

The Multiple-Computations Theorem and the physics of singling out a computation

Meir Hemmo[†], Orly Shenker[‡]
Forthcoming in: *The Monist*

Abstract

The problem of multiple-computations discovered by Hilary Putnam presents a deep difficulty for functionalism (of all sorts, computational and causal). We describe in outline why Putnam's result, and likewise the more restricted result we call the Multiple-Computations Theorem, are in fact *theorems* of statistical mechanics. We show why the mere interaction of a computing system with its environment cannot single out a computation as the preferred one amongst the many computations implemented by the system. We explain why non-reductive approaches to solving the multiple-computations problem, and in particular why computational externalism, are dualistic in the sense that they imply that non-physical facts in the environment of a computing system single out the computation. We discuss certain attempts to dissolve Putnam's unrestricted result by appealing to systems with certain kinds of input and output states, as a special case of computational externalism, and show why this approach is not workable without collapsing to behaviourism. We conclude with some remarks about the non-physical nature of mainstream approaches to both statistical mechanics and the quantum theory of measurement with respect to the singling out of partitions and observables.

1. Introduction

A very influential theory of mind in contemporary philosophy as well as science is the computational theory of the mind, according to which the relations between the mind and the brain is one of software to hardware, i.e., of a computation and the physical system on which it is implemented. If the mind (or certain parts of it) is a computation, the task of science is to discover what is the computation and how precisely it is implemented on the brain. A well-known problem faced by this view is (what we call in this paper) "*the multiple computations theorem*", according to which the any given any relevant degree of freedom in the brain – down to its microscopic physical details, as it evolves according to the physical equations of motion during a given time interval – implements more than one computation, and the task for the computational theory of the mind is to find out which of them is the mind. We examine several proposals for solving this problem, show why they are not satisfactory, and propose our own solution to it. We analyse the notion of computation in terms of physics, within a reductive physicalist framework, and our solution is given in purely physical terms.

[†] Philosophy Department, University of Haifa; meir@research.haifa.ac.il

[‡] Philosophy Department, The Hebrew University of Jerusalem; orly.shenker@mail.huji.ac.il

Hilary Putnam was perhaps the first thinker who proposed what is called today the functionalist theory of the mind, namely the idea that mental states are functional states, in particular computationally characterized states (see Putnam 1967; 1991; for a general overview, see Rescorla 2017). Putnam proposed this view as an alternative to the so-called identity theory according to which mental states are identical with physical-chemical states,¹ which he thought was wrong because it is quite implausible. But twenty years later Putnam (1988) realised that his functionalist approach to the mind faces a serious, perhaps devastating, difficulty:

“The difficulty with this claim, and with all such claims, is not that physically possible organisms don't have functional organizations, but that they have too many. A theorem proved in the Appendix to this book shows that there is a sense in which everything has every functional organization. When we are correctly described by an infinity of logically possible "functional descriptions," what is the claim supposed to mean that one of these has the (unrecognizable) property of being our "normative" description? Is it supposed to describe, in some way, our very essence?” (Putnam 1988, p. xiv-xv)

Here Putnam talks about what we call “*the multiple-computations theorem*”, according to which our brain, as it evolves, implements during the same time interval more than a single computation. Since according to the computational theory of the mind our mind is a computation implemented by the brain, the theorem might entail that have many minds (see e.g., Shagrir 2012); We call this “*the many-minds problem*”². To avoid the many minds problem one need so show that only one of the implemented computations is preferred in that it is the mental one: the theory has to offer a criterion for *singling out that computation*. Now, the computational-functionalist view is usually taken to be a version of non-reductive *physicalism*, according to which everything is *fundamentally* physical; and then the question arises whether preference of one computation as the mental one can be done on the basis of physics. This is the question addressed in this paper.

The paper is structured as follows. In Section 2 we describe in a bit more detail the multiple-computations theorem and the problem of singling out a computation. In Section 3 we give in outline our argument explaining why the multiple-computations theorem is a theorem of physics, being a consequence of statistical mechanics. In Section 4 we argue that the view called computational externalism, according to which it is the physical interaction of a computing system with its environment that singles out one computation as preferred, fails because it leads to a regress that cannot be halted. In Section 5 we discuss attempts to dissolve Putnam's unrestricted version of the theorem by appealing to systems with certain kinds of input and output states, and we argue that this view is a special case of computational externalism and that, as noted by Putnam himself, it implies behaviourism. We conclude in

¹ See different presentations of reductive type identity physicalism in e.g., (Smart 2017; Papineau 1993, 2002; Hemmo and Shenker 2019a, 2019b, 2021a, 2021c; Shenker 2017c).

² To be distinguished by the “many minds” interpretation to quantum mechanics described in (Albert and Loewer 1988).

Section 7 with some remarks about the non-physical nature of some mainstream approaches to both classical statistical mechanics and quantum mechanics with respect to the singling out of partitions and observables, and discuss what a physical solution of the problem should look like.

2. The problem of singling out a computation

According to Putnam's most influential and strong formulation of the idea of multiple-computations (1988, p. 121):

“[e]very ordinary open system is a realization of every abstract finite automaton.”³

Putnam illustrated this very strong theorem as follows. Consider a system that, following the laws of physics, evolves from time 12:00 to time 12:07, changing its *physical state* every minute, such that its states at these moments are $S_0, S_1, S_2, S_3, S_4, S_5, S_6$ and S_7 . If one *assigns the value* “1” to the disjunction S_0 or S_2 or S_4 or S_6 and the value “0” to the disjunction S_1 or S_3 or S_5 or S_7 , then the sequence of physical states may be seen as implementing the sequence of symbols 10101010. To see the nature of this theorem, notice that a different *value assignment*, in which the value “1” is assigned to the disjunction S_0 or S_1 or S_2 or S_3 and the value “0” to the disjunction S_4 or S_5 or S_6 or S_7 , results in the system implementing the sequence of symbols 11110000, as it undergoes *exactly the same physical evolution* as before. As we can see from the above quotation, Putnam's idea is that the value assignment is not dictated by the system's sequence of states, and in this sense as far as the computing system is concerned all of the value assignments are on a par, and none is preferred, that is: by looking at the computing system itself, in all of its details, it is impossible to single any one computation as preferred. none of them is more actual or more natural (etc.) than the others; all the computations, corresponding to all the value assignments, exist simultaneously: not only potentially but in actuality.

Putnam's example may give the impression that the number of computations that may be simultaneously implemented by a given dynamical system is finite, but if we emphasize its physical basis, we can see that this is not the case. The states S_i are coarse-grained states and since the state space of a physical system is *continuous*, there are infinitely many ways of partitioning it to coarse-grained states such as the S_i sets, and therefore, even a simple mechanical system, consisting of a single particle constrained to a finite region of physical space and to a finite time interval, is subject to Putnam's strong result.⁴ Unless additional constraints on the notion of physical implementation are introduced, the system implements

³ Requiring certain physical input and output adds some constraints, but even in that case Putnam's claim is quite strong; see Section 4 about this issue.

⁴ On coarse graining of the state space, see Sklar 1993; Albert 2000; Uffink 2007; Frigg 2008; Hemmo and Shenker 2012; Shenker 2017a, 2017b; Goldstein et al. 2020; on the multiplicity of partitions in the classical state space, see Hemmo and Shenker 2016 (“Ludwig's problem”). On applying this result for the multiplicity of value assignments see Hemmo and Shenker 2019a, 2021a.

infinitely many computations corresponding to the infinitely many partitions of the state space into macrostates and corresponding value assignments. Searle's (1992) famous example of the wall behind him implementing the "Wordstar" software is a point concerning this general applicability of Putnam's result.

The multiple computations theorem raises two closely related problems.

One is the many minds problem, mentioned above. Since a single brain undergoing some specific microphysical process during some specific period of time implements several computations during that time interval, and since each of these different computations is associated with some different cognitive process, the same brain seems to have several minds. But empirically it seems that we have a single whole mind. How can this be, and how can this be accounted for in terms of *physics*? Notice that the result is that there are several distinct *whole* minds implemented by the *whole* brain; this is not the unproblematic case in which different computations are implemented by different parts of the brain.

The other (closely related) problem concerns physical implementations of computations on systems that we use as computers, a paradigmatic case being the laptop on which this paper is being written. It is a fact that we perceive this physical system as carrying out one specific computation, for example some particular word processing, rather than any other computation that is it actually being carried out on the same system at the same time, according to the multiple computations theorem. How do we come to perceive the particular computation that we do? How can we account for this on the basis of *physics*?

Moreover, the two problems can be combined to form a third problem. Since without adding constraints, almost every physical process implements every finite computation, systems that we normally do *not* take to be cognitive systems at all seem to carry out, amongst the computations that they implement, also computations that are associated with cognition.

To solve all three problems we focus here on the following question that gives rise to all three problems. Since (according to the theorem) any physical system can be said to implement simultaneously infinitely many computations, can one single out on the basis of physics a single computation that is actually implemented? In other words, what are the *physical* facts that determine which computation (if any) is actually implemented? This is first and foremost a question of fact (or an ontological question) concerning the way in which physical computations come about.

The attempts in the literature to solve these problems are along two lines of thinking:

1. One is to object to the multiple-computations theorem and argue that it is mistaken in the sense that most (except perhaps one) of the processes are not "computations", if one understands this concept properly.
2. The other line of thinking is to concede the theorem, and find criteria for preferring one computation as the one that is associated with the mind or with the computation implemented on this laptop.

We now say a few words about the first line of thinking, and in the rest of the paper focus on the second. Critics of Putnam's result have argued that Putnam's notion of the physical implementation of a computation is much too liberal, and that constraints have to be imposed on the physical-to-computational mapping, if we are to say that a system implements a computation. For example, in order for a system to implement a computation it is argued that the system has to be associated with a certain causal structure; see Chrisley (1994), Melnyk (1996), Chalmers (1996, 2011, 2012); others have argued that the system has to satisfy some dispositional constraints, see Klein (2008); yet others have put forward mechanistic conditions, see Piccinini (2008, 2015); Milkowski (2013); for modal constraints, see Chalmers (1996, 2011, 2012); Copeland (1996); Scheutz (1999, 2001); and for pragmatic constraints, see Egan (2012). For an overview of this issue, see (Piccinini 2017).

Since our aim is to address the second line of thinking, we do not purport to defend any of these different notions of physical computation. *As a working hypothesis* we shall use the so-called "simple mapping" account of computation which figures in Putnam's (1988) analysis. In the context of the multiple-computations theorem, even if one adds constraints on the mapping, there are examples in the literatures showing that relatively simple dynamical systems, that are nevertheless complex enough for implementing computation, implement more than one computation as they undergo some micro-dynamical process. (according to any of the above-mentioned or other criteria), implement simultaneously more than a single computation (see e.g., Copeland 1996; Shagrir 2001, 2012; Sprevak 2010; Piccinini 2015; and Hemmo and Shenker 2019a; 2021a). For our purpose in this paper, this modest version is all we need. Hereafter, we use the name the "*Multiple-Computations Theorem*" to refer to this modest version of Putnam's result. We take this result to be uncontroversial.⁵ And so we turn to study the above-mentioned second line of thinking that aims to solve the problems raised by this theorem.

In the context of this second line of thinking, our aim in this paper is two-fold.

(I) We first wish to generalise Putnam's result and show in outline that the multiple-computations theorem is a theorem of physics, specifically: it is a theorem of statistical mechanics (classical or quantum).

(II) Secondly, we will offer a solution to the problem of how to single out a computation based on the physicalist type-identity theory of mind and brain. In particular, we will show that the

⁵ Moreover, we have shown (see Hemmo and Shenker 2019a, 2021a) that according to statistical mechanics a physical system undergoing a given micro-dynamical evolution can compute at the same time interval two different computations: one of these computations is *necessarily thermodynamically minimally dissipative* according to Landauer's Principle in physics, whereas the other computation need not be dissipative (on Landauer's Principle, see Landauer 1961, 1992; Bennett 1982, 2003; on Landauer's Principle and physical computation, see Ladyman 2009; and Hemmo and Shenker 2019a, 2021a).

externalist approach for trying to do so, by which the physical interaction of a computing system with its environment select a computation, doesn't work, since it is subject to an infinite regress. We will argue that the only way to avoid the regress and single out a computation on the basis of physics is to adopt the physicalist identity theory.

Remark. The problem of singling out a computation should be clearly distinguished from two another issues concerning the functional theory of the mind that Putnam was worried about. (1) One is Putnam's argument that apparently a given *mental* kind may be realised by heterogeneous *functional* kinds, a result that eventually let Putnam to reject the functional theory of the mind (see Shagrir 2005)⁶. (2) Another is the argument that the same functional kind can be multiply realised by heterogeneous *physical* kinds. By contrast to the previous argument, this one was seen by Putnam (1967) as correct and as one of the strongest arguments in favour of functionalism, and remains so to this day (see Polger and Shapiro 2016). We have argued in detail elsewhere (2019b, 2021a) that multiple-realizability of mental kinds by physical kinds is *incompatible* with physicalism, since it implies that each and every mental state (token state) has *non-physical* features.⁷ We set this point aside in the present paper.

From now on we focus on the question of how to single out a computation on the basis of physics alone, and we will show that this can be done only if mental kinds are strictly identical with physical kinds (so that multiple realizability is ruled out).

3. Multiple-computations and the physics of computation

Let us start by specifying explicitly our working hypotheses.

⁶ Multiple-realization seems to be compatible with supervenience of mental kinds on functional kinds (or on physical kinds). But the multiplicity of functions (e.g., computations) given a single physical process seems to be incompatible with supervenience (see e.g., the many-minds problem mentioned above).

⁷ Here is the argument in outline. Consider a microstate M of the entire universe. What facts make it the case that M realizes (or is a token of) one high-level property L rather than another L'? If these facts are in M (or about M) By assumption, multiple-realizability means that there is no physical fact in M (or about M, i.e. there is no macrovariable of M; see below), that can provide an answer to our question. The only remaining option is that there is some list of all the physical microstates that realize the high-level kind L, so that an item that belongs to the list is of kind L. Here is a dilemma. Either the list is part of physics (horn 1), or it isn't (horn 2).

Horn 1: If the list is physical, then we should add it to the microstate M, in which case multiple realization no longer holds with respect to the full state (call it M*), since now (after adding the list to M) there is a physical fact about M* that answers our question.

Horn 2: If the list is not physical, then the *non-physical fact* of the existence of the list must hold in each and every microstate M, so as to determine whether M belongs to L or L'. And so multiple-realization, if genuine, is a case of token dualism.

Another way to see this is by invoking Laplace's Demon. Suppose that the Demon has access to all possible and actual fundamental physical facts and only to physical facts. If so, the Demon can know whether M is of kind L only if the fact that makes M be of the kind L is a macrovariable of M. But if, as in Horn 2 of the dilemma, the fact that makes M a case of L is not in M, then it is not accessible even to Laplace's Demon, and this is a case of token dualism. Note that the same argument applies (*mutatis mutandis*) if M and M* are replaced with temporal sequences of microstates (for more details, see e.g., Hemmo and Shenker 2019b).

Working Hypothesis 1: As stated above, in what follows we shall work with the so-called “simple mapping” account of computation which figures in Putnam’s (1988) analysis.

Working Hypothesis 2: As noted above, some complex enough systems, that are suitable for implementing computation (according to any of the above-mentioned accounts of computation) implement more than one computation as they undergo some micro-dynamical process.

Working Hypothesis 3: Our third hypothesis is that everything is fundamentally physical. In particular, computations are implemented by physical processes as described by contemporary physics (classical or quantum).⁸

We will now show that the multiple-computations theorem is a special case of a more general theorem of statistical mechanics, in which a similar problem arises of whether and how there are preferred partitions of the states of a physical system to macrostates. We will first describe the origin of the partitions problem in statistical mechanics and then make the connection with computations.

The core idea of statistical mechanics is this. A system is, at each moment, in a given *microstate*, which is its precise state, and this microstate evolves according to the equations of motion. Every microstate is different, but different microstates can be partially identical, that is, they can be identical in some aspect of them (in some so-called *macrovariable*) given by their partial description. A set of microstates that share the same macrovariable, that are an equivalence set relative to that macrovariable, is called a *macrostate*. Consider for example macrovariables that are energy distributions: the microstates of a system are partitioned into sets, in each of which the energy is distributed among the particles in a different way. A famous macrostate is the Maxwell-Boltzmann energy distribution (see Sklar 1993; Albert; Uffink 2007; 2017; and Hemmo and Shenker 2012). What makes it special? It happens to be the case that if the energy is distributed among the particles of an ideal gas in that way, then the gas satisfies certain regularities, for example the ideal gas law. But the states of a physical system can be partitioned into macrostates in infinitely many ways (actually a continuous infinity) corresponding to infinitely many possible partial descriptions of the microstates. For example, the microstate of a system S can be measured by various measuring devices, each of which is sensitive to a different *aspect* (or *macrovariable*) of S’s microstates, and (correspondingly) provides a different partial description of it. Each such partial description brings about a different partition of the microstates of S into macrostates. One device can measure (for example) the position of S’s particles (or the volume of S) while another device can measure the voltage, and the outcome of each measurement will entail that the same microstate belongs to a different equivalence set of microstates. All of those macrostates-sets and corresponding macrovariables, are on a par: as it were, they all “exist” in the same sense,

⁸ We don’t address here Hempel’s dilemma concerning the empirical adequacy and meaning of the hypothesis of physicalism; see our view on this issue in (Firt, Hemmo and Shenker 2021).

since all of them are nothing but aspects (given by partial descriptions) of the microstate, which according to physicalism is all there is in the fundamental reality. Since all of these sets are brought about by partial descriptions of the same microstate, this same microstate belongs to all of these sets simultaneously. Moreover: as the microstate of the system evolves according to the laws of nature, its different aspects evolve, and some of them may exhibit regularities.

Despite all of this, it is a fact that *we* experience only some of the aspects of the microstates of our environment, including some that exhibit interesting and useful regularities. How is this fact to be explained? In a physicalist framework the explanation is quite simple: We are physical systems that interact physically with our environment, and are therefore physically sensitive to certain aspects of our environment and not others (See the “Ludwig scenario” in Hemmo and Shenker2016).

Some of the macrovariables to which we (as physical system) are sensitive, exhibit regularities. This in itself doesn’t make them more “real” or more “natural” than other macrovariables. It may be that other macrovariables, to which we are not sensitive, exhibit regularities as well, and we are not aware of those regularities precisely because we are not sensitive to those macrovariables. As our science evolves we may come to conjecture that some macrovariables of our environment, to which we are not sensitive, behave regularly and may therefore be useful to us, and we build measuring devices that are sensitive to them and “translate” them to macrovariables to which we are sensitive.

The fact that the choice of a preferred partition that has physical significance depends on which macrovariables *we* happen to be able to measure or sense (via our sense organs) may raise the suspicion that a non-physical element of an “agent” or a “mind” or a “self” etc. is being introduced to the theory. Is that the case? If one accepts the physicalist approach one is *committed* to the answer in the negative. And then the only way to account physically for the fact that a partition is preferred relative to an observer is to accept a fully physical account of the observer, that is, to accept type-identity reductive physicalism of the mind. (We have shown elsewhere that non-reductive approaches are committed to token dualism, see Shenker 2017 and Hemmo and Shenker 2019b, 20201c, 2021d).⁹

And here are the consequence for the theory of physical computation.

According to the minimal “simple mapping” account of computation, a computation is fixed by two requirements: dynamics plus value assignment (seeHemmo and Shenker2019a, 2021a).

⁹ The *same* argument (mutatis mutandis) applies in quantum statistical mechanics. Moreover, in pure Hilbert space quantum mechanics there is an additional problem of multiplicity, which does not arise in classical mechanics, namely the problem of how to single out a preferred basis in Hilbert space; see Hemmo and Shenker 2020b, 2021b.

The dynamics is, in terms of physics, the evolution of the complete microstate of the computing system, according to the equations of motion. Building a physical computer means building a system that behaves in a certain way.

Value assignment is, in terms of physics, focusing our attention on certain aspects (macrovariables) of that microstate and calling them “1” and “0”; in other words, a value assignment consists in (i) *selecting as preferred* a partitioning the set of microstates of the computing system into certain sets, in each of which the microstates share a certain aspect; (ii) a mapping from one such set of microstates (that share a certain aspect) to the value “1” and from another such set of microstates (that share another aspect) to the value “0”.

If we wish to implement a computation of a certain logical functions, we need to make sure that there be *harmony* between the dynamics and the value assignment, and here there are two ways to go. (1) we can start by identifying a certain micro-dynamics of our computing system and then we choose certain macrovariables as representing the computational states 0 or 1 in such a way that the computation we are interested in will be carried out as the microstate evolves. (2) The other way is to start with a value assignment, namely, with a mapping between the 0 and 1 and certain macrovariables of the computing system, and then construct a dynamical process that the system will undergo such that as those mapped macrovariables evolve they will give us the desired computation. This second way is perhaps better suited to describe the construction of standard computers as we know them. The macrovariables are selected so that we, humans, are sensitive to them, for example: a part of the computing system consists of the output device in which pixels are of the size and colours to which our eyes are sensitive.

