



# The Brain-Computer Metaphor Debate Is Useless: A Matter of Semantics

Blake A. Richards<sup>1,2,3,4\*</sup> and Timothy P. Lillicrap<sup>5</sup>

<sup>1</sup> Mila, Montreal, QC, Canada, <sup>2</sup> Department of Neurology and Neurosurgery, Montreal Neurological Institute, McGill University, Montreal, QC, Canada, <sup>3</sup> School of Computer Science, McGill University, Montreal, QC, Canada, <sup>4</sup> Learning in Machines and Brains Program, CIFAR, Toronto, ON, Canada, <sup>5</sup> DeepMind Inc., London, United Kingdom

## OPEN ACCESS

### Edited by:

Giorgio Matassi,  
FRE3498 Ecologie et dynamique des  
systèmes anthropisés (EDYSAN),  
France

### Reviewed by:

Marcos Cramer,  
Technical University Dresden,  
Germany  
Michael Levin,  
Tufts University, United States

### \*Correspondence:

Blake A. Richards  
blake.richards@mila.quebec

### Specialty section:

This article was submitted to  
Theoretical Computer Science,  
a section of the journal  
Frontiers in Computer Science

**Received:** 06 November 2021

**Accepted:** 17 January 2022

**Published:** 08 February 2022

### Citation:

Richards BA and Lillicrap TP (2022)  
The Brain-Computer Metaphor  
Debate Is Useless: A Matter of  
Semantics.  
Front. Comput. Sci. 4:810358.  
doi: 10.3389/fcomp.2022.810358

It is commonly assumed that usage of the word “computer” in the brain sciences reflects a metaphor. However, there is no single definition of the word “computer” in use. In fact, based on the usage of the word “computer” in computer science, a computer is merely some physical machinery that can in theory compute any computable function. According to this definition the brain is literally a computer; there is no metaphor. But, this deviates from how the word “computer” is used in other academic disciplines. According to the definition used outside of computer science, “computers” are human-made devices that engage in sequential processing of inputs to produce outputs. According to this definition, brains are not computers, and arguably, computers serve as a weak metaphor for brains. Thus, we argue that the recurring brain-computer metaphor debate is actually just a semantic disagreement, because brains are either literally computers or clearly not very much like computers at all, depending on one’s definitions. We propose that the best path forward is simply to put the debate to rest, and instead, have researchers be clear about which definition they are using in their work. In some circumstances, one can use the definition from computer science and simply ask, what type of computer is the brain? In other circumstances, it is important to use the other definition, and to clarify the ways in which our brains are radically different from the laptops, smartphones, and servers that surround us in modern life.

**Keywords:** neuroscience, psychology, computer science, brains, computers, Turing machines, parallel distributed processing

## 1. INTRODUCTION

Computation has been a central feature of research in the brain sciences (neuroscience, psychology, and cognitive science) for decades. Papers in the brain sciences are full of references to algorithms, coding, and information processing (Diamant, 2008; Maass, 2016; Oteiza et al., 2017). At the same time, there is a long and continuing history of debate around these words (Maccormac, 1986; West and Travis, 1991; Smith, 1993; Vlasits, 2017). According to many scientists and philosophers, computers are used as a metaphor to understand brains and this metaphor can be misleading or counter-productive (Carello et al., 1984; Cisek, 1999; Epstein, 2016; Cobb, 2020). Throughout the history of the brain sciences over the last 80 years, one can find researchers who comfortably use

computational theory and language to explore and understand brains (Marcus, 2015), as well as researchers who reject the use of such concepts for use with brains (Epstein, 2016). Indeed, the early dream of cognitive science in the second half of the twentieth century depended on the links between brain sciences and artificial intelligence (AI) (Newell, 1980; Simon, 1980; Pylyshyn, 1984; Hunt, 1989), yet the failure to make good progress in AI in the 1970's, 80's, and 90's, and the inability to connect such systems convincingly to the brain sciences, led some researchers to conclude that the "metaphor of the brain as a computer" was broken at its foundations (Dreyfus and Hubert, 1992; Van Gelder, 1998). To this day, one can still find in equal measure both brain scientists who use theories from computer science (Kwisthout and van Rooij, 2020) and brain scientists who argue against the brain as a computer metaphor (Brette, 2018).

However, closer inspection of the debates on this topic reveal a fundamental misunderstanding between the participants regarding the definition of the word "computer". Indeed, many of the entries in these debates do not grapple concretely with the definition of the word "computer" before declaring either way that the brain is or is not well-explained with computational theory. To actually resolve this debate, it is helpful to bring the definition of "computer" into clear focus.

Here, we argue that closer examination of the manner in which both computer scientists and non-computer scientists use the word "computer" indicates that there are at least two distinct definitions in operation: (1) A definition from computer science rooted in the formal concepts of computable functions and algorithms. (2) A definition from outside of computer science based on the electronic devices we use on a regular basis and how they operate. To make matters worse, some neuroscientists, cognitive scientists, and psychologists have a mixed familiarity with the formal concepts from computer science that underpin the first definition. This means that semantic debates stemming from misaligned definitions are particularly apt to emerge in the context of the brain sciences, leading to proponents on either side who seem irreconcilable.

