TWO ROADBLOCKS OF COMPUTATIONALISM

Napoleon M. Mabaquiao, Jr.
De La Salle University, Manila

With its use of the powerful technology of computer, the computational theory of mind or computationalism, which regards minds as computational systems, has been widely hailed as the most promising theory that will carry out the project of explaining the workings of the mind in purely scientific terms. While it continues to serve as the primary framework for scientifically inclined theorizing and investigations about the nature of minds, especially in the area of cognitive science, it, however, continues to face strong objections from its critics. And with the growing complexity and sophistication of the arguments used to promote and reject the theory, the debate has become intractable. It has become quite difficult to assess which side of the dispute is gaining the upper hand. Such difficulty may be due to a variety of reasons. In this essay, I critically examine two of such reasons. The first concerns the ambiguity of the theory's intended scope of application: whether it is limited to the mind's cognitive features only or it also includes the mind's phenomenal features. The second concerns the vagueness of how the so-called computer modelling of human cognitive processes is able to duplicate such processes. Accordingly, if insufficiently addressed, they remain as two roadblocks to the entire project of computationalism.

INTRODUCTION

With the continuous development of computer technology, computationalism remains as the dominant and most promising framework for *naturalizing the mind*, referring to the project of explaining the workings of the mind in purely scientific terms (see Jerry Fodor 1991, 489; David Chalmers 1993, 18). While most scientifically inclined investigations done on the mind, especially those done in the areas of cognitive science and artificial intelligence, continue to have it as their overarching framework, computationalism, nonetheless, remains controversial as a philosophical theory of mind. It continues to face strong objections from its critics, which its supporters relentlessly respond to (see, for instance, Milkowski 2017). Now with the growing complexity and sophistication of the arguments of both proponents and critics of the theory, it has

become quite difficult to assess which side of the dispute is gaining the upper hand. Inthis essay, among the possible reasons for this difficulty, I critically examine the following two. The first concerns the intended scope of application of the theory, whether it includes all features of the mind or only some of them. As a philosophical theory of mind, it seems natural to suppose that it is an account of the mind in general, or, at least, of the essential features of the mind. But this is far from clear. There is, in fact, an ambiguity whether what is claimed to be computational about the mind is limited only to the mind's cognitive features or likewise includes the mind's phenomenal features. The second concerns the disagreement on what the computer modelling of human cognitive processes is able to accomplish in so far as establishing the thesis of computationalism is concerned. In particular, the computationalist claim that the computer simulation of human thought processes duplicates such processes remains controversial.

On the whole, I will try to show in this essay that as long as these two reasons remain insufficiently addressed, the computationalist thesis that minds are computers will be controversial as ever. They, in effect, will remain as two roadblocks to the entire project of computationalism. I shall divide my discussion into three parts. To put things in perspective, I shall provide a brief background of the computational theory of mind in the first part. In the second part, I shall examine how much of the mind's capacities are being claimed to be computational by computationalism. In the third, I shall deal with the alleged role of the computer modelling of human cognitive process in advancing the theses of computationalism.

COMPUTATIONALISM AND THE PHILOSOPHY OF MIND

The computational theory of mind, or computationalism for short, is a philosophical theory of mind that claims that minds are computers or computational systems (Rescoria 2017, 1; Milowski 2018b, 1). To better understand this theory, let us situate it in the general discussion on the nature of the mind's existence. Views on the nature of mind's existence are generally divided into two kinds: the *non-materialist views*, which attribute a non-physical existence to minds; and *materialist views,* which attribute a physical existence to minds. Non-materialist views include *idealist* and *(substance) dualist* views about the mind. On the other hand, materialist views are generally divided into *non-realist (or irrealist) materialist views*, which reject the idea that mental states have separate existence or distinct reality from the physical states of either the brain or the body; and *realist materialist views,* which maintains said idea.

Under non-realist materialism are the following views: the *(mind-brain) identity theory*, which reduces mental states to brain states; *behaviorism*, which, in its strong version, reduces mental states to behavioral dispositions; *eliminative materialism,* which claims that the theory that posits the existence of minds is outdated and mis-taken; and *instrumentalism*, which claims that the attribution of mental states to an entity is purely metaphorical being a mere convenient device for predicting the entity's behavior. Under realist materialism, on the other hand, are the following views: *functionalism,* which regards mental states as states occurring on the level of a system's

functional organization and as definable in terms of their causal roles; *computationalism*, which regards mental states as computational states or as states of a computational system; *biological naturalism*, which regards mental states as higher-level biological states; and the *quantum view of consciousness*, which regards mental states as quantum states in the brain. (See Mabaquiao 2012, chap. 2.)

Computationalism, in its philosophical form, is a development from functionalism. For this reason, let us look into some of the details of functionalism. Functionalism regards mental states as functional states. But what exactly are functional states; and how are they different from the physical states of the brain or the body? Functional states have three key features. First, they refer to the physical states of a physical system on the level of the system's functional organization (see Putnam 1991), which in turn refers to the specific way that the materials making up the system are arranged to perform certain functions. An example is the way the parts of a clock are arranged to perform the function of telling time. With regard to the human brain, its functional organization refers to the way the brain's neurons are arranged in order for the brain to perform its various functions. The brain's functional states thus do not refer to the physical states of the brain's neurons but to the physical states of its functional organization.

