IS THE BRAIN A DIGITAL COMPUTER? RETHINKING A BINARY QUESTION

Yasemin J. Erden

Is the brain a digital computer? What about your own brain? This article will examine these questions, some possible answers, and what persistent disagreement on the topic might indicate. Along the way we explore the metaphor at the heart of the question and assess how observer relativity features in it. We also reflect on the role of models in scientific endeavour. By the end you should have a sense of why the question matters, what some answers to it might be, and why your preferred answer is highly pertinent to any discussion.

Introduction

Between 2012 and 2015 I gave a series of conference papers to interdisciplinary audiences on the topics of personal and digital identity, artificial intelligence (AI) and metaphor. At the start of the endeavour I naively thought the claims were uncontroversial. I didn't expect the talks to generate significant debate, even if I hoped for discussion. In fact, many of the talks polarised the audience, and debate was sometimes fierce. Though papers kept to time, questions stretched far into breaks. It is significant if a debate can distract tired and hungry participants from the lure of coffee and food. The most passionate of these debates concerned the relation between brains and digital computers. Those papers hinged on Searle's question: *Is the brain a digital computer?* It seemed to me that the metaphor contained in the question was deeply problematic, but it became apparent that others

doi:10.1017/S1477175620000329
Think 57, Vol. 20 (Spring 2021)

were fully committed to it. This article offers an overview of some arguments I offered in those papers, as well as reflections on what the persistent disagreement might show.

I explore these ideas in three sections. First, by dissecting the question and considering metaphors and models. Second, by examining the role of observers in this process. Finally, by exploring what some answers and associated expectations can tell us about language-users. In this way I make two related claims. First, that problems arise when terminology is unthinkingly adopted in discourse on brains, especially as modelled in science and technology. Second, these problems extend beyond the theoretical and academic. I don't discuss the relation between minds and brains in this article. There isn't space, and much good work has been written on that already.

Is the Brain a Digital Computer?

a. Why the Answer Is No: In Searle's 1990 Presidential Address to the American Psychological Association he asked, 'Is the brain a digital computer?' His answer, which was negative, is and was widely disputed (cf. Chrisley 1995 and Chalmers 1996). In this section I examine the metaphor at the heart of the question.

The idea of the brain as a computer builds on a long tradition of *body as machine*, as found in Aristotle and Descartes among others. But what does it mean when we think of a person mechanistically, or when we compare a brain to a computer? For instance, is the point that the brain *is the same as* a digital computer, or is the point that they are *functionally equivalent?* In fact, the metaphor allows for a variety of such meanings, and this is part of the difficulty. In Wittgensteinian terms, we have yet to understand the language-game in which such terms are meaningful. A language-game on Wittgenstein's account brings 'into prominence the fact that the speaking of language is part of an activity, or of a life-form' (Wittgenstein 2001: §23).

To unpick the metaphor, let's consider some of the ways in which it can be meaningful. For instance, we can draw comparisons between computers and brains. Some materialist accounts in philosophy already do this when they show similarities between brain states and processes, and digital computation. There is something compelling about this idea: both brains and computers take in information, which they process, and which leads to outputs, including behaviour. In each there is calculation, assessment, and analysis, as well as the selection and discarding of more or less relevant data. A lot of this happens outside the awareness of the person whose brain or computer it is.

On top of which, modern computers can do a lot of things that humans do. They make use of external data via cameras, rather like our vision; they take information from microphones, like our hearing; we can also speak to them, through programs like Siri or Alexa. Both digital computers and human brains make careful assessments by weighing up what is or isn't useful for the completion of a task, and very clever computers can also examine past efforts so as to adapt and change. In this way, future success in a task becomes more likely. This kind of weak artificial intelligence (AI) offers computation the chance to learn and improve, in a manner not necessarily or obviously different from what humans seem to do. To the person thinking at this point, 'OK, but computers need to be *programmed* to do all this', the response can be, and often is (especially at AI conferences), '*so do humans*'. These are just a few ways in which a brain might be considered akin to a digital computer.

Yet in all of the above, the comparisons are by analogy. It wouldn't be true to say that digital computers do exactly the same thing as our organic brain matter in any of the above states or processes. Calculation as undertaken in neuronal activity is not the same as calculation as found in binary programming. Even highly complex, contemporary computation, like neural nets, which might mimic brain states or behaviours to some extent, relies on numbers not biology. Should this change, then the question that Searle

https://doi.org/10.1017/S1477175620000329 Published online by Cambridge University Press

asked would also need to change, but for the moment it's safe to say that number crunching *is not the same as* neuronal activity *except by comparison*.

