ELSEVIER

Stud. Hist. Phil. Sci. 35 (2004) 811–833

Studies in History and Philosophy of Science

www.elsevier.com/locate/shpsa

Functionalism, computationalism, and mental states

Gualtiero Piccinini

Department of Philosophy, Washington University, Campus Box 1073, One Brookings Dr., St. Louis, MO 63130-4899, USA

Received 25 November 2003; received in revised form 24 February 2004

Abstract

Some philosophers have conflated functionalism and computationalism. I reconstruct how this came about and uncover two assumptions that made the conflation possible. They are the assumptions that (i) psychological functional analyses are computational descriptions and (ii) everything may be described as performing computations. I argue that, if we want to improve our understanding of both the metaphysics of mental states and the functional relations between them, we should reject these assumptions.

Keywords: Functionalism; Computationalism; Computational functionalism; Mental states; Computational theory of mind; Functional analysis

1. Introduction

> Particularly striking in retrospect was the widespread failure to distinguish the computational program in psychology from the functionalist program in metaphysics . . . (For an instance where the two are run together, see Fodor 1968[b]). (Fodor, 2000, p. 105 n. 4)

Fodor is right: some philosophers of mind have conflated functionalism and computationalism about mental states. To a first approximation, functionalism is the metaphysical view that mental states are individuated by their functional relations with mental inputs, outputs, and other mental states. Functionalism *per se* is

E-mail address: gpiccini@artsci.wustl.edu (G. Piccinini).

0039-3681/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.shpsa.2004.02.003

neutral on how those functional relations should be characterized. Specifically, functionalism is not committed to the view that the functional relations that individuate mental states are computational. Computationalism, instead, is precisely the hypothesis that the functional relations between mental inputs, outputs, and internal states are computational. Computationalism *per se* is neutral on whether those computational relations constitute the nature of mental states. Throughout this paper, I will assume that functionalism and computationalism are distinct doctrines, neither one of which entails the other.[1] If neither view entails the other, how did they get conflated?

This question is interesting for both historical and diagnostic reasons. On the one hand, reconstructing this episode of recent philosophy reveals the great influence that Hilary Putnam and Jerry Fodor's writings in the 1960s have exerted on contemporary philosophy of mind. On the other hand, doing so unearths two unwarranted assumptions that have affected—and still affect—debates in the philosophy of mind. The two assumptions are that psychological functional analyses are computational descriptions and that everything performs computations. This paper tells the story of the conflation between functionalism and computationalism and calls for the rejection of the assumptions that led to it. As a result, it helps pave the way for an improved understanding of the metaphysics of mental states, the functional relations between them, and how those two issues relate to each other. At the same time, it also helps appreciate the positive contributions contained in Putnam and Fodor's early philosophy of mind in a fresh and uncluttered way.

I will focus on philosophers who introduced and discussed functionalism in the context of *scientific* theories of mind. Other philosophers introduced and discussed functionalism in the context of *folk* theories of mind (Lewis, 1966, 1972, 1980; Armstrong, 1970) or alleged *analytic truths* about the mind (Shoemaker, 1984). The latter philosophers did not give a computational formulation of functionalism, nor did they imply that functionalism entailed computationalism, and therefore their views are irrelevant to the present topic.

2. The brain as a Turing Machine

Modern computationalism was formulated by Warren McCulloch in the 1930s, and published for the first time by him and his collaborator Walter Pitts in the 1940s (McCulloch & Pitts, 1943).[2] Roughly speaking, McCulloch and Pitts held that the functional relations between mental inputs, outputs, and internal states were computational, in the sense rigorously defined a few years earlier by Alan Turing in terms of his Turing Machines (Turing, 1965; first published 1936–1937). McCulloch and Pitts also held that specific mental phenomena could be explained

[1] For an explicit defense of these formulations of functionalism and computationalism, and of their logical independence, see Piccinini (2003a), Ch. 8.

[2] For a detailed analysis of McCulloch and Pitts's theory, see Piccinini (forthcoming). For a more detailed reconstruction of the early history of computationalism, see Piccinini (2003a), Chs. 2–6, and (2003b).

by hypothesizing specific computations that could bring them about. According to McCulloch and Pitts, the computations postulated by their theory of mind were performed by specific neural mechanisms. McCulloch and Pitts offered rigorous mathematical techniques for designing neural circuits that performed those computations. Finally, they held that by explaining mental phenomena in terms of neural mechanisms, their theory solved the mind–body problem, but they did not formulate an explicit solution to the mind–body problem.

Computationalism was initially picked up by a group of neurophysiologists, mathematicians, and engineers, whose core members—besides McCulloch and Pitts—were Julian Bigelow, Arturo Rosenblueth, Norbert Wiener, and John von Neumann. Among his important contributions to computationalism, von Neumann studied how computations like those hypothesized by McCulloch and Pitts could be performed by mechanisms—such as neural mechanisms—whose components had a probability of malfunctioning higher than zero. Von Neumann's study—together with other contributions to computationalism and computability theory—was published in a collection edited by Claude Shannon in collaboration with one of von Neumann's students, John McCarthy (Shannon & McCarthy, 1956).

Computationalism attracted the attention of two philosophers. In a long paper, which offered a broadly type–type reductionist picture of the world, Paul Oppenheim and Hilary Putnam addressed the reduction of mental phenomena—such as 'learning, intelligence, and perception'—to cellular activity, and specifically to the activity of networks of neurons (Oppenheim & Putnam, 1958, pp. 18–19).

Oppenheim and Putnam argued that Turing's analysis of computation 'naturally' led to the hypothesis that the brain was a Turing Machine (TM):

> The logician Alan Turing proposed (and solved) the problem of giving a characterization of *computing machines* in the widest sense—mechanisms for solving problems by effective series of logical operations. This naturally suggests the idea of seeing whether a 'Turing machine' could consist of the elements used in neurological theories of the brain; that is, whether it could consist of a network of neurons. Such a nerve network could then serve as a hypothetical model for the brain. (Oppenheim & Putnam, 1958, p. 19; their emphasis)

Then, Oppenheim and Putnam pointed to McCulloch and Pitt's theory (McCulloch & Pitts, 1943), von Neumann's model of reliable computation from unreliable components (von Neumann, 1956), and related work (for example, Shannon & McCarthy, 1956) as theories of the brain that were capable of reducing mental phenomena to neural activity.[3]

[3] They added: 'In terms of such nerve nets it is possible to give hypothetical micro-reductions for *memory, exact thinking, distinguishing similarity or dissimilarity in stimulus pattern, abstracting* of "essential" components of a simulus pattern, recognition of shape regardless of form and of chord regardless of pitch . . ., *purposeful behavior* as controlled by negative feedback, *adaptive behavior*, and *mental disorders*' (Oppenheim & Putnam, 1958, p. 20; their emphasis).

Although Oppenheim and Putnam did not explicitly mention the mind-body problem and did not comment on McCulloch and Pitts's statement that their computational theory solved it, Oppehneim and Putnam implicitly presented computationalism as a type–type physicalist solution to the mind–body problem. The conceptual step from there to the conclusion that the mind was the abstract computational organization of the brain (as described by its TM table) was relatively short. This suggests that Putnam's later computationalism about the mind derived at least in part from the belief that the brain was a TM.

In subsequent papers published in the early 1960s, Putnam retained the view that the brain might be a (probabilistic) TM, which he sometimes called a probabilistic automaton. In 1961 Putnam wrote, 'I would suggest that there are many considerations which point to the idea that a Turing machine plus random elements is a reasonable model for the human brain' (Putnam, 1975a, p. 102). In that paper, Putnam did not say what those considerations were. In a later paper of 1964 (Putnam, 1975b), in a similar context, he mentioned McCulloch and Pitts's theory and the mechanical mice created by Shannon (1952).

3. The analogy between minds and Turing Machines

Putnam addressed the mind-body problem explicitly in a paper published in 1960, where he argued that there was an 'analogy between man and Turing machine' (Putnam, 1960, p. 159). According to Putnam, mental states were analogous to the internal states of a TM as described by its abstract machine table, whereas brain states were analogous to the physical states of a hardware realization of a TM.

Putnam used his analogy to construct an analogue of the mind-body problem for TMs. The existence of the TM analogue problem, in turn, was meant to show that the mind-body problem was a purely 'verbal' or 'linguistic' or 'logical' problem, which did not demand a solution any more than its TM analogue demanded a solution. Here we won't dwell on Putnam's critique of the mind-body problem, but only on the relationship between Putnam's analogy and the doctrine of functionalism. Throughout his paper, Putnam stressed that the analogy between minds and TMs could not be stretched too far. He explicitly rejected the following claims: that the mind–body problem literally arises for TMs (ibid., p. 138), that machines think or that humans are machines (ibid., p. 140), and that machines can be properly said to employ a language (ibid., p. 159).

Putnam also applied his analogy between men and TMs to psychological theories:

> It is interesting to note that just as there are two possible descriptions of the behavior of a Turing machine—the engineer's structural blueprint and the logician's 'machine table'—so there are two possible descriptions of human psychology. The 'behavioristic' approach ... aims at eventually providing a complete physicalistic [fn: In the sense of Oppenheim & Putnam, 1958 ...] description of human behavior. This corresponds to the engineer's or physicist's

> description of a physically realized Turing machine. But it would also be possible to seek a more abstract description of human mental processes, in terms of 'mental states' (physical realization, if any, unspecified) and 'impressions' (these play the role of symbols on the machine's tapes)—a description which would specify the laws controlling the order in which the states succeed one another, and the relation to verbalization . . . This description, which would be the analogue of a 'machine table,' it was in fact the program of classical psychology to provide! (Ibid., pp. 148–149)

In this passage, Putnam pointed out that psychological theories could be formulated in two ways: one described behavioral dispositions and physiological mechanisms, the other described 'mental states' and 'impressions'. Then Putnam suggested that *if* it were possible to formulate an 'abstract' psychological theory in terms of mental states, then that theory would stand to a psychological theory describing physiological mechanisms in the same relation that TM programs stood to descriptions of their physical realizations. Putnam's suggestion was offered without argument, as an analogy with descriptions of TMs.

After drawing the analogy between psychological theories and TMs, Putnam explicitly *denied* that 'abstract' psychological theories could be formulated, because functionalism about minds did not hold:

> Classical psychology is often thought to have failed for methodological reasons; I would suggest, in the light of this analogy, that it failed rather for empirical reasons—the mental states and 'impressions' of human beings do not form a causally closed system to the extent to which the 'configurations' of a Turing machine do. (Ibid., p. 149)

By hypothesizing that human minds 'do not form a causally closed system', Putnam suggested that there could be no theory of minds in terms of 'mental states', and *a fortiori* no complete functional description of minds. This view is the denial of functionalism, and it shows that in his first paper on minds and TMs, Putnam was not yet a functionalist. In that paper, his only goal was to convince the reader that there was enough positive analogy between humans and TMs to show that the mind–body problem was a pseudo-problem.

According to Putnam, the positive analogy between minds and TMs included an important element that later became part of functionalism. That element was multiple realizability, namely the view that the same mental states could be physically realized in different ways. Putnam wrote as follows:

> The functional organization (problem solving, thinking) of the human being or machine can be described in terms of the sequences of mental or logical states respectively (and the accompanying verbalizations), without reference to the nature of the 'physical realization' of these states. (Ibid., p. 149)

Another element of Putnam's positive analogy was going to play an important role in computationalism. This was the view that TM computations are like human thought processes in that that they admit of degrees of *rationality*: 'In the case of rational thought (or computing), the "program" which determines which states follow which, etc., is open to rational criticism' (ibid., p. 149). This was presumably because TM tables were algorithms, and algorithms were a (more or less appropriate) rationale for solving problems (cf. ibid., p. 143). These elements would become integral parts of Putnam's mature formulation of functionalism. We'll come back to them later.

4. Functionalism

According to Putnam (1980, p. 35 n. 2), he first formulated the doctrine of functionalism about minds in his paper 'The mental life of some machines', delivered to the Wayne State University Symposium in the Philosophy of Mind in 1962, and published as Putnam (1967). In that paper, Putnam again introduced his analogy between minds and TMs, this time with the purpose of analyzing mentalistic notions such as 'preferring', 'believing', and 'feeling'. He pointed out that in many ways, the sense in which TMs were said to have beliefs, preferences, or feelings, was different from the sense in which human beings were said to have such states. But he submitted that these differences were irrelevant to his argument (ibid., pp. 177–178). His strategy was to apply mentalistic descriptions to appropriate TMs and argue that for these machines, all traditional metaphysical doctrines about minds (materialism, dualism, and logical behaviorism) were incorrect (ibid.). This constituted a shift from the 1960 article, where the same analogy between minds and TMs was used to argue that the mind–body problem was a purely verbal problem.

Still, Putnam explicitly argued that 'it is somewhat unlikely that either the mind or the brain is a Turing Machine' (ibid., p. 184). His reasons had mainly to do with the probabilistic nature of neural or mental events: '[r]easoning *a priori* one would think it more likely that the interconnections among the various brain states and mental states of a human being are probabilistic rather than deterministic and that time-delays play an important role' (ibid.). Putnam immediately added that it might be difficult to discover the actual functional organization of the mind or the brain empirically:

> [A]n automaton whose states are connected by probabilistic laws and whose behavior involves time-delays can be arbitrarily well-simulated by the behavior of a Turing Machine. Thus, in the nature of the case, mere empirical data cannot decide between the hypothesis that the human brain (respectively, *mind*) is a Turing Machine and the hypothesis that it is a more complex kind of automaton with probabilistic relations and time-delays. (Ibid.)

Unfortunately, Putnam did not justify or elaborate on his statement that automata connected by probabilistic laws and whose behavior involves time-delays 'can be arbitrarily well-simulated by the behavior of a Turing Machine', nor did he give any references. He also did not say how, from that, it followed that empirical data could not decide between the hypothesis that the mind or brain was functionally organized like a TM and the hypothesis that it had a different functional organization. In the absence of more details on these statements of Putnam's, it is difficult to evaluate their plausibility. But regardless of whether he was justified in concluding that the two hypotheses were empirically equivalent, Putnam explicitly distinguished between the two hypotheses.

