, nerve action, and the all-or-none principle. *Osiris*, 9, 208–235.
- Freeman, W. J. (2001). *How brains make up their minds*. New York: Columbia University Press.
- Gallistel, C. R. (1990). *The organization of learning*. Cambridge, MA: MIT Press.
- Gazzaniga, M. S. (Ed.). (2000). *The new cognitive neurosciences*. Cambridge, MA: MIT Press.
- Gerard, R. W., & Duyff, J. W. (Eds.). (1962) *Information processing in the nervous system: Volume III proceedings of the international union of physiological sciences (XXII International Congress, Leiden, 1962)*. Amsterdam: Excerpta Medica Foundation.
- Globus, G. G. (1992). Towards a noncomputational cognitive neuroscience. *Journal of Cognitive Neuroscience*, 4(4), 299–310.
- Goldman, A. I. (1997). Science, publicity, and consciousness. *Philosophy of Science*, 64, 525–545.
- Grush, R. (2001). The semantic challenge to computational neuroscience. In P. Machamer, R. Grush, & P. McLaughlin (Eds.), *Theory and method in the neurosciences* (pp. 155–172). University of Pittsburgh Press.
- Grush, R. (2004). The emulation theory of representation: Motor control, imagery, and perception. *Behavioral and Brain Sciences*, 27, 377–442.
- Jeffress, L. A. (Ed.). (1951). *Cerebral mechanisms in behavior*. New York, Wiley.
- Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference and consciousness*. New York: Cambridge University Press.
- Keeley, B. (2000). Shocking lessons from electric fish: The theory and practice of multiple realizability. *Philosophy of Science*, 67, 444–465.

- Lloyd, D. (2002). Functional MRI and the study of human consciousness. *Journal of Cognitive Neuroscience*, 14, 818–831.
- Lycan, W. (1987). *Consciousness*. Cambridge, MA: MIT Press.
- Machtey, M., & Young, P. (1978). *An introduction to the general theory of algorithms*. New York: North Holland.
- McClelland, J. L., & Patterson, K. (2002). Rules or connections in past-tense inflections: What does the evidence rule out? *Trends in Cognitive Science*, 6, 465–472.
- McCulloch, W. S. (1949). The brain as a computing machine. *Electrical Engineering*, 68, 492–497.
- McCulloch, W. S., & Pitts, W. H. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 7, 115–133.
- Miller, G. A., Galanter, E. H., & Pribram, K. H. (1960). *Plans and the structure of behavior*. New York: Holt.
- Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs: Prentice-Hall.
- Newell, A., & Simon, H. A. (1976). Computer science as an empirical enquiry: Symbols and search. *Communications of the ACM*, 19, 113–126.
- Peacocke, C. (1999). Computation as involving content: A response to Egan. *Mind and Language*, 14(2), 195–202.
- Perkel, D. H. (1990). Computational neuroscience: Scope and structure. In E. L. Schwartz (Ed.), *Computational neuroscience* (pp. 38–45). Cambridge, MA: MIT Press.
- Piccinini, G. (2003). Data from introspective reports: Upgrading from commonsense to science. *Journal of Consciousness Studies*, 10(9–10), 141–156.
- Piccinini, G. (2004). The First computational theory of mind and brain: A close look at McCulloch and Pitts's 'logical calculus of ideas immanent in nervous activity'. *Synthese*, 141(2), 175–215.
- Piccinini, G. (forthcoming a). Computationalism, the Church–Turing Thesis, and the Church–Turing Fallacy. *Synthese*.
- Piccinini, G. (forthcoming b). Computational modeling vs. computational explanation: Is everything a Turing machine, and does it matter to the philosophy of mind? *Australasian Journal of Philosophy*.
- Piccinini, G. (forthcoming c). Computational explanation and mechanistic explanation of mind. In M. De Caro, F. Ferretti, & M. Marraffa (Eds.), *Cartographies of the mind: The interface between philosophy and cognitive science*. Dordrecht: Kluwer.
- Piccinini, G. (forthcoming d). Computation without representation. *Philosophical Studies*.
- Pinker, S., & Ullman, M. (2002). The past-tense debate: The past and future of the past tense. *Trends in Cognitive Science*, 6, 456–463.
- Pylyshyn, Z. W. (1984). *Computation and cognition*. Cambridge, MA: MIT Press.
- Revonsuo, A. (2001). On the nature of explanation in the neurosciences. In P. Machamer, R. Grush, & P. McLaughlin (Eds.), 2001, *Theory and method in the neurosciences* (pp. 45–69). University of Pittsburgh Press.
- Rumelhart, D. E., McClelland, J. M., and the PDP Research Group (1986). *Parallel distributed processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.
- Searle, J. R. (1992). *The rediscovery of the mind*. Cambridge, MA: MIT Press.
- Shagrir, O. (2001). Content, computation and externalism. *Mind*, 110(438), 369–400.
- Smith, B. C. (2002). The foundations of computing. In M. Scheutz (Ed.), *Computationalism: New directions* (pp. 23–58). Cambridge, MA: MIT Press.
- Stich, S. (1983). *From folk psychology to cognitive science*. Cambridge: MIT Press.
- Thomson, E., Lutz, A., & Cosmelli, D. (2005). Neurophenomenology: An introduction for neurophilosophers. In A. Brook, & K. Akins (Eds.), *Cognition and the brain: The philosophy and neuroscience movement* (pp. 40–97). New York: Cambridge University Press.
- Varela, F. (1996). Neurophenomenology: A methodological remedy for the hard problem. *Journal of Consciousness Studies*, 3(4), 330–349.
- Wolfram, S. (2002). *A new kind of science*. Champaign, IL: Wolfram Media.