Minds & Machines (2012) 22:183–190
DOI 10.1007/s11023-012-9271-5

# Computationalism: Still the Only Game in Town

## A Reply to Swiatczak's "Conscious Representations: An Intractable Problem for the Computational Theory of Mind"

**David Davenport**

Received: 27 July 2011 / Accepted: 4 February 2012 / Published online: 18 February 2012


**Abstract** Mental representations, Swiatczak (Minds Mach 21:19–32, 2011) argues, are fundamentally biochemical and their operations depend on consciousness; hence the computational theory of mind, based as it is on multiple realisability and purely syntactic operations, must be wrong. Swiatczak, however, is mistaken. Computation, properly understood, can afford descriptions/explanations of any physical process, and since Swiatczak accepts that consciousness has a physical basis, his argument against computationalism must fail. Of course, we may not have much idea how consciousness (itself a rather unclear plurality of notions) might be implemented, but we do have a hypothesis—that all of our mental life, including consciousness, is the result of computational processes and so not tied to a biochemical substrate. Like it or not, the computational theory of mind remains the only game in town.

**Keywords** Computational theory of mind · Computationalism · Consciousness · Computation · Mental representation

## Introduction

Advances in neuroscience have led to a resurgence in articles that claim computationalism is wrong. Swiatczak (2011) is just the latest in a long line of such articles that goes all the way back to Dreyfus (1972) "What computers can't do" and include Searle's infamous Chinese room argument (Searle 1980). A more recent example from Minds and Machines is Bishop (2009) "Why computers can't feel pain". Such papers attempt to falsify computationalism by arguing that some aspect of our mental life—awareness, emotions, feelings, pain, understanding,

D. Davenport (✉)
Computer Engineering Department, Bilkent University, 06800 Ankara, Turkey
e-mail: david@bilkent.edu.tr

etc.—is impossible for machines (computers) and so the computational theory of mind (CTM) must be fundamentally mistaken. In the following paragraphs I take issue with some of the points raised in Swiatczak's article and try to show why it too fails to make the case against computationalism, why—to (mis)quote Fodor (1975)—the computational theory of mind is still "…the only game in town".

Swiatczak bases his argument against CTM on biochemical and neuropsychological evidence, which claims to show that humans employ two kinds of representation—mental and non-mental—and that the activity of mental representations crucially depends on consciousness. He argues that this creates an insurmountable problem for a theory based on multiple realisability and purely syntactic operations over structured representations; "…this project (CTM) cannot succeed because mental representations do not resemble sentences and their operations do not depend on their syntactic properties. Rather, mental representations have biochemical features and their operations depend on, among other things, consciousness" (Swiatczak 2011, p 20).

Before looking at the details of Swiatczak's argument, it is perhaps appropriate to say a few words about the notion of consciousness. Indeed, a moment's thought should suffice to realise it is not one notion, but many. The Stanford encyclopaedia of philosophy entry on Consciousness (Van Gulick 2011) provides an excellent overview of the voluminous research in this area. For the purposes of this paper only three forms are of concern (1) creature consciousness, (2) functional mental-state consciousness, and (3) subjective mental-state consciousness. The first, creature consciousness, essentially equates to the creature (agent) being "alive", aware of and responsive to its environment. Mental-state consciousness, on the other hand, relates to the creature's internal (representational) states. The functional form is concerned with how such states can be intentional representations, that is, how they can relate to things in the world and acquire meaning, and so guide the agent's behaviour. This is contrasted with the subjective form of mental-state consciousness (what Chalmers (1995a) called the "hard problem"), which is concerned with how (physical) representations can give rise to the qualitative "feeling" associated with sensations. These two forms of mental-state consciousness are also known as (access) A-consciousness and (phenomenal) P-consciousness (Block 1995). With Swiatczak, we leave phenomenal consciousness for another day and concentrate here only on the "easier" problems.

## Understanding Consciousness

The evidence Swiatczak presents for the two kinds of representation, what he calls mental and non-mental, is not particularly new. It comes from the pioneering work of Nobel laureate Eric Kandel and his team at Columbia University (summarized in Kandel 2001), and from Larry Squire and his colleagues at the University of California (see in particular, the review by Clark et al 2002.) Both neuroscience and psychology have long distinguished between declarative memory (that allows for the recollection of facts and events) and procedural memory (that supports performance skills). Procedural memory is known to depend on the cerebellum and

the brainstem, while declarative memory also requires the hippocampus and neocortex.