Later in the paper, instead of the traditional metaphysical mind-body doctrines, Putnam offered functionalism about minds:

> It seems that to know for certain that a human being has a particular belief, or preference, or whatever, involves knowing something about the functional organization of the human being. As applied to Turing Machines, the functional organization is given by the machine table. A description of the functional organization of a human being might well be something quite different and more complicated. But the important thing is that descriptions of the functional organization of a system are logically different in kind either from descriptions of its physical–chemical composition or from descriptions of its actual and potential behavior. (Ibid., p. 200)

In this passage, Putnam explicitly proposed that knowing minds was the same as knowing the functional organization of human beings, in analogy with the functional organization of TMs that is given by a TM table. But he immediately added that the description of the functional organization of a human being might be 'something quite different and more complicated' than a TM description. In saying 'something quite different and more complicated', Putnam probably meant a mechanism similar to TMs but with probabilistic state transitions and time-delays, which he took to be empirically indistinguishable from TMs. At any rate, Putnam left his formulation of functionalism open-ended as to the functional organization of the mind, and said that 'the important thing is that descriptions of the functional organization of a system are logically different in kind either from descriptions of its physical–chemical composition or from descriptions of its actual and potential behavior'. This shows that Putnam's original doctrine of functionalism was not committed to computationalism. That is, Putnam initially formulated the doctrine that mental states were functional states, in analogy with the way that TM states are functional states, without assuming that the functional relations between mental states and inputs, outputs, and other internal states were computational.

5. Psychological theories as functional analyses

Putnam's (1960) theme of psychological theories was picked up by his student Jerry Fodor, and through Fodor's work it became an important motivation for a computational formulation of functionalism. Fodor offered his analysis of psychological theories in an article that had 'a particular indebtedness' to a book by psychologist J. A. Deutsch (1960) and to Putnam's 1960 article (Fodor, 1965, p. 161). Fodor's paper was also influenced by psychologist Stuart Sutherland's work (for example, Sutherland, 1960) on octopus vision (Fodor, personal correspondence).

Fodor's account of psychological theories was largely borrowed from Deutsch (1960, especially pp. 10–15). According to Fodor, psychological theories were developed in two logically distinct phases. Fodor called phase one theories *functional analyses*, and phase two theories *mechanical analyses*. Fodor explicated the distinction between functional and mechanical analysis by the example of internal combustion engines. Functional analysis identified the functions of engine parts, namely their contribution to the activities of the whole engine. For an internal combustion engine to generate motive power, fuel must enter the cylinders, where it detonated and drove the cylinders. In order to regulate the flux of fuel into the cylinders, functional analysis said there must be valves that were opened by valve lifters. Valve lifters contributed to the activities of the engine by lifting valves that let fuel into the cylinders. Given a functional analysis of an engine, mechanical analysis identified physical structures that corresponded to the functional analysis of the engine. In certain engines, camshafts were identified as physical structures that functioned as valve lifters, although in other engines the same function may be performed by other physical structures. By the same token, according to Fodor, phase one psychological theories identified psychological functions (functional analysis), whereas phase two psychological theories identified physiological structures that performed those functions (mechanical analysis). From his notion of functional analysis as explicated by his example of engines, Fodor inferred that psychological theories had indefinitely many realizations or 'models'; that is, different mechanisms could realize a given functional analysis (Fodor, 1965, pp. 174–175). Fodor also argued that the relationship between phase one (functional analysis) and phase two (mechanical analysis) psychological theories was not a relation of reductive 'microanalysis' in the sense of Oppenheim and Putnam (1958), in which there was a type–type correspondence between the predicates of the two descriptions (ibid., p. 177). In all of this, Fodor was following quite closely the account of psychological theories proposed by Deutsch, who also inferred that there was a 'theoretically infinite variety of counterparts' of any type of mechanism postulated by phase one psychological theories (Deutsch, 1960, p. 13).

Prima facie, Fodor could have proposed his analysis of psychological theories independently of Putnam's (1960) analogy between TMs and psychological descriptions. On the one hand, TMs are individuated by a finite number of internal *state* types and what state transitions must occur under what conditions, regardless of what component types must make up the system that realizes the TM or what their functions must be. On the other hand, functional analyses—based on Fodor's examples and Deutsch's formulation—are specifications of mechanism types, consisting of different *component* types and their assigned functions without specifying the precise state types and state transitions that must occur within the analyzed system. A functional analysis of an engine does not come in the form of a TM table, nor is it obvious how it could be turned into a TM table or whether turning it into a TM table would have any value for explaining the functioning of the engine. TM tables can be analyzed into subroutines, and subroutines can be analyzed into sequences of elementary operations, but this is not a functional analysis in Deutsch's sense. Even the fact that both TM tables and functional analyses can be multiply realized seemed to originate from different reasons. A functional analysis can be multiply realized because systems with different physical properties can perform the same concrete function (for example, generate motive power), whereas a TM table can be multiply realized because systems with different physical properties, no matter what functions they perform, can realize the same abstractly specified state transitions. At the very least, the thesis that the two are related requires analysis and argument. So, *prima facie* there is no reason to think that giving a functional analysis of a system is equivalent to describing that system by using a TM table. In fact, neither Fodor nor his predecessor Deutsch mentioned computers or TMs, nor did they infer, from the fact that psychological descriptions are functional analyses, that either the mind or the brain was a TM.

Nevertheless, when Fodor described phase one psychological theories in general, he departed from his example of the engine and from Deutsch's view. Fodor described psychological functional analyses as postulations of internal *states*, that is, as descriptions that closely resembled TM descriptions: 'Phase one explanations purport to account for behavior in terms of internal states' (Fodor, 1965, p. 173). This way of describing functional analyses was clearly under the influence of Putnam's (1960) analogy between minds and TMs. In a later work, where Fodor repeated his two-phase analysis of psychological theories, he explicitly attributed his formulation of phase one psychological theories to Putnam's (1960) analogy (Fodor, 1968a, p. 109).[4] In his 1965 article, however, Fodor did not discuss TMs explicitly, nor did he explain how his formulation of phase one psychological theories in terms of states squared with his example of the engine.

Thus, Fodor (1965) introduced in the philosophical literature both the notion that psychological theories were functional analyses and the notion that psychological theories were like TM tables in that they were descriptions of transitions between state types. Both themes would be very successful in philosophy of psychology.

Fodor's analysis of psychological theories may have helped Putnam accept the possibility—which he had rejected in his 1960 article—of formulating a psychological theory of human behavior. In a paper of 1964 that appeared in print before Fodor's, Putnam seemed to endorse Fodor's (1965) elaboration of Putnam's (1960) view of psychological theories:

> Psychological theories say that an organism has certain states which are not specified in 'physical' terms, but which are taken as primitive. Relations are specified between these states, and between the totality of the states and sensory inputs ('stimuli') and behavior ('responses'). Thus, as Jerry Fodor has remarked (Fodor, 1965), it is part of the 'logic' of psychological theories that (physically) *different* structures may obey (or be 'models' of) the *same* psychological theory. (Putnam, 1975b, p. 392; original emphasis)

[4] Fodor didn't mention that Putnam (1960) claimed that there couldn't be a psychological theory of this type because minds were not causally closed systems; Fodor implicitly disagreed with Putnam on this point.

Later in the same paper, Putnam offered a conditional formulation of a computational version of functionalism. He did not explicitly endorse the view that the brain was a probabilistic TM, but he wrote that *if* the human brain were such a device, then any physical realization of the same TM table would have the same *psychology* that humans had:

> [I]f the human brain is a 'probabilistic automaton,' then any robot with the same 'machine table' will be psychologically isomorphic to a human being. If the human brain is simply a neural net with a certain program, as in the theory of Pitts and McCulloch, then a robot whose 'brain' was a similar net, only constructed of flip-flops rather than neurons, would have exactly the same psychology as a human. (Ibid., pp. 394–395)

This shows that in the mid-1960s, Putnam was still taking seriously McCulloch and Pitts's theory that the brain was a computing mechanism, and concluded from it that a robot realizing the appropriate computational description would have the same psychology that humans did. Soon thereafter, Putnam finally embraced computationalism about the mind.