What this shows us is that any comparison is made by analogy, and this includes the question that Searle poses above. We can therefore interpret his question to mean *how alike are the brain and a digital computer?* To this we can expand Searle's answer to mean *not nearly enough for the comparison to stand up to careful scrutiny*. In this way we can see the idea of the brain as a digital computer as akin to a metaphor, and what we need to ask next is whether the metaphor is any good. We already know that Searle thinks not, but let's consider why.

We need to start this analysis by first noting that to call something *metaphorical* does not diminish the value it can have, including potential for creativity. Instead it means not taking for granted any hidden values. For instance, we can see the metaphor as creative and enriching on the one hand, and on the other why it might be limiting or even problematic. Interestingly, *computation* is itself metaphorical given that it first described human actions and was only later applied to machines. This offers an indication of the fluency of meaning that comes from creative employment of metaphors.

A metaphor allows words, concepts, and images to stand in place of, or in relation to, other things so as to surprise, explain, or predict, among other aims. Metaphors, and analogies generally, are crucial for understanding and making sense of the world, as well as for expanding thinking. They can also be deceptive. They can narrow important gaps between disparate things, or disguise and oversimplify crucial differences. The metaphor of a mind or personality as akin to a river for example, with associated values of depth and shallowness, is just such an example. Ideas of depth and shallowness play into reductive stereotypes of people as one dimensional. In fact, human brains are more likely to have capacity for both depth and shallowness, especially as brains can adapt in ways that stationary rivers

https://doi.org/10.1017/S1477175620000329 Published online by Cambridge University Press

cannot. We could extrapolate further, but the point should be clear. A metaphor is powerful, but we cannot take for granted the values it offers or diminishes.

Another way to think of the analogy of the brain as digital computer is to think of the comparison in terms of a *model*. To explain this further, let's consider one kind of modelling, namely *simulations.* Simulations are found in weather forecasting, for instance, and they stand in place of events about which we have beliefs or predictions (likely or otherwise). In this way, simulations function in a manner akin to analogies and metaphors, except for one key difference. Our *expectations* of a simulation are different when compared to those of a metaphor. Whereas the former should contain a degree of accuracy and precision, the latter need not. It would not do to predict sunny weather when in fact a storm is coming. Analogies are valued differently. We ask of an analogy: is it helpful or interesting, appropriate or acceptable? Does it show, clarify, or explain something? Conversely, does it make something more confusing? A metaphor is sometimes taken to have little content beyond its literary powers, even though all kinds of analogies, including metaphors, pervade language. They offer ways to understand a world that is sometimes incomprehensible, sometimes overwhelming, and to view new aspects or insights in accessible ways.

Decisions always need to be made in the selection of data for comparison, including for modelling. For example, to model the lifecycle of a butterfly we can identify processes by which an egg is laid, hatches as a caterpillar, and (via the pupa stage) becomes a butterfly. In so doing we identify features we think crucial to the development of this creature. Experience plays a role because this view of the butterfly emerges from sensory perception. There are, however, other ways to model its development. Genetically or hormonally, for instance, as a way to map the division of genes, and other internal changes unobservable to the naked eye. A second model of this kind need not contradict the first.

https://doi.org/10.1017/S1477175620000329 Published online by Cambridge University Press

Other coexisting models may contradict each other, or seem to, yet not be mutually exclusive. Nanomaterials are one such example. Nanomaterials can have properties that differ from the same material at the macroscale. For instance, at the nanoscopic scale (nanoscale) both gold and silver have active properties, though these materials appear inert at ordinary human levels of perception. This difference has not meant we have to completely rethink the models by which we map our world of ordinary perceptual phenomena. Gold as *we generally perceive it* is inert. That the model of gold and silver at the nanoscale differs does not engender a contradiction. Instead, we recognize that antimicrobials, perceptible at one level, imperceptible at another, are part of the complexity of this material. This can of course be said of any material – though some may turn out to be less surprising.

Let's consider one further example: human health. Since this is more personal than either nanoparticles or butterflies, models might seem more accessible or intuitive. But here as well decisions must be made. If we offer a model of human health that focuses on nutrition only, this would be rather limited. If we add to the model information about exercise, it would be more expansive but not yet holistic. What about mental health, or occupational health, both of which impact on a person's ability to function in everyday life? The process by which we determine emphasis and thereby the utility of a model does not end, since any new information or perspective can challenge, support, or enhance the model.