The second key feature of functional states is that they are definable in terms of their causal roles as input, output, and intervening internal states of a physical system (see Block 1991, 211-212). On this point, functionalism converges with the *causal theory of mind* developed by David Lewis (1991a, 1991b) and David Armstrong (1968, 1991) in the course of justifying the (mind-body) identity theory. Causally defined, a functional state is caused by an input state B, is causally interacting with an intervening internal state C, and causes an output state D.

The third key feature of functional states is the principle of *multiple realizability*, according to which the same functional states can occur in physical systems consisting of different materials but having the same functional organization. This makes functional states *substrate neutral*, in that they are not affected by the nature of the materials that make up the system. And being functional states, mental states are, thus, substrate neutral in that while they occur in brains, they can also occur in other physical systems having the same causal or functional organization as the human brain.

Functionalism has developed into several forms. For our purposes, let us distinguish between the following two. The first version is *general functionalism*, which is non-committal as to the kind of physical system deemed appropriate for mental states. It does not specify the particular type of system that mental states can occur. The second is *computer functionalism,* advanced by Hilary Putnam (1991), which specifies that the particular type of physical system appropriate for mental states to occur is one that instantiates the *Turing machine*. The Turing machine was conceived by Alan Turing (1936) originally to resolve foundational issues in mathematics but which later on also served as the abstract model of the modern-day digital computer. In this consideration, what machine functionalism in effect claims is that mental states can occur only incomputing machines, which then makes the human brain as a kind of computer. Needless

to say, it was this version of functionalism that paved the way for the development of computationalism (Rescoria 2017). In this consideration, computationalism is sometimes referred to as "computer functionalism" or "computational functionalism," implying that computationalism is a version of functionalism.

Computationalism contends that the functional states with which functionalism identifies mental states are of the computational kind. Being so, computational states also have the key features of functional states but with some modifications. First, while functional states are the physical states of a physical system on the level of the system's functional organization, computational states, in its classical form, are the physical states of a computing system on the level of software. The mind is treated as the software of the computational system, while the brain as its implementing hardware. Mental states, on this view, specifically refer to the states of the mind software when being run or implemented by the brain hardware. Second, as functional states can be defined in terms of the causal relations of inputs, intervening internal states, and outputs, so are computational states. And third, as functional states are multiply realizable, so are computational states. But while in functionalism this means that two physical systems consisting of different materials but having the same functional organization will have the same functional states; in computationalism, this would mean that two computers or computing systems different in their hardware materials (though the same in sophistication in terms of software-implementing capacities) but the same in their software will have the same computational states.

In addition to the machine functionalism of Putnam, computationalism further developed when Jerry Fodor (1979), in collaboration with Zenon Pylyshyn (1990), linked it with his *representational theory of mind* (also known as the *language of thought* hypothesis). According to the representational theory of the mind, the brain has an inherent language or system of representation that has a language-like structure called the "language of thought," which serves as the vehicle or medium of the mind in performing its computations or thinking operations. This language of thought consists of mental symbols or internal representations which the mind manipulates according to rules inherent in the brain.

While all this was happening in the discipline of philosophy, a parallel development was occurring in the discipline of artificial discipline, a branch of computer science devoted to the construction of intelligent machines. In the discipline of artificial intelligence, a distinction was made between the general and neutral claim that computers are powerful tools for understanding the workings of the human mind and the specific and bold claim that human minds are themselves computer programs. The former has been called "Weak AI," while the latter "Strong AI" (such terms are due to John Searle 1980). A major influence in the development of Strong AI was the *physical symbol system hypothesis* introduced by Allen Newell and Herbert Simon, according to which "a physical symbol system has the necessary and sufficient means for general intelligent action" (Newell and Simon 1976, 116). On this view, intelligent action is possible only for a physical system that computes in terms of symbols. Human minds and computers are such physical symbol systems, which explain their capacity for intelligent actions.

Computationalism, in this regard, has then been closely associated, if not equated, with Strong AI.

Computationalism has evolved in different forms. The kind of computationalism that we have described above, the initial version of computationalism, has been called *symbolic* or *classical computationalism*. It is described as symbolic since it defines the computing that occurs in thinking as a process of manipulating symbols or representations according to some rules. It is described as classical, on the other hand, in recognition of the fact that it was in this form that computationalism became a popular and dominant theory of mind. The general idea of computationalism, that thinking is a kind of computing, may have its roots to the ideas of philosophers before classical computationalism (such as Leibniz), but it was during the introduction of classical computationalism that computationalism became an influential theory of the mind.

While having its roots also in pre-classical computationalism, another influential form of computationalism developed partly as a reaction to classical computationalism, namely the view called the *artificial neural network approach,* also known as *connectionist computationalism* or simply as *connectionism.* This approach uses a computer model that is based on the neural networks of human brains to explain the computing processes of the mind. It explains such processes as interactions among the units of networks. There are also variations within this approach, and one popular version is the *parallel distributed processing network* developed by James McClelland and David Rumelhart, among others. Furthermore, if for the symbolic approach computing is serial (that is, a step-by-step process and hence one operation occurring at a time), for connectionism it is parallel (that is, simultaneous operations occurring at the same time) (see Crane 1995, 154-162). There is an on-going debate among the proponents of each approach on which of these two is the correct or superior approach to computationalism (see, for instance, Fodor and Pylyshyn 1990, Smolensky 1993, Rumelhart 1990, and McClelland and Rumelhart 1993). Some prominent AI scientists, such as Marvin Minsky (1995, 649), however, has called for a synthesis of these two approaches. In either form, however, the basic thesis of computationalism remains, that thinking is a species of computing and that mental states are computational states.