6. Computational functionalism

Putnam upgraded from functionalism to *computational* functionalism in his best known paper on the subject, published in 1967 (Putnam, 1999). In that paper, he stated his functionalist doctrine directly in terms of probabilistic TMs:

> A Description of S where S is a system, is any true statement to the effect that S possesses distinct states S_1, S_2 . . ., S_n which are related to one another and to the motor outputs and sensory inputs by the transition probabilities given in such-and-such a [Turing] Machine Table. The Machine Table mentioned in the Description will then be called the Functional Organization of S relative to that Description, and the S_i such that S is in state S_i at a given time will be called the Total State of a system relative to that Description. It should be noted that knowing the Total State of a system [is] relative to a Description. Description involves knowing a good deal about how the system is likely to 'behave,' given various combinations of sensory inputs, but does *not* involve knowing the physical realization of the S_i as, e.g., physical–chemical states of the brain. The S_i, to repeat, are specified only *implicitly* by the Description—i.e., specified *only* by the set of transition probabilities given in the Machine Table.
>
> The hypothesis that 'being in pain is a functional state of the organism' may now be spelled out more exactly as follows:

1. All organisms capable of feeling pain are Probabilistic Automata.
2. Every organism capable of feeling pain possesses at least one Description of a certain kind (i.e., being capable of feeling pain is possessing an appropriate kind of Functional Organization).

3. No organism capable of feeling pain possesses a decomposition into parts which separately possess Descriptions of the kind referred to in (2).
4. For every Description of the kind referred to in (2), there exists a subset of the sensory inputs such that an organism with that Description is in pain when and only when some of its sensory inputs are in that subset. (Putnam, 1999, p. 30; original emphasis)

Clause 2 says that there is a particular TM table that describes creatures with mental states. Clause 4 says that a creature is in pain if and only if that particular TM table, when applied to the creature, indicates that the creature is in the type of state that corresponds to pain. Since this is supposed to hold for every mental state, mental states can be individuated through a TM table. (Clause 3 is an *ad hoc* addition to avoid attributing mental states to collective individuals, for example, bee swarms (ibid., p. 31).)

In the same paper, Putnam argued that his proposal was a more plausible 'empirical hypothesis' about mental states than either type-identity materialism or logical behaviorism. Putnam's arguments about both doctrines went as follows.

First, type-identity materialism. For present purposes, type-identity materialism is the thesis that any type of mental state is identical to some type of brain state. Putnam argued that type-identity materialism was committed to finding the same types of brain states in all organisms that realized the same types of mental states, whether they were mammals, reptiles, mollusks, or Martians. He also argued that this was implausible, because these organisms had widely different brains, or perhaps no brains at all. These multiple realizability considerations were similar to those he made in previous (pre-computationalism) papers (Putnam, 1960, 1967); they were also analogous to the multiple realizability considerations made by Fodor (1965) employing the distinction between the functions and component types of a mechanism, which did not rely on the notion of computation. Unlike type-identity materialism, Putnam's functionalist doctrine was only committed to those organisms having the same functional organization, not to their having the same types of brain states. Hence, Putnam concluded that functionalism was more plausible than type-identity materialism (Putnam, 1999, pp. 31–32; first published 1967). The assumption that the functional organization of minds was given in terms of TM tables played no role in Putnam's comparison of the plausibility of functionalism and type-identity materialism, and was not even mentioned by Putnam while making that comparison.[5]

Second, logical behaviorism. For present purposes, logical behaviorism is the thesis that mental states are sets of behavioral dispositions. About logical behaviorism, Putnam offered a number of considerations similar to those he offered in previous papers of 1963 and 1967 (Putnam, 1980, 1967). The main consideration

[5] For a more systematic list of multiple realizability arguments, none of which establishes the conclusion that the mind is computational, see Piccinini (2003a), Ch. 8.

hinged on the premise that, contrary to logical behaviorism, the same set of behavioral dispositions may correspond to different mental conditions. One of Putnam's examples involved two animals whose motor nerves were cut. These animals had the same set of behavioral dispositions, namely, they were paralyzed. But intuitively, if the first animal had intact sensory nerves whereas the second animal had cut sensory nerves, under appropriate stimulation the first animal would feel pain whereas the second animal wouldn't. So, contrary to logical behaviorism, being in pain was not a behavioral disposition. Putnam's functionalist doctrine escaped this objection because it individuated the different states of the two animals in terms of the effects of sensory inputs (or lack thereof) on their internal states. Hence, Putnam concluded that functionalism was more plausible than logical behaviorism (1999, pp. 32–33). Again, the assumption that the functional organization of minds was given by TMs played no role in Putnam's comparison of functionalism and logical behaviorism, and it was not mentioned by Putnam during his discussion.

So, what was Putnam's reason for formulating his functionalist solution to the mind–body problem as entailing computationalism about the mind? The most important part of the answer is probably rooted in the history of science. In Section 2 we saw that at the time of Putnam's writing, the question of whether minds were machines and thus whether machines could think had become a hot topic of debate among scientists. By the late 1940s, members of the cybernetics movement proposed the construction of intelligent machines and popularized the idea that machines could think. Their view was that brains were computing mechanisms, and that the appropriate kind of computing mechanisms could think like brains. And in the late 1950s, the new discipline of Artificial Intelligence (AI) was created with the explicit purpose of programming digital computers to produce behavior that was ordinarily considered intelligent. Some members of the AI community explicitly construed their research program as the construction of psychological theories of human thinking in the form of computer programs.[6] As a result, the debate around the relationship between mind-brains and machines centered on digital computers, whose theoretical model is the universal TM. MIT, where both Putnam and Fodor were working in the early 1960s, was one of the central institutions for both cybernetics and AI. It is likely that this historical background made it natural for Putnam, Fodor, and later for other philosophers to assume that any machine capable of thinking, be it natural or artificial, had to be a computing machine (which, by the Church–Turing thesis, could be modeled by a TM).

Putnam appeared to have read some of the cybernetics and AI literatures. In those literatures, he was likely to come across the statement—which was made quite frequently—that anything that can be precisely described can be simulated by a computer program, or equivalently, by a TM (for example, von Neumann, 1951). When he formulated functionalism, Putnam stated that a brain or mind with probabilistic state transitions and time-delays could be arbitrarily well simulated by a

[6] On the history of classical artificial intelligence, see McCorduck (1979), Gardner (1985), Crevier (1993).

TM (Putnam, 1967, p. 184). When he formulated his computational version of functionalism, Putnam made the even stronger pancomputationalist assumption that 'everything is a Probabilistic Automaton under some Description' (Putnam, 1999, p. 31; first published 1967. 'Probabilistic Automaton' was Putnam's term for probabilistic TM). Putnam added that, since everything was a TM, his clause (1)—that organisms capable of feeling pain were TMs—was 'redundant', 'empty', and was 'only introduced for expository reasons' (ibid., p. 31).