In this article, we clarify these two distinct definitions. We show that if one adopts the definition from computer science, then the question is not whether computers are a good metaphor for brains, because brains arguably are *literally* computers based on this definition. In contrast, if one adopts the definition from outside of computer science then brains are not computers, and arguably, computers are a very poor metaphor for brains. Thus, the argument over whether or not computers are a good or bad metaphor for brains is actually just a matter of semantics. Under one definition, brains are literally computers, whereas under another, they are clearly not. There is, therefore, little utility in continuing these debates. We close on a prescription for the brain sciences. We suggest that the question for scientists should instead be: if we adopt the definition from computer science, then what kind of a computer are brains? For those using the definition from outside of computer science, they can be assured that their brains work in a very different way than their laptops and their smartphones—an important point to clarify as we seek to better understand how brains work.

## 2. MEANING AS USE

Before we discuss the different definitions of the word "computer", it is important that we clarify our approach to the definitions and meanings of words. In this paper, we adopt a perspective that focuses on the *use* of words for understanding their meaning, and thus, their definition. Therefore, we will avoid telling the reader that, for example, "computers are formally defined as X, and everyone must adopt this definition". Instead, we will draw the reader's attention to the ways in which the word "computer" *is in fact used* in contexts inside and outside of computer science, and proceed from there.

Briefly, the idea that we can best understand the meaning of a word by looking at its use in context has a long history in philosophy, perhaps best exemplified by the works of Ludwig Wittgenstein. Wittgenstein argued in the *Philosophical Investigations* that "*in most cases, the meaning of a word is its use*" (Wittgenstein, 1953). This idea flies in the face of many of our intuitive notions about how words work; like the young Wittgenstein, many of us tend to think about meaning in terms of correspondence, i.e., that "*individual words in language name objects and sentences are combinations of such names*" (Wittgenstein, 1953). But, in fact, meaning is crucially modified by context and use, rather than corresponding to particular objects, so the meaning of most words are fuzzy and impossible to write down precisely and uniquely. Wittgenstein showed how much confusion is generated by failing to pay attention to how words are used in context and we believe that much of the confusion around the question "Is the brain (like) a computer?" results from just this sort of confusion. In particular, the "computer as brain" debate often devolves into a semantic disagreement generated by a mismatch in expectations between two uses of the word "computer", which we will clarify below.

Of course, it should be noted that there can be many working definitions of the word "computer", but only two that are prominent and important in our context. The definition used by computer scientists is important because it underpins work in computational neuroscience and AI. And, at the same time, the definition used by academics outside of computer science is important because it's the one that most writers in the brain sciences intuitively reach for during these debates. As we'll see, someone operating with the computer science definition who says that the "brain is a computer" is certainly correct. Simultaneously, someone using the definition from outside of computer science who says that "the brain is not a computer and computers are not a good metaphor for brains" is also correct. Thus, unless the time is taken to clear up the question of usage, there's bound to be disagreement with little ground given by either side. As such, we must first explore these two distinct uses.

## 3. THE USE OF THE WORD "COMPUTER" INSIDE COMPUTER SCIENCE

Here we will provide an overview of the definition of the word "computer" based on the use of the word in computer science. As we will describe, this use-based definition partly

relies on the formal definition of the word “algorithm”. However, the definition of “computer” derived solely from the formal definition for “algorithm” is actually so broad as to be nearly meaningless. Nonetheless, the use of the word “computer” in computer science shows that computer scientists generally mean something more restrictive than the formal definitions would indicate. As we will show, the more restrictive, use-based definition is still applicable to brains.

### 3.1. The Formal Definition of “Algorithm” and The Church-Turing Thesis

Within computer science the formal definition for the word “algorithm” dates back to the early twentieth century, before the invention of modern computers and the discipline of computer science as it exists today. Back then, what would become computer science was essentially a branch of mathematics. Many mathematicians at the time were concerned with questions about a class of mathematical tools that they called “effective methods”. An effective method is a finite recipe that one can follow mechanically to arrive at an answer to some mathematical problem (Copeland, 2020), e.g., long division is an effective method for solving division problems with arbitrarily large numbers. Today, we refer to effective methods as “algorithms”. The intuitive definition of an algorithm is therefore as above: a finite recipe that one can follow mechanically to arrive at an answer to some problem (Cormen et al., 2009). But, we also have a formal definition thanks to the work of those early mathematicians. For example, in 1900, the mathematician David Hilbert put forward a set of 23 problems to be solved in the twentieth century, the 10<sup>th</sup> of which was “*Can we develop an algorithm for determining whether the roots of a polynomial function are integers?*” (Hilbert, 1902). Later, these types of questions were expanded in scope to larger questions such as the *Entscheidungsproblem*, which asks whether there is an algorithm for determining whether any given statement is valid within an axiomatic language (Hilbert and Ackermann, 1999).

It turned out that these mathematicians had stumbled onto a very deep set of problems. As they began to explore algorithms more and more, they started to wonder whether some problems in mathematics, including Hilbert’s 10<sup>th</sup> problem and the *Entscheidungsproblem*, did not in fact have any solution. The way this is sometimes phrased is, are there problems that are not “decidable”? A problem is “decidable” if and only if there exists an algorithm for solving it (Cormen et al., 2009), and there was a growing realization that some problems were likely not decidable. Of course, mathematicians being mathematicians, they desired a proof that an algorithm didn’t exist in such cases. The problem was that at the time the definition of an algorithm was the informal, intuitive definition above. Without a formal definition of the word “algorithm” it was impossible to prove that some problems were, in fact, not decidable.