Another theory of mind that has recently emerged that is associated with computationalism is the so-called *computational neuroscience*, which claims that it is the brain, not the mind, that is a computer (Rescoria 2017, 7). Computational neuroscience is somewhat of a merger between the identity theory and computationalism. As the mind is nothing but the brain, then it is the brain that is more appropriately a computer. Some, however, believe that computational neuroscience more of a method than a theory—as Wein and Pona (2015, 2) states: "Computational neuroscience is a branch of neuroscience that uses computer simulations to study the brain; it is not a theory of mind, but a method of studying the brain." In any case, acknowledged key figures in the development of computer neuroscience includes Patricia Churchland (1986), Paul Churchland (1995, 2007), and Chris Eliasmith (2013). This theory specifically claims that the human brain processes themselves, not the physical states of the functional organization of the brain nor the software allegedly being implemented by the brain hardware, are the

computational states. Unlike the classical and connectionist computational theories, computational neuroscience, by grounding computationalism in biology, abandons the thesis of multiple realizability (Wein and Pona 2015, 2). This means that their findings cannot be made as bases for constructing artificial or machine intelligence or a silicon-based creature. More importantly, it is susceptible to the same objections levelled against the identity theory, especially the charge of "neuro-chauvinism"—that mentality is exclusive to creatures whose brains consist of neurons.

Computationalism has been adopted as the main framework of *cognitive science,* referring to the projected ultimate science of the mind. While multidisciplinary in approach for utilizing the methods and findings of several disciplines that include philosophy, computer science, psychology, anthropology, biology, neuroscience, and linguistics, cognitive science uses computationalism as its overarching framework for its scientific investigations about the mind. As Milkowski (2018b, 1) writes: "It is generally assumed that CTM is the main working hypothesis of cognitive science." And as Freidenberg and Silverman (2006, 2) explain: "In order to really understand what cognitive science is all about we need to know what its theoretical perspective on the mind is. This perspective centers on the idea of computation, which may alternatively be called information processing" (see also Howard Gardner 1985, 6-7, 384-85; and Robert Harnish 2002, 2-3).

THE SCOPE OF THE COMPUTATIONAL MIND

How much of the mind's capacities are claimed to be computational by computationalism? To facilitate our analysis, let us distinguish among three kinds of mind: the *general, cognitive,* and *phenomenal* minds. The general mind is the mind construed in terms of both its cognitive and phenomenal features. The cognitive mind is the mind construed in terms of its cognitive features only. And the phenomenal mind is the mind construed in terms of its phenomenal features only. When computationalists claim that the mind is a computer, which among these minds are they referring to? Some believe that it should refer to the general mind. As Wein and Pona (2015, 2), referring to computationalism as CTM, states: "CTM is supposed to be a universal theory of mind: it should be possible to explain every aspect of the mind by providing a suitable computational description …." But some believe that it only refers to the cognitive mind, thereby excluding the phenomenal mind. As Milkowski (2018b, 2) notes: "The generic claim that the mind is a computer may be understood in various ways, depending on how the basic terms are understood. In particular, some theorists claimed that only cognition is computation, while emotional process are not computational (Harnish 2002, 6), yet some theorists explain neither motor nor sensory processes in computational terms (Newell and Simon 1972)."

The classic objections to functionalism contend that functionalism fails as a theory of mind because it leaves out certain fundamental features of the mind. Such arguments, for instance, point out that functionalism leaves out the mind's phenomenal features. These arguments include the *China brain argument* (see Block 1991, 211-229)*, knowledge argument* (see Jackson 1991, 291-295)*,* and *inverted qualia argument* (see Schoemaker

1982, 357-381). As computationalism is a variant of functionalism, these objections equally apply to it. Searle's *Chinese room argument* (see Searle 1980, 417-457)*,* on the other hand, points to the failure of computationalism to account for an important feature of the cognitive mind, namely its inherent intentionality. The previous arguments all point out that computationalism fails because it leaves out the phenomenal mind. Searle, however, in effect shows that even when limited to the cognitive mind, computationalism still fails because it leaves out the intentional mind. The presupposition of these objections is that computationalism, to be successful, should account for the general mind.

Defenders of computationalism usually respond to these arguments in a variety of ways. One is by exposing the flaws in the reasoning of these arguments. An example of this kind of reaction is Chalmers's *fading* and *dancing qualia arguments*, which are intended to show the implausibility of the absent and inverted qualia arguments. Another is by rejecting the assumption that certain features of the mind, such as qualia and intentionality, are really essential to mentality. A classic example of this kind of approach is Dennett's instrumentalist theory of mind, according to which the attribution of intentionality, or consciousness in general, is merely metaphorical for it is done solely for its usefulness in predicting future behaviors of organisms (see Dennett 1989, 1991). Still another is by clarifying that computationalism is not really intended to account for the general mind (which includes the intentional mind) or the phenomenal mind, for it is only intended to account for the cognitive mind. In this respect, some computationalists contend that these "objections fail because they make computationalism a straw man" (Milkowski 2018a, 58).

