 Mind & Matter Vol. 18(1), pp. 61–72

The Ontological Impossibility of Digital Consciousness

Riccardo Manzotti
Department of Philosophy
IULM University, Milan, Italy
and
Gregory Owcarz
Department of Philosophy
Syracuse University Strasbourg, France

Abstract

In the field of consciousness studies, a recurrent approach has consisted in explaining consciousness as an emergent property of information or as a special kind of information. The idea is that the central nervous system processes information and that under the right circumstances information is responsible for the emergence of phenomenal experience. Many consider information to be more akin to the mental than to its raw physical underpinnings. If information had such ontological status, it would be conceivable to realize consciousness in digital systems, either by creating artificial consciousness, or by uploading and preserving human consciousness, or both. Unfortunately, this is not a viable possibility since information so construed simply does not exist and thus cannot be a case of consciousness nor be the underpinnings of consciousness. In this paper we will show that information is only an epistemic shortcut to refer to joint probabilities between states of affairs among physical events. If information as an entity beyond those relations is not part of our ontology, then digital consciousness is impossible.

1. Digital Consciousness, Consciousness, and Information

Many thinkers have considered whether consciousness might be the outcome of information and, thus, whether it might be possible to replicate consciousness in a computer. Such an hypothesis, referred to here as digital consciousness, is also a theory of consciousness insofar as it implies that the necessary and sufficient conditions for being conscious are having a certain kind of information processing, the specific kind determined by the particular brand of digital consciousness one endorses.

This idea has been popular with scholars and laymen alike, perhaps for the following reasons. First, information is seen as a level of reality more immaterial than physical processes as such. Thus, many view

information as an intermediate step toward the mental, but a step that appears consistent with a physicalist picture of reality. Among respected fundamental entities, information seems less material than others in the physical world. Computers don't have minds, but they do have information. Or so it seems.

Second, there remains the ever-popular analogy between the brain and a computer, highly suggestive of an analogy between the mind and information (or between software and hardware). This analogy has been strengthened by recent successes in artificial intelligence that have promised, but have yet to achieve, something similar to the human mind.

Third, and less readily admitted, digital consciousness seems to offer a technological means for gaining immortality. If our lives were nothing more than sophisticated software running a simulation on a computer, we could in theory download such software from our biological hardware prone to decay and then upload it to a permanent machine format, as depicted in science-fiction movies like *San Junipero* and television episodes of *Black Mirror*.

In the field of artificial consciousness, the possibilities taken into consideration consist of top-down higher-level modelling of cognitive functions that try to mirror recognized functional structures of the brain, or of bottom-up, more piece-meal, processing efforts. The difference between the two strategies is not always clear. The former seems to support a more traditional behavioral-functional stance, but since there doesn't seem to be empirical evidence that a certain functional structure leads to the emergence of something as unexpected as consciousness, behavioral-functional theorists often adopt an inherently anti-reductionist stance. The latter often posit some fundamental emergent properties that might then be scaled up to provide consciousness to the whole system. However, here too no evidence or explanation is presented as to why that should happen.

The prevailing attitude among scientists and a good many philosophers is no different from that of the layman. Particularly in neuroscience, the tendency to see the brain as a computational device is widespread (Piccinini 2016, 2020, Maley and Piccinini 2014, Chalmers 2011). All such models hinge on a substantive notion of information. By substantive information we mean something that might physically underpin natural phenomena. We will argue that the notion of information in these models has been reified and conclude that, if information is not a substance, such models cannot work.

We need not present a detailed survey of all such approaches; given their similarities, one prominent example will suffice. Consider Giulio Tononi (2004), who has suggested the existence of a special kind of information, dubbed integrated information, or Φ, which by means of some unexplained special power takes on or transforms into something with phenomenal character. What is critical in this theory is that no explanation

whatsoever is offered as to why integrated information should become anything different from and indeed more than information. In other words, even if the kind of information that gives an improved description of what goes on in a brain is integrated information (likewise an empirical claim lacking evidence), why should we assume it produces anything like our experience of the world? The solution offered by supporters of this view is that it is a brute fact of nature that information is like our phenomenal experience. Instead of an explanation we have the mere assumption of either top-down idealism or bottom-up panpsychism.

The criticism we will put forward is much more radical. We question all informational approaches to consciousness by arguing that information as such does not exist. We will show that information is only an epistemic shortcut to refer to conditional probabilities between events. If this is the case, information cannot be used as physical underpinnings of another physical phenomenon such as consciousness, because information is not a real thing with any ontological weight so to speak, but is rather a fictional entity introduced for good practical reasons. Information is more like a center of mass than like electricity. Thus information cannot be used as the foundation for consciousness and no digital version of consciousness is possible.

The outline of the argument is the following:

- Consciousness is a real physical phenomenon though its nature remains uncharted.
- All real physical phenomena must have physical underpinnings.
- Information is not a physical phenomenon.
- Thus, information cannot underpin, produce, or become consciousness.

2. What Is Information?

Is information a physical constituent of reality? There is no reason to assume that it is, regardless of theorists who make that claim (Fresco 2014, Tononi *et al.* 2016, Landauer 1991). The traditional definition of information, as put forward by information theory pioneer Claude Shannon, is not about a physical quantity, but about the probability that a symbol or a sequence of symbols is received accurately at the end of a communication channel (Capurro and Hjøland 2005, Shannon and Weaver 1949, Shannon 1948, Adriaans and van Benthem 2008). It is important to stress that the language we use in describing a physical phenomenon is not neutral to our ontological commitments.

Describing a physical structure as a "communication channel" is already a case of biased terminology that borders on anthropomorphism.

The phrase already assumes there is something that is "communicated" as in transferred from one end of the channel to the other. The word "channel" is also misleading. It metaphorically suggests something like water running through it; and this is the way most people have used it. The layman sees a communication channel as the medium that carries information. According to Wikipedia,

> A channel is used to convey an information signal, for example a digital bit stream from one or several senders (or transmitters) to one or several receivers. A channel has a certain capacity for transmitting information, often measured by its bandwidth in Hz or its data rate in bits per second.

Finally, it is commonly held that information can be stored as a kind of material. This fills the metaphor of water flow to the brim. Not only does the channel transmit information, but such information, just like water, flows at a certain rate from the source to the destination and can be stored in tanks in the form of memory devices.

This is of course only an impressionistic sketch of a more subtle scientific picture of information. It captures some widespread intuitions about the nature of information, but upon reflection, it fails to make much sense. Information is not some stuff that travels from one end to the other end of a channel the way water flows down a pipe. Is there really anything going from the source to the destination as the popular metaphor has it? Is there really anything moving from the sender to the receiver? In contrast with other obvious examples of channels, e.g. electricity or water, in this case there is nothing literally flowing. While certain communication channels require something moving along the channel – traditional mail carriers cart actual envelopes from sender to receiver – channels as understood in modern communications require no necessary transfer of matter.

Suppose I send a printed page in an envelope to a friend of mine in California. My letter, a physical object, a sealed package of matter and energy, will travel from Italy to the US carried in a variety of ways, physically being transferred from here to there. Is there anything additional being transferred, namely information, in this case, or is it just the sheet of paper that is transferred? The argument is that it is only this bit of matter that is carried from my Italian desk to my friend's Californian mailbox. Inside the envelope, there is no information over and above the paper and the ink. Or consider the case of the optical nerve. When light rays hit the retina, our cones and rods release rhodopsin that triggers a cascade of physical events that proceed from the retina towards the cortex. Some of these events are the release of acetylcholine at the synapses and others are the propagation of ions inside axons and dendrites. This is all well-known and described in detail by neurologists. However, the

question remains: Is there anything here besides ion densities and acetylcholine levels? We need not add anything else, unless our aim is merely to create a short-hand way to describe such events. If something physical, either matter or energy, was transferred from A to B, why should we need to add an extra element called information?

Finally, let us put to good use a common toy example. Suppose one had a set of mechanized domino tiles that spring back into their upright position after a fixed time interval. Such a set of dominoes, if properly arranged, may be used as an ideal device to "transmit information". If there were enough of them, they might be distributed in long lines and used to "send signals" between people at a distance. Based on some preliminary agreement, the tumbling dominoes at one end of any such line may accurately indicate what is taking place at the other end. So we may use such mechanized tiles as a standard, if rather slow, information channel.

Clearly such cumbersome devices are formally equivalent to what is routinely done using copper wires, optical fibers, or wireless communications. Yet wires and wireless communications seem a bit mysterious. We cannot see what is going on, and for all but the electronic engineer, naive and misleading notions may arise as to what's inside those wires or what's transmitted across that wireless network. The naive observer might be tempted to believe that there is something – fittingly called information – that is flowing along. The mechanized dominoes, however, take all the mystery away. They reveal a macroscopic picture that leaves no doubt: there is nothing more than dominoes tumbling down and springing back up. Of course, the physical systems that we describe in terms of information do consume energy. But from a purely physical perspective there is nothing in such systems besides matter and energy, least of all a purported form of information. It is just an epistemic tool to address more complex states of affairs in concise useful ways.

3. Debunking The Notion of Information as a Physical Entity

We turn now to current models that suggest some degree of ontological autonomy and provide updated arguments (causal overdetermination, epistemic role, etc.) that eliminate information/computation of any actual role in the physical world. The bottom line will be that consciousness cannot be the result of any digital domain of existence – no consciousness that comes from bits.

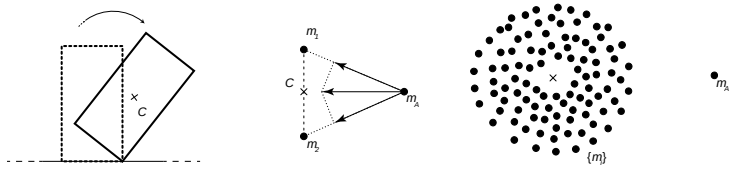
In physics there are many merely fictional entities that are extremely useful. They are conventional constructs whose goal is that of providing efficient summaries of complex states of affairs. Consider the case of the

center of mass of a physical body or set of bodies. It is a very useful notion that has a clear mathematical formulation:

$$\vec{x}_c = \frac{\sum m_i \vec{x}_i}{\sum m_i}, \quad m_c = \sum m_i \tag{1}$$

The center of mass is a position $\vec{x}_c$ defined relative to a system of objects with individual positions $\vec{x}_i$ and individual masses m_i whose total mass is m_c. It is the average position of all the parts of the system, weighted according to their masses. Its significance is given by the fact that, relative to various circumstances, it is as though the entire mass were concentrated in one point. We may feel as though we are attracted by the center of mass of the earth, for instance, but of course this is not the case. We are attracted by every particle of the planet and none of them is literally at the center of the planet. But this is a very useful simplification.

The notion of center of mass in many cases allows us to simplify the description of the physical world significantly. Imagine three common cases:



In all three cases it is much easier to predict what will happen by assuming that all masses are concentrated in a point. All three bodies behave as though all the masses of m_i were concentrated in the center of mass. Yet in the location of the center of mass there is nothing. It is only a computational shortcut, an epistemic invention, to cope with our limited computational resources. Actually, ever since computers have become powerful enough, people have stopped using such tricks and nowadays a gravitational simulation of a galaxy has no need to appeal to the notion of a center of mass. It is always possible to achieve a general solution, just as precise if not more, by computing the gravitational pull of each of the particles.

