International Journal of Machine Consciousness
Vol. 4, No. 2 (2012) 401–420

DOI: 10.1142/S1793843012400239

THE COMPUTATIONAL STANCE IS UNFIT FOR CONSCIOUSNESS

RICCARDO MANZOTTI

Institute of Consumption, Communication and Behavior,
IULM University, Via Carlo Bo, 8, 16033 Milano, Italy
riccardo.manzotti@iulm.it

It is customary to assume that agents receive information from the environment through their sensors. It is equally customary to assume that an agent is capable of information processing and thus of computation. These two assumptions may be misleading, particularly because so much basic theoretical work relies on the concepts of information and computation. In similarity with Dennett's intentional stance, I suggest that a lot of discussions in cognitive science, neuroscience and artificial intelligence is biased by a naïve notion of computation resulting from the adoption of a computational stance. As a case study, I will focus on David Chalmers' view of computation in cognitive agents. In particular, I will challenge the thesis of computational sufficiency. I will argue that computation is no more than the ascription of an abstract model to a series of states and dynamic transitions in a physical agent. As a result, computation is akin to center of masses and other epistemic shortcuts that are insufficient to be the underpinnings of a baffling-yet-physical phenomenon like consciousness.

Keywords: Information; computation; agents; consciousness; AI.

1. The Computational Stance

In computer science and artificial intelligence (AI), it is often held that an agent may collect, acquire, store, transmit, generate, and compute information. Although the notion is vague and frequently used with totally different meanings [Floridi, 2009], there is a widespread belief that information is somehow real and thus it may be used to generate further properties in agents — for instance consciousness. An example is offered by Giulio Tononi's theory of information integration that suggests that consciousness is the expression of a property of information [Tononi, 2004]. In a different yet conceptually similar way, Chalmers [1996] suggested that information has a dual aspect — quantitative and phenomenal.

A notion like information is, of course, useful. Yet it may be misleading to take information to be similar to phenomena like electricity, causation, or digestion. Most of the arguments presented here have been already put forward in different contexts (for instance, consider Searle [1980, 1983]; Brooks [1991]; Chrisley [1994, 2003]; Harnad [1994]; Chrisley and Ziemke [2002]; Fresco [2012] and Shagrir [2012]) and yet it is surprising to observe that in many fields like cognitive sciences, consciousness studies, AI and neuroscience, something akin to the old notion of information keeps resurrecting.

To understand the implications of this paper, it is useful to place it in the context of other related studies. The fact is that there is some vagueness as to the boundaries between consciousness and cognition [Baars, 1988; Adams, 2003; Adams and Aizawa, 2008; Shanahan, 2010; Fekete and Edelman, 2011]. As a result, most strong theses as to the foundations of cognition are taken to be a foundation for consciousness as well. This is particularly true in the case of information. However, the notion of information may be unfit to support a phenomenon like consciousness. Information and consciousness may belong to totally different ontological domains. In particular "information" may not refer to a physical entity but rather may be a fictitious entity introduced as a result of an epistemic (as opposed to ontological) strategy, and thus may not refer to anything at all [Manzotti, 2011c]. Information may be the result of a particularly successful descriptive strategy, which I will dub the computational stance. In this respect, as I will argue in the following, information would not be a natural entity — something on which one may build further entities like consciousness — but an epistemic entity — something useful to describe reality.

A useful analogy to highlight the shortcomings of information is intentionality. I will skip here the obvious connections between the two concepts. Intentionality has been considered either as something intrinsic in agents or as an epistemic notion. The latter option has been championed by Dennett. Consider Dennett's [1987] introduction of the intentional stance:

> "*I will describe the strategy, which I call the intentional strategy or adopting the intentional stance. To a first approximation, the intentional strategy consists of treating the object whose behavior you want to predict as a rational agent with beliefs and desires and other mental states exhibiting what Brentano and others call intentionality.*"

Dennett strives to debunk the notion of intentionality. According to him, intentionality is not intrinsic to agents. Intentionality is not *really* real. It is a way to describe what an agent does. Dennett's intentional stance shifts the discussion from an ontological/physical/biological level to an epistemic one. Is it not it possible that information is akin to Dennett's intentionality? After all, so far, it has been very difficult to single out what information is in physical terms. In this regard, Floridi [2004] complains that

> "[*What is information?*] *is the hardest and most central question in* [*philosophy of information*]. *Information is still an elusive concept. This is a scandal not by itself, but because so much basic theoretical work relies on a clear analysis and explanation of information and of its cognate concepts.*"

It is precisely because it is unclear what information is that here it is worth questioning its usefulness in providing a constitutive level for consciousness in agents. As to the nature of information, Floridi [2004] provides three possible cases:

> "...*information as reality* (*e.g.*, *as patterns of physical signals*, *which are neither true nor false*), *also known as ecological information*; *information about reality* (*semantic information*, *alethically qualifiable*); *and information for reality* (*instruction*, *like genetic information*)."

It is obvious that all three cases are relevant for AI and indeed match with different approaches to cognition. However, since information is a weak concept loosely defined and apt to provoke endless counterarguments, in the field of cognitive science and AI, information has often been linked to the notion of computation. In part because of the influx of computer science and in part because computation seems a more autonomous notion, many authors have tried to take advantage of it. In fact, in this paper I will frequently swing back and forth between information and computation since the two concepts are deeply interwoven and connected together.

The point I will defend here is that information and computation are the result of adopting the computational stance. This is similar to the thesis that intentionality is the result of adopting Dennett's intentional stance. Information and computation do not correspond to a physical phenomenon but to a way of describing certain physical phenomena in relation to an agent's attitudes and behaviors. Taking advantage of Dennett's terminology, to a first approximation, I suggest that the computational stance consists of treating the object whose behavior you want to predict as a series of states and dynamic transitions isomorphic to an abstract model of computation. One deals with the object states and transitions *as if* they were a computation, since one may trust its behavior to a sufficient extent. In short, I will defend the view that computations are not real like mass or electric charge and thus they cannot be responsible for the occurrence of something like consciousness or cognition that I take to be baffling-yet-real-physical phenomena. No physical phenomena may arise from purely epistemic stances.

Of course, it may be objected that if computations are real enough to get rockets to the moon or process text, then why cannot they be real enough for cognition or consciousness? There is a difference between computations and mass that holds both for the realist and for the anti-realist. So far, mass is primary [Bitbol, 2008]. This is not to say that we are sure that mass is real. In the future, we may discover that mass is *really* something else. However, whatever mass will be, something will always play that role that now we assign to mass. On the other hand, computation is not primary. We can do without computations when we describe the events taking place in a brain or in a computer. Even in an anti-realist description of the universe, computations and mass occupy a different position. Mass comes first.

As an ideal case study for the purpose at hand, I will focus on Chalmers' proposal to use computation as a foundation for information and, in turn, to link consciousness to a certain informational structure. My aim is to show that such a strategy is faulty because the suggested notions exist only in the eye of the beholder. They are the result of adopting the computational stance for agent-related physical processes. Furthermore, I will argue that the computational stance is often adopted because many authors consider the mind as something internal to the agent's cognitive machinery.

2. Chalmers' Notion of Computation and Cognition

Chalmers [1993/2012] puts forward a computational foundation for cognition. In order to achieve this goal, he strives to flesh out a theory of computation that singles out computation as a natural phenomenon (here I refer to natural phenomena as something opposed to either artificial or fictitious phenomena like centers of mass, country borders, or meridians). A center of mass does not exist. Its mode of existence is that of a handy shortcut in our mathematical formulas. He suggests that "a system implements a computation if the causal structure of the system mirrors the formal structure of the computation" [Chalmers, 1993/2012]. In turn, computation is suggested to be the foundation for *mentality*, which is taken to be "rooted in such [causal] patterns" [Chalmers, 2010]. Furthermore, he maintains that there are invariant causal structures that, if implemented, are instances of computation. Chalmers' crucial claim is that such instances of computation are sufficient for the occurrence of the mind (or some aspect of it) — a claim he subsequently backs up with the theory of the double aspect of information [Chalmers, 1996]. Eventually, he provides a series of counter-arguments against well-known objections to computation. Since Chalmers is definitely not an eliminativist as to the mind [Chalmers, 1996], it follows that computation has to be a natural kind in order to be able to be the foundation of mentality. Chalmers [1993/2012] claims that:

> "*Computation is central to the foundations of modern cognitive science, but its role is controversial. Questions about computation abound*: *What is it for a physical system to implement a computation? Is computation sufficient for thought? What is the role of computation in a theory of cognition?* [...] *I give* [*an analysis of implementation*], *based on the idea that a system implements a computation if the causal structure of the system mirrors the formal structure of the computation. This account can be used to justify the central commitments of artificial intelligence and computational cognitive science*: *the thesis of computational sufficiency, which holds that the right kind of computational structure suffices for the possession of a mind* [...]."

Yet, are his arguments convincing? Why should computation be taken as a natural phenomenon? Pace Chalmers, I suspect that the answer is likely to be negative. I will try to show that:

(a) A computation arises from the descriptive strategy applied to certain phenomena.
(b) A computation is not a natural kind.

A curious result of this line of argument is that a computer is not really performing computation, no more than a comic character is really saying something. A computer is a huge collection of transistors and electric gates. From a physical standpoint, a computer is an electric machine — it is not doing computations, it is transferring electricity from one place to another and dissipating a huge amount of it. Computations and information are not inside the computer. They are a way to describe what the computer is doing as a result of the adoption of the computational stance. The same argument holds for speaking of "gates" and "transistors". We may describe the physical object named "computer" in many different ways. Our different descriptive strategies do not bring into existence different entities though.

I will maintain that Chalmers' paper shows that computation is nothing but a way to describe actual causal relations. Of course, if the argument presented in this paper has any merit, it follows that computation cannot be an ontological foundation for mentality. As a result the thesis of computational sufficiency cannot possibly hold.

Does Chalmers' view offer any support for considering computation anything but a useful way to describe causal processes? The notion of computation is epistemically useful, but it may refer to nothing more than causation. It must be stressed that the view advanced here is compatible both with the existence of computing devices and with a useful and sound theory of computation. It is also compatible with the thesis of computational explanation. However, the question of whether computation is a natural phenomenon has to be taken seriously since the answer may be a negative one with dire consequences for many current discussions as to machine consciousness.

In the following, I will address three main objections to Chalmers' thesis. First, I reject the thesis that implementation is anything more than an epistemic tool. I will argue that Chalmers' argument shows only that a causal structure *may be taken to be* an implementation — Chalmers' argument does not show that a causal structure *is* an implementation. Second, I will consider whether the notion of computation refers to something more than causation. Third, I will analyze the notion of isomorphism between abstract structure and causal patterns as applied here. In the end, I will suggest that computation is nothing but a way to describe causal processes. As such, computation is unfit to be the foundation for the mind. Eventually, I will briefly discuss the relation between computation and mind and, once the informational stance will be set aside, I will outline some alternative theoretical possibilities.

3. Is Implementation Anything More than an Epistemic Tool?

A crucial step in Chalmers' argument is the claim that a physical system, under the right conditions, implements a given computation. He states that:

> "*A physical system implements a given computation when the causal structure of the physical system mirrors the formal structure of the computation.*" [Chalmers, 1993/2012]

However, this is left largely unexplained. Rather, Chalmers outlines a criterion to attribute computational meaning to a physical system. The fact that there is an unambiguous criterion to associate formal computational structures with causal structures does not entail that the causal structure either embodies or implements the computational structure. At most, under the right conditions, an observer may attribute a computation to a physical system. Alternatively, an observer may use the physical system as an implementation of that computation. But, is there any autonomous relation between the physical system and the computation?

Computational structures are very efficient models for describing certain kinds of causal interactions. In particular, computational structures are efficient ways to express functional structures. Of course, this does not entail that functionalism and computationalism overlap. In fact, "functionalism and computationalism need not go together. Functionalism may be combined with a non-computational theory of mind, and computationalism may be combined with a non-functionalist metaphysics". [Piccinini, 2010]. Yet, is there any real need to assume that the underlying causal structure either is different, or has a new property, or embodies an abstract structure? Are we not downloading arbitrarily our epistemic tools into the physical world? At best, the computational stance is epiphenomenal with respect to the underlying causal processes. It may be objected that the attribution of epiphenomenality is ill-judged since Chalmers includes an explicit causal requirement for attribution. Yet Chalmers' causal requirement will hold for centers of mass, too. Centers of mass have a causal role too. Yet, such a causal role is completely drained by entities that are more fundamental. The issue to be discussed here is whether the ascription of computation has any causal role. I would say that computations *as computations* have no causal efficacy. The causal processes that we label as computations have causal efficacy, of course. However such causal efficacy is totally independent of their computational role.