In sum: A physical implementation of a given computation by a given physical system is a particular *harmony* between the *micro-dynamical evolution* of the system and a *partition* of the microstates of the system into macrostates, such that if those macrostates are associated with computational states then the dynamics yields a sequence of macrostates that corresponds to the desired sequence of computational states, i.e., to the desired computation.

In statistical mechanics such harmony is the way to describe all macroscopic evolution, such as those described by the Second Law of thermodynamics (see Hemmo and Shenker 2012, Ch. 7.5 and 7.6). Our result is that this idea of harmony between dynamics and partition to macrostates is the physical underpinning of Putnam’s (1988) theorem. According to our analysis, the *value assignment* in implementing a certain desired computation is nothing but a *partition of the states of the computing system into the “right” macrostates*: these macrostates are the “right ones”, since given the dynamics they yield the sequence of computational states, that fits the desired computation. Conversely, given a certain partition, a micro-dynamics will be the “right one” if it yields the right sequence of computational states.

4. The physics of singling out a computation

The fact that we experience certain macrovariables but not others is explained, then, by our physical interaction with the environment. But this is not the end of the story because this idea raises several problems that need to be addressed.

We just mentioned that, despite the multiplicity of macrovariables of each microstate and hence the multiplicity of the partitions to macrostates, the observer may be sensitive to only one of those macrovariables, ignoring the rest for all practical purposes at that moment, and so that macrovariable is preferred relative to that observer. Talk about observers may have non-physical connotations, that we want and can avoid. To do so is convenient to start by thinking in terms of *measuring devices* (and address human observers later). Along the lines mentioned above, each possible measuring device is sensitive to only one macrovariable of the measured system, and therefore *objectively and physically relative to that device* that macrovariable is preferred. Of course, *other* measuring devices will “see” *other* macrovariables as preferred, and that too is an objective physical fact. (In the context of value assignment, as we shall see, this means that every measuring device would fix a value assignment relative to it.)

However, here the following problem arises: a measuring device does *not* select a preferred macrovariable of the measured system, for the following reason. (In the context of value assignment, this will have implications for the externalist approach to the multiple-computations problem.)

We start (in this paragraph) with the most general and abstract argument, and then (in the following paragraphs) describe the idea in more concrete but perhaps a bit more restricted context. Once the measuring device and the measured system are coupled by the interaction they form a unified system, and from then on the trajectory of its microstate is best described in the state of the combined system. And now the problem addressed above re-appears: the microstate of the combined system has infinitely many aspects (macrovariables), corresponding to infinitely many ways to partition the (combined) state space into macrostates-sets. Under some of these descriptions the measuring device will appear to have the properties and regularities we are used to seeing in our experience, and these may nicely correspond to the relative aspects (macrovariables) that we expect to see in the measured system. Different descriptions, however, that is: different partitionings of the combined system, will not yield such nice looking properties and regularities, and will not reveal these facts of which we are familiar from experience. The properties and behaviours when seen under these partitionings will look cumbersome. As far as physics is concerned, this fact does not make them less “real” nor less “existent” nor even less “natural” than the nice-looking partitionings; so it doesn’t yield a criterion for preferring one partitioning over others (or one value assignment over others). And this fact is precisely what we want to explain: *why* is that that we see the world in terms of the “nicely behaving” macrovariables and not the “cumbersome” ones, given that from the perspective of fundamental physics all of the

macrovariables are equally real.¹⁰ Can physics provide a criterion for such preference? This is our problem and task.

Here is a more concrete (but perhaps more limited) way of thinking about the same problem. A measuring device is part of the environment of the measured system; for simplicity, let us assume for a moment that the measuring device is the entire environment of the measured system. The idea that the interaction with the environment selects the preferred computation (or partition) is sometimes called in the literature *computational externalism* (see e.g., Harbecke and Shagrir 2019; Shagrir 2018; Piccinini 2015). The idea is that the environment is sensitive to certain properties (aspects, macrovariables) of the computing system and not to others; and so relative to that environment these macrovariables are preferred and give rise to the value assignment, and hence selects one computation as preferred, thus solving the multiple-computations problem. The environment of this laptop, for example, is sensitive to certain properties (macrovariables) of the laptop and not to others; and so relative to this environment, is preferred and fixes the computation that it actually carries out (that it carries out Word rather than Searle's Wordstar, for example). A particularly important part of this laptop's environment is the user, in particular the human observer. The engineers build our laptops with our physical making in mind, so that the way that we interact with the laptop induces a value assignment that, given the dynamics – that is built to be in harmony with this value assignment! – the desirable computations appears. And this seems to be an objective-physical basis for preferring a certain computation over others!

Unfortunately, this idea of how to single out a computation by means of a physical interaction with the environment doesn't work. To see why, we need to think for a moment about the physical nature of measuring devices (which is the environment). Consider a measuring device E. This measuring device, being itself a physical system which is, is in some physical microstate at each moment, and its microstate has its own aspects (macrovariables, give by a partial description of that microstate). In fact, when our measuring device E interacts with our system of interest S (thus allegedly fixing one of its macrovariables as preferred relative to E), it is not the *entire* microstate of E that interacts with S: only some macrovariables (aspects) of the microstate of E get correlated with some macrovariables (aspects) of the microstate of S. Suppose that, as E interacts with S, the macrovariable ME1 of E gets correlated with macrovariable MS1 of S. The environment E has other macrovariables as well, and it may be (and is often the case!) that another macrovariable of E's microstate, say ME2, interacts with a different macrovariable of S's microstate, say MS2. For example, the electromagnetic interaction between E and S brings about correlations between certain aspects (macrovariables) of their microstates, and the gravitational interaction between E and S brings about correlation between other aspects of these microstates. And so, in virtue of these two simultaneous interactions (of ME1 with MS1 and of ME2 with MS2) the *same* environment E picks out *two* distinct macrovariables of S; and again, none of them is preferred over the

¹⁰ In QM a counterpart of this idea undermines the Many Worlds interpretations even when decoherence obtains, see Hemmo and Shenker 2020b.

other. In order to prefer MS1 over MS2 (or the other way around of course), we need to prefer ME1 over ME2 (or vice versa). But nothing in E, nor in S, dictates this preference. Needless to say, introducing another environment in order to prefer ME1 over ME2 leads to a useless (possibly infinite) regress. And so, we are back to square one: the multiple-computations problem cannot be solved by introducing the interaction with the environment, because we now have a *multiplicity* of system-environment correlations.