If everything was a TM, then minds were TMs too, and Putnam's clause (1) was justified. But this justification came at a price. First, the assumption that everything was a TM was itself in need of justification, and Putnam didn't give any. Upon scrutiny, pancomputationalism turns out to be true only in a trivial and philosophically uninteresting sense. For the only sense in which everything is unquestionably a TM is the sense in which everything's behavior can be simulated by some TM to some degree of approximation. It is far from true that everything can be 'arbitrarily well simulated' by a TM; on the contrary, the behavior of most complex physical systems diverges exponentially from any computational simulation.[7] But even if everything can be simulated by a TM to some degree, it doesn't follow that everything is functionally organized as a TM, or that everything performs computations in the sense in which TMs do. If pancomputationalism has to carry the nontrivial implication that everything performs computations, then it must be construed in a more stringent way. As I argue elsewhere, a nontrivial pancomputationalism must be construed as the claim that everything has the function of producing output strings of symbols in response to input strings of symbols in accordance with a (Turing-computable) rule that applies to all inputs and outputs and depends on the inputs for its application. But this version of pancomputationalism is patently false: most things do not have functions (for example, hurricanes), or if they do their functions do not involve the manipulation of strings in accordance with general rules (for example, stomachs), or if they do they do not involve the relevant kind of rule (for example, genuine random number generators). Furthermore, the question of whether all possible computing mechanisms compute Turing-computable functions is at present still open.[8]

Second, the assumption that everything was a TM trivialized Putnam's analogy between minds and TMs, which had originally motivated his introduction of TMs into discussions of the mind–body problem. An important reason for Putnam's analogy was that TMs were open to rational criticism. If everything was a TM, then everything should be open to rational criticism. But most philosophers, even if they agree that there is a sense in which TMs are open to rational criticism, would probably reject the conclusion that everything is open to rational criticism. By importing pancomputationalism into the philosophy of mind, Putnam created more problems than he solved. Nevertheless, variations on the pancomputationalist

[7] See, for example, Strogatz (1994).

[8] For a detailed defense of these statements, see Piccinini (2003a), Chs. 7 and 8. Cf. also Copeland (2000, 2002).

theme can still be found at work behind the views of many contemporary philosophers.[9]

Let us recapitulate Putnam's path to computational functionalism. In 1960, when Putnam first formulated the analogy between minds and TMs, he denied that the mind was a closed causal system (Putnam, 1960). Later, perhaps influenced by Fodor's (1965) analysis of psychological theories—which he seemed to endorse in 1964 (Putnam, 1975b)—he formulated the doctrine of functionalism, namely, the doctrine that the mind is a closed causal system of functionally individuated states (Putnam, 1967). Finally, he added the further *computationalist* thesis that functional descriptions of minds are TM tables (Putnam, 1999; first published 1967). The transition between functionalism and computational functionalism was made without argument, though it was implicitly supported by Putnam's pancomputationalism (which, in turn, was stated without argument). In arguing in favor of computational functionalism, and against both type-identity materialism and logical behaviorism, Putnam used arguments that made no appeal, either as a premise or as a conclusion, to the computationalist thesis. This is unsurprising, because those arguments were already formulated in Putnam's previous papers of 1960, 1963, and 1967, where the computationalist thesis was not present (Putnam, 1960, 1980, 1967).

7. Functional analysis and explanation by program execution

Fodor's (1965) description of phase one psychological theories as having the same form as TM tables paved the way for the later *identification* of psychological functional analyses and computer programs. In a paper published a few years later (Fodor, 1968b), Fodor repeated his view, already present in Fodor (1965), that psychological theories provided descriptions of psychological functions.

But this time, he added that psychological theories were canonically expressed as lists of instructions: 'the paradigmatic psychological theory is a list of instructions for producing behavior' (Fodor, 1968b, p. 630). Although this flies in the face of the history of psychology, which is full of illustrious theories that are not formulated as lists of instructions (for example, Freud's psychoanalysis, or Skinner's behaviorism), Fodor did not offer evidence for this statement.[10] Fodor said that each instruction in a psychological theory could be further analyzed in terms of a list of instructions, which could also be analyzed in the same way. This did not

[9] For example: 'a standard digital computer . . . can display any pattern of responses to the environment whatsoever' (Churchland & Churchland, 1990, p. 26); 'For any object there is some description of that object such that under that description the object is a digital computer' (Searle, 1992, p. 208); 'the laws of physics, at least as currently understood, are computable, and . . . human behavior is a consequence of physical laws. If so, then it follows that a computational system can simulate human behavior' (Chalmers, 1996a, p. 329). Similar views are expressed by Block & Fodor (1972), p. 250; Chalmers (1996b), p. 331; Scheutz (1999), p. 191.

[10] At the time of Fodor's writing, though, some psychologists did propose such a view of psychological theories (Miller, Galanter, & Pribram, 1960), and at least according to Gilbert Harman (personal correspondence), their work influenced Fodor.

lead to infinite regress because for any organism, there was a finite list of elementary instructions in terms of which all psychological theories for that organism must ultimately be analyzed. This type of explanation of behavior, based on lists of instructions, was explicitly modeled by Fodor on the relationship between computers, computer programs, and the elementary instructions in terms of which programs were ultimately formulated.[11]

Fodor did not distinguish between functional analysis and the analysis of capacities in terms of lists of instructions, nor did he discuss the relationship between the two. On the contrary, he discussed the two as if they were the same. However, in light of Section 5, I reserve the term 'functional analysis' for an analysis that partitions a system into components and ascribes functions to the components. The analysis of capacities in terms of lists of instructions may be called *task analysis*. There are good reasons to keep these notions distinct. A functional analysis postulates a set of component types and their functions, which in turn can be given a functional analysis, and—unlike a task analysis—is not committed to analyzing the behavior of the system into sequences of elementary operations of finitely many types. On the other hand, a task analysis explains a capacity of a system as the execution of a sequence of operations of finitely many types by that system, which need not be analyzed into components and their functions. When a capacity of a system is explained by appealing to the causal role of a series of instructions constituting a task analysis of that capacity, like in ordinary computers, they are given an explanation by program execution.

Fodor's goal in this paper was to defend intellectualist explanations in psychology against Ryle's (1949) criticisms. In Fodor's formulation, intellectualist explanations were explanations of psychological capacities that appealed to tacit knowledge of rules. In his view, psychological theories formulated as lists of instructions offered intellectualist explanations of psychological capacities. He argued that, contra Ryle, intellectualist explanations were not methodologically flawed. Fodor analyzed intellectualist explanations as lists of instructions that were executed in the sense in which computer programs were executed, and used this analysis to argue that intellectualist explanations involved no infinite regress. But in order for these explanations to be properly intellectualistic, that is, different from the kind of causal process that Ryle would accept as an explanation of behavior, Fodor needed to stress that intellectualist explanations were significantly different from explanations in terms of generic causal processes:

> [T]he intellectualist is required to say not just that there are causal interactions in which the organism is unconsciously involved, but also that there are unconscious processes of learning, storing, and applying rules which in some sense 'go on' within the organism and contribute to the etiology of its behavior. (Fodor, 1968b, p. 632)

[11] Notice that from this point on in the literature, computer programs tended to replace TM tables in formulations of functionalism, without a distinction being drawn between them. As I point out elsewhere, they are not the same kind of description, and they presuppose very different functional analyses (Piccinini, 2003a, Ch. 10).

Fodor's distinction between psychological explanations that appealed to mere causal interactions and those that appealed to genuine tacit knowledge, or rule following, or program execution, would be accepted in the literature as what makes computational theories of mind different from theories that are not computational.