Analogies and models exist in conceptual space connecting more or less disparate ideas or objects. They can be fruitful, but they can also trick us. The idea of a model in particular connotes the scientific, especially as it gestures towards the verifiable, quantifiable, and objective. Those who would reject the idea of a model as a form of metaphor sometimes do so by pointing to the status of the latter as ambiguous, uncertain, and unscientific. Despite this, however, a model is itself analogous. We can even call it a technical metaphor. It is technical because it presents the possibility for testing and experimentation, but it remains analogous because even in this state it remains a process by which one thing *is used in place of another*. Rather like a map.

A map does not claim to be the territory that it represents but stands instead as a model for that which it more or less accurately depicts. It remains dependent on scaling, interpretation, and (political) boundaries and is affected by the selection or omission of data deemed more or less pertinent to a particular need or situation. Models and maps are useful in scientific discourse to aid understanding or prediction, but they can offer false promises about objectivity. The fallout from bias in science and technology, including computation, bear this out.

The process of selection can be less problematic in literary comparisons, such as similes. To *wander lonely as a cloud*, for instance, does not constitute a truth claim about one's passage through space or time, nor of the intrinsic qualities of people or clouds. Yet all comparative depictions, whether in metaphors, models or maps, incorporate a particular account of being in the world, one mediated by symbiotic processes of sensory experience and cognitive processing. A model cannot posit *all* the ways to view an event or idea, in the same way that a map cannot show all that it gestures towards. Instead, models, maps, and metaphors show areas *identified by observers* as important, whether to highlight relations, similarities, or differences. This requires interpretation and judgment, without which map-making and reading, model construction, and metaphor writing, would be impossible and pointless.

b. Observer Relativity: When Searle (1990) denied that the brain is a digital computer, he did so by citing what he describes as an *observer standpoint* with regard to computation. He argued that it is impossible to identify even the most rudimentary ideas of computation without also recognizing those who do the identifying, i.e. us. Were there to be no metaphor-writing, question-posing, interpretative beings who can identify patterns and systems, there could not be any judgements, whether about computation or anything else.

On Searle's account, the identification of a process *as* computational does not mean we identify intrinsic features in physics. Instead he describes it as an 'observer relative characterization'. To recognize that a thing functions *as* something is to use the *as* in just the sort of comparative or metaphorical sense that I describe in the section above. This works in practical terms too, since to use my hand *as a plate* means little more than that I'm eating food either from or above it. In this way the hand functions like a plate in certain circumstances. The comparison between them is relative to my having observed the relation. When I offer the comparison, others can understand what I mean, or at least imaginatively engage with what I might mean. Thus, we avoid multiple, inaccessible, first-person accounts of the world.

The above fits Kant's explanation in the *Critique of Pure Reason*. There he describes the synthetic unity of consciousness as an objective condition of all knowledge. He explains this in terms of lines. In creating a line, the observer is active in the drawing, and in the recognition, of the line. This is a condition for any knowledge we can have of lines generally. Without human experience, there is no knowledge of lines. We can apply this principle to Searle's account of computational states: they are such because of our characterization of them as such. For these reasons, computation is relative to the observer who interprets and explains the physical features that are identified. This applies even where such knowledge is mapped into a model and even where the model allows us to accurately predict or manipulate matter. This is true for computation, for lines, and for particles of matter at the nanoscale. We cannot assume that a model gives us access to *how things stand*, unmediated by the observer who examines and interprets relations.

https://doi.org/10.1017/S1477175620000329 Published online by Cambridge University Press

To recognize the similarities of brain and computational processing requires interpretation, and this relies in turn on how we define terms like computation. To illustrate the idea further, let's consider Kant's ideas about *phenomena* and *noumena*. On his account objects are represented as *objects of experience* via our senses, including any connections between them. These are phenomena. This is not the same as what those objects *might* be apart from our engagement with them, namely *as objects of the pure understanding*. That would be noumena. Objects of pure understanding are unknowable because we simply cannot remove ourselves from the experience of *knowing*. In other words, there is no way in which we can remove ourselves, with our interpretations of things and our understanding of systems, such that we can *know* what things or systems simply *are* in any objective sense of that term. In this way, our perceptions of the world are of phenomena, and this cannot offer an account of noumena, nor guarantee unquestionable accuracy even of the phenomena as we perceive it. Against what would we measure this, after all?

Kant's account demonstrates one way in which it is impossible to identify ideas of computation without locating the identifier. It also has implications for the metaphor of the brain as a digital computer. This means we are now left with the question of why one person believes that a brain is or functions like a digital computer, while someone else does not. Kant suggests that as long as an individual judgement can be given sufficient ground, and so long as we can ensure it is valid for all human reason, we can consider this judgement to be true. In other words, understanding relies on judgements and interpretations that may be agreeable to others. The problem here is that the idea of a brain as a digital computer does not have agreement, but perhaps this too can tell us something. What does it say about a human brain when it has a conception of itself that alienates other human brains? I suggest it tells a story of complexity which cannot be captured by one metaphor alone.

https://doi.org/10.1017/S1477175620000329 Published online by Cambridge University Press

c. Different Language-Games and Seeing As: In his *Philosophical Investigations*, Wittgenstein (2001: pp. 170–1) suggests:

> The concept of 'seeing' makes a tangled impression. Well, it is tangled. – I look at the landscape, my gaze ranges over it, I see all sorts of distinct and indistinct movements; this impresses itself sharply on me, that is quite hazy. After all, how completely ragged what we see can appear! And now look at all that can be meant by 'description of what is seen'. – But this just is what is called description of what is seen. There is not one genuine proper case of such description – the rest being just vague, something which awaits clarification, or which must just be swept aside as rubbish.

A tangle suggests connections and comparisons that can shift and change, without definitively *right* ways to view something. This need not imply that such connections are completely arbitrary. As Wittgenstein (2001) notes elsewhere, we share a common *form of life* and as such, we are apt to make similar sorts of connections and comparisons. Or at the very least many of us are able to understand even those which appear radically different from our own. This allows for shared judgements about meanings such that agreement is possible. Meaningful metaphor use relies on experience-rich contexts being embedded in language-use, such as described by Wittgenstein (2001) in terms of language-games. The notion of a language-game evokes a sense in which understanding language requires involvement, just as we come to understand a game through watching and playing. We can no more view our language from an objective perspective than we can slip out of our skin. Meaningful language-games require not only a successful following of rules, but also the willing of participants to try to understand what is being communicated.

Understanding and agreement do not always come easily, however. Instead, different language-users may view each language-game slightly differently. This is especially likely where the language-game involves subjective or qualitative details, such as those about brains. There are *qualia* or *things it is like* to have a brain, whether those qualia are apparent to each person or not. Just as there are qualia about all kinds of mundane and familiar experiences, even when their mundane quality means they do not regularly feature in conscious experience. Such as, for instance, the taste of water, which may only become apparent when it doesn't taste as it usually does. The qualia of subjective experience represent an unavoidable feature in language, especially as we try to make ourselves understood.

Understanding the metaphor of brains as digital computers therefore requires, at least partly, that we understand the language-user. To do this means accepting some seemingly contradictory claims. First, that words can be meaningful without context, intentionality or the presence of, or origin in, a specific language-user. Otherwise dictionaries wouldn't work, and words would change meaning whenever someone misinterpreted them. Second, that the *expectation* of a context, intention, or speaker is part of the way in which we find meaning in words, including in metaphors. Put another way, to understand what someone *means* can involve more than just paying attention to the meaning of each individually uttered word. The apparent contradiction is resolved when it is accepted that some meaningful language, including metaphors, says as much about expectation and a need for meaning as about the words themselves. Understanding what is said is thus partly about understanding the speaker. If the metaphor at the heart of this article makes sense to someone, then this offers us an insight into *how* they think as well as *what* they think.

Brains are necessarily embodied and embedded, and though our linguistic experiences are shared, it is not a

https://doi.org/10.1017/S1477175620000329 Published online by Cambridge University Press

given that the multiple experiences of having a brain can be easily reduced to one account. This is not to say that shared understanding is impossible, but rather that such understanding needs to take into account the qualitatively different range of experiences. What it is to suffer from memory loss, or even aphasia (where one's ability to understand or produce speech is affected), for instance, are qualitatively unique experiences of people with brains. There are many other examples of cognitive impairment and brain damage to add to this list, and each is rich in qualia.