The solar system implements the most precise conceivable computation of a solar calendar. However, this qualification would not add anything to what we know about the solar system. Computations do not have causal powers beyond those of the entities that ought to implement them. These entities already exhaust the causal work. Chalmers' causal requirement do not really suggest that computations as computations have a causal power. At the best of my understanding, he claims that a physical system implements a computation if the causal structure of the system mirrors the formal structure of the computation. At best, mirroring is definitely an epiphenomenal relation.

Of course, it is not epiphenomenal with respect to the efficiency of our description of what is going on in an agent.

As a useful analogy, consider a center of mass. It is a respectable and useful concept with many possible applications. However, centers of mass do not exist literally. They are not natural phenomena. They are just epistemic tools. In fact, although it appears that we are attracted by the center of mass of the earth, if we would dig a hole inside our planet, it may turn out that the center of the earth is hollow. There is no need for the existence of anything at the exact location of a center of mass because a center of mass is not a physical entity; rather, it is nothing but an epistemic tool that describes the earth's distribution of mass. We are attracted by every piece of matter composing the earth, yet it is *as if* the mass of all such pieces were concentrated in the center. A center of mass is just an epistemic shortcut to describe a much more complex state of affairs — centers of mass live only in physicists' minds and textbooks. They are the result of a mathematical stance we adopted since Galileo's time. Centers of mass would be the wrong kind of entities to support something like the mind, since they do not really exist. If scientists could easily reason in terms of volumetric integrals, the very notion of center of mass may never have been introduced. The computational stance may be used to single out those physical systems that have a certain (abstract causal) property. However, "abstract causal property" is an oxymoron. If properties were abstract they could not be causally actual. If they were causal, they could not be abstract. Chalmers is appealing to a further level of reality that cannot have any actual causal power. Such abstract properties are in the eye of the beholder and thus cannot support a physical phenomenon like consciousness. Perhaps naïvely, but consistently, I keep assuming that the conscious mind is a physical phenomenon [Manzotti, 2006, 2011a, 2011b].

Is the notion of computation akin to the notion of center of mass? Is computation just an epistemic tool? Is it a fictitious entity? Indeed, I argue that this is the case, at least for what matters here. Chalmers' argument shows, *pace* Searle and Putnam, that not all physical systems implement every possible computation. I do agree: not all physical systems implement every computation. In fact, the argument that Chalmers puts forward is about the conditions that have to be met for a system *to be taken as* an implementation of a given computation. However, Chalmers does not say anything about the real occurrence of computations in the physical world. More precisely, it seems to me that the suggested criterion for implementation is just an efficient mapping rule between abstract models and physical states of affairs. The mapping rule does not add anything to what is there nor does it single out any special level. The mapping rule is just a convenient way to describe a certain system. As a counterexample, consider two different ways to implement a dam on a river. In the first case, one actually builds a 100-m-high dam with tons of concrete and steel. The stream of water is blocked and a lake is physically implemented. Yet, what has been built is not a dam because of any mapping rule. It is a dam for what it is. One may even blissfully ignore any "dam" concept. There is not a physical underpinning and implementation of a dam. There is something that prevents the stream of water from flowing. What is there is made of water, concrete, steel, and rocks. On the contrary, in computer science it is customary to implement abstract structures (logic gates, computations, memories, and so forth). They are taken to be physical instantiations of abstract concepts — a very practical way of dealing with them. However, talking in this way suggests that the electronic components have a sort of double life.

Implementation looks a bit like the pineal gland of the theory of computation since it ought to link the ethereal world of abstract computations with the physical world. Do not all sciences use formal, non-causal, abstract objects (e.g., numbers) in order to characterize concrete, physical things? Yes. However, these abstract objects (for instance the speed x of a particle p with a mass m) refer to some portion of physical reality that, although one may never be sure has been correctly characterized, has some kind of concrete existence. On the contrary, what do implementations of computations refer to? Do they refer to anything like a portion of the physical world no matter how temporarily singled out? Chalmers may rebut that this argument seems to consider the physical system and the computation as two different entities. Thus, it may be objected that there is nothing like the computation to be put in relation with the physical system. Yet, such a counter-argument would defeat itself by supporting the conclusion I would like to reach in the end — namely, that there are only physical processes. The computational level does not exist outside of observers' minds and thus it cannot be the foundation for anything.

4. Computation as such is Epiphenomenal

Computations seem to exist because a human subject adopts the computational stance to describe causal patterns. Computations do not seem to add anything to the physical world. In fact, it does not make any difference whether a physical system occurs without anyone considering it an instance of computation. It would take place exactly in the same way. Does a gear behave differently whether we consider it as part of a computational device? Does a transistor behave differently whether we do not know that it is part of a micro-processor? In all these cases, the computational level is only in the eye of the beholder. In intergalactic space, isolate particles of matter move along their trajectory blissfully unaware of the existence of any center of mass. Likewise, electronic stuff or mechanical gears operate unaware of the fact that one may have a computational model of their interactions.

Consider a Chinese mechanical calculator built in such a way that the correspondence between the symbols engraved on its keys and digits is unknown. We may even ignore that it is a calculator. However, we may reverse-engineer the mechanism and formulate a causal description of what happens when someone presses any key. The important thing is that a complete causal description of the object may be formulated without having the slightest idea of its computational nature. It may be objected that the same goes for all modes of explanation (chemistry, biology, etc.) above the lowest physical level. However, in science there are two conflicting views. On one hand, there are scientists who defend the epistemological (sometimes even ontological) autonomy of their domain. On the other hand, there is the attempt to achieve the theory of everything in which everything, at least in principle, will collapse to one level of explanation. Apart from my personal preference for the latter, it is fair to maintain that the contrast between these two views has not reached any kind of definitive outcome. Furthermore, is there any convincing example of a physical phenomenon (apart from consciousness) that is not in the process of being explained in terms of its physical underpinnings? I am not aware of any.

Chalmers himself outlines a Dennett-like computational stance that suggests a practical way to deal with certain physical systems. To some extent, he implicitly anticipates this paper's main argument: computation arises only at an epistemic level. As mentioned above, Dennett [1987] singles out various explanatory stances tuned to different situations — a physical stance, a design stance, and the famous intentional stance. The idea that a causal structure may be described using a computation is akin to Dennett's design stance. The same idea may be reapplied here introducing the computational stance. Once one assumes that a mechanism has been built to match a logical and mathematical structure, one can skip many physical details as irrelevant. Indeed such a descriptive strategy is obviously profitable. To recap, computations are not the outcome of a computational-level of reality in the same way in which blueprints are not the outcome of a design-level of reality. Assuming that causal structures implement computations is analogous to assuming that mechanisms implement blueprints. There are no blueprints waiting to be implemented. Rather, there are mechanisms that may be conveniently described adopting a design stance by means of blueprints and the like. Likewise, it may be argued that not all mechanisms implement all possible blueprints. Once again, there is no vacuity in the relation between mechanisms and blueprints. Yet, it is clear that there is no need to assume that the design-level corresponds to a natural level. Blueprints are just epistemic entities.

What about the fact that the aspects of a physical system that make it correctly describable via a particular blueprint are objective, intrinsic, and causally efficacious? This is irrelevant to the argument defended so far. The point at stake is whether such description adds something to our understanding of what is going on. Is the computational level real or is it just a useful tool? Consider once more the example of the center of mass. The individuation of my center of mass depends on objective, intrinsic and causally efficacious conditions. Yet, there is no way to know from a set of particles whether the notion of center of mass will be of any use. It depends on who is describing the movement of the particles. A much more analytically oriented mind than that of human beings may prefer to keep track of all the individual particles. No need to introduce anything like the center of mass. Thus, the aspects of a physical system that make it correctly describable in terms of a particular blueprint are indeed objective, intrinsic, and causally efficacious. However, what is neither objective and intrinsic nor causally efficacious is the necessity to consider such aspects as the implementation of a computation.

The computational level may be removed from one's description of reality and everything will keep behaving in the same way — the computational level has no causal efficacy. To a certain extent, the same argument would hold for chemistry, biology, and so forth. There is no space here for a not-even-remotely adequate analysis of such a huge issue. However, two sketchy considerations may be listed.

First, if chemistry (or any other disciplines) makes strong ontological commitments, there must be strong reasons to support them. I do not see many of them. We all know the success story of thermodynamics. Is it not the same for most of chemical features that, at least in principle, may be reduced to quantum processes? Second, chemistry (and other disciplines) may encounter some resistance (perhaps temporarily, perhaps not) to further reduction. However, these disciplines deal with empirical facts that are the result of observations. Computations are not observables in the same sense in which the effect of physical aspects it is. It is true that no physical aspect is observable *per se*. However, its causal efficacy and thus its existence are suggested because of some perturbation in the expected flow of events. You measure a light ray and you see that it bends. A hidden mass is then suggested as an explanation for the unexplainable. Clearly, there is no intrinsic access to the nature of mass. With computations, the approach is different. We already have a working causal explanation of the flow of events. Nothing unexpected happens. Yet, we may choose to use an explanation that — like the center of mass — is shorter, clearer, and more generative of similar conditions.

Computations are epistemic tools that summarize certain causal patterns and match them with certain logical constructs human beings have learned to use. Furthermore, computations have no ontological status since they do not refer to something that is intrinsic to physical systems. They tell us much more about the cognitive processes of the observers than about the mechanism itself.

There is a further analogy between the information/symbol dichotomy and the computation/causal structure dichotomy. Computation is dynamic while information is often depicted as a static pattern. However, both are history-dependent (which is not to say that you may use every sign to represent every item of information). If you find a sign on a rock there is no way to get its meaning except by referring to the history of that sign. Similarly, the knowledge of the causal structure does not tell one whether that structure implements a particular computation. Of course, you can always resort to a reverse-engineering that maps computational structures to causal structures. Indeed it could be a useful way to represent the causal structure. Yet it would be an arbitrary choice that does not tell us anything intrinsic about the physical nature of the mechanism. Additional knowledge would be needed as to whom built the device and to which purpose.

5. Isomorphism between a Physical System and an Abstract Structure?

The third issue I would like to address is the suggested isomorphism that links computation and physical world: "a system implements a computation if the causal structure of the system mirrors the formal structure of the computation" [Chalmers, 1993/2012].

The notion of isomorphism is not as simple as it seems. It is somehow related to the notion of similarity and thus it is plagued by more or less the same shortcomings.

In logic and math, an isomorphism is a mapping between entities that have a correspondence between their parts/properties/internal relations. In a certain sense, isomorphic structures are structurally identical and minute differences are ignored. However, it ought to be added that, in the above sense, two entities are isomorphic if and only if they subsume to the same category of entities. For instance, two paintings are isomorphic if they show some geometrical correspondence. Furthermore, they are isomorphic because they both subsume to the category of physical objects.

Consider a sequence of notes and a sequence of colors — they may hardly have any obvious correspondence although a mapping rule may always be suggested. The problem is that such a mapping rule becomes more and more arbitrary. In fact, is a higher musical note really closer to the blue or to the red? We may arbitrary choose any correspondence rule we like. For instance, in the past, various matching rules between colors and sounds have been proposed. However, these rules required an appeal either to shared phenomenal aspects or physical properties. And yet such an arbitrary correspondence is more obvious than the one between an abstract computational structure and a causal one. In fact, these two entities do not share anything but what a beholder arbitrarily chooses them to share.

Chalmer describes the implementation of a combinatorial-state automaton (CSA) in these terms:

> "*A physical system P implements a CSA M if there is a vectorization of internal states of P into components* $[s_1, s_2, \ldots]$, *and a mapping f from the substates* s_j *into corresponding substates* S_j *of M, along with similar vectorizations and mappings for inputs and outputs, such that for every state-transition rule* $([I_1, \ldots, I_k], [S_1, S_2, \ldots]) \geq ([S'_1, S'_2, \ldots], [O_1, \ldots, O_l])$ *of M: if P is in internal state* $[s_1, s_2, \ldots]$ *and receiving input* $[i_1, \ldots, i_n]$ *which map to formal state and input* $[S_1, S_2, \ldots]$ *and* $[I_1, \ldots, I_k]$ *respectively, this reliably causes it to enter an internal state and produce an output that map to* $[S'_1, S'_2, \ldots]$ *and* $[O_1, \ldots, O_l]$ *respectively.*" [Chalmers, 1993/2012]

In such a formulation there is a not so innocent trick — the causal states of the system and the formal states are presented using the same symbols and the same terminology. The definition assumes that there are rules, states, and values both in the combinatorial-state automata and in the physical system. Is it really so? Indeed S and M seem similar to the extent that the issue of implementation appears almost trivial. Yet, there remains a troublesome series of correspondences to explain (Table 1).

For one, consider the correspondence between the state-transition rule and the reliable cause relation. The first is deterministic and *a priori* defined. It cannot change. On the other hand, the *reliable* cause relation does depend on a huge list of causal circumstances that are external to the physical system considered. For instance, if someone would switch off the power supply of my computer, the expected reliable causal connection would not hold any longer. They would abruptly fail.

Table 1. Correspondences between a CSA and a physical system.

CSA M	Physical system P
Formal state S	Internal state s
State-transition rule (no letter provided)	Reliable causal relation (no letter provided) between states
Value of formal state S	Physical event defining state s
Input I	Input i
Output O	Output o

Should we consider the power supply as part of the system? What about the earth's rotation, the atmospheric pressure, the environmental temperature, Jack's Luddite propensity to smash computing devices, and so forth? Should they all be included in order to have a truly reliable causal relation? The problem of extended causal circumstances are well known and do not need to be extensively listed here. The fact is that, in order to have a reliable causal relation, we need to change the boundaries of the physical system we started with. And we are not sure we will ever have anything like a reliable causal connection. On the other hand, since this strategy is dubious, the alternative is to admit that we are not dealing with a real physical system but with an idealization of it. But this would mean that the physical system we are dealing with is nothing but an abstract formalization of a physical system. Thus it would be easy to claim that what Chalmers accomplished is an isomorphism between abstract models rather than between a formal system and a physical one.

Another problem related with causal relations has to do with their existence when they are not actual. In fact, a formal combinatorial-state automaton possesses all its transition rules in some abstract domain. Transition rules are analytical while causal relations are contingent. Is Chalmers thinking of causal relations akin to Davidson's contingent and actual causal relations or rather of normative causal relations [Davidson, 1970a, 1970b; Lewis, 1973]? The difference is important since Davidson's contingent causal relations do not entail the existence of those relations that are not actually occurring. This is true to such an extent that it could be argued that the set of reliable causal transitions does not exist. Current actual causal relations are all that is available. On the contrary, Chalmers needs to appeal to *reliable* causal relations since he needs to apply a correspondence between the elements of two sets: the set of state-transition rules and the set of reliable causal relations. They have to be reliable causal relations and not simply contingent causal relation because the latter are contingent. Therefore, they are actual when and only when they take place. Since they may happen or not it means that when they are not actually taking place one cannot appeal to them. This would imply that the computation could never exist as a whole. Unfortunately, the set of reliable causes, from a contingent perspective, does not exist. Therefore, the correspondence rule cannot be applied since the two sets have a different number of elements.

Third, is there any real similarity between causal relations and state transitions? A state transition is defined in a timeless domain. In fact you can imagine a timeless map with all the required state-transitions mapped in a timeless fashion. On the other hand, an actual causal relation is something totally different [Salmon, 1998; Dowe, 2000]. When things go as expected, there is a change in something (or, sometimes, an expected change does not occur). Yet the change is usually much more fine-grained than the description we provide of it. Consider a so-called change of state inside a neuron: how many sub-events really take place? Consider a change of state inside a transistor. From a microscopic perspective, how many atomic and subatomic events have to take place to allow the transistor to change its behavior? If we watch the neuron or the transistor at such microscopic scale, what is the correspondence with the simplified idea of a transition rule? Once again, the scale at which the system is a computation is more a matter of choice by the observer than an intrinsic property of the system.

A further problem arises as to the nature of states. While the nature of a formal state s is straightforward, it is far from being clear what an internal state s is in physical terms. There is a difference between the state and its content which is not obvious in the case of the physical system. The state is the set of aspects that fix the system. In this context, the content is the actual value of such set of aspects. The state is the speed as a variable and the content is the value of that variable. In a formal system, the variable and its content are separate entities. In a physical system, there is no difference. There is just an event (or a series of events). They are, at the same time, the state and the content of the state. The notion of physical state is plagued by further puzzles that will be only briefly mentioned here. First, there is the issue of the arbitrariness of the definition of the notion of state (why certain properties? What about properties like acceleration that is a second-order property)? Second, is a state a natural kind or not? Third, are the values of a state a natural kind or are they arbitrary conventions? Of course, all these problems are easily solved in practice whenever it is possible to stipulate conventionally that a certain range of values are defined as corresponding to a certain content of a certain state. Yet this is purely arbitrary. Once again, the notion of state seems to be an epistemic only, observer-dependent notion.

Again, it is important to stress here that I am not arguing against either the usefulness of developing computational models or the possibility to associate efficiently a computation to a causal structure. Rather, I question whether such epistemic devices have any ontological weight as Chalmers' argument seems to argue for.

6. Individuation is Neither a Problem nor a Solution

It is well known that various authors claimed that computational states are universally (or at least multiply) realizable — namely any physical system may be interpreted as implementing any computational characterization [Putnam, 1988; Searle, 1992; Harnad, 1994]. They argued that this conclusion is fatal to the computational thesis or to any attempt to ground the mind on computation. Leaving aside the differences between their individual arguments, they seem to converge on the idea that if individuation between computational states and physical systems fails then computation cannot be the basis for the mind. Furthermore, they agree that the connection between a given physical system and a given formal computation is the result of interpretation by an external cognitive agent. Therefore, the fact that a particular system realizes a particular automaton is allegedly a vacuous one and thus it does not say anything about the physical system but rather about the relations between the physical system and the external cognitive agent.

To rebut these arguments, many proponents of the computational thesis — and Chalmers is one of them — proposed causal and counterfactual notions to restrict the notion of computation so that it may be impervious to the charge of vacuity and multiple realizability. Other authors went further. For instance, Chrisley [1994] put forward a strong argument suggesting that "computation would not necessarily be an explanatorily vacuous notion even if it were universally realizable".