Rather than using Eq. (1) to compute the gravitational force of a system with mass m_c at the center of mass $\vec{x}_c$ acting upon a body with mass M,

$$\vec{F} = G\frac{Mm_c}{||\vec{x}_c - \vec{x}||},$$

a modern computer will compute

$$\vec{F} = G\sum \frac{Mm_i}{||\vec{x}_i - \vec{x}||}.$$

The notion of center of mass has disappeared, for it has after all never really existed.

Likewise, the notion of information is completely epiphenomenal. The main argument is one of causal overdetermination, for there is no need of anything like information to explain what happens, just as with centers of mass. Once one explains what happens in terms of causality, or by Bayesian probability, everything is explained. There are no additional facts to be given an account.

However, the notion of information as something that goes from A to B is more intuitive and has significant epistemic attraction. As this case shows, the physical structure of the domino set as previously arranged dictates that the probability P to go from X to Y (read "probability that Y given X") is given by

$$P(Y|X) \cong 1$$

But of course there is nothing generated in X that is carried to the other end and finally consumed in Y. This is both common sense and the outcome of more rigorous arguments, such as causal overdetermination that rules out the existence of multiple causally redundant causes of phenomena. Obviously we can apply Ockham's razor here, since introducing the notion of information does not explain any existing fact. On the contrary it introduces an additional entity that itself must be explained. It is a blatant example of the ancient *obscurus per obscurium* fallacy, explaining something difficult or mysterious by means of something even more mysterious.

The actual case is much easier to explain without introducing the notion of information. Due to the structure of the physical system, $P(Y|X) \cong 1$. Is that a common case? Or is it something special that, because it is so rare, has induced people to associate it with a special invisible stuff called information?

In short, the world is such that unless events are in causal proximity (the cue ball hitting the eight), given two random events X and Y, it is highly likely that $P(Y|X) \cong 0$ and that $P(X|Y) \cong 0$, too. This is generally the case across the universe. There are notable exceptions though: biological organisms, nervous systems, and information/computation/communication devices. All such systems have been selected or built in such a way as to maximize $P(Y|X)$ for the relevant coupling of events.

With the development of more complex nervous systems, the causal distance (in terms of causal hops) has also increased in terms of space and time. As X and Y grow farther and farther apart, when the distance between them becomes eminently spatial, causal links are referred to as a "transmission of information" between them. When the more remarkable

dimension is temporal, the phenomenon is called "memory"; and when it is both, as in astronomical distances, people simply tend to be confused.

4. What Does Information Measure?

Information does not measure a kind of "stuff" that is inside neurons and circuits, but it expresses the number of events that, by means of a physical structure, may be rendered into strong probabilistic correlations. In more colloquial terms, it expresses the number of yes/no answers to questions that can be given about the device under examination. So if a device is said to have a capacity of 1024 bytes = 1024 x 8 = 8192 bits, this means that the device may allow at its ends to have 8192 independent pairs of events (X, Y) with $P(Y|X) \cong 1$. Each "so-called" bit is nothing else than a small channel that puts into causal connection events at its ends. So information does not measure something inside the physical world in the same sense in which we measure how many apples a tree has produced or how many molecules are in a rock or how much force two masses exert upon each other by means of gravitational pull. Information is a conventional way to estimate certain portions of the physical world. Information is not inside a physical device, be it a neuron or a circuit, but it estimates what might be done, in the right context, by that physical device. Thus, information cannot be stipulated outside the context.

Consider the traditional definition of information given by Shannon (1948):

$$H = -\sum p_i \log p_i$$

The point is that each of the probabilities p_i is a conditional probability relative to the system and the context in which the system works. Each probability p_i is the conditional probability relative to some previous event. Probabilities are not real physical entities unless one accepts a non-reductionist ontology for dispositions. Probabilities are a clear way to summarize inductive evidence based on historical data. This is a useful epistemic endeavor in many circumstances, though without any ontological ramifications. Probabilities are not an additional physical stuff over and above the physical events. Probabilities are like useful historical centers of masses to refer to historical series.