This set the stage for the development of modern computer science as we know it today. A pair of mathematicians, Alonzo Church and Alan Turing, independently decided to try to develop a formal definition for “algorithm” for the sake of developing proofs related to the *Entscheidungsproblem* and decidability more

broadly (Church, 1936a,b; Turing, 1936). Church invented a formal logical system he called lambda calculus, and defined an algorithm as anything that could be done with lambda calculus (Church, 1936a,b). Turing invented a mathematical construct known as a Turing machine, and defined an algorithm as anything that could be done with Turing machines (Turing, 1936). Both researchers used their definitions to show that there was no solution to the *Entscheidungsproblem*. As well, while the two researchers had developed what looked like very different definitions, they turned out to be mathematically equivalent (Turing, 1937). Continued work in computability theory, the branch of computer science and mathematics concerned with the study of decidable problems, has suggested that any attempt to formalize the intuitive definition of algorithm will end up being equivalent to lambda calculus and Turing machines (Cook, 1992, 2014; Copeland, 2020). As such, computer scientists today largely accept an idea known as **the Church-Turing thesis**, which states (very roughly), that any algorithm can be implemented *via* a Turing machine, i.e., it proposes that we accept Church and Turing’s definitions as given (Copeland, 2020). Thus, when people seek a proof that there is no algorithm for some problem, they often do so by proving that you can’t solve the problem with a Turing machine (Cook, 1992).

Importantly for the discussion here, the formal definition for an “algorithm” also gives rise to a formal conception of the word “computer”. Specifically, computer scientists define a “computable function” to be any function whose values can be determined using an algorithm. A “computer” is then formally physical machinery that can implement algorithms in order to solve computable functions (though one may also take a slightly more expansive approach Copeland, 1997). It’s worth noting that this conception of what a computer is makes no reference to human made artifacts, or electronics, or silicon chips, etc. And, if we think back to the use of the word “computer” at the point in history when Church and Turing were working, this makes a lot of sense: “Computers” at this time were people whose job was to sit down with pencil and paper and use effective methods (i.e., algorithms) to solve various problems (e.g., to integrate equations) (Grier, 2001). Clearly, these people were computers according to the definition above, because they were solving computable functions, even though they were of course not human-made artifacts. Thus, the formal definition of the word “algorithm” rests on the Church-Turing Thesis, and this in turn provides us with a formal definition of “computable functions”, which is what “computers” solve. And, none of this has anything to do with the physical characteristics or internal workings of the computer, only with its ability to physically implement computable functions.

### 3.2. Limiting the Scope of the Formal Definition in Practice in Computer Science

If we consider the definition above for “computer”, a problem arises: this definition can be applied to almost any object in the universe. Consider for a moment the fact that the movement of objects in the world can be described by computable functions, e.g., the parabolic curve of a thrown ball. As such,

the definition that rests on the formal conception of algorithms and decidability, when applied directly, tell us that all objects in the world are computers, since they are physically implementing computable functions. Put another way, if you wanted to calculate a parabolic curve you could throw a ball and simply track its movement, so in some sense, you could use the ball to solve your mathematical problem, and it is thus a “computer” solving your parabolic curve. Though this is formally correct it is conceptually unsatisfying. What use is it for us to define “computer” in this manner if it trivially renders most of the universe and everything within it a computer?

In this instance, the *use* of the word becomes critical. Despite the formal definitions, computer scientists rarely refer to thrown balls as “computers”. Is that because computer scientists only use the word to refer to electronic devices like our laptops and smartphones? No, there are clear examples of references in computer science to computers that are very different from the typical digital computers we’re all familiar with, including analog computers, quantum computers, stochastic computers, DNA computers, and neuromorphic computers (Gaines, 1967; Rubel, 1993; Adleman, 1994, 1998; Beaver, 1995; Paun et al., 2005; Van Noort and Landweber, 2005; Elbaz et al., 2010; Ladd et al., 2010; Furber, 2016; Schuman et al., 2017; Tsividis, 2018; van de Burgt et al., 2018; Shastri et al., 2021). None of these forms of computer operate like a laptop or smartphone; they can use analog signals, stochastic operations, parallel calculations, biological substrates, etc. And yet, the usage of the word “computer” in such articles does not appear to be intended as a metaphor. So, what then renders something a “computer” in computer science, according to the way the word is used?

What we can see in research papers is that computer scientists generally use the word “computer” to refer to any physical machinery that can, *in theory*, implement *any* computable function (per the definition above), i.e., a physical system that in principle can serve as a “universal” computation device (Beaver, 1995; Van Noort and Landweber, 2005; Ladd et al., 2010). For example, when Adleman (1994) closed his paper on DNA-based computation he said, “*One can imagine the eventual emergence of a general purpose computer consisting of nothing more than a single macromolecule conjugated to ribosome like collection of enzymes that act on it*”. Here, the key point is the words “general purpose”. It is the potential for general purpose computation with DNA that, we argue, makes computer scientists inclined to talk about “DNA computers”, despite the fact that a macromolecule conjugated to a ribosome like collection of enzymes would engage in calculation in a very different manner than a modern silicon chip.

Note also our use of the phrase “*in theory*”, above. Many of the systems that computer scientists refer to as “computers” cannot in practice implement any computable function due to size, memory, time, noise, and energy limitations. So, for example, quantum computers are not yet capable of computing any computable function, but in theory they could, and so we refer to them as “computers”. And, of course, a laptop is a “computer” because it can be shown that the operations it utilizes could theoretically implement any computable function, though in reality some functions would take too long or require too much

memory (e.g., calculating the number of prime numbers less than  $10^{10^{10}}$ ). In contrast, a thrown ball is limited to implementing only those functions that describe its movement through space. Thus, when computer scientists use the word “computer”, they generally use it to refer only to physical machinery that could, in theory, compute any computable function, which is by no means applicable to most things (Adleman, 1998; Elbaz et al., 2010; Ladd et al., 2010; van de Burgt et al., 2018).

### 3.3. Applying the Definition From Computer Science to Brains

Given the use-based definition above (physical machinery that can implement any computable function in theory), are brains computers? The answer for most scientists should be *yes*. First, though there is disagreement in philosophy as to whether brains are purely physical systems and whether their operations rely solely on physical machinery, the perspective of physicalism is widely accepted by brain scientists and we are not aware of any brain scientists who doubt that the operations of the brain are fundamentally physical. Second, with the aid of a pencil and paper, a human brain can in theory implement any program that one could implement with modern digital computers. The only limits would be time and energy, which as noted, also apply to other computers, like laptops. Even without pencil and paper, the only real limit to a person implementing any computer program is again the limits on their memory, time, and energy, not their general capabilities, *per se*. Conceptually, we can perform all of the same operations specified by the languages that we program our laptops with. Third, and perhaps more importantly, if one is concerned with practical implications for the brain sciences, real neural circuits are in theory, likely capable of implementing all of the functions that artificial neural networks (ANNs) can, if not more. And, computer scientists have shown that ANNs can implement any computable function (Hornik, 1991; Siegelmann and Sontag, 1995). In other words, as long as real brains have the same or greater capabilities than ANNs (again ignoring memory, time, and energy constraints), then they are surely capable of implementing any computable function.

Therefore, according to the use-based definition of “computer” in computer science, brains are literally computers. *There is no metaphor*. The claim here is not that brains work anything like our laptops and smartphones. But the use-based definition of “computer” in computer science isn’t “something that works like a laptop or smartphone”—DNA and quantum bits are very different from silicon chips. The definition of “computer” is physical machinery that, in theory, can implement any computable function, and brains meet this definition at least as well as many of the other devices that we all refer to as “computers” on a regular basis with no complaint and no hint of a metaphor.

We should address here a few of the common misconceptions that lead people to object to this line of logic. First, one of the most common points of confusion is that some people think that the formal definition based on the Church-Turing thesis implies that to be a computer an object’s internal machinery itself must operate in a similar manner to Turing

machines (Fodor, 1981; Copeland, 2020). But, this is simply a misunderstanding, as many types of computers (e.g., analog computers or neuromorphic computers) do not operate like a Turing machine. This misunderstanding may derive from the fact that modern digital computers bear some resemblance to the Turing machine formalism. But importantly, Turing machines are just mathematical constructs—they are sets of rules, not physical machines. Your laptop computer is no more a Turing machine than it is a lambda calculus. Nothing about the way computer scientists use the word “computer” demands that the object work like a Turing machine—the object in question must simply be capable of implementing the same functions as Turing machines.

Second, another reasonably common claim is that brains can't be computers because they can solve problems that are not decidable (Penrose, 1989; Siegelmann, 1995). We note that no one has ever convincingly demonstrated that brains can actually do this. However, importantly, this claim speaks to the question of whether brains are literally computers, not whether computers are a good metaphor for brains. As such, though it is an interesting objection that warrants consideration, it does not change our fundamental point, which is that there is no metaphor in play when we apply the definition of the word “computer” from computer science to brains. Finally, another source of confusion can enter into the discussion when simulations are discussed. We can, of course, simulate aspects of how neural circuits work using digital computers. And so, it has sometimes been believed that the claim that brains are computers derives from our ability to simulate them, and in turn, it has been (rightly) pointed out that the ability to simulate something with a computer does not make that thing a computer (Brette, 2018), e.g., we can simulate a ball bouncing but that does not make a ball a computer. But, as outlined above, it is not our ability to simulate neural circuits that makes brains computers, it is their theoretical ability to implement any computable function. Hence, the question of simulation is actually irrelevant to the question of whether brains are computers or not. The only relevant question is: Can brains implement any computable function in theory? And we argue that the answer is certainly “yes”.

#### 4. THE USE OF THE WORD “COMPUTER” OUTSIDE OF COMPUTER SCIENCE

All of this may be a bit surprising to many readers, because the definitions of “computer” given above is not how the average person, nor the average academic outside of computer science, understands and uses this word. As such, we may ask for an alternative definition of “computer”, one that aligns better with the usage of people outside of computer science.

When most people speak of a “computer” today, they use the word to refer to human-made electronic devices that can perform complex mathematical calculations, display multimedia content, and communicate with other similar devices. According to this usage, a computer can be defined as something like “an electronic appliance that we can use for calculation, communication, and

entertainment”. Obviously, this definition does not apply to brains, nor would it serve as a particularly good metaphor either.