But even if we grant the construal of computationalism as a theory of the cognitive mind, thereby restricting mentality to intelligence, there is still the problem of how much of the cognitive mind computationalism is intended to account for. We earlier saw that this was the main point of Searle's Chinese room argument, showing that the intentional mind is left out by computationalism. Let us further elaborate on this. When computationalists speak of the mind, they usually mean intelligence or the cognitive aspect of the mind. What they claim to be computational are only the mind's cognitive states (the propositional attitudes). This means that emotions or the affective aspect of mentality, at least originally, are not part of the computationalist project. Now, in the case of humans, we normally understand intelligence both in terms of functionality and consciousness. The functionality of intelligence generally refers to abilities or capacities to perform certain tasks such as solving problems, performing operations, answering questions, and following rules. We say, for instance, that someone is intelligent if she can solve certain mathematical problems or perform certain mathematical operations. Its consciousness, on the other hand, generally refers to awareness of having certain mental states or processes such as understanding and reasoning. We normally think that both aspects should be present for someone to be truly intelligent. If someone claims to understand something but cannot perform the associated task, we doubt whether she is really intelligent. Or she actually performs the task but solely because she has been conditioned to do so, we also doubt whether she is really intelligent.

The kind of intelligence, however, that computationalists speak about is a general one, one that does not just apply to humans but to machines as well. It seems obvious that machines, *in principle,* can share the functional aspect of human intelligence; but it is not clear whether they can also share its conscious aspect. The consideration of the possibility of machine intelligence, thus, puts the fundamentality of consciousness in the definition of intelligence into question. Three views can be distinguished concerning the relative importance of consciousness and functionality to the nature of intelligence or mentality in general. The first is the default view, which we shall refer to as the *ordinary view,* which regards functionality and consciousness as equally fundamental in defining intelligence. The second, which we shall call the *purely functional view*, regards the mind's functionality as adequate in defining mentality. The third, which we shall call the *purely conscious view*, regards the mind's consciousness as adequate in defining intelligence. In the contemporary scene, the debate is mainly between the purely functional view and the ordinary view. The purely conscious view, which can be attributed to the idealists and substance dualists, is generally no longer regarded as a strong contender. For as contemporary philosophy of mind is geared towards the naturalization of the mind, the fundamentality of functionality is taken as a given and the only question is whether or not consciousness is equally fundamental.

Consequently, we can adopt either a generalized construal or a specialized construal of the claims of computationalism. We adopt a generalized construal of these claims when we attribute a normal view to these claims; that is to say, we understand these claims as applying to both functionality and consciousness of minds. On the other hand, we adopt a specialized construal of these claims when we attribute a purely functional view to these claims; that is to say, we understand these claims as only applying to the functionality of minds. The inadequacy arguments obviously work only under a generalized construal of the claims of computationalism; they are misplaced under a specialized construal of the computationalist claims.

Understandably, most proponents of computationalism take a specialized construal. In a footnote in his book *The Mind Doesn't Work That Way*, Jerry Fodor (2000, 1), for instance, writes: "This is not to claim that CTM is any of the truth about consciousness, not even when the cognition is conscious. There are diehard fans of CTM who think it is; but I'm not of their ranks" (CTM refers to the Computational Theory of Mind). Herbert Simon and Craig Kaplan (1990, 1-2) also write: "Intelligence is to be judged by the ability to perform intellectual tasks, independently of the nature of the physical system that exhibits this ability." Still, Roger Schank and Peter Childers (1984, 51) write: "When we ask *What is intelligence?* we are really only asking *What does an entity, human or machine, have to do or say for us to call it intelligent.*"

On the other hand, most of the critics of computationalism, such as John Searle and Roger Penrose, prefer a generalized one. Penrose (1989, 525-26), for instance, writes: "There is also the question of what one means by the term 'intelligence'. This, after all, is what the AI people are concerned with, rather than the perhaps more nebulous issue of 'consciousness'.... In my own way of looking at things, the question of intelligence is a subsidiary one to that of consciousness. I do not think that I would believe that true

intelligence could be actually present unless accompanied by consciousness." In his Chinese room argument, Searle (1980) objects to the claim that machines that are able to simulate the intelligent behavior of humans are genuinely intelligent themselves. Searle explains that in the case of humans, there is awareness of what their mental states mean or represent in the world; while in the case of machines they just manipulate symbols according to the rules of the given program without any awareness of what these symbols mean. This means, for Searle, consciousness (in the form of intentionality) is as fundamental to intelligence as the mind's functionality.

In some cases, however, supporters of computationalism themselves are not clear which construal they adopt as they seem unsure about the relevance of consciousness to the computationalist project. In her informal survey, Drew McDermott (2007) shows that most serious AI researchers are ambivalent on the importance of consciousness in the computationalist project. She (2007, 119) writes: "Although one might expect AI researchers to adopt a computationalist position on most issues, they tend to shy away from questions about consciousness." Interestingly, she herself, as a supporter of computationalism, seems ambivalent on the subject matter. She counts herself as one of those who regard consciousness as unimportant to the computational project; thus she (2007, 119) writes: "When it comes to the problem of phenomenal consciousness, however, the AI researchers who care about the problem and believe that AI can solve it are a tiny minority … I count myself in that minority …." But then she (2007, 119) believes that despite the various objections hurled against it, "the basic computationalist working hypothesis survived intact: that the embodied brain is an 'embedded' computer, and that a reasonably accurate simulation of it would have whatever mental properties it has, including phenomenal consciousness." So it seems that consciousness is important after all; otherwise, why emphasize the point that consciousness would be part of what would be duplicated in a computer simulation of the brain.

THE ROLE OF THE COMPUTER MODELLING OF THE MIND

Milkowski (2013, vii) writes: "The mind can be explained computationally because it *is* computational…. My central claim reflects my adherence to realism: a computational account of the mind can constitute a genuine explanation *only insofar as the mind is itself computational.*" If the mind can be explained computationally and the only way that it can be done, aside from the sufficient sophistication of the needed technology to carry it out, is if the mind itself is computational, then a computational explanation of the mind would constitute a powerful argument for the computationality of the mind. There are, however, questions here. One, how does one establish that the only way one can computationally explain the mind is that if the mind itself is computational? Isn't it possible that the mind is non-computational despite the fact that it lends itself to a computational explanation? Isn't it possible that the computational explanation of the mind is just a convenient and practical way (a convenient stance, , following the language of Dennett) of accounting for the various activities and manifestations of the mind?

Another way of stating the view that the only way a computer simulation of the mind can be done is if the mind itself is computational is that the computer simulation of the mind *duplicates*, not just merely simulates or models, the mind itself. As Herbert Simon (1995, 676) straightforwardly puts it: "a computer simulation of thinking thinks." Simon (1995, 676), in what follows, explains how this is so by distinguishing between a computer simulation of digestion which does not duplicate digestion and a computer simulation of the mind which duplicates the mind: "The materials of digestion are chemical substances, which are not replicated in a computer simulation. The materials of thought are symbols—patterns, which can be replicated in a great variety of materials (including neurons and chips), thereby enabling physical symbol systems fashioned of these materials to think." In gist, the reasoning of Simon is that a computer simulation of human thought duplicates human thought because the materials of human thought are symbols which can be duplicated by a computer. Put in another way, the computer's symbol-manipulating process can duplicate human thought because human thought is itself a symbol-manipulating process. With some computationalists who are not inclined to understand thinking in terms of symbol manipulation but in terms of information processing, the argument of Simon can put as follows. Computers process information, and so do minds. When computers simulate the information-processing activities of minds, these simulations are themselves information-processing activities.

What enables, it shall be observed, the computer simulation of thinking to duplicate thinking is that the thinking process is multiply realizable, in that it is substrate neutral or that is independent of the physical medium in which it occurs. It is unlike the process of digestion which is biologically dependent in that it cannot occur without the necessary biological elements. Referring to the principle of multiple realizability as the principle of *organizational invariance*, Chalmers (2010, 37-38), in what follows, seems to echo Simon in explaining that what a computer simulation duplicates about human thought includes consciousness:

> In general, if a property is not an organizational invariant, we should not expect it be preserved in a computer simulation (a simulated rainstorm is not wet). But if a property is an organizational invariant, we should expect it be preserved in a computer simulation (a simulated computer is a computer). So given that consciousness is an organizational invariant, we should expect a good enough computer simulation of a conscious system to be conscious, and to have the same sorts of conscious states as the original system.

Accordingly, if one denies the principle of multiple realizability, arguing that thinking is not substrate neutral, then the computer simulation of minds can merely simulate minds, never duplicate them. Wein and Pona (2015, 2) state this point clearly as follows:

> It is uncontroversial that computation as understood in computer science can be implemented via different media. Thus, if mind is Turing-style computation, it can be realized not only in brain, but also using, for instance,

silicon chips. Clearly, for Artificial Intelligence only cognitive science theories that imply Multiple Realizability are of interest. Later we will see that in relation to consciousness, the rejection of Multiple Realizability thesis is one of the main argument strategies against CTM.

And precisely, this is Searle's line of reasoning when he argues that a computer simulation of minds only simulates and never duplicates minds. Searle (1980, 29) contends as the computer simulation of digestion is not itself a process of digestion (or the computer simulation of a hurricane is not itself a hurricane), the computer simulation of human thinking process is not itself a thinking process. In short, he, contra Simon, argues that a computer simulation of thinking does not think, or, contra Chalmers, a computer simulation of a conscious entity is not itself conscious. And this is because Searle does not believe that thinking or minds are substrate neutral. In his biological naturalism, he argues that consciousness is a higher-level biological phenomenon, making the biological elements of consciousness essential to the occurrence of consciousness. Be this as it may, the issue then of whether a computer simulation of minds merely simulate or duplicate minds is a crucial, if not the crucial, issue to deal with in assessing the thesis of computationalism that minds are computers. For if such simulations do not duplicate minds, then it may be the case that while minds lend themselves to such simulations they are non-computational. Though his point about Turing is controversial, Rescoria (2017, 3) expresses the tension regarding the role of computer modelling as applied to minds in what follows:

> Formalization and computation are thus closely related, and together yield the result that reasoning that can can be formalized can also be duplicated (or simulated) by the right type of machine. Turing himself seems to have been of the opinion that a machine operating this way would literally be doing the same things that human performing computations is doing—that it would be 'duplicating' what the human computer does. But other writers have suggested that what the computer does is merely a 'simulation' of what the human computer does….

