Efficient Mechanisms

Jorge I. Fuentes

Please cite from https://www.tandfonline.com/doi/full/10.1080/09515089.2023.2193216

Abstract

A distinguishing feature of neural computation and information processing is that it fits models that describe the most efficient strategies for performing different cognitive tasks. Efficiency determines a distinctive sense of teleology involving optimal performance and resource management through a specific strategy. I articulate this kind of teleology and call it *efficient teleological function*. I argue that efficient teleological function is compatible with mechanistic explanation and, most likely, neural computational mechanisms are efficiently functional in this sense. They are members of a distinctive class of computational mechanisms whose efficiency is intertwined with their functionality. This is illustrated by widely discussed approaches to mind, such as Barlow's efficient coding hypothesis or the ones associated with the so-called "predictive mind", which propose that the brain employs very efficient coding strategies to save energy resources that are critical to the organism's survival.

Keywords: teleological functions, neural mechanisms, neural computation, optimality explanations, efficient coding, predictive mind.

1. **Introduction**

Functional mechanisms (Garson 2013) – those assigned with a purpose, as when we say that ribosomes' end is to synthesize proteins out of mRNA's information – can perform their teleological functions (teleofunctions from now on) more or less efficiently. In this paper, I argue that efficiency considerations may enrich mechanistic explanations by delivering what I call *efficient mechanistic explanations*.

Crucial hypotheses in cognitive neuroscience assume that the brain's computational teleofunctions are performed in an optimal manner, namely, in a manner that optimizes a tradeoff between variables of performance and variables of resources. Examples include Barlow's efficient coding hypothesis, design explanations, and the predictive mind approach to cognition. Some philosophers (most notably, Chirimuuta 2014, 2018) argue that the new mechanism cannot account for the central role that optimality explanations play in cognitive neuroscience. Chirimuuta calls these explanations *why explanations* since they account for the reason a function is performed using a determinate strategy and contrasts them with the *how explanations* that mechanisms can provide. By contrast, Wajnerman Paz (2017) claims that every optimality explanation in cognitive neuroscience requires a mechanistic explanation of how the function is performed. Otherwise, we could not assess its optimality. Wajnerman Paz concludes that every explanation in neuroscience has at least one *how* part and that *why* explanations can be integrated with that mechanistic part in a yet-to-be-elucidated way. I will extend Wajnerman Paz's reasoning and argue that this integration aims to account for two different aspects of functional mechanisms: (1) how the function is carried out (mechanistic aspect), and (2) why the function is carried out in that way (teleofunctional aspect). Thus, efficiency considerations are not only compatible with teleofunctional mechanistic explanations in cognitive neuroscience, *contra* Chirimuuta, but an integral part of the latter.

Furthermore, neurocognitive functions are widely supposed to be computational in a non-trivial way, and the notion of teleofunction is pivotal in the mechanistic approach to physical computation (e.g., Piccinini 2015, 2020; Coelho Mollo 2018, 2019). Thus, the connection between teleofunctions and physical computation has been previously established by Piccinini (2007) as a strategy for avoiding limited pancomputationalism, that is, the view that every physical system implements at least one computation. Thus, Piccinini (2007, 111-12) claims that

> [M]ost systems, such as planetary systems and the weather, are not functional systems—they are not subject to mechanistic explanation in terms of the functions performed by their components. Computational explanation does not apply to any of those, because computational explanation applies only to a subclass of functional systems.

I agree with Piccinini that this strategy avoids limited pancomputationalism and is the best available strategy for this purpose[1]. Of course, this strategy alone does not yet delimit

---

[1] Piccinini distinguishes between limited and unlimited pancomputationalism and argues that to rule out the latter, in which every physical system is said to implement any computation, a causal account of computation

computational mechanisms. Given the wide variety of functional mechanisms (e.g., the spleen, the kidneys, a car, the visual system, a laptop, etc.), we still need to determine which are computational and in virtue of what. Piccinini's goal-contribution approach to teleofunctions maintains that a functional mechanism is one that regularly contributes to a goal of organisms through the exercise of its activities. Additionally, Piccinini (2020, 144) claims that when those activities can be characterized as the manipulation of medium-independent variables according to a rule defined over those variables, we have a mechanism with a computational function.