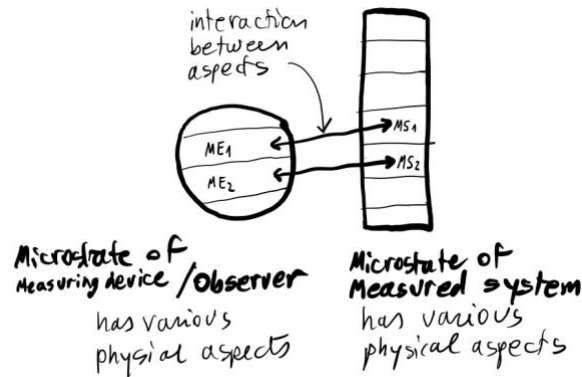


Figure 1

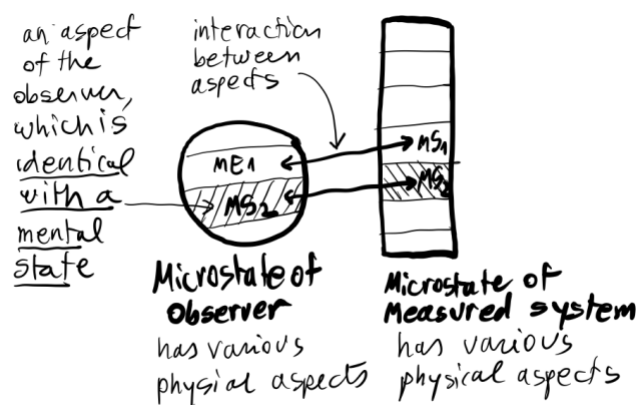
Here is another way to look at this. The laptop on which this paper is written implements more than a single computation, and the fact that you see it as implementing one particular computation (e.g., word processing) is relative to the fact that a particular macrovariable in you (call it ME1, alluding to the previous discussion) interacts with a particular macrovariable of the laptop (call it MS1, alluding to the previous discussion). This solves only part of the problem, but the main problem remains: what selects within your sense organs or brain or both, the macrovariable ME1 rather than ME2? Both exist in your brain, so it might have been the case that the interaction between macrovariable ME2 (rather than ME1), with its interaction with the laptop's macrovariable MS2 (rather than MS1) would fix the computation that you experience in the computer. The problem is that, as far as we could see until now, there is no physical fact that prefers ME1 over ME2 within your brain, without repeating the problem by some sort of an infinite regress.¹¹ (Additionally, it also seems problematic to think that your experience is fixed by something or somebody observing you, so that if that observer is replaced or disappears your experience would change or disappear. But we skip this problem now.)

Nevertheless, despite the above regress, it seems to us that this externalist approach is in the right direction, although it misses a crucial point, as we shall immediately see. Our conjecture is that the regress can be stopped by a *further special* kind of step along the chain of interactions.

¹¹ The same regress arises in quantum statistical mechanics for essentially the same reason, where in quantum mechanics there is an additional problem of introducing new physical facts that should break the basis-symmetry of Hilbert space; see Hemmo and Shenker 2020b, 2021b.

The idea is this. The fact that we have a single mind, as well as the fact that we see our laptop as implementing a certain particular computation, are facts about our experience of ourselves and of our environment. And the solution is to understand this experience within reductive type-identity physicalism, as follows. Among the many macrovariables (aspects) of the observer, one is special in that it is *identical with* the observer's mental experience (see Figure 2). The fact that, whenever the observer is in microstate that has aspect (macrovariable) MS2, for example (as in Figure 2), obtains regardless of the environment in which this observer is situated and regardless of the fact that the observer's brain has other aspects (such as MS1, in Figure 2); those aspects exist, no less than MS2, and qua aspects (or macrovariables) of the microstates of the observer S they are on a par with the macrovariable MS2; the only difference is that aspect (macrovariable) MS2 is identical of the observer's mental experience. And this is the reason why in the experience of this observer there is preference to MS2 and correspondingly to ME2 of the environment.¹²

In this way the mind-brain identity theory stops the regress. If one accepts an identity theory, it follows that there is a fact of the matter concerning which macrovariables of the brain (perhaps together with our sense organs and body) are identical with which mental states. The central hypothesis of this theory is that every mental state of every mental type M at any time *t* just is (*identical to*) a state in which some macrovariable M of the brain obtains at *t*. If this hypothesis is true, the regress along the chain of interactions is stopped whenever the mental state of an observer becomes correlated with a certain partition of the states of the computing system.



¹² Non-reductive approaches reject identity and accept instead a metaphysical dependence relation of supervenience, or realization or grounding, etc. We show elsewhere (see Hemmo and Shenker 2019a, 2019b, 2021a) that all these approaches entail dualism, because the metaphysical relation that fixes which physical kinds (or states) belong to which mental kinds adds *non-physical* facts into reality. For a non-reductive approach to statistical mechanics, see Albert (2000), (2014); Loewer (2020); and our criticism in Hemmo and Shenker 2021e. In Hilbert space quantum mechanics an additional problem arises because of the basis symmetry of Hilbert space with respect to the equation of motion and the physical state of the system; see Hemmo and Shenker (2020b, 2021b).

5. Computational externalism and behaviourism

Up to now we have argued that the interaction with the environment is not enough to select a preferred partition (or computation), and therefore that computational externalism faces a regress. In particular we argued that the only way to stop the regress and select a computation is by *identifying* certain physical *internal* states of the system which are identical to an observer's mental states. We shall now argue, following Putnam but within our present framework, that any attempt to solve the multiple-computations problem and stop the regress by appealing to some external input and output states lead to straightforward behaviourism.

The externalist view (to which we object here) is sometimes motivated by observations such as the following.¹³ We know from observations that many living organisms respond to external stimulations by moving their bodies. And we know from brain science (for example) that such motion is a result of muscle contractions and relaxations, which are brought about by signals received from nerve endings at neuromuscular junctions. In short, spike trains are the main triggers of behaviour. The thought is that the neuromuscular junction is the (part of the) local environment of the computing system (which is mostly inside the brain), and this local environment fixes the relevant neuro-computational states in the computing system. Let us describe this thought in terms that are more closely connected to our present discussion. The behaviour is the output of the computing system, and therefore neuromuscular junctions are the relevant environment, which selects the preferred computation as the one which is identical to our mental states. In other words: *because* the behaviour – which is the evolution of a *certain* set of aspects (macrovariables) in the body (that is, in the neuromuscular junctions environment) – is the output, the *relative* macrovariables of the computing systems fix the computation. A different set of aspects in the computing system would not only fix a different computation, but would also fix a different set of aspects (macrovariables) in the body, which are not the “behaviour” as we experience it and seek to explain it. Let us call the set of macrovariables of the neuromuscular junctions that are correlated with standard notions of “behaviour”, and are correlated with certain macrovariables in the brain, “the *nice* set of macrovariables”; and other sets are “*cumbersome*” since they correspond to facts in the world that we cannot clearly call “behaviour”, and that in general we don’t even perceive as “kinds” since they are not correlated with the macrovariables to which our sense organs are sensitive and that appear in our theories. The question now is this: what sort of fact makes it the case that the “*nice*” set of macrovariables are “*real*”, so that this is “*really*” the computation carried out in the brain? In particular, is this sort of fact *physical*? This is the problem parlayed into the terms of our present discussion.

¹³ We thank an anonymous reviewer for raising this point.

As already argued by Putnam (1988, Appendix), the multiple-computations theorem itself entails that *if one fixes (for whatever reasons) certain input and output states of a desired computation*, then the interaction between the systems entails that there always are relative physical states of the computing machine that will be mapped to the logical states of the computation and will stand in the 'right' causal relations to one another and to the input and the output states. We join Putnam (1988 p. 124-125) in concluding that if this argument stops the regress it does so at the cost of straightforward behaviourism:

“Imagine, however, that an object S which takes strings of "I"s as inputs and prints such strings as outputs behaves from 12:00 to 12:07 exactly as if it had a certain description D. That is, S receives a certain string, say "111111," at 12:00 and prints a certain string, say "11," at 12:07, and there "exists" (mathematically speaking) a machine with description D which does this (by being in the appropriate state at each of the specified intervals, say 12:00 to 12:01, 12:01 to 12:02, . . . , and printing or erasing what it is supposed to print or erase when it is in a given state and scanning a given symbol). In this case, S too can be interpreted as being in these same logical states A,B,C, ... at the very same times and following the very same transition rules; that is to say, we can find physical states A,B,C, . . . which S possesses at the appropriate times and which stand in the appropriate causal relations to one another and to the inputs and the outputs. The method of proof is exactly the same as in the theorem just proved (the unconstrained case). Thus we obtain that the assumption that something is a "realization" of a given automaton description (possesses a specified "functional organization") is equivalent to the statement that it behaves as if it had that description. In short, "functionalism," if it were correct, would imply behaviourism! If it is true that to possess given mental states is simply to possess a certain "functional organization," then it is also true that to possess given mental states is simply to possess certain behavior dispositions!” (Putnam 1988, 124-125)

In the above externalist motivation, the preference of certain sets of macrovariables (for what we called the “nice” set of macrovariables) is due to their being *behavioural* states, since it is only facts accessible to us about the external behaviour of the organism that filter out all other computations that are equally implemented in the brain according to the theorem.

Thus, we can phrase our argument in the form of a dilemma for computational externalism and the computational theory of the mind: Given the multiple-computations theorem, there are only two ways to single out a preferred computation:

Horn 1: Adopt computational functionalism, and fix the computation by the interaction of the computing system with its environment, thus endorsing behaviourism, as Putnam says;¹⁴

¹⁴ The multiple-computations theorem and all of our above arguments equally apply to the causal version of functionalism; see Hemmo and Shenker (2021f).

Horn 2: Or adopt a full-fledged type-identity theory in which the internal macrovariables of the brain identical to mental states fix the preferred computation.

Here the account of the mental is type-identical, and not computational. One might, for reasons of convenience, choose to describe the evolution of the macrovariables that are identical to the mental states and processes in terms of computation, and to the extent that this is convenient we have no objection to it; but then the computational description is not significant, and *multiple realisation is ruled out*. The fact that the mind is described as implementing a computation is secondary and not essential to the fact that some physical states of our brain (and perhaps body) are (identical with) mental states. What makes physical kinds mental is the type-identity of mind and brain, rather than some computational identity. This is the internal feature that stops the regress and breaks the symmetry of all simultaneously implemented computations implied by Putnam's theorem.¹⁵ (It also fixes the computation by this laptop as the computation that we physically see due to the interaction between our physical-mental macrovariables and those of the laptop.)

6. Conclusion

We have shown in this paper that the only way to solve the multiple-computations problem and single out a computation is by appealing to the brain states of observers. Here it is instructive to cite von Neumann (1932), who wrote about the quantum-mechanical theory of measurement:

[W]e must always divide the world into two parts, the one being the observed system, the other the observer. In the former, we follow up all physical processes (in principle at least) arbitrarily precisely. In the latter, this is meaningless. The boundary between the two is arbitrary to a very large extent... That this boundary can be pushed arbitrarily deeply into the interior of the body of the actual observer is the content of the principle of the psycho-physical parallelism – but this does not change the fact that in each method of description the boundary must be put somewhere, if the method is not to proceed vacuously, i.e., if a comparison with experiment is to be possible. Indeed, experience only makes statements of this type: an observer has made a certain (subjective) observation; and never any like this: a physical quantity has a certain value. (von Neumann 1932, Ch. 6, p. 420)

¹⁵ Single out a computation on the basis of physics alone, other proposals for singling out a computation (see e.g., Piccinini 2015; Coelho Mollo 2018, 2019, 2020; Dewhurst 2018a, b; Schweizer 2016, 2019; Millhouse 2019; Shagrir 2018, 2021; Fresco and Milkowski 2020) should be compatible with type-identity of physical kinds and mental kinds. We argued elsewhere that any theory which is committed to mind-brain supervenience but allows multiple-realizability is *token-dualist* in the sense that it is committed to the existence of *non-physical* facts on *every* occasion in a which there is a fact about the mental (see Hemmo and Shenker 2015, 2019a, 2019b, 2021a, 2021c, 2021d; Shenker 2017c).

Applying von Neumann’s line of thought to the computational theory of the mind and the problem of selecting the “mental” computation *which is our mind* according to this theory, the many-minds problem is solved by realizing that the preferred macrovariable, that gives rise to a preferred partition to macrostates, is the one *identical* with our relevant mental states; and this preferred partition, in turn, leads to the preference of certain macrovariables in our environment, for example of this laptop, and hence to the experience that the laptop carries out one computation rather than another.

It is crucial to note that here identity physicalism is absolutely necessary for solving the problem on the basis of physics. One can easily see from our argument that if the mental states of observers are not strictly identical (in the sense of type-identity) to certain macrovariables of the brain, then von Neumann’s division of the world into observer and observed along the measurement chain will introduce into the chain *non-physical* facts. In this sense, only identity physicalism accounts for *both* the environment and our mental states by the *same* physics, that is by stipulating that the mental states of observers are nothing but physical states, using the same physics in which one describes the observed parts of the world.

Indeed, it seems to us that our proposal of type-identity physicalism is a necessary condition for solving not only the multiple-computations problem, but also the measurement problem in quantum mechanics. In the quantum case, one can show that the objective identification of mental states of observers with features of the observers' brains plays a crucial role in solving the quantum version of the multiple-computations problem in the two major theories currently on the table in the foundations of (non-relativistic) quantum mechanics: these are Bohm’s pilot-wave (1952) theory (see Hemmo and Shenker 2013) and the theory of spontaneous state reduction by Ghirardi, Rimini and Weber (GRW, 1986; see Bell 1987). Both are correctly taken to be theories of “quantum mechanics *without* an observer”, since in both theories mental states are not added to the primitive ontology, but rather they are identical with features of the brain that these theories attempt to describe (see Allori 2013). However, we showed elsewhere (see Hemmo and Shenker 2020b, 2021b) that our solution is *not* open for the Everett (1957) (many-worlds) interpretation of quantum mechanics. The reason is that in the Schrödinger evolution - which is all there is according to this interpretation! – there are no physically preferred bases unless one adds something to the Hilbert space structure: in particular, in states of quantum entanglement, nothing makes the brain states of observers - which by stipulation are identical to mental states in the *absence of entanglement* - physically preferred. This holds even if the entangled brain states of the observers are subject to decoherence.

Acknowledgement

We thank David Papineau for valuable comments on an earlier draft of this paper. This research is supported by the Israel Science Foundation, grant number 1148/2018.

References:

1. Albert, David (2000) *Time and Chance*, Cambridge, MA: Harvard University Press.

2. Albert, David (2014) *After Physics*, Cambridge, MA: Harvard University Press.
3. Allori, Valia (2013) "Primitive Ontology and the Structure of Fundamental Physical Theories", in: D. Albert, A. Ney (eds.), *The Wave Function: Essays in the Metaphysics of Quantum Mechanics*, pp. 58-75, Oxford University Press.
4. Bennett, Charles (1973) "Logical Reversibility of Computation", *IBM Journal of Research and Development*, 17, 525-532.
5. Bennett, Charles (1982) "The Thermodynamics of Computation: A Review," *International Journal of Theoretical Physics*, 21, 905–940.
6. Bennett, Charles (2003) "Notes on Landauer's Principle, Reversible Computation, and Maxwell's Demon", *Studies in History and Philosophy of Modern Physics*, 34, 501–510.
7. Bohm, David (1952) "A Suggested Interpretation of the Quantum Theory in Terms of 'Hidden Variables', Parts I + II, *Physical Review*, 85, 166-79; 180-93.
8. Bub, Jeffery (2001) "Maxwell's Demon and the Thermodynamics of Computation", *Studies in History and Philosophy of Modern Physics*, 32, 569–579.
9. Chalmers, David (1996) "Does a Rock Implement Every Finite-State Automaton?" *Synthese*, 108, 309-333.
10. Chalmers, David (2011) "A Computational Foundation for the Study of Cognition", *Journal of Cognitive Science*, 12, 323–357.
11. Chalmers, David (2012) "The Varieties of Computation: A Reply", *Journal of Cognitive Science*, 13, 211-248.
12. Coelho Mollo, Dimitri (2018) "Functional Individuation, Mechanistic Implementation: The Proper Way of Seeing the Mechanistic View of Concrete Computation", *Synthese*, 195, 3477–97.
13. Coelho Mollo, Dimitri (2019) "Are There Teleological Functions to Compute?", *Philosophy of Science*, 86, 431–52.
14. Coelho Mollo, Dimitri (2020) "Against Computational Perspectivalism", *The British Journal for the Philosophy of Science*, axz036.
15. Chrisley, R. L. (1994) "Why Everything doesn't Realise Every Computation", *Minds and Machines*, 4, 403–420.
16. Copeland, B. Jack (1996) "What is Computation?" *Synthese*, 108, 335–359.
17. Craver, Carl, F. (2013) "Functions and Mechanisms: A Perspectivalist View", in P. Huneman (ed.), *Functions: Selection and Mechanisms*, Dordrecht: Springer.
18. Dewhurst, Joe (2018a) "Individuation Without Representation", *The British Journal for the Philosophy of Science*, 69 (1), 103-116.
19. Dewhurst, Joe (2018b) "Computing Mechanisms Without Proper Functions", *Minds and Machines*, 28 (3), 569-588.
20. Egan, Frances (2012) "Metaphysics and Computational Cognitive Science: Let's Not Let the Tail Wag the Dog", *Journal of Cognitive Science*, 13, 39–49.
21. Egan, Frances (2017) "Function-theoretic Explanation and the Search for Neural Mechanisms", in David M. Kaplan (ed.), *Explanation and Integration in Mind and Brain Science*, pp. 145-163, Oxford University Press.
22. Everett, Hugh III (1957) "'Relative state' formulation of quantum mechanics," *Reviews of Modern Physics* 29, 454-462.

23. Feline, Laura (2018) "Quantum Theory Is Not Only About Information," *Studies in History and Philosophy of Modern Physics*, available online 22 June 2018.
24. Firt, Erez, Hemmo, Meir and Shenker, Orly (2021) "Hempel's Dilemma: Not Only for Physicalism", forthcoming.
25. Fresco, Nir and Milkowski, Marcin (2020) "Mechanistic Computational Individuation without Biting the Bullet", *British Journal for the Philosophy of Science*, forthcoming.
26. Frigg, Roman (2008) "A Field Guide to Recent Work on the Foundations of Statistical Mechanics," in D. Rickles (ed.), *The Ashgate Companion to Contemporary Philosophy of Physics*, pp. 99-196, London: Ashgate.
27. Fuchs, Chris and Schack, Ruediger (2014) "QBism and the Greeks: Why a Quantum State does not Represent an Element of Physical Reality", *Physica Scripta*, 90 (1), 015104.
28. Ghirardi, G., Rimini, A. and Weber, T. (1986) "Unified Dynamics for Microscopic and Macroscopic Systems," *Physical Review*, D, 34, 470-479.
29. Godfrey-Smith, P. (2009) "Triviality Arguments Against Functionalism", *Philosophical Studies*, 145, 273-295.
30. Goldstein, Shelly, Lebowitz, Joel, Tumulka, Roderich, and Zanghì, Nino (2020) "Gibbs and Boltzmann Entropy in Classical and Quantum Mechanics", in Valia Allori (ed.), *Statistical Mechanics and Scientific Explanation: Determinism, Indeterminism and Laws of Nature*, forthcoming.
31. Harbecke, Jens and Shagrir, Oron (2019) "The Role of the Environment in Computational Explanations", *European Journal for Philosophy of Science* 9, 37.
32. Hemmo, Meir and Shenker, Orly (2012) *The Road to Maxwell's Demon*, Cambridge: Cambridge University Press.
33. Hemmo, Meir and Shenker, Orly (2013) "Probability Zero in Bohm's Theory", *Philosophy of Science*, 80 (5) (Chicago University Press), 1148-1158.
34. Hemmo, Meir and Shenker, Orly (2015) "The Emergence of Macroscopic Regularity," *Mind & Society*, 14 (2), 221-244.
35. Hemmo, Meir and Shenker, Orly (2016) "Maxwell's Demon", *Oxford Handbooks Online*, Oxford University Press.
36. Hemmo, Meir and Shenker, Orly (2019a) "The Physics of Implementing Logic: Landauer's Principle and the Multiple-Computations Theorem", *Studies in History and Philosophy of Modern Physics*, 68, 90-105.
37. Hemmo, Meir and Shenker, Orly (2019b) "Two Kinds of High-Level Probability", *The Monist*, 102, 458-477.
38. Hemmo, Meir and Shenker, Orly (2020b) "Why the Many Worlds Interpretation of Quantum Mechanics Needs More Than Hilbert Space Structure", in: Rik Peels, Jeroen de Ridder, René van Woudenberg (eds.), *Scientific Challenges to Common Sense Philosophy*, pp. 61-70, New York: Routledge, Taylor & Francis.
39. Hemmo, Meir and Shenker, Orly (2021a) "A Challenge to the Second Law of Thermodynamics from Cognitive Science and Vice Versa," *Synthese*, forthcoming.
40. Hemmo, Meir and Shenker, Orly (2021b) "Why Decoherence Does Not Solve the Preferred Basis Problem in the Many-Worlds Interpretation of Quantum Mechanics," *Synthese*, under review.
41. Hemmo, Meir and Shenker, Orly (2021c), "Flat Physicalism", *Theoria*, forthcoming.