Let us look more closely at what transpires in the so-called computer simulation of minds? According to Anders Sandberg (2003, 3): "Simulations are processes that mimic the relevant features of target processes… A computer simulation is an attempt to model a particular system by creating a software representation that represent objects, relations and dynamics of the system in such a way that relations between objects in the simulation map onto relations between equivalence classes of objects in the original system." Now what about the computer simulation of human thought or intelligence? How is it done? But first, what is really being simulated in this kind of simulation? There seems to be three possible answers: first, thinking itself or the thought processes themselves; second, the brain activities that correlate with thinking processes; and third, intelligent behaviors or behaviors regarded as manifestations of thinking.

It seems apparent that thinking itself, being epistemically subjective (that is, being directly knowable only by its bearer), is never directly simulated. Mind simulation is merely inferred from either brain simulation or behavior simulation, which are the ones that can be done directly. Simon (1984, 25), speaking in behalf of the computationalists, wrote: "we all believe that machines simulate human thought." Observe, however, how he, along with Newell in their 1963 essay "GPS, A Program that Simulates Human Thought," explained how simulation of human thought proceeded in their research: "We…conceive of an intelligent program that manipulates symbols in the same way that our subject does—by taking as inputs the symbolic logic expressions, and producing as outputs a sequence of rule applications that coincides with the subject's…. If the fit of such program were close enough to the overt behavior of our human subject…then it would constitute a good theory of the subject's problem solving" (Newell and Simon 1995, 419-450). The main point seems to be that if a computer program enables a machine to perform certain actions which when performed by humans are considered intelligent, such as solving logic problems, then such a program has allegedly succeeded in simulating the thinking process of humans when performing the said actions.

We can gather from the explanation of Newell and Simon above that there are two levels of simulation that take place. On the first level is the computer simulation of human intelligent behavior; on the second level is the computer simulation of human thought. The first level is a simulation of external outputs (the external outputs of a computer simulate the behavioral outputs of the human mind) while the second level is a simulation of internal processes (the computer software simulates the "human software"). The first level is a direct kind of simulation while the second level is an inferred one—inferred from the first level. What happens here is similar to our knowledge of other minds. We cannot directly know the mental states of other persons; but we can infer them based on the similarity of their behavior with ours when we are expressing our own mental states. In any case, the point here is that human thought is never directly simulated; its simulation is inferred, in the case of Newell and Simon, from a behavioral simulation.

But before we deal with simulation on the level of intelligent behavior, let us look into the simulation on the level of brain activities. It may be thought that the way to computationally simulate human thought is by a computer simulation of the human brain. If the brain activities are constitutive of mental activities, and brain activities are computational (the main claim of computational neuroscience), then the computer simulation of brain activities is a direct simulation of mental activities (see the discussion of Milkowski on how neuroscience studies brain computation, Milkowski 2018a, 530-32). It is the brain itself that is the computer or that does the computing, unlike in the classical model where the brain is just the implementing device of the mind software that does the computing. As McDermott (2007, 145) writes: "What I argue is that the essence of computationalism is to believe (a) that brains are essentially computers; and (b) digital computers can simulate them in *all important respects*, even if they aren't digital at all." [My italics] The problem here is in treating the brains as the computers (not their mental software) this will abandon the principle of multiple realizability for it will make

the cognitive process or information processing substrate relative in that it will be biologically determined. Since computer modelling is done by a machine, which is not of the same biological stuff as the human brain, how can then we claim that the computer simulation done by the machine of the human biological brain duplicates the computational process of the brain? If the computational process of the biological brain is not multiply realizable, then it cannot be duplicated by a non-biological system.

Let us now examine the computer simulation done on the level of intelligent behavior. Roger Schank and Peter Childers contend that the reasoning behind the attribution of intelligence to machines would be basically the same when we are to determine whether extraterrestrials or aliens are capable of understanding or intelligence, for in both cases we just have to rely on behaviors. Referring to aliens (extraterrestrial beings), they (1984, 55) write: "We would have no understanding of their civilization or their physiology, and the sheer problem of communication with such entities would limit sharply the accuracy of our assessment of their intelligence. Our basis for evaluating their intelligence would rest solely on the outputs we received from them, and their understanding of us would rest solely on our output to them."

It shall be observed that the main point of Schank and Childers basically follows the reasoning behind the Turing test (see Turing 1950). The Turing test uses a machine simulation of human intelligent behavior, in the form of answering questions, to determine whether machines can be said to be intelligent. This is precisely what we earlier identified as the first level of simulation involved in the alleged computer simulation of human thought. The goal of the machine is to mimic human intelligent behavior in ways such that an interrogator would be unable to distinguish between a machine and a human respondent. And in order to make the judgment of the interrogator objective, the features of the machine and human respondents that are irrelevant for intelligence attribution— such as physical features and sound of voice—the interrogator is separated from these respondents by a wall. Here the interrogator's only access to the respondents is through textual communication via a teletype machine. The point of the test is consistency in the attribution of intelligence: if the human respondent is considered intelligent through his/her answers, the machine respondent whose answers cannot be distinguished from those of the human respondent must therefore likewise be regarded as intelligent.