Within academia, there are also people in the brain sciences and philosophy who are more knowledgeable about computers (and brains) but who are still only partially familiar with the ideas from computer science presented above. For these people, the usage of the word “computer” often still centers on the human-made electronic devices we are all familiar with, but it includes some more details of how those devices work. Specifically, the vast majority of modern digital computers are extensions of the “Von Neumann architecture”, first developed by the polymath John Von Neumann in the 1940's (Von Neumann, 1993). Though there have been changes to Von Neumann's original design (Godfrey and Hendry, 1993), some of his ideas are still central to modern digital computers. These include the use of a central processing unit (CPU) for sequential operations of arithmetic logic, a control unit in the CPU that stores the sequence of instructions for the CPU to perform, a random access memory (RAM) module for storing intermediate calculations, and an external memory (or “hard drive”) for long-term storage of information. It's interesting to note that Von Neumann's designs are reminiscent of how we define Turing machines, with an internal state, and a step-by-step processing of input symbols to produce output. Given this apparent similarity, many writers use the word “computer” to mean something like “human-made machines that have the qualities of Von Neumann architecture machines, and which resemble aspects of Turing machines” (Cisek, 1999; Epstein, 2016; Cobb, 2020). Hence, one can find articles where people refer to computers and computation as being necessarily sequential, or discrete, or restricted to passive processing of a stream of inputs using a step-by-step program (Van Gelder, 1998; Cisek, 1999; Brette, 2018, 2019; Cobb, 2020). For example, Cisek (1999) notes the importance of control for brains and animals, which he argues is ignored by the computer metaphor for the brain, because it instead presupposes that “...*perception is like input, action is like output, and all the things in-between are like the information processing performed by computers.*” His point here is that brains are not simply taking inputs and producing outputs based on some internal state (akin to the formalism of Turing machines), but rather, they are constantly engaged in adaptive interactions for controlling the body and the world in order to achieve specific ends. However, control is something that people in computer science would happily say computers can do (Arnăutu and Neittaanmäki, 2003). Thus, Cisek (1999)'s concern is less about “computers” as they are defined in computer science, and more about “computers” as they are defined by those outside of computer science.

With the definition from outside of computer science in hand, are brains computers? Most certainly not. Brains do not use sequential processing—quite the opposite they use massively parallel processing (Rumelhart et al., 1988). Brains do not use discrete symbols stored in memory registers—they operate on high-dimensional, distributed representations stored *via* complex and incompletely understood biophysical dynamics (Jazayeri and Ostojic, 2021). And, brains do not passively process inputs to generate outputs using a step-by-step program—they control an embodied, active agent that is continuously interacting

with and modulating the very systems that generate the sensory data they receive in order to achieve certain goals (Cisek, 1999; Brette, 2019). Thus, with the definition from outside of computer science we can say not only that brains are not computers, we can also say that computers are poor metaphors for brains, since the manner in which they operate is radically different from how brains operate.

There are some complications to this that should be noted. First, brains are capable of some forms of more traditional tasks that our digital computers are good at, i.e., various forms of discrete, sequential processing (Fodor, 1981; Marcus, 2015). For example, people can do long-division, symbolic logic, list sorting, etc. So, we might say that computers (according to the definition from outside of computer science) can serve as reasonable metaphors for *some types* of human cognition. Moreover, modern digital computers are rapidly evolving to incorporate more parallel, distributed, dynamic operations (Shukur et al., 2020), and some engineers are actively trying to explicitly mimic the operations of brains using “neuromorphic” chips (Furber, 2016; Schuman et al., 2017; van de Burgt et al., 2018; Shastri et al., 2021). These more modern forms of human-made computers present some complications for the use-based definition of “computer” from outside of computer science. Nonetheless, if we are committed to the concept of use-based meaning, then we can say that when some authors dismiss the brain-computer metaphor (Carello et al., 1984; Cisek, 1999; Brette, 2018) they are using the word “computer” to mean something more like traditional, Von Neumann architecture machines, not neuromorphic chips, etc. And, as noted, such authors are correct, brains are not very much like these traditional digital computers.

## 5. DISCUSSION

Tying our two different threads together, we can conclude that the question of whether brains are computers (or like computers) is really a matter of semantics: it depends on which definition you are using. If you adopt the definition of “computer” based on how computer scientists use the word (to refer to physical machinery that can theoretically engage in any decidable computation), then brains are literally computers. Alternatively, if we adopt the definition of “computer” based on the usage from outside of computer science (to refer to devices that sequentially and discretely process inputs in a passive manner), then brains are not computers, and at best, computers serve as a weak metaphor for only a limited slice of human cognition. The message that we are providing here to the brain sciences community is, we hope, very clear: brains are either literally computers, or really not much like computers, depending on the definition we employ. Thus, it is ultimately a matter of semantics, and arguably, debates about the “brain-computer metaphor” are not productive. We can simply stop engaging in them.

It is worth noting that our argument here rests on an important stance vis-à-vis the philosophy of science. Specifically, we are assuming that scientists can and do use words and

concepts in a literal manner. This is in contrast to a potential perspective that views all concepts as metaphors (Lakoff and Johnson, 1980). Putting aside the larger philosophical debate that would be possible on this matter, we wish here simply to clarify and recognize that our perspective very much so rests on the idea that there are non-metaphorical uses of words and concepts in science.