The experiments Swiatczak recounts clearly show that being conscious of (aware of) sensory input, makes it more likely that memory of it is retained[1] (something most teachers know only too well). Details of the methodology and interpretation of the results need not concern us here, but there are two points regarding the experiments that bear emphasizing. First, awareness seemingly proves necessary in trace conditioning only when the input stimuli are separated over periods greater than the normal maximum response period of the cerebellum (less than a few hundred milliseconds), beyond which the hippocampus and neocortex—the presumed seat of consciousness—generally need to be involved. Second, the method used to determine whether subjects were "conscious" of the relevant stimuli varied considerably. For example, in some cases human subjects were surveyed after the experiment was over, to determine whether they had realised there was a relationship between the conditioned and unconditioned stimuli, while in other cases attempts were made to try to determine this through questioning during the experiment, and in yet others alternative conditioned stimuli were used. In the case of animals, such as the mice in Kandel's experiments, completely different strategies had to be adopted.

The fact that we have no clear-cut notion of what consciousness is and the obvious difficulty of establishing whether a subject is actually in the appropriate state, surely makes it unsuitable as a means to distinguish between the two kinds of representation, as Swiatczak suggests we do. It might be much better to say simply that certain representations, those Swiatczak refers to as mental representations, involve additional areas of the brain as compared with non-mental representations. Obviously this is not as attractive as referring to them as "conscious representations" (which may be why Swiatczak doesn't do so), though it does make it clear that they may have a natural causal explanation.

That mental representations would appear to have a causal explanation is important. While on occasion Swiatczak seems to view consciousness as something almost dualist in nature, this is evidently not his intention.

> …Does this mean that consciousness is a ghostlike force with the power to influence the effectiveness of synaptic transmission between neurons? The answer is no. There is no gap in the biochemical activities of neurons that could be filled by supernatural forces. Whatever the mechanism, it must involve natural powers. One possible hypothesis is that weakly emergent properties of the human nervous system have the power to influence the activity of the neurons and molecules engaged in trace conditioning (Swiatczak 2011).

Swiatczak spends some time discussing the ontological status of representations. He decides they must be concrete entities since they clearly have causal powers and "it is difficult to understand how something unreal and non-causal could influence behaviour." He argues that, "[if t]he molecular and cellular mechanisms of implicit

[1] Notice, that, in effect, creature consciousness is affecting mental-state consciousness.

memory storage are preserved across evolution …what is so special about human memory that the representations it involves are part of this person's mental life? What is it that makes representations of the same kind mental in one species and non-mental in another? It must be external to those representations. … [the] external factor that determines the status of representations is consciousness." He goes on to say that "[w]hat appears to make a difference between an ordinary representation and a mental representation is the fact that one's awareness can causally influence the latter without being able to change the former. Therefore, mental representations appear to be those that can be changed by consciousness." Squire's results "lead to the conclusion that consciousness has the power to change the physical properties of neural representations."

Having accepted that consciousness has a natural physical (non-dualist) basis and involves additional regions of the brain (those in the hippocampus and neocortex), Swiatczak's claim that "conscious" mental representations are inexplicable from the computationalist perspective, now appears unfounded. It comes down to an acknowledgment that one physical "thing" can affect another and so change the causal behaviour of the system. This, of course, is unsurprising. The details are irrelevant to the argument, but might involve influences due to top-down mechanisms commonly assumed to operate between different brain modules/layers.

Does Swiatczak imply something more, something related to the way in which the change is achieved? Kandel's work shows neuronal learning in considerable detail from the short to long-term. Part of this mechanism appears to invoke genes which literally create new synaptic connections. Perhaps it is this plasticity, this rewiring of the brain, to which he is referring. But again, from the computational and, of course, the agent's perspective, all that matters is the outcome—the resulting causal behaviour—not how it is achieved. The biochemistry is very clever, but relevant only insofar as it enables the biological agent to achieve its goals. Perhaps there are other solutions using different biochemical substances that could be substituted and still have the same effect. Indeed, it is quite possible that Nature has solved the same problems in different ways in other species (and even other parts of our brains). The real question is whether the nature of consciousness would change if the specific mechanism/chemistry changed? There is no evidence to suggest that it might.

The human nervous system is currently the only example we would all agree is actually conscious. But this doesn't mean consciousness has to be fundamentally biological or has to be limited to humans. Clearly, not everything is conscious. We may well agree that vending machines and laptop computers (currently) aren't—though we often treat them as if they were and would probably welcome more intelligence in them too. We would also probably accept that the giant sea snail (Aplysia—used in Kandel's experiments) isn't conscious, though this may be a more arbitrary decision. But what of dogs and cats and dolphins; how can we decide their status in the consciousness stakes? And, by the same token, were we to encounter some non-carbon based alien intelligence, apparently equal in every way to our own, wouldn't we have to ascribe consciousness to them? And what of the increasingly sophisticated robots now being built? Indeed, on what basis do we assign consciousness to other human beings (or even to ourselves)?

If it is not to be a purely arbitrary distinction, then we must look at how and why consciousness arises. The problem of what constitutes consciousness cannot be answered by assuming that only humans are conscious and so begging the question as Swiatczak seems to do. It is also important to note that consciousness is almost certainly not an all-or-nothing affair. It depends, amongst other things, on the overall architecture of the agent: the sensory inputs it has, the range of actions available to it, and the extent and nature of its memory. Some creatures lack appropriate long-term memory functions and so can only live "in the moment", yet can be aware of their environment in a way (or at least to a degree) that vending machines cannot. Domesticated dogs and cats clearly have a more developed cognitive ability, and presumably a greater degree of consciousness, itself the result of longer-term memory functions, yet still not on a par with the self-awareness found in humans. And, of course, even those agents that can be considered conscious, are not always so. They may be asleep, or unconscious ("knocked out"), or in a coma. Sometimes, they may be unaware of their environment, yet fully conscious, attending to some internal mental puzzle.[2]

## Understanding Computation

So far, so good then; there is little here to upset the computationalist apple cart. We have only a limited idea what consciousness is and so try to explain it in terms of things we do understand. Indeed, the CTM is the hypothesis that all of our mental life, including consciousness, somehow results from primitive representations of the sort produced when certain computations are implemented. It is still an open question, but as a research program it appears coherent and has made considerable progress. Surprisingly, though, in trying to spell out the relation between mind and machine, it has proved necessary to clarify concepts previously thought to be—reasonably—clear; notions such as symbol, representation and computation itself (not to mention semantics, information, truth, attention, goals, ethics, free will and determinism).

Swiatczak takes Jerry Fodor's statement that "cognitive mental processes are operations defined on syntactically structured mental representations that are much like sentences" (Fodor 1975), as the basis of the computationalist theory of mind. Swiatczak repeatedly states, but never seems to argue or provide evidence for the brain not having sentence-like structures. The only indication Swiatczak does provide is that, "mental representations have biochemical features and their operations depend on, among other things, consciousness", which is no help at all.

Does the brain have sentence-like representational structures? Yes (and no!) Clearly, DNA is such a structure and is found in the brain, but this is the wrong "level". The representations DNA is concerned with relate to the construction of the agent, including the basic wiring of the brain. Brain function, however, is related to sensory inputs, and it is representations at this cognitive level that we are concerned

[2] Such distinctions correspond roughly to the transitive (aware-of) and intransitive (awake) versions of creature consciousness.

with. Fodor points out that at some level of description there must be representations that equate to our (sentence-like) linguistic utterances, the Mentalese terms he sees as being part of our language of thought (LOT). Presumably Swiatczak is implying that no such representations have been, or will be, found, and if he expects something similar to the DNA case, i.e. chemicals or molecules being quite literally replicated and glued together whenever the agent thinks, then he is probably correct. But Fodor is correct too. The problem lies in our understanding of computation and the ways in which it can be implemented.

Although as yet no consensus exists, there is movement towards the view that a computation is "an abstract specification for a causal structure" (Chalmers 1995b) and that its purpose is prediction. Any physical state that can causally affect the evolving sequence of system states might be taken to be a representation. These can be combined in many different ways, concatenation into sentence-like structures being just one form; linking, as in neural network-like structures, being another (Davenport 2012). Representations are created and gain meaning (for the agent itself) as a result of interaction with the environment (Bickhard and Terveen 1995; Davenport 2000). Through sequences of trial and error interactions, such representations form multiple, increasingly correct and detailed models of the environment, at least sufficient for the agent's purposes (usually survival related). In order to achieve the longer-term predictive capacities needed for human-level performance, sensor input must eventually be decoupled from the environment. This implies that some representations within the system (agent) will be close to, and vary purely in accord with, sensory input, whilst others will be completely decoupled and so available to perform "what-if" simulations. That representations form some sort of hierarchy in which both bottom-up and top-down processing takes place is also well-known. There is an obvious similarity here, with Swiatczak's two kinds of representation—non-mental and mental—and the role he sees consciousness playing.