The Turing test, however, is only a test for machine intelligence; and not for the computationality of human thought or of thinking in general (see Mabaquiao 2014). If a machine passes the test, it is intelligent; but it does not mean that intelligence is computational in nature. Take again the case of an alien, if the alien passes the Turing test then it is intelligent. But should we suppose that the alien thinks like a computer? We really do not know; all we know is that it manifests behaviors which when manifested by humans are considered intelligent. The point of the test is simply to show that the inner mechanism of a certain system is irrelevant to the ascription of intelligence to it. Again, when Turing asked whether machines could be intelligent, he was not after whether intelligence is computational or whether human intelligence is also governed by some cognitive algorithm. He was only after whether given what machines can do, we can legitimately say they are intelligent. And the test that he developed to determine

machine intelligence is in fact not exclusive to computing machines; for it can very much apply to any kind of entity suspected of being intelligent such as aliens. Given this, the fact that computing machines are governed by computer programs is irrelevant to the ascription of intelligence to these machines. What is relevant is simply whether they can perform the intelligent tasks. It is thus possible for a non-computational system to exhibit intelligent behavior, pass the Turing test, and consequently be regarded as intelligent. In other words, the machine needs only to simulate human intelligent behavior to be considered intelligent, it does not need to duplicate human thought processes. The computational simulation of human minds, on this level of simulation, does not require that human minds be computational.

CONCLUSION

Two issues, one having to do with scope and the other with method, serve as roadblocks to the project of computationalism. On the issue of scope, it is unclear how much of the mind's capacities computationalism intends to account for. The issue occurs not just between computationalists and their critics but also among computationalists themselves. There is a standing disagreement on whether computationalism should account for the mind's phenomenal and cognitive features or merely for its cognitive features. As regards the mind's cognitive features, there is a further disagreement on whether it should account for the functionality of intelligence only or also of the consciousness of intelligence. On the other hand, on the issue of method, it is not clear how the duplication of minds can be arrived at through the routes of computer simulations of brain processes and intelligent behavior. The computer simulation of brain processes duplicates thought processes only if brain processes are taken to be constitutive of thought processes. This, however, leads to the rejection of a necessary condition for this duplication to occur: that thought processes be substrate neutral and thus be multiply realizable. Taking the route of computationally simulating intelligent behavior is likewise problematic, mainly for failing to sufficiently establish that minds are computational. Given this, no duplication of minds can be sufficiently established.

REFERENCES

Armstrong, David M. 1991. The causal theory of mind. In *The nature of mind.* Edited by David M. Rosenthal. Oxford: Oxford university press.

Block, Ned. 1991. Troubles with functionalism." In *The nature of mind,* edited by David Rosenthal, 211-229. Oxford: Oxford University Press.

Chalmers, David. 1993. A computational foundation for the study of cognition. https://www.ida.liu.se/divisions/hcs/seminars/cogsciseminars/Papers/ Chalmers_Computa   tional_foundations.pdf. [*Journal of Cognitive Science* 2012 (12): 323-357]. Accessed 9 June 2014.

_____. 2010. The singularity: A philosophical analysis." *Journal of Consciousness Studies* (17): 7-65. http://consc.net/papers/singularity.pdf. Accessed 4 October 2012.

Churchland, Paul. 1995, *The engine of reason, the seat of the soul*. Cambridge: MIT Press.

Churchland, Patricia. 1986, *Neurophilosophy*. Cambridge: MIT Press.

Crane, Tim. 1995. *The mechanical mind: A philosophical introduction to minds, machines and mental representation.* London: Penguin Books Ltd.

Dennett, Daniel. 1989. *The intentional stance.* Massachusetts: The MIT Press

_____. 1991. Three kinds of intentional psychology. In *The nature of mind,* edited by David Rosenthal, 613-625. Oxford: Oxford University Press.

Eliasmith, Chris.,2013, *How to Build a Brain*, Oxford: Oxford: University Press.

Fodor, Jerry. 1979. Th*e language of thought.* Cambridge: Harvard university press.

_____. 1991. Methodological solipsism considered as a research strategy in cognitive psychology." In *The nature of mind,* edited by David Rosenthal, 485-498. Oxford: Oxford University Press.

_____. 2000. *The mind doesn't work that way: The scope and limits of computational psychology.* Massachusetts: The MIT Press.

Freidenberg, Jay & Silverman, Gordon. 2006. *Cognitive science: An introduction to the study of mind.* California: Sage Publications, Inc.

Gardner, Howard. 1985. *The mind's new science: A history of the cognitive revolution.* U.S.A.: BasicBooks-A Division of HarperCollinsPublishers.

Harnish, Robert. 2002. *Minds, brains, computers: A historical introduction to the foundations of cognitive science.* Oxford: Blackwell Publishers.

Jackson, Frank. 1991. What Mary didn't know. In *The nature of mind,* edited by David Rosenthal, 291-295. Oxford: Oxford University Press.

Mabaquiao, Napoleon. 2014. Turing and computationalism. *Philosophia: An International Journal of Philosophy* 15 (1): 50-62.

_____. 2012. *Mind, science and computation.* Manila: De La Salle University Publishing House and Vibal Foundation, Inc.

_____. 2011. Computer simulation of human thinking: An inquiry into its possibility and implications." *Philosophia: An International Journal of Philosophy* 40(1): 76-87.

Milkowski, Marcin. 2013. *Explaining the computational mind.* Massachusetts: The MIT Press.