The natural question that emerges from the realization that the brain-computer metaphor debate is actually just a semantic disagreement is to ask whether it matters which definition of “computer” we adopt? Does it affect the brain sciences in any meaningful way to adopt one definition or the other? In particular, should the field be concerned with the definition from computer science at all, given that it is not terribly intuitive and not what most people in the brain sciences think of when they hear the word “computer”?

We would argue that the definition we adopt is very important, and both definitions should be considered. The usage of “computer” in computer science can actually be very useful for the brain sciences in some circumstances. The reason is that when one realizes that brains are literally computers (in the computer science sense of the word) then much of the theory about computation from computer science is applicable to brains. This connection is what opens up space in computational neuroscience to explore the brain using conceptual tools from computer science and AI, which has produced both important insights in neuroscience (Richards et al., 2019) and advances in AI (Hassabis et al., 2017). Indeed, asking the question, “*What sort of computer is the brain?*”, is arguably the underpinning of modern neural networks (Rumelhart et al., 1988), which have been very useful for the brain sciences. Asking this question is how we arrive at core concepts in computational neuroscience such as parallel processing, content addressable memory, and spike-based computation. Similarly, consider the question of randomness in computation. Thanks to our understanding that the brain is a computer we can apply concepts from computer science, such as convergence and constraint satisfaction, to better understand the normative importance of stochastic vesicle release in neurons (Maass and Zador, 1999; Habenschuss et al., 2013). Similarly, concepts from compression theory help us to understand the nature of representations in the brain (Olshausen and Field, 1996) and dynamic programming concepts used in reinforcement learning help us to understand memory replay (Mattar and Daw, 2018). More broadly, the inter-disciplinary intersection between AI and the brain sciences depends on the computer science definition of the word “computer”, and so, if we reject this definition outright we risk shutting the door on a very active field of research that has proven very fruitful for both the brain sciences and AI. At the same time, it is worth being vigilant and clear that brains do not work like our laptops and smartphones, and these devices serve as a poor metaphor for brains. So, depending on the audience and the purpose of the work, sometimes we should adopt the definition from outside of computer science, as long as we are clear on what that definition of “computer” actually implies. There is no single correct definition for “computer”—but we all must be clear on what we mean when we write and speak. On this

point, the vast majority of researchers across all disciplines must surely agree.

## AUTHOR CONTRIBUTIONS

BR and TL co-wrote the paper. All authors contributed to the article and approved the submitted version.

## REFERENCES

- Adleman, L. M. (1994). Molecular computation of solutions to combinatorial problems. *Science* 266, 1021–1024. doi: 10.1126/science.7973651
- Adleman, L. M. (1998). Computing with DNA. *Sci. Am.* 279, 54–61. doi: 10.1038/scientificamerican0898-54
- Arnăutu, V., and Neittaanmäki, P. (2003). *Optimal Control from Theory to Computer Programs*. Berlin: Springer Science & Business Media. doi: 10.1007/978-94-017-2488-3
- Beaver, D. (1995). A universal molecular computer. *DNA Based Comput.* 27, 29–36. doi: 10.1090/dimacs/027/03
- Brette, R. (2018). *What Is Computational Neuroscience? Is the Brain a Computer?* Available online at: <http://romainbrette.fr/what-is-computational-neuroscience-xxx-is-the-brain-a-computer/>
- Brette, R. (2019). Is coding a relevant metaphor for the brain? *Behav. Brain Sci.* 42, e243. doi: 10.1017/S0140525X19001997
- Carello, C., Turvey, M. T., Kugler, P. N., and Shaw, R. E. (1984). “Inadequacies of the computer metaphor,” in *Handbook of Cognitive Neuroscience*, ed M. S. Gazzaniga (Boston, MA: Springer), 229–248. doi: 10.1007/978-1-4899-2177-2\_12
- Church, A. (1936a). A note on the Entscheidungsproblem. *J. Symbol. Logic* 1, 40–41. doi: 10.2307/2269326
- Church, A. (1936b). An unsolvable problem of elementary number theory. *Am. J. Math.* 58, 345–363. doi: 10.2307/2371045
- Cisek, P. (1999). Beyond the computer metaphor: behaviour as interaction. *J. Conscious. Stud.* 6, 125–142.
- Cobb, M. (2020). *Why Your Brain Is Not a Computer*. The Guardian. London. Available online at: <https://www.theguardian.com/science/2020/feb/27/why-your-brain-is-not-a-computer-neuroscience-neural-networks-consciousness> (accessed February 27, 2020).
- Cook, S. A. (1992). “Computability and complexity of higher type functions,” in *Logic from Computer Science*, ed Y. N. Moschovakis (New York, NY: Springer), 51–72. doi: 10.1007/978-1-4612-2822-6\_3
- Cook, S. A. (2014). Conversations: from Alan Turing to NP-completeness. *Curr. Sci.* 106, 1696. doi: 10.18520/cs/v106/i12/1696-1701
- Copeland, B. J. (1997). The broad conception of computation. *Am. Behav. Sci.* 40, 690–716. doi: 10.1177/0002764297040006003
- Copeland, B. J. (2020). “The Church-Turing thesis,” in *The Stanford Encyclopedia of Philosophy*, ed E. N. Zalta (Metaphysics Research Lab, Stanford University).
- Cormen, T. H., Leiserson, C. E., Rivest, R. L., and Stein, C. (2009). *Introduction to Algorithms*. Cambridge, MA: MIT Press.
- Diamant, E. (2008). Unveiling the mystery of visual information processing in human brain. *Brain Res.* 1225, 171–178. doi: 10.1016/j.brainres.2008.05.017
- Dreyfus, H. L., and Hubert, L. (1992). *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge, MA: MIT Press.
- Elbaz, J., Lioubashevski, O., Wang, F., Remacle, F., Levine, R. D., and Willner, I. (2010). DNA computing circuits using libraries of DNAzyme subunits. *Nat. Nanotechnol.* 5, 417–422. doi: 10.1038/nnano.2010.88
- Epstein, R. (2016). *Your Brain Does Not Process Information and It Is Not a Computer*. Available online at: <https://aeon.co/essays/your-brain-does-not-process-information-and-it-is-not-a-computer>
- Fodor, J. A. (1981). The mind-body problem. *Sci. Am.* 244, 114–123. doi: 10.1038/scientificamerican0181-114
- Furber, S. (2016). Large-scale neuromorphic computing systems. *J. Neural Eng.* 13, 051001. doi: 10.1088/1741-2560/13/5/051001