Part of the issue in answering the question that Searle poses is therefore the simplicity that the question presupposes. To the question *is the brain a digital computer*, the answer might still be no, but to the question, *what similarities are there between brains and computers*, we can offer a more interesting range of answers. So long as we take care not to reduce important complexities and differences. In so doing we can also acknowledge that complexities will matter more to one person, while simple comparisons are fruitful for another. This is akin to recognizing that one model of health works well for one person, but not for another. To understand a brain is partly to make sense of the tools, and the words, that each person uses to explain and describe their brain, including the comparisons that they draw.

Ignoring complexity can and does cause harm to individuals, including by underplaying variety in human thinking, by ignoring diversity in brains, and by setting up unrealistic expectations for apparently predictable patterns in human behaviour. Depression is a pertinent example, especially as it is sometimes treated as an illness with a singular cause, including as an illness of the brain. Alcoholism has suffered from similar reductivism, with associated reductive attempts to cure or to treat those who suffer with it. Yet contemporary research into both depression and alcoholism have shown that simple 'one size fits all' treatments are rarely successful, and that what may be a cure for some can be a trigger for others. Depression can have multiple disparate overlapping triggers, such that it can be impossible to identify any one singular 'cause', and though alcoholism is sometimes treated as a disease of addiction, successful treatment sometimes comes from a shift of focus to treating the circumstances instead of the disease. This can include addressing issues of loneliness or disconnection. People can be harmed when there is a disregard for context in these and other complex illnesses. For instance, in locating an illness in either the body, the mind, or the brain, rather than thinking about the person holistically and as embedded in an environment. The metaphor of the brain as a digital computer lends itself well to such simplification, and belies the fact that machines lack this sort of complexity. Especially if we expect human problems to have the same causal simplicity as mechanical failures.

For these reasons, the question of whether and how the brain might be akin to a digital computer is as relevant now as it ever was, and a failure to challenge it can bring harm. Yet to ignore that the metaphor makes sense for some people is to ignore the strength of its appeal as well as what it says about those to whom it appeals. I suggest that if the debate shows us anything it is that the brain as a digital computer is on the one hand too simple a metaphor for the complexity of human thinking and being. On the other hand, the metaphor is meaningful (at least to some) such that it cannot be dismissed. Especially not as casually as I tried to do some years ago. Conversely, those others cannot assume that a metaphor being meaningful for them means it can or should be applied to all language-users. It is for these reasons that Searle's closed question is so dissatisfying as well as polarizing.

Conclusion

This article has sought to recognize why the metaphor of the brain as a digital computer *can* be illustrative, useful, or even attractive, but it has also noted that such reductive accounts can limit understanding and even expectations about brains and computers. Particularly if we discard those elements that do not fit the model we construct. To illustrate this the article offered examples that show what can be lost if one model is privileged over others. The *brain as digital computer* analogy remains a dominant model, and this has had ramifications, not least for how we understand ourselves. As discussed above, we might consider deficient a model of health which focussed on nutrition alone, yet there are similarly simplistic claims made about illnesses and brains, with associated simplistic 'cures'.

To say that something *can be described as* does not indicate a *failure to be accurate*, but rather shows the limitations of analogies and of models. The fact is that something being described in one way does not preclude it *also* being described in another. In the same way, a person 'can be' described as tall relative to the standard of tallness in one culture, whereas they may be described as of average height according to the standards of another. This article therefore offers a middle way for thinking about the metaphor that calls for shared understanding across language-games and between disparate language-users. To do otherwise is to fail to recognize the limitations inherent in interpretation and understanding, and to leave ourselves open to being seduced by metaphors and models into seeing the world in simplistic ways. Just as people were once tricked by political maps that showed the continent of Africa as significantly smaller than it is. As the diversity of responses to my conference papers about these topics show, there is very little that is simple about humans, their brains, and how they see themselves.

Yasemin J. Erden is based in the Department of Philosophy at the University of Twente. y.j.erden@utwente.nl

https://doi.org/10.1017/S1477175620000329 Published online by Cambridge University Press

References

Chalmers, David J. (1996) 'Does a Rock Implement Every Finite-State Automaton?', *Synthese* 108(3): 309–33.

Chrisley, Ron L. (1995) 'Why Everything Doesn't Realize Every Computation', *Minds and Machines* 4(4): 403–20.

Searle, John R. (1990) 'Is the Brain a Digital Computer?', *Proceedings and Addresses of the American Philosophical Association* 64: 21–37.

Wittgenstein, Ludwig (2001) *Philosophical Investigations*, trans. G. E. M. Anscombe (Oxford: Blackwell).

https://doi.org/10.1017/S1477175620000329 Published online by Cambridge University Press