_____. 2017. Objections to computationalism. A short survey. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society. Computational Foundations of Cognition* (pp. 2723–2728). Presented at the 39th Annual Meeting of the Cognitive Science Society, London: Cognitive Science Society. https://mindmodeling.org/cogsci2017/papers/0515/index.html.

_____. 2018a. From computer metaphor to computer modeling: The evolution of computationalism. *Minds and Machines* (28): 515-541.

_____. 2018b. The computational theory of mind. The Internet Enccyclopedia of Philosophy,   ISSN 2161-0002, https://www.iep.utm.edu/compmind/. Accessed 31 October 2018.

McDermott, Drew. 2007. Artificial intelligence and consciousness. In *The Cambridge*

handbook of consciousness, edited by Philip D. Zelano, Morris Moscovich, and Evan Thompson, 117-150. Cambridge: Cambridge University Press.

McClelland, James and David Rumelhart. 1993. On learning the past tenses of English verbs. In *Readings in philosophy and cognitive science.* Edited by Alvin Goldman. Cambridge: The MIT press.

Minsky, Marvin. 1995. Logical versus analogical or symbolic versus connectionist or neat versus scruffy. In *Computation and intelligence: Collected readings.* Edited by George Luger. Cambridge: The MiT press.

Newell, Allen & Simon, Herbert. 1961. Computer simulation of human thinking. *Science, New Series,* 134 (3495): 2011-2017.

_____. 1976. Computer science as empirical inquiry: Symbols and search. *Communications of the ACM (Association for Computing Machinery)* 19 (3): 113-126.

_____. 1995. GPS, a program that simulates human thought. In *Computation and intelligence: Collected readings,* edited by G. Luger, 415-428. Cambridge: The MIT Press.

Penrose, Roger. 1989. *The Emperor's new mind: Concerning computers, minds, and the laws of physic.* Oxford: Oxford University Press.

Peschl, Markus & Scheutz, Matthias. 2001. Some thoughts on computation and simulation in cognitive science." In *Proceedings of the Sixth Congress of the Austrian Philosophical Society.* Online: https://hrilab.tufts.edu/publications/scheutzpeschl00linz.pdf.Accessed February 14, 2017.

Putnam, Hilary. 1991. The nature of mental states. In *The nature of mind.* Edited by David M. Rosenthal. Oxford: Oxford University Press.

Pylyshyn, Zenon. 1990. Computing in cognitive science. In *Foundations of cognitive science.* Edited by Michael Posner. Cambridge: The MIT Press.

Rescoria, Michael. 2017. The computational theory of mind. *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2017/entries/computational-mind/>. Accessed 4 March 2018.

Rumelhart, David. 1990. The architecture of mind: A connectionist approach. In *Foundations of cognitive science.* Edited by Michael Posner. Cambridge: The MIT Press

Sandberg, Anders. 2013. Feasibility of whole brain emulation. In *Theory and philosophy of artificial intelligence,* edited by V. Müller. Berlin: Springer. http://shanghailectures.org/sites/default/files/uploads/2013_Sandberg_Brain-Simulation_34.pdf. Accessed 8 May 2014.

Sandberg, Anders & Bostrom, Nick. 2008. *Whole brain emulation: A roadmap.* Oxford University: Technical Report, Future for Humanity Institute.http://www.fhi.ox.ac.uk/Reports/2008-3.pdf. Accessed 8 October 2012.

Schank, Roger & Childers, Peter. 1984. *The cognitive computer.* Reading: Addision-Wesley Publishing Company, Inc.

Searle, John. 1980. Minds, brains, and programs. *Behavioral and Brain Sciences* 3 (3):

417-457.http://www.cogsci.soton.ac.uk/bbs/Archive/bbs.searle2.html. Accessed 8 October 2011.

_____ . 1990. Is the brain's mind a computer program?" *Scientific American,* 26-31. http://www.cs.princeton.edu/courses/archive/spr06/cos116/ Is_The_Brains_Mind_A_Computer_Program.pdf. Accessed 3 January 2011.

Schoemaker, Sydney. 1982. The inverted spectrum. *The Journal of Philosophy* 79 (7): 357-381.

Simon, Herbert. 1995. Machine as mind. In *Computation and intelligence: Collected readings*, edited by George Luger, 675-692. Massachusetts: The MIT Press.

_____ . 1984. Why should machines learn. In *Machine learning: An artificial intelligence approach,* edited by R. Michelski, J. Carbonell J, and T. Mitchell, 25-37. Berlin: Springer.

Simon, Herbert & Kaplan, Craig. 1990. Foundations of cognitive science. In *The foundations of cognitive science,* edited by Michael Posner, 1-47. Massachusetts: The MIT Press.

Smolensky, Paul. 1993. On the proper treatment of connectionism. In *Readings in philosophy and cognitive science.* Edited by Alvin M. Goldman. Cambridge: The MIT press.

Sun, Ron. 2008. Introduction to computational cognitive modeling. In *The Cambridge handbook of computational psychology,* edited by Ron Sun, 3-20. Cambridge: Cambridge University Press, 2008.

Turing, Alan. 1950. Computing machinery and intelligence. *Mind* 59 (236): 433-460.

Wein, TU and Pona. 2015. Computational theory of consciousness. *Seminar in artificial intelligence.* https://www.logic.at/lvas/SemAI/Consciousness.pdf. Accessed 10 October 2018.

Submitted: 25 August 2018; revised: 27 April 2019