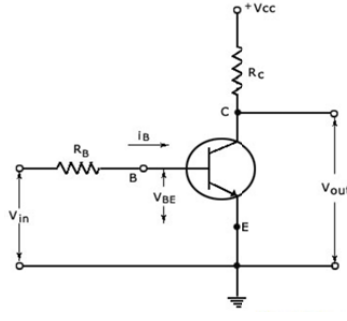
Likewise, H does not measure a physical quantity in a system. It provides an estimate of how much a system's behavior, given certain conventional assumptions, is a consequence of its history. Other notions such as mutual information or even integrated information are just more sophisticated versions of the same (Tononi *et al.* 2016; Peressini 2013, Aaronson 2017). They provide useful mathematical outlines for the degree of correlations between historical series of events (rather than a single series of

events, as in Shannon's original formula). At bottom they do not address anything existing physically.

We can return to our mechanized domino example to clarify the point of the causal overdetermination of information. Consider a row of dominoes, from A to Z, ready to topple into each other from left to right. If the first one falls, the rest will fall in ordered fashion too. Do we need anything beyond the tiles and causation to describe what is happening? If A falls into B and B into C and so on, Z will eventually fall too. Nothing is shuttled from A all the way to Z; it's just a chain of separate events. The point of an information system, or of a communication channel, is only that of maximizing the conditional probability of Z given A. There is nothing more.

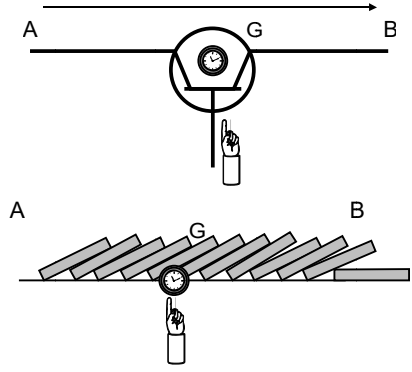
The notion of information is completely causally determined by the falling of the dominoes. Based on the Eleatic principle (Merricks 2001, Kim 2005, 1998, Alexander 1920), we can rule out any entity that does not have any irreducible causal role. Information is in this category and we can conclude that information thus construed does not exist. In fact the dominoes can be used to model all known cases of information devices.

Memory cells, for instance, are nothing but dominoes waiting to fall. When we read them their internal time delay is reset and so we can see whether they are waiting to fall. They are not static containers; they are not boxes of stuff. They are decidedly not "cells of memory storing a bit of information", but rather a channel through which a wave is passing. And the channel has a gate that allows the wave to go through only when one wants it to.



The preceding figure is the classic picture of a transistor, the basic building block of any computer. It has been drawn countless times in this manner for historical reasons. However it is hard to detect its logic, which is actually quite simple. To grasp it, rotate it 90 degrees counterclockwise and compare it side by side with a series of dominoes.

They have just the same structure. There is a causal chain from A to B and a Gate G that allows the causal chain to propagate. There is no



need to add any extra ingredients. Again, there is no information shuttled from A to B. A computer consists only of such transistors. No matter how many transistors there are, they do not contain any extra elements, no extra juice, no spirit, no *elan vital*, and no information. Our computers are nothing but toppling dominoes.

5. Conclusion

Taking information to be a real thing is committing the fallacy William James called "vicious abstractionism" or "vicious intellectualism" in various places, especially to criticize Kant and Hegel for their idealistic philosophies. Perhaps the scholars who take information as if it were real are the Kants and Hegels of today. They bamboozle us into believing in the physical existence of something abstract. Here are the words of James (1909, pp. 135f, italics added):

> We conceive a concrete situation by singling out some salient or important feature in it, and classing it under that; then, instead of adding to its previous characters all the positive consequences which the new way of conceiving it may bring, we proceed to use our concept privately; reducing the originally rich phenomenon to the naked suggestions of that name abstractly taken [...] and acting as if all the other characters from out of which the concept is abstracted were expunged. *Abstraction, functioning in this way, becomes a means of arrest far more than a means of advance in thought.*