Without this distinction, computational theories of mind cannot be distinguished from neurophysiological theories that have nothing intellectualistic about them, because they do not appeal to the possession of knowledge, rules, or other 'mentalistic' constructs. This distinction is reminiscent of Putnam's (1960) point that the analogy between men and TMs is such that both are open to rational criticism. This can be true only if the sense of rule following that is being used in the analogy is something more than an ordinary causal happening; it has to be a sense in which rule following is a process that admits of degrees of rationality. Similar distinctions would be used in the literature on computationalism by supporters and critics alike to identify the difference between genuinely computational theories of mind and theories that could not be distinguished from 'purely' neurophysiological or functional theories.[12]

Of course, the distinction between explanation by program execution and other forms of explanation does not entail by itself that minds are programs, nor did Fodor suggest that it does. The mind would be a program only if all psychological capacities, states, and processes could be correctly explained by appeal to program execution. For Fodor, whether this was the case was presumably an empirical question. On the other hand, Fodor's paper firmly inserted into the philosophical literature a thesis that begged the question of whether all psychological capacities could be explained by program execution: the thesis that psychological theories were canonically formulated as lists of instructions for producing behavior. This left no room for alternative explanations to be considered. In particular, the general type of functional analysis identified by Deutsch (1960), which explained mental capacities by postulating types of components and their functions, was transformed, through Fodor's (1965) reinterpretation, into explanation by program execution. After that, the mongrel of psychological functional analyses and computer programs remained in the literature on functionalism and psychological explanation, where there are many statements to the effect that psychological theories are functional analyses, and that psychological functional analyses (or sometimes all functional analyses) are computer programs or some other form of computational description.[13]

[12] For instance, see Fodor (1975), p. 74 n. 15; Dreyfus (1979), pp. 68, 101–102; Searle (1980), pp. 37–38; Searle (1992), p. 208.

[13] For example, similar conflations can be found in works by Dennett (1975, 1978), Cummins (1975, 1983), Marr (1982), and Churchland & Sejnowski (1992). Even Harman, who criticized Fodor and Putnam's construal of functional analysis, followed their lead in this respect. Harman called Fodor's explanations by functional analysis *narrow* because they attempted to explain an organism only in terms of its internal states, inputs, and outputs, without reference to how the system interacted with its environment. Harman argued that a complete explanation of an organism requires a *wide* functional story, that is, a story that took into account the relation between the organism and its environment. Nevertheless, even Harman identified the narrow functional story about an organism with the description of the organism by a program (Harman, 1988, esp. pp. 240–1).

In summary, Putnam and Fodor formulated their analyses of minds and psychological theories in a way that made it very natural to think that minds were programs or that mental states were states of a program. This was because they construed psychological theories as functional analyses, and psychological functional analyses as programs. However, it should be clear by now that none of the considerations offered by Putnam and Fodor in support of their analysis of minds and psychological theories constitute a reason to believe that the mind is a program or that the functional relations between mental states are computational, that is, a reason in favor of computationalism.

8. Later developments of functionalism

Functionalism has been extensively discussed in the philosophical literature. Much of the discussion has centered on functionalism in relation to folk psychology and reductionism or on attempts to refute functionalism *a priori*. These topics are irrelevant here. This section addresses elaborations of functionalism by philosophers who maintained Putnam's motivation: to give a metaphysics of states that were ascribed to individuals by scientific psychological theories. Later writers discussed difficulties faced by functionalism in dealing with certain aspects or paradoxes of mentality. While weakening or elaborating on Putnam's functionalism, these authors still maintained that their versions of functionalism were computational. But like Putnam, they gave no reason to believe that functional relations between mental inputs, outputs, and internal states are computational. The goal of this section is to show that their views do not affect the conclusion that functionalism *per se* provides no motivation to believe that minds are computational. As examples, the views of Block and Fodor (1972) and Lycan, 1981, 1987) are examined.

Block and Fodor (1972) argued for a version of functionalism that was inspired by, but slightly weaker than, Putnam's 1967 version (Putnam, 1999). As we saw in Section 6, in his 1967 paper (Putnam, 1999) Putnam identified mental states with states defined by certain TM tables. Block and Fodor discussed a number of problems that arose specifically from identifying mental states with states of TM tables.[14] Moreover, Block and Fodor seemed to be aware that computational

[14] Block and Fodor enumerated the following problems: (i) a difficulty in drawing the distinction between dispositional mental states (beliefs, desires, inclinations, and so on) and occurrent mental states (sensations, thoughts, feelings, and so on) (Block & Fodor, 1972, pp. 242–243), (ii) a difficulty in representing a set of mental states as occurring simultaneously (ibid., pp. 243–244), (iii) a difficulty in ascribing the same mental state to two organisms unless they have identical TM tables (or relevant portions thereof) even though it was natural to ascribe the same mental state to two organisms in the presence of slight differences, say, in their behavioral dispositions associated to that state (ibid., pp. 245–246), (iv) the fact that by definition there were only finitely many states of any TM table, but persons could be in infinitely many type-distinct psychological states, hence the two could not be put in one-to-one correspondence (ibid., pp. 246–247), and (v) a difficulty in representing structural relations among mental states (for example, that believing that *A* was a constituent of believing that *A & B*) (ibid., pp. 247–248). Block and Fodor also discussed the problem of qualia, arguing that qualia may not be states that can be functionally individuated at all. But, they noted, if qualia could not be individuated functionally, qualia were irrelevant to the proper formulation of functionalism, so they ignored this problem (ibid., pp. 244–245).

functionalism as formulated by Putnam was a version of the more general doctrine of functionalism, according to which mental states are individuated by their functional relations with inputs, outputs, and other mental states.[15] They ended their paper saying that identity conditions for psychological states would be given by psychological laws when psychologists discover them, and it was a mistake to try to restrict such identity conditions by reference to things like states of TM tables, just as it was a mistake to restrict them to states falling under behavioral or neurophysiological laws. Psychological states should only be individuated by psychological laws (Block & Fodor, 1972, pp. 248–249).

Nonetheless, in formulating their version of functionalism, far from reverting to a general formulation of functionalism that left it open for psychologists to devise their own functional descriptions, Block and Fodor restricted functional descriptions of mental states to what they call 'computational states of automata', where a computational state was 'any state of the machine which is characterized in terms of its inputs, outputs, and/or machine table states' (ibid., p. 247). It is remarkable that, despite their appeal to psychological laws and their rejection of philosophical strictures on the identity of mental states, Block and Fodor still maintained that mental states were computational states, and concluded their paper saying that '[i]t may be both true and important that organisms are probabilistic automata' (ibid., p. 249). They gave no reason for this conclusion, and just as Putnam gave no reason for why mental states should be states of TM tables, Block and Fodor gave no reason for their restriction of functional descriptions to computational descriptions.

Like Block and Fodor, Lycan, 1981, 1987) offered a version of functionalism inspired by Putnam's 1967 version (Putnam, 1999). But unlike Block and Fodor, Lycan was concerned with the relationship between different levels of organization in the mind–brain. He argued that without some restriction on what counts as a realization of a TM table, a system realized any TM table that was realized by the system's proper parts.[16] He found this putative consequence of Putnam's functionalism unacceptable. The reason was that, according to Lycan, it was conceivable that a proper part of an organism had mental states of its own. These premises, combined with the functionalist doctrine that having mental states was the same as realizing a certain TM table, entailed that an organism had the mental states of all of its proper parts.[17] Lycan rejected this conclusion on intuitive grounds, and used this rejection to motivate a restriction on the notion of realization (Lycan, 1987, pp. 28–30). Lycan's restriction, introduced in Chapter 4 of his book, was a teleological requirement having to do with what a functional state did *for* an organism.