## FUNDING

This work was supported by NSERC (Discovery Grant: RGPIN-2020-05105; Discovery Accelerator Supplement: RGPAS-2020-00031), CIFAR (Canada CIFAR AI Chair and Learning in Machines and Brains Program), and Healthy Brains, Healthy Lives (New Investigator Award: 2b-NISU-8).

- Gaines, B. R. (1967). Stochastic computing, in *Proceedings of the April 18-20, 1967, Spring Joint Computer Conference AFIPS '67 Spring* (New York, NY: Association for Computing Machinery), 149–156. doi: 10.1145/1465482.1465505
- Godfrey, M. D., and Hendry, D. F. (1993). The computer as von Neumann planned it. *IEEE Ann. History Comput.* 15, 11–21. doi: 10.1109/85.194088
- Grier, D. A. (2001). Human computers: the first pioneers of the information age. *Endeavour* 25, 28–32. doi: 10.1016/S0160-9327(00)01338-7
- Habenschuss, S., Jonke, Z., and Maass, W. (2013). Stochastic computations in cortical microcircuit models. *PLoS Comput. Biol.* 9, e1003311. doi: 10.1371/journal.pcbi.1003311
- Hassabis, D., Kumaran, D., Summerfield, C., and Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron* 95, 245–258. doi: 10.1016/j.neuron.2017.06.011
- Hilbert, D. (1902). Mathematical problems. *Bull. Am. Math. Soc.* 8, 437–479. doi: 10.1090/S0002-9904-1902-00923-3
- Hilbert, D., and Ackermann, W. (1999). *Principles of Mathematical Logic*, Vol. 69. Providence, RI: American Mathematical Society.
- Hornik, K. (1991). Approximation capabilities of multilayer feedforward networks. *Neural Netw.* 4, 251–257. doi: 10.1016/0893-6080(91)90009-T
- Hunt, E. (1989). Cognitive science: definition, status, and questions. *Annu. Rev. Psychol.* 40, 603–629. doi: 10.1146/annurev.ps.40.020189.003131
- Jazayeri, M., and Ostojic, S. (2021). Interpreting neural computations by examining intrinsic and embedding dimensionality of neural activity. *Curr. Opin. Neurobiol.* 70, 113–120. doi: 10.1016/j.conb.2021.08.002
- Kwisthout, J., and van Rooij, I. (2020). Computational resource demands of a predictive Bayesian brain. *Comput. Brain Behav.* 3, 174–188. doi: 10.1007/s42113-019-00032-3
- Ladd, T. D., Jelezko, F., Laflamme, R., Nakamura, Y., Monroe, C., and O'Brien, J. L. (2010). Quantum computers. *Nature* 464, 45–53. doi: 10.1038/nature08812
- Lakoff, G., and Johnson, M. (1980). *Metaphors We Live By*. Chicago, IL: University of Chicago Press.
- Maass, W. (2016). Searching for principles of brain computation. *Curr. Opin. Behav. Sci.* 11, 81–92. doi: 10.1016/j.cobeha.2016.06.003
- Maass, W., and Zador, A. M. (1999). Dynamic stochastic synapses as computational units. *Neural Comput.* 11, 903–917. doi: 10.1162/089976699300016494
- Maccormac, E. R. (1986). “Men and machines: the computational metaphor,” in *Philosophy and Technology II* (ed C. Mitcham and A. Huning (Berlin: Springer), 157–170. doi: 10.1007/978-94-009-4512-8\_11
- Marcus, G. (2015). *Opinion - Face It, Your Brain Is a Computer*. New York, NY: The New York Times.
- Mattar, M. G., and Daw, N. D. (2018). Prioritized memory access explains planning and hippocampal replay. *Nat. Neurosci.* 21, 1609–1617. doi: 10.1038/s41593-018-0232-z
- Newell, A. (1980). Physical symbol systems. *Cogn. Sci.* 4, 135–183. doi: 10.1207/s15516709cog0402\_2
- Olshausen, B. A., and Field, D. J. (1996). Natural image statistics and efficient coding. *Netw. Comput. Neural Syst.* 7, 333. doi: 10.1088/0954-898X\_7\_2\_014
- Oteiza, P., Odstrcil, I., Lauder, G., Portugues, R., and Engert, F. (2017). A novel mechanism for mechanosensory-based rheotaxis in larval zebrafish. *Nature* 547, 445–448. doi: 10.1038/nature23014
- Paun, G., Rozenberg, G., and Salomaa, A. (2005). *DNA Computing: New Computing Paradigms*. Berlin: Springer Science & Business Media.
- Penrose, R. (1989). *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*. New York, NY: Oxford University Press. doi: 10.1093/oso/9780198519737.001.0001