The point of this paragraph is to shift the focus of discussion. I suspect that the issue of individuation is neither a problem nor a solution. Suppose it is possible to find a perfect law connecting the set of physical systems with the set of formal computational structures. So, given a physical system you have one and only one computational structure and vice versa. This is very unlikely, of course. However, it would be an ideal situation for the proponent of the computational thesis. No multiple realizability whatsoever and apparently no need to resort to an external interpretation. Yet, would this perfect law have any explanatory value as to the nature of computations? Would the existence of this law provide any justification as to the fact that computational structure are indeed implemented into the corresponding physical systems? I do not see why. Such a bridge law — that would allow perfect individuation — may or may not exist and its existence would not tell me anything at all as to the nature of what it puts into correspondence.

For instance, one may show that there is a perfect correspondence between odd and even numbers. What does that mean? Are odd numbers the implementation of even ones? I do not see why. Or consider a bridge law that links each person on the earth with a formal code. It may be done. Would it mean that people are the implementation of these formal codes? Very unlikely.

Of course, when it is suggested that there is a bridge law between formal computations and physical systems — perhaps flanked by causal and counterfactual conditions — the thesis seems a lot more plausible. And yet why? I suspect that there is some pre-theoretical assumption as to the fact that computations have to be physically realized since causal processes in cognitive agents appears like the right kind of stuff for computation. In other words, there is an implicit aura of plausibility that backs up the computational thesis. Computations seem to be there, inside computers and brains. So it is a matter of logic to find the waterproof causal and counterfactual conditions for their attribution. Yet attribution is not identity. Individuation does not explain why computation should be there. Why should it be so important to show that there is a unique computational description for every physical system? Even if there were, it would not necessarily tell me anything about the computational character of the physical system. It would only show that there is indeed a strong bridge law. So what?

On the other hand, even if computation were universally realizable, as Chrisley observed, computation may still be a fact. In this regard, he stated that "computational explanations [...] are actually satisfying in a large number of cases [...] it is very useful to understand many physical systems [...] in terms of computational properties; and there are many more systems for which such an understanding is *not* useful" [Chrisley, 1994]. Yet, this view does not support any ontological commitment as to the nature of computation. Rather, Chrisley put forward very clearly the point of the computational stance. Dennett could not have used better words to outline the intentional stance. To convince yourself, imagine using the same words to speak of centers of mass. Of course, there are certain systems for which it is useful to introduce the notion of center of mass and other systems for which it is *not* useful.

In short, a working bridge law (and it is questionable whether there is any) is not what we need. We would need a constitutive law capable of explaining why a certain causal process should be a computation independently of external attributions and interpretations. Such a constitutive law ought to explain the difference between a simple causal process and a computational process. We may concede that computation is spread universally. Yet if this were the case, the constitutive law should at least explain why the notion of computation has any use.

7. Does Computation Entail Dualism between Information and Physical World?

Chalmers' paper is not metaphysically innocent once one considers his subsequent works [Chalmers, 1996, 2007] although in other works of his a different view is advocated. Loosely speaking, his suggestion that computation may be the foundation of the mind seems to entail that information is the stuff of the mind. In fact, he claims that information may have a dual nature. In his writings, information is very close to the status of a natural entity. While he admits that "information spaces are abstract spaces, and information states are abstract states, they are not part of the concrete physical or phenomenal world"; he is confident that "it seems intuitively clear that information spaces and states are realized throughout the physical world" [Chalmers, 1996]. Unfortunately, most of the examples he presents (thermostat, book, telephone lines, light switches) seem to be dependent on the arbitrary choices of observers. Even natural signs like smoke or tree rings are related with cognitive agents' goals. He is not the only who does this. Most writings about information have tackled the relation between agents and physical world [Dretske, 1981; Adams, 2003; Floridi, 2009]. If information were realized by states of things, what are we to make of the states of gerrymandered objects like mereological sums [Paul, 2010]? Would not this lead to some kind of informational universalism?

It is surprising that Chalmers sometimes seems to deal with information as an ontological constituent of reality. Furthermore, he claims that there is physically realized information: "Physical realization is the most common way to think about information embedded in the world" [Chalmers, 1996]. It is a curious way to formulate the relation between information and the world as if they were two ontologically related levels. He claims that:

> "*It is natural to suppose that this double life of information spaces corresponds to a duality at a deep level. We might even suggest that this double realization is the key to the fundamental connection between physical processes and conscious experience.* [...] *We might put this by suggesting as a basic principle that information* (*in the actual world*) *has two aspects*, *a physical and a phenomenal aspect*." [Chalmers, 1996]

I do not question the issue of phenomenal content here. Rather, I stress that Chalmers presents again and again information as part of the fundamental ontology of the world: information is embedded, is realized, is stored, and has multiple aspects. It is difficult to resist the conclusion that Chalmers considers information as something substantial and indeed he ought to present arguments supporting his ontological commitments [Chalmers, 1996]. Other authors have recently considered the same possibility defending either implicitly or explicitly the idea that information is real [Landauer, 1992; Wolfram, 2002; Tononi, 2004; Fekete and Edelman, 2011]. Of course, the topic of the ontological status of information cannot be wholly addressed here. Yet, it unavoidably leads to possible severe misunderstanding and wrong premises especially in those fields like cognitive science and AI whose foundations are still vague. As to the mind, computation does not seem to score any better than information. Notions like "realization", "implementation", and "embodiment" play a suspect role in encouraging covert dualistic views of reality. Suggesting that abstract computational structures are implemented or realized in the physical world is an ambiguous claim. Is it just a metaphorical talk with no ontological commitment, or do we have to take at face value the idea that x is implemented in y — both x and y being something real?

Other domains in science are more innocent. Consider the formal molecular relationship of water. Does it exist above and beyond the physical things that realize it? I would, perhaps naïvely, maintain that either the formal molecular relationship brings into existence some new causal power (or unexpected events) or not. In the first case, it is a case of an emergent property and one should give a proper explanation in term of emergence. So far, emergentism does not look to be a particularly promising direction. In fact, it has been argued that all cases of emergence are either brute emergence (and thus akin to miracles) or apparent emergence (and thus reducible to the underling physical underpinnings) [Bitbol, 2008; Strawson, 2008]. In the second case, the formal molecular relationship is nothing more than a historically located way to speak of what happens when two atoms of hydrogen and one atom of oxygen share their orbitals. Therefore, in the second case, also such formal structure is merely metaphorical or fictional akin to the center of mass.