[15] For example, at some point they referred to 'functionalism in the broad sense of that doctrine which holds that the type-identity conditions for psychological states refer only to their relations to inputs, output, and one another' (ibid., p. 245).

[16] Putnam's clause 3, cited in Section 6, was designed to block this inference, but Lycan rejected it as unjustified.

[17] Incidentally, Lycan's argument can be modified to apply to Block and Fodor's version of functionalism.

In formulating his version of functionalism, Lycan stopped mentioning TMs altogether. Building on the ideas of Attneave (1961); Fodor (1968b), and Dennett (1975), Lycan argued that the best way to understand minds was to break them down into subsystems, each of which contributed to the activities of the whole by performing some intelligent task, such as solving a problem or analyzing a perceptual representation. Each mental subsystem could be further analyzed into sub-subsystems, and each sub-subsystem could be further analyzed until one reached components that had no interesting 'mentalistic' properties. He called this picture *teleological* functionalism. Lycan defended his view at length, arguing that it was an appealing metaphysics of the mind and was not affected by his objection to Putnam's version of functionalism.

In the rest of the book, Lycan offered an account of intentionality and qualia based on his teleological version of functionalism. In formulating his teleological functionalism, he made no use of the notion of TM or any other computational notion. His discussion was couched in terms of systems and subsystems, and what they contributed to each other's activities.[18] Despite this, and despite the flaws that Lycan found in Putnam's computational formulations of functionalism, he stated that his version of functionalism was still a computational theory of mind.[19] He did not explain what it was about teleological functionalism that made it a *computational* theory, or what metaphysical reasons there were to think that mentation had to do with computation. Instead, he stated that '[f]or the foreseeable future, computation is our only model for intelligence' (Lycan, 1987, p. 149). But how intelligence can be modeled is an empirical question, with no immediate relation with functionalism as a metaphysical doctrine about the individuation of mental states.

9. Conclusion

The literature on functionalism examined in this paper offers no metaphysical reason for the computationalist component of computational functionalism. In retrospect this is not surprising, because functionalism and computationalism are logically independent. None of the traditional arguments for functionalism or for the view that psychological theories are functional analyses establishes computationalism. The only premise that goes to some length in supporting computationalism, namely the pancomputationalist premise that everything can be described as

[18] Although Lycan did not cite Deutsch (1960), his metaphysics of mind was reminiscent of Deutsch's view of phase one psychological theories. This is not surprising, because Deutsch inspired Fodor (1965), whose ideas carried over in Fodor (1968b), and Fodor (1968b) was among the main acknowledged sources of Lycan's view. I argued above that Fodor tailored Deutsch's view to suit Putnam's analogy between minds and TMs; Lycan can be seen as undoing Fodor's tailoring so as to recover a view very close to Deutsch's original formulation.

[19] For example, cf. the following passage: 'an articulate computational theory of the mind has also gained credence among professional psychologists and philosophers. I have been trying to support it here and elsewhere' (Lycan, 1987, p. 128).

performing computations, has two serious shortcomings: it is hardly true in its nontrivial variants and, even if it were true, deriving computationalism about the mind from it would trivialize computationalism about the mind into a thesis with no explanatory force.

The lack of arguments for the computational component of computational functionalism raises the question of why, as a matter of historical fact, Putnam formulated functionalism in its computational variety and why his followers received Putnam's formulation without requiring a justification. A relevant fact—seen in Section 3—is that Putnam reached his functionalist doctrine by developing an analogy between minds and TMs. Furthermore, some of the other theses present in the functionalist literature contributed to the impression that minds are computational. One of those theses is Putnam's problematic thesis—discussed in Section 6—that everything can be described as a TM; another is Fodor's conflation—mentioned in Section 5—of functional analysis and descriptions in terms of state transitions, which prepared the terrain for conflating psychological functional analyses and computational descriptions (cf. Section 7).

Perhaps the combination of these views contributed to create the philosophical illusion that computational functionalism could solve the mind–body problem, explain the mind in terms of computation, and be based on some simple conceptual point such as multiple realizability or the thesis that everything can be described as a TM. Once these different ideas are pulled apart, we see that there is nothing left to the idea that functionalism as a solution to the mind–body problem entails computationalism as a substantive thesis about mind–brains. If my reconstruction is correct, computational functionalism can't be all those things at once. Either it incorporates an empirical hypothesis about the character of the functional relations between mental inputs, outputs, and internal states, which is independent of functionalism as a solution to the mind–body problem and must be supported on independent grounds, or it is a trivial thesis that has no role to play in the philosophy of mind.

The conclusion that computationalism is an empirical hypothesis about the functional organization of the mind is not new. In the 1970s, Jerry Fodor himself was among the first philosophers to defend computationalism on empirical grounds (Fodor, 1975). Starting from the assumption that mental processes respect the semantic properties of mental states, Fodor argued that our only mechanistic explanation for such mental processes is a computing mechanism that manipulates (in an appropriate way) symbols whose syntax mirrors the semantic properties of mental states.[20] By then, Fodor and some others explicitly distinguished between functionalism and computationalism. Nevertheless, the logical independence of these two doctrines has not been fully recognized. Perhaps this is because the two assumptions that helped early Putnam and Fodor to conflate functionalism and computationalism—namely, that everything can be described as performing com-

[20] The conceptual relationships between functionalism, computationalism, and mental contents are more complex than they may appear. I discuss this issue in detail in Piccinini (2004).

putations and that functional analyses are computational descriptions—have been retained by many mainstream philosophers (of both classicist and connectionist convictions) who are interested in mechanistic theories of the mind.[21] Neither of these assumptions is justified, so they should be replaced by a deeper understanding of, on the one hand, the ways in which things may or may not be said to perform computations, and, on the other hand, the nature of functional analysis.[22]

Many philosophers of mind still find it difficult to envision mechanistic theories of mind that are not computational, even though there are classes of possible theories that remain largely untapped. There is no room here to do justice to this topic, so I will only mention one possibility recently discussed by Jack Copeland (2000). Copeland suggests that instead of an ordinary computing mechanism, the mind might be a hypercomputer—a mechanism that computes functions that are not computable by Turing Machines. I doubt that the mind is a hypercomputer, but regardless of whether it is, the point remains that philosophers of mind should remain open to the possibility of mechanistic theories of mind according to which the functional organization of the mind is not computational.

Acknowledgements

A version of this paper was presented at the 2003 APA Pacific Division in San Francisco. I am grateful to the audience and commentator, Ray Elugardo, for their feedback. I also thank Eric Brown, Carl Craver, Robert Cummins, Paul Griffiths, Peter Machamer, Diego Marconi, Oron Shagrir, two anonymous referees, and anyone else who gave me comments on previous versions of this paper.

References

Armstrong, D. M. (1970). The nature of mind. In C. V. Borst (Ed.), *The mind/brain identity thesis* (pp. 67–79). London: Macmillan.

Attneave, F. (1961). In defense of homunculi. In W. Rosenblith (Ed.), *Sensory communication* (pp. 777–782). Cambridge, MA: MIT Press.

Block, N., & Fodor, J. A. (1972). What psychological states are not. *Philosophical Review*, *81*(2), 159–181.

Chalmers, D. J. (1996). *The conscious mind: In search of a fundamental theory*. Oxford: Oxford University Press.