42. Hemmo, Meir and Shenker, Orly (2021d) "Why Functionalism is Token-Dualism", in Gal Vishne, Stavros Ioannidis, Meir Hemmo and Orly Shenker (eds.), *The Multilevel Structure of Reality*, Springer, forthcoming.
43. Hemmo, Meir and Shenker, Orly (2021e) "Is the Mentaculus the Best System of Our World? In B. Loewer, B. Weslake and A. Winsberg (eds.), *Time's Arrow and the Origin of the Universe: Reflections on Time and Chance – Essays in Honor of David Albert's Time and Chance*, Harvard University Press, forthcoming.
44. Hemmo, Meir and Shenker, Orly (2021f) "Is the Mind in the Brain in Contemporary neuroscience?" Unpublished manuscript. Under review in *Nous*.
45. Klein, C. (2008) "Dispositional Implementation Solves the Superfluous Structure Problem", *Synthese*, 165, 141–153.
46. Ladyman, James, A.C. (2009) "What does it Mean to Say that a Physical System Implements a Computation?" *Theoretical Computer Science* 410, 376–383.
47. Landauer, Rolf (1961) "Irreversibility and Heat Generation in the Computing Process", *IBM Journal of Research and Development* 3, 183–191.
48. Landauer, Rolf (1992) "Information is Physical", in *Proceedings of PhysComp, Workshop on Physics and Computation*; IEEE, pp. 1–4, Computers Society Press: Los Alamitos, CA, USA.
49. Landauer, Rolf (1996) "The Physical Nature of Information", *Physics Letters A*, 217, 188–193.
50. Leff, H.S. and Rex, A.F. (1990), *Maxwell's Demon: Entropy, Information, Computing*, Princeton, New Jersey: Princeton University Press.
51. Leff, H.S. and Rex, A. (2003) *Maxwell's Demon 2: Entropy, Classical and Quantum Information, Computing*, 45–72, Institute of Physics Publishing: Bristol, UK.
52. Loewer, Barry (2020) "The Mentaculus Vision", in: Valia Allori (ed.), *Statistical Mechanics and Scientific Explanation: Determinism, Indeterminism and Laws of Nature*, forthcoming
53. Maroney, Owen J.E. (2005) "The (Absence of a) Relationship between Thermodynamic and Logical Reversibility", *Studies in History and Philosophy of Modern Physics*, 36, 355–374.
54. Maroney, Owen, J.E. (2009a) "Information Processing and Thermodynamic Entropy", in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2009 Edition).
55. Maroney, Owen J.E. (2009b) "Generalising Landauer's Principle", *Physical Review, E*, 79, 031105.
56. Massimi, Michela (2018) "Four Kinds of Perspectival Truth", *Philosophy and Phenomenological Research*, 96, 342–59.
57. Melnyk, Andrew (1996) "Searle's Abstract Argument Against Strong AI", *Synthese*, 108, 391–419.
58. Miłkowski, Marcin (2013) *Explaining the Computational Mind*, MIT Press.
59. Millhouse, Tyler (2019) "A Simplicity Criterion for Physical Computation", *British Journal for the Philosophy of Science* 70, 153–178.
60. Papineau, David (1993) "Physicalism, Consciousness and the Antipathetic Fallacy," *Australasian Journal of Philosophy* 71, 169–183.
61. Papineau, David (2002) *Thinking About Consciousness*, Oxford University Press.

62. Piccinini, Gualtiero (2008) "Computation without Representation", *Philosophical Studies*, 137, 205–241.
63. Piccinini, Gualtiero (2015) *Physical Computation*, Oxford: Oxford University Press.
64. Piccinini, Gualtiero (2017) "Computation in Physical Systems", in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2017 Edition).
65. Putnam, Hilary (1967) "Psychological Predicates," in: William H. Capitan and Daniel D. Merrill (Eds.), *Art, Mind and Religion*, pp. 37–48, Pittsburgh, PA: University of Pittsburgh Press. Reprinted as "The Nature of Mental States," in Putnam, H., *Mind, Language and Reality*, pp. 429–40, Cambridge: Cambridge University Press (1975).
66. Putnam, Hilary (1988) *Representation and Reality*, Cambridge, Mass: MIT Press.
67. Rescorla, Michael (2017) "The Computational Theory of Mind", in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition).
68. Schuetz, Mathias (2012) "What is it Not to Implement a Computation: A Critical Analysis of Chalmers' Notion of Implementation", *Journal of Cognitive Science*, 13, 75-106.
69. Schweizer, Paul (2016) "In What Sense Does the Brain Compute?", in V. C. Müller (ed.), *Computing and Philosophy: Selected Papers from IACAP 2014*, pp. 63–79, New York: Springer.
70. Schweizer, Paul (2019) "Computation in Physical Systems: A Normative Mapping Account", in D. Berkich and M. V. d'Alfonso (eds), *On the Cognitive, Ethical, and Scientific Dimensions of Artificial Intelligence: Themes from IACAP 2016*, pp. 27–47, Berlin: Springer.
71. Searle, John (1992) *The Rediscovery of the Mind*, Cambridge, Mass: MIT Press.
72. Shagrir, Oron (2001) "Content, Computation and Externalism", *Mind*, 110, 369-400.
73. Shagrir, Oron (2005) "The Rise and Fall of Computational Functionalism", in Yemima Ben Menahem (ed.), *Contemporary Philosophy in Focus: Hilary Putnam*, pp. 220-250, Cambridge University Press.
74. Shagrir, Oron (2012) "Can a Brain Possess Two Minds?" *Journal of Cognitive Science*, 13, 145-165.
75. Shagrir, Oron (2020) "In Defense of the Semantic View of Computation", *Synthese*, 197, 4083–4108
76. Shagrir, Oron (2021) *The Nature of Physical Computation*, Oxford: Oxford University Press, forthcoming.
77. Shenker, Orly (2017a) "Foundations of Statistical Mechanics: Mechanics by Itself", *Philosophy Compass*, 12 (12): e12465.
78. Shenker, Orly (2017b) "Foundations of Statistical Mechanics: the Auxiliary Hypotheses", *Philosophy Compass*, 12 (12): e12464
79. Shenker, Orly (2017c) "Flat Physicalism: Some Consequences", *Iyyun: The Jerusalem Philosophical Quarterly*, 66, 211–225.
80. Shenker, Orly (2020) "Foundations of Quantum Statistical Mechanics", in: Eleanor Knox and Alastair Wilson (eds.), *Routledge Companion to the Philosophy of Physics*, Oxford: Routledge, forthcoming.
81. Sklar, Lawrence (1993) *Physics and Chance*, Cambridge: Cambridge University Press.
82. Sprevak, Mark (2010) "Computation, Individuation, and the Received View on Representation", *Studies in History and Philosophy of Science*, 41, 260–270.

83. Sprevak, Mark (2018) “Triviality Arguments about Computational Implementation”, in M. Sprevak and M. Colombo (eds.), *Routledge Handbook of the Computational Mind*, pp. 175–91, London: Routledge.
84. Stoljar, Daniel, "Physicalism", in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition).
85. Szabo, Laszlo, E. (2012) “Mathematical Facts in a Physicalist Ontology”, *Parallel Processing Letters*, 22: 1240009.
86. Uffink, Jos (2001) “Bluff Your Way in the Second Law of Thermodynamics”, *Studies in History and Philosophy of Modern Physics*, 32, 305–394.
87. Uffink, Jos (2007) “Compendium to the Foundations of Classical Statistical Physics”, in: J. Butterfield, and J. Earman (eds.), *Handbook for the Philosophy of Physics, Part B*, pp. 923–1074, London: Elsevier.
88. Uffink, Jos (2017) "Boltzmann's Work in Statistical Physics", in Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2017 Edition).
89. Von Neumann, John ([1932] 1955) *Mathematical Foundations of Quantum Mechanics*, Princeton, NJ: Princeton University Press. Reprint. Translated by R. T. Beyer. Originally published as *Mathematische Grundlagen der Quantenmechanik*, Berlin: Springer (1932).