8. Conclusion

The notion of computation gained respectability and momentum in cognitive science, philosophy of mind, AI and neuroscience. However, this well-deserved reputation does not necessarily entail that computation has to be part of the natural ontology as implicitly suggested in Chalmers' target paper since he aims at singling out a computational foundation for cognition. In particular, Chalmers' suggestion to ground computation in the physical world by means of implementation and causal isomorphism sounds unconvincing. The former appears to be an abstract matching rule and the latter does not seem to apply. Rather, I defend the view that computation is nothing more than a useful way to describe causation — computation appears only at an epistemic level. As a result, both computation and information cannot be the foundation for either cognition or consciousness insofar as they are real phenomena.

Of course, in the future it may turn out that certain physical processes bring into existence special properties (for instance consciousness). This is not ruled out by this discussion. However, at present, there are no convincing arguments that show that a certain kind of computational structure ought to produce something unexpected. It would be as if I would attribute, to a list of printed words in alphabetical order, a special property intrinsic to the list itself. It is a fact that the familiarity with the alphabet leads convincingly for many to believe that the alphabetical order intrinsic to lists of printed words. Yet, it would be a mistake to do so. Words and the alphabet exist only in the observer's mind. In the physical world there are only black and white patches.

The ubiquity of computational devices has led many to believe that information is physical and that it is stored and transmitted as if it were possible to locate it. Of course, one may cry, my files are inside my pen-drive which is safely in my pocket! Where else could they be? Yet such a counterargument would fail to see that the physical existence of information in one's pocket is not neutral to a complex chain of causal relations that link the file with antecedent and subsequent events. I have argued elsewhere that information is not an intrinsic property of a state of matter [Manzotti, 2011c]. Information is spread in the environment.

Likewise, computation is not a generator of information since between the two notions there is circularity. In short, a physical process is a computation if it involves information. However, a physical state bears information insofar as it is involved directly or indirectly with a computation. A tree ring bears information only when an agent takes advantage of it to perform a computation (for instance computing the average yearly rainfall). Until then, the tree ring is just a ring-shaped collection of dead cells. Or consider a series of puddles and the set of pits on a CD. Why does the former have no information and the latter does? Is it because of some intrinsic difference, or because in the case of the CD there is a human agent taking advantage of them to perform a computation (Figs. 1 and 2)? I have here defended the latter option.

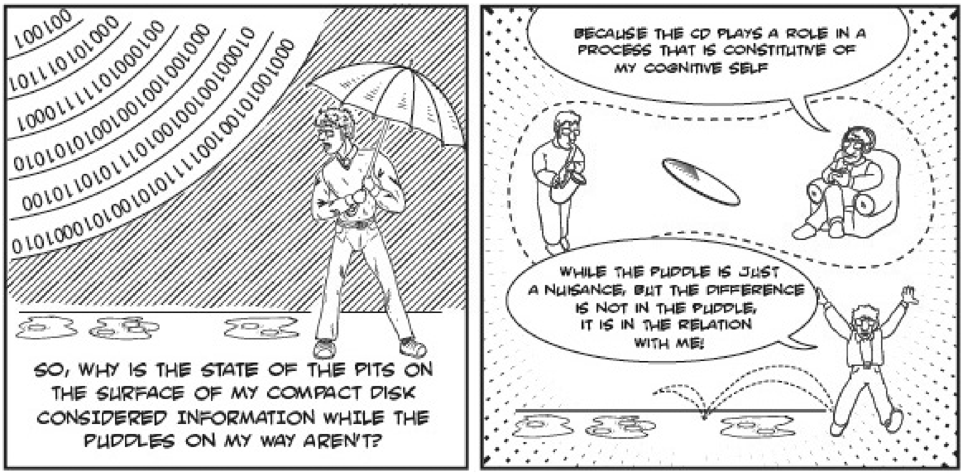

Fig. 1. Why do physical states have information? *Courtesy*: *APA Newsletter on Philosophy and Computers*.

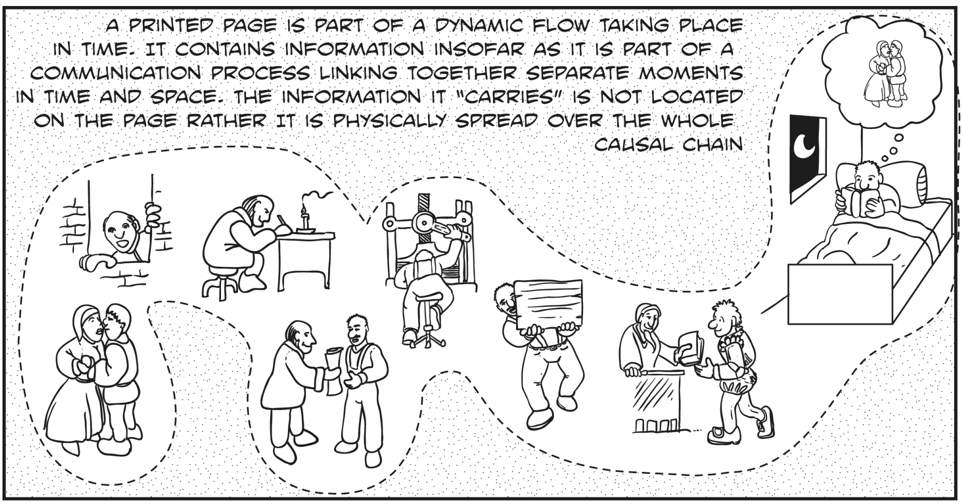

Fig. 2. The spread nature of computation. *Courtesy*: *APA Newsletter on Philosophy and Computers*.

However, it would be unwise, for the case of computation, to resort to some kind of special properties inside human beings as has been done for intentionality. I do not believe in intrinsic intentionality anymore than I believe in intrinsic information or intrinsic computation. Rather I would consider a relational picture of these notions. For instance, could not information be physically spread over all the causal processes involved? Consider the meaning of a printed page. Is that meaning literally located where the page is, or is it spread over the whole and complex causal chain leading from the facts it refers to up to the author passing through the printing process and ending in the reader's brain (Fig. 1)? Is not information the whole process? I consider seriously this option that leads to a totally different understanding of the physical underpinnings of the mind — a view that is likely to find encouragement from externalist views as to the nature of the mind [Clark and Chalmers, 1998; Manzotti, 2006, 2011a, 2011b; O'Regan, 2011]. Too often we slip into the habit of considering information and computation akin to food and digestion, so to speak. Computation and information are affected by circularity. Neither of them may be used to ground the other.

My argument is simple: computation is a useful epistemic tool in the same sense in which centers of mass are useful tools in astronomy; however, it does not constitute a natural phenomenon on which something like the mind may be founded. Computation is the result of adopting the computational stance to describe the underlying causal structure of reality.

References

Adams, F. [2003] "The informational turn in philosophy," *Mind Mach.* **13**, 471–501.

Adams, D. and Aizawa, K. [2008] *The Bounds of Cognition* (Blackwell Publishing, Singapore).

Baars, B. J. [1988] *A Cognitive Theory of Consciousness* (Cambridge University Press, Cambridge).

Bitbol, M. [2008] "Is consciousness primary?" *Neuroquantology* **6**(1), 53–72.

Brooks, R. A. [1991] "Intelligence without representation," *Artif. Intell.* **47**, 139–159.

Chalmers, D. J. [1993/2012] "A computational foundation for the study of cognition," *J. Cogn. Sci.* **12**(4), 323–357.

Chalmers, D. J. [1996] *The Conscious Mind: In Search of a Fundamental Theory* (Oxford University Press, New York).

Chalmers, D. J. [2007] "Naturalistic dualism," in *The Blackwell Companion to Consciousness*, eds. Velmans, M. and Schneider, S. (Blackwell, Oxford), pp. 359–368.

Chalmers, D. J. [2010] "The singularity: A philosophical analysis," *J. Conscious. Stud.* **17**, 7–65.

Chrisley, R. [1994] "Why everything doesn't realize every computation," *Mind Mach.* **4**, 403–420.

Chrisley, R. [2003] "Embodied artificial intelligence," *Artif. Intell.* **149**, 131–150.

Chrisley, R. and Ziemke, T. [2002] "Embodiment," in *Encyclopedia of Cognitive Science* (Macmillan, London).

Clark, A. and Chalmers, D. J. [1998] "The extended mind," *Analysis*, **58**, 10–23.

Davidson, D. [1970a] "Causal relations," in *Essays on Actions and Events*, ed. Davidson, D. (Blackwell, Oxford), pp. 149–162.

Davidson, D. [1970b] "Events as particulars," in *Essays on Actions and Events*, ed. Davidson, D. (Blackwell, Oxford), pp. 181–188.

Dennett, D. C. [1987] *The Intentional Stance* (MIT Press, Cambridge, MA).

Dowe, P. [2000] *Physical Causation* (Cambridge University Press, Cambridge).

Dretske, F. [1981] *Knowledge & the Flow of Information* (MIT Press, Cambridge, MA).

Fekete, T. and Edelman, S. [2011] "Towards a computational theory of experience," *Conscious. Cogn.* **20**, 807–827.

Floridi, L. [2004] "Open problems in the philosophy of information," *Metaphilosophy* **35**(3).

Floridi, L. [2009] "Philosophical conceptions of information," in *Formal Theories of Information*, ed. Sommaruga, G. (Springer-Verlag, Berlin), pp. 13–53.
Fresco, N. [2012] "The explanatory role of computation in cognitive science," *Mind Mach.*
Harnad, S. [1994] "Computation is just interpretable symbol manipulation; cognition isn't," *Mind Mach.* **4**, 379–390.
Landauer, R. [1992] "Information is physical," *Proc. Workshop on Physics and Computation PhysComp* (IEEE Comp. Sci. Press, Los Alamitos).
Lewis, D. [1973] "Causation," *J. Philos.* **70**, 556–567.
Manzotti, R. [2006] "A radical externalist approach to consciousness: The enlarged mind," in *Mind and Its Place in the World. Non-Reductionist Approaches to the Ontology of Consciousness*, eds. Batthyany, A. and Elitzur, A. C. (Ontos-Verlag, Frankfurt), pp. 197–224.
Manzotti, R. [2011a] "The spread mind. Is consciousness situated?" *Teorema* **30**(2), 55–78.
Manzotti, R. [2011b] "The spread mind. Seven steps to situated consciousness," *J. Cosmol.* **14**, 4526–4541.
Manzotti, R. [2011c] "Where is information?" *APA Newslett. Philos. Comput.* **10**(2), 14–21.
O'Regan, K. J. [2011] *Why Red Doesn't Sound Like a Bell. Understanding the Feel of Consciousness* (Oxford University Press, Oxford).
Paul, L. A. [2010] "Mereological bundle theory," in *The Handbook of Mereology*, eds. Burkhardt, H., Seibt, J. and Imaguire, G. (Philosophia Verlag, Munich).
Piccinini, G. [2010] "The mind as neural software? Understanding functionalism, computationalism, and computational functionalism," *Philos. Phenomenol. Res.* **LXXXI**(2), 269–311.
Putnam, H. [1988] *Representation and Reality* (MIT Press, Cambridge, MA).
Salmon, W. C. [1998] *Causality and Explanation* (Oxford University Press, New York).
Searle, J. R. [1980] "Minds, brains, and programs," *Behav. Brain Sci.* **1**, 417–424.
Searle, J. R. [1983] *Intentionality, an Essay in the Philosophy of Mind* (Cambridge University Press, Cambridge, MA).
Searle, J. R. [1992] *The Rediscovery of the Mind* (MIT Press, Cambridge, MA).
Shagrir, O. [2012] "Computation, implementation, cognition," *Mind Mach.* **22**(2), 137–148.
Shanahan, M. [2010] *Embodiment and the Inner Life. Cognition and Consciousness in the Space of Possible Minds* (Oxford University Press, Oxford).
Strawson, G. [2008] *Real Materialism and Other Essays* (Clarendon Press, Oxford).
Tononi, G. [2004] "An information integration theory of consciousness," *BMC Neurosci.* **5**, 1–22.
Wolfram, S. [2002] *A New Kind of Science* (Wolfram Media, Champaign, IL).