Chalmers, D. J. (1996). Does a rock implement every finite-state automaton? *Synthese*, *108*, 310–333.

Churchland, P. M., & Churchland, P. S. (1990). Could a machine think? *Scientific American*, *CCLXII*, 26–31.

Churchland, P. S., & Sejnowski, T. J. (1992). *The computational brain*. Cambridge, MA: MIT Press.

[21] Of course, there are also several philosophers who oppose computationalism. The current argument pertains to their view only in so far as they subscribe to functionalism. But typically, functionalist philosophers who oppose computationalism have no mechanistic theory of mind on offer in alternative to computationalism. Furthermore, some of them retain one or both of the above assumptions.

[22] On the former, see Piccinini (2003a), Ch. 8. On the latter, see Craver (2001).

Copeland, B. J. (2000). Narrow versus wide mechanism: Including a re-examination of Turing's views on the mind-machine issue. *Journal of Philosophy*, *XCVI*(1), 5–32.

Copeland, B. J. (2002). Hypercomputation. *Minds and Machines*, *12*, 461–502.

Craver, C. (2001). Role functions, mechanisms, and hierarchy. *Philosophy of Science*, *68*, 53–74.

Crevier, D. (1993). *AI: The tumultuous history of the search for Artificial Intelligence*. New York: HarperCollins.

Cummins, R. (1975). Functional analysis. *Journal of Philosophy*, *72*(20), 741–765.

Cummins, R. (1983). *The nature of psychological explanation*. Cambridge, MA: MIT Press.

Dennett, D. C. (1975). Why the law of effect will not go away. *Journal of the Theory of Social Behavior*, *5*, 169–187. (Reprinted in Dennett, 1978,71–89)

Dennett, D. C. (1978). *Brainstorms*. Cambridge, MA: MIT Press.

Deutsch, J. A. (1960). *The structural basis of behavior*. Chicago: University of Chicago Press.

Dreyfus, H. L. (1979). *What computers can't do*. New York: Harper & Row.

Fodor, J. A. (1965). Explanations in psychology. In M. Black (Ed.), *Philosophy in America*. London: Routledge and Kegan Paul.

Fodor, J. A. (1968). *Psychological explanation*. New York: Random House.

Fodor, J. A. (1968). The appeal to tacit knowledge in psychological explanation. *Journal of Philosophy*, *65*, 627–640.

Fodor, J. A. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.

Fodor, J. A. (2000). *The mind doesn't work that way*. Cambridge, MA: MIT Press.

Gardner, H. (1985). *The mind's new science: A history of the cognitive revolution*. New York: Basic Books.

Harman, G. (1988). Wide functionalism. In S. Schiffer, & S. Steele (Eds.), *Cognition and representation* (pp. 11–20). Boulder: Westview.

Lewis, D. K. (1966). An argument for the identity theory. *Journal of Philosophy*, *63*, 17–25.

Lewis, D. K. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, *50*, 249–258.

Lewis, D. K. (1980). Mad pain and Martian pain. In N. Block (Ed.), *Readings in philosophy of psychology, Vol. 1* (pp. 216–222). Cambridge, MA: MIT Press.

Lycan, W. (1981). Form, function, and feel. *Journal of Philosophy*, *78*, 24–50.

Lycan, W. (1987). *Consciousness*. Cambridge, MA: MIT Press.

Marr, D. (1982). *Vision*. Cambridge, MA: MIT Press.

McCorduck, P. (1979). *Machines who think: A personal inquiry into the history and prospects of Artificial Intelligence*. S. Francisco, CA: Freeman.

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.

Oppenheim, P., & Putnam, H. (1958). Unity of science as a working hypothesis. In H. Feigl, M. Scriven, & G. Maxwell (Eds.), *Concepts, theories, and the mind–body problem* (pp. 3–36). Minnesota studies in the philosophy of science, Vol. II. Minneapolis: University of Minnesota Press.

Piccinini, G. (2003a). *Computations and computers in the sciences of mind and brain*. Ph.D. thesis, University of Pittsburgh. (Available at http://etd.library.pitt.edu/ETD/available/etd-08132003-155121/)

Piccinini, G. (2003b). Alan Turing and the mathematical objection. *Minds and Machines*, *12*(1), 23–48.

Piccinini, G. (forthcoming). 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*.

Piccinini, G. (2004). Functionalism, computationalism, and mental contents. *Canadian Journal of Philosophy*, *34*(3), 375–410.

Putnam, H. (1960). Minds and machines. In S. Hook (Ed.), *Dimensions of mind: A symposium* (pp. 138–164). New York: Collier.

Putnam, H. (1967). The mental life of some machines. In H. Castañeda (Ed.), *Intentionality, minds, and perception* (pp. 177–200). Detroit: Wayne State University Press.

Putnam, H. (1975a). Some issues in the theory of grammar. In idem, *Philosophical papers, Vol. 2. Mind, language and reality* (pp. 85–106). Cambridge: Cambridge University Press. (First published in *Proceedings of Symposia in Applied Mathematics, 12* (1961), 25–42)

Putnam, H. (1975b). Robots: Machines or artificially created life? In idem, *Philosophical papers, Vol. 2. Mind, language and reality* (pp. 386–407). (First published in *Journal of Philosophy*, LXI (1964), 668–691)

Putnam, H. (1980). Brains and behavior. In N. Block (Ed.), *Readings in philosophy of psychology, Vol. 1* (pp. 24–36). London: Methuen. (First published in R. J. Butler (Ed.), *Analytical philosophy* (pp. 1–20). New York: Barnes and Noble, 1963)

Putnam, H. (1999). The nature of mental states. In W. Lycan (Ed.), *Mind and cognition: An anthology* (2nd ed.) (pp. 27–34). Malden: Blackwell. (First published as Psychological predicates. In H. Putnam, *Art, philosophy, and religion* (pp. 37–48). Pittsburgh: University of Pittsburgh Press, 1967)

Ryle, G. (1949). *The concept of mind*. London: Hutchinson.

Scheutz, M. (1999). When physical systems realize functions .... *Minds and Machines*, *9*, 161–196.

Searle, J. R. (1980). Minds, brains, and programs. *The Behavioral and Brain Sciences*, *3*, 417–457.

Searle, J. R. (1992). *The rediscovery of the mind*. Cambridge, MA: MIT Press.

Shannon, C. E., & McCarthy, J. (1956). *Automata studies*. Princeton: Princeton University Press.

Shannon, C. E. (1952). Presentation of a maze solving machine. In H. von Foerster, M. Mead, & H. L. Teuber (Eds.), *Cybernetics: Circular causal and feedback mechanisms in biological and social systems. Transactions of the Eighth Conference* (pp. 169–181). New York: Macy Foundation.

Shoemaker, S. (1984). *Identity, cause and mind*. Cambridge: Cambridge University Press.

Strogatz, S. H. (1994). *Nonlinear dynamics and chaos*. Cambridge, MA: Perseus.

Sutherland, N. S. (1960). Theories of shape discrimination in octopus. *Nature*, *186*, 840–844.

Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. In M. Davis (Ed.), *The undecidable* (pp. 116–154). Ewlett: Raven. (First published 1936–1937)

Von Neumann, J. (1951). The general and logical theory of automata. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 1–41). New York: Wiley.

Von Neumann, J. (1956). Probabilistic logics and the synthesis of reliable organisms from unreliable components. In C. E. Shannon, & J. McCarthy (Eds.), *Automata Studies* (pp. 43–98). Princeton: Princeton University Press.