On this broader understanding of computation, the brain clearly is a computer, albeit one that has an architecture and implementation quite different from that of the familiar desktop machine. That our mental life results from computations (the operations of a computer) thus still seems a perfectly reasonable hypothesis.

## Concluding Remarks

According to Swiatczak, experimental evidence suggests that awareness can directly affect the causal properties of some representations, a finding that seems to conflict with a Computationist vision where only syntax matters. But this is misleading, at best. For one thing, consciousness is an outcome of the purely physical substrate; it does not "cause" anything as such, rather it is a way to talk about certain patterns of activity in a more abstract manner (in the same way that chemistry and biology abstract from physics). That consciousness has a non-magical, physical explanation and that, in the relevant experiments, additional regions of the brain were involved, suggests that they might somehow be responsible for whatever representational changes did occur. Moreover, given a newer, broader conception of computation

that sees it as a way of specifying causal structures—suitable for modelling and prediction—there really is no conflict at all. The computational theory of mind is the hypothesis that mental life can be explained in computational terms. It has already proved quite successful, providing insights into creature consciousness and functional mental-state content consciousness and, so far, there is no reason to think that it will not also help with the hard problem of phenomenal mental-state consciousness. At the very least, there is currently no real alternative.

Cognitive science, computer science and neuroscience, are still comparatively young and developing fields. They each have their own literature and terminologies, which often makes it difficult for practitioners trained in one field to interpret and integrate results from other areas, especially more established disciplines like philosophy. Reexamining foundational beliefs into, for example, the nature of computation (Denning 2007), information, and truth (Floridi 2009), is sparking new and exciting discussions concerning our own nature and place in the universe. Some of these are highly controversial, in particular the idea that we humans are "machines" and therefore not as unique as we have long presumed ourselves (Floridi 2010; Demir 2010). That all the biochemicals come together to form a living being is certainly astounding. That they become self-aware and ultimately appear able to unravel their own inner mechanisms, seems little short of a miracle. But that those mechanisms are computational now appears certain. Like it or not, computationalism, *properly understood*, really still is (and will remain)… the only game in town.

## References


Bickhard, M. H., & Terveen, L. (1995). *Foundational issues in artificial intelligence and cognitive science: Impasse and solution*. Amsterdam: Elsevier.

Bishop, M. (2009). Why computers can't feel pain. *Minds and Machines, 19*, 507–516. doi: 10.1007/s11023-009-9173-3.

Block, N. (1995). On a confusion about the function of consciousness. *Behavioral and Brain Sciences, 18*, 227–247.

Chalmers, D. (1995a). Facing up to the problem of consciousness. *Journal of Consciousness Studies, 2*, 200–219.

Chalmers, D. (1995b). On implementing a computation. *Minds and Machines, 4*, 391–402.

Clark, R. E., Manns, J. R., & Squire, L. R. (2002). Classical conditioning, awareness, and brain systems. *Trends in Cognitive Sciences, 6*(12), 524–531.

Davenport, D. (2000). Computationalism: The very idea. *Conceptus studien 14*, Fall.

Davenport, D. (2012). The two (computational) faces of AI. In V. M. Müller (Ed.), *Theory and Philosophy of Artificial Intelligence*. SAPERE, Berlin: Springer.

Demir, H. (2010). The fourth revolution: Philosophical foundations and technological implications. *Knowledge Technology Policy, 23*(1–2), 1–6.

Denning, P. J. (2007) Computing is a natural science. *Communications of the ACM*, *50*(7), 13. http://portal.acm.org/citation.cfm?.

Dreyfus, H. (1972). *What computers can't do*. New York: Harper and Row.

Floridi, L. (2009). Outline of a theory of truth as correctness for semantic information. *Erkenntnis, 74*(2), 147–175.

Floridi, L. (2010) The digital revolution as a fourth revolution. http://www.philosophyofinformation.net/massmedia/pdf/bbc-1.pdf

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

Kandel, E. R. (2001). The molecular biology of memory storage: A dialogue between genes and synapses. *Science, 294*(5544), 1030–1038. doi:10.1126/science.1067020.

Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(03), 417–424. doi: 10.1017/S0140525X00005756.

Swiatczak, B. (2011) Conscious representations: An intractable problem for the computational theory of mind. *Minds and Machines*, *21*, 19–32. http://dx.doi.org/10.1007/s11023-010-9214-y. doi: 10.1007/s11023-010-9214-y.

Van Gulick, R. (2011) Consciousness. In Zalta E. N. (Ed.), *The Stanford Encyclopedia of Philosophy* (summer 2011 edn).