- Pylyshyn, Z. W. (1984). *Computation and Cognition: Towards a Foundation for Cognitive Science*. Cambridge, MA: MIT Press.
- Richards, B. A., Lillicrap, T. P., Beaudoin, P., Bengio, Y., Bogacz, R., Christensen, A., et al. (2019). A deep learning framework for neuroscience. *Nat. Neurosci.* 22, 1761–1770. doi: 10.1038/s41593-019-0520-2
- Rubel, L. A. (1993). The extended analog computer. *Adv. Appl. Math.* 14, 39–50. doi: 10.1006/aama.1993.1003
- Rumelhart, D. E., McClelland, J. L., and PDP Research Group (1988). *Parallel Distributed Processing*, Vol.1. Cambridge: MIT Press. doi: 10.1016/B978-1-4832-1446-7.50010-8
- Schuman, C. D., Potok, T. E., Patton, R. M., Birdwell, J. D., Dean, M. E., Rose, G. S., et al. (2017). A survey of neuromorphic computing and neural networks in hardware. *arXiv [Preprint] arXiv:1705.06963*.
- Shastri, B. J., Tait, A. N., Ferreira de Lima, T., Pernice, W. H. P., Bhaskaran, H., Wright, C. D., et al. (2021). Photonics for artificial intelligence and neuromorphic computing. *Nat. Photon.* 15, 102–114. doi: 10.1038/s41566-020-00754-y
- Shukur, H., Zeebaree, S. R., Ahmed, A. J., Zebari, R. R., Ahmed, O., Tahir, B. S. A., et al. (2020). A state of art survey for concurrent computation and clustering of parallel computing for distributed systems. *J. Appl. Sci. Technol. Trends* 1, 148–154. doi: 10.38094/jastt1466
- Siegelmann, H., and Sontag, E. (1995). On the computational power of neural nets. *J. Comput. Syst. Sci.* 50, 132–150. doi: 10.1006/jcss.1995.1013
- Siegelmann, H. T. (1995). Computation beyond the turing limit. *Science* 268, 545–548. doi: 10.1126/science.268.5210.545
- Simon, H. A. (1980). Cognitive science: the newest science of the artificial. *Cogn. Sci.* 4, 33–46. doi: 10.1207/s15516709cog0401\_2
- Smith, C. U. (1993). The use and abuse of metaphors in the history of brain science. *J. History Neurosci.* 2, 283–301. doi: 10.1080/09647049309525577
- Tsividis, Y. (2018). Not your Father's analog computer. *IEEE Spectrum* 55, 38–43. doi: 10.1109/MSPEC.2018.8278135
- Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungsproblem. *J. Math.* 58, 5.
- Turing, A. M. (1937). Computability and  $\lambda$ -definability. *J. Symbol. Logic* 2, 153–163. doi: 10.2307/2268280
- van de Burgt, Y., Melianas, A., Keene, S. T., Malliaras, G., and Salleo, A. (2018). Organic electronics for neuromorphic computing. *Nat. Electron.* 1, 386–397. doi: 10.1038/s41928-018-0103-3
- Van Gelder, T. (1998). The dynamical hypothesis in cognitive science. *Behav. Brain Sci.* 21, 615–628. doi: 10.1017/S0140525X98001733
- Van Noort, D., and Landweber, L. F. (2005). Towards a re-programmable DNA computer. *Nat. Comput.* 4, 163–175. doi: 10.1007/s11047-004-4010-3
- Vlasits, A. (2017). *Tech Metaphors are Holding Back Brain Research*. Wired. (San Francisco, CA). Available online at: <https://www.wired.com/story/tech-metaphors-are-holding-back-brain-research/> (accessed June 12, 2017).
- Von Neumann, J. (1993). First draft of a report on the EDVAC. *IEEE Ann. History Comput.* 15, 27–75. doi: 10.1109/85.238389
- West, D. M., and Travis, L. E. (1991). The computational metaphor and artificial intelligence: a reflective examination of a theoretical falsehood. *AI magazine* 12, 64–64.
- Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Basil Blackwell. Available online at: <https://static1.squarespace.com/static/54889e73e4b0a2c1f9891289/t/564b61a4e4b04eca59c4d232/1447780772744/Ludwig.Wittgenstein.-.Philosophical.Investigations.pdf>

**Conflict of Interest:** TL is employed by DeepMind Inc., and BR works as an external consultant for DeepMind Inc.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Richards and Lillicrap. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.