In best-case scenarios, scientists adopt a construct as a hypothetical explanatory variable that is not directly observable. The concept of center of mass in physics is such a non-directly observable construct. As we have seen the analogy with information is very strong. By and large, the degree to which a construct is useful and accepted as part of the current

paradigm in a scientific community depends on empirical research that has demonstrated that a scientific construct has validity. Yet this does not justify adding it to the building blocks of the physical world. Constructs are not made of building blocks.

Information is not physical because:

- Information is neither measured empirically nor observed (there are no information-scopes or information-meters).
- Information is estimated, computed, stipulated, but never measured, because it depends on assumptions about the role of a physical system.
- Information is epiphenomenal.
- Information is causally redundant (thus suffering from causal overdetermination).

Therefore, there are no reasons to take information any more seriously than we do our centers of mass. If information is not a real physical phenomenon, there is no hope to use it as a foundation for mental states, and thus there is no chance a computer will ever show any genuine digital consciousness.

References

Aaronson S. (2017): Why I am not an integrated information theorist (or, the unconscious expander). *Shtetl Optimized. The Blog of Scott Aronson*, accessible at `www.scottaaronson.com/blog/?p=1799`, with a commentary by D. Chalmers at `www.scottaaronson.com/blog/?p=1799#comment-107443`.

Adriaans P. and van Benthem J., eds. (2008): *Philosophy of Information*, Elsevier, Amsterdam.

Alexander S. (1920): *Space, Time and Deity*, MacMillan, London.

Capurro R. and Hjørland B. (2005): The concept of information. *Annual Review of Information Science and Technology* **37**, 343–411.

Chalmers D.J. (2011): A computational foundation for the study of cognition. *Journal of Cognitive Science* **12**, 322–357.

Fresco N. (2014): *Physical Computation and Cognitive Science*, Springer, Dordrecht.

James W. (1909): *The Meaning of Truth*, Harvard University Press, Harvard.

Kim J. (1998): *Mind in a Physical World*, MIT Press, Cambridge.

Kim J. (2005): *Physicalism, or Something Near Enough*, Princeton University Press, Princeton.

Landauer R. (1991): Information is physical. *Physics Today* **44**(5), 23.

Maley C. and Piccinini G. (2014): Neural representation and computation. In *Handbook of Neuroethics*, ed. by J. Clausen and N. Levy, Springer, Berlin, pp. 79–94.

Merricks T. (2001): *Objects and Persons*, Clarendon Press, Oxford.

Peressini A. (2013): Consciousness as integrated information. *Journal of Consciousness Studies* **20**(12), 180–206.

Piccinini G. (2016): The computational theory of cognition. In *Fundamental Issues of Artificial Intelligence*, ed. by V.C. Muller, Springer, New York, pp. 203–222.

Piccinini G. (2020): The myth of mind uploading. In *The Mind-Technology Problem – Investigating Minds, Selves and 21st Century Artefacts*, ed. by R. Clowes, K. Gartner and I. Hipolito, in press. Accessible at `www.academia.edu/36762354/The_Myth_of_Mind_Uploading`

Shannon C.E. (1948): A mathematical theory of communication. *Bell System Technical Journal* **27**, 379–423, 623–656.

Shannon C.E. and Weaver W. (1949): *The Mathematical Theory of Communication*, University of Illinois Press, Urbana.

Tononi G. (2004): An information integration theory of consciousness. *BMC Neuroscience* **5**, 1–22.

Tononi G., Boly M., Massimini M., and Koch C. (2016): Integrated information theory: From consciousness to its physical substrate. *Nature Reviews Neuroscience* **17**, 450–461.

Received: 19 May 2020
Accepted: 25 May 2020

Reviewed by Antonio Chella and Teed Rockwell