The goal-contribution approach to functionality has several advantages over previous accounts. Specifically, it allows us to make functional assignments that do not depend on evolutionary history, which in most cases is unknown to the researcher. Also, it can provide a unified concept of teleofunction despite the differences between the selective processes underlying biological systems and manufactured artifacts. Defining a unified concept of function is of great importance for our purposes since the notion of computation can be equally applied to artificial and biological cognitive systems. The latter will allow us to state a computational theory of neural processing that is not vacuous. I will argue that this approach can be further enhanced by adding efficiency constraints that are important when specifically considering neural computation.

---

suffices. However, a normative sense of function is required to rule out the limited version. Otherwise, a planetary system could be said to be computing the solution to Newtonian differential equations. Only once we have formulated this normative sense can we state that planetary systems lack the function to compute.

As emphasized by many, including both Barlow (1961) and Chirimuuta (2014, 2018), one of the most salient qualities of neural processing is that it fits models that set out the most efficient strategies for performing different cognitive tasks. As I will define it, an *efficient teleofunction* depends not only on the disposition to contribute to a goal but also on doing so through an optimal strategy. The optimized quantities of an efficient teleofunction generally result from a tradeoff between resource and performance variables.

I claim that it is very likely that neural computational mechanisms are functional in this particular sense, i.e., they have efficient teleofunctions. This could be exemplified by what Sterling and Laughlin (2015) call “principles of neural design”. Additionally, this is the notion of functionality presupposed by widely endorsed approaches to neurocognitive processes, such as those associated with the so-called “predictive mind”. Predictive mind approaches to cognition postulate highly efficient coding strategies that minimize the energy resources employed in each of the brain’s tasks.

The paper will proceed as follows. In section 2, I will critically analyze some accounts of teleofunction that precede Piccinini’s approach and argue that his approach is indeed a step forward. This critical assessment will mainly focus on the usefulness of these accounts for delimiting computational mechanisms in cognitive neuroscience. Subsequently, in section 3, I will provide evidence that efficient design explanations are ubiquitous in neurocomputational explanations. Next, in section 4, I will characterize a notion of efficient teleofunctional mechanism and efficient computational mechanism based on optimality considerations. Additionally, I shall argue that this approach provides a solid basis for identifying the distinctiveness of computational neural teleofunctions. Finally, in section 5, I will address three objections to the claim that teleofunctions can be used to avoid pancomputationalism.

2. The accounts of functionality that preceded the goal-contribution approach

A very influential philosophical approach to function attribution is the causal role theory (e.g., Cummins 1983; Craver 2007, 2013). According to this theory, the attribution of functions depends on the researcher's interest in a particular causal role or phenomenon produced by a system. This causal role theory is a kind of function perspectivalism. Perspectivalism implies, on the one hand, a triple synonymy between function, explanandum phenomenon, and behavior of a mechanism. On the other hand, it entails that the concept of function is not normatively charged. By contrast, normative or teleological accounts about functions are a family of accounts that provides non-subjective and epistemically transparent norms to identify cases of correct functioning of a mechanism as opposed to cases of malfunctioning and to distinguish what the mechanism is supposed to produce from its by-products.

To be clear, the debate concerning normativity that I am going to be analyzing here occurs inside the new mechanical philosophy plus its philosophical antecedents. Among mechanists, it is a common assumption that the concepts of mechanism and phenomenon are intimately linked. It is claimed that mechanisms explain phenomena, and every mechanism is the mechanism of a phenomenon (e.g., Craver 2007, Glennan 1996, Garson 2017). Thus, a researcher focuses her attention on a specific phenomenon for which she wishes to find a scientific explanation, e.g., gene expression, and performs experiments that allow her to discover the entities involved (DNA, mRNA, ribosomes, amino acids, etc.), their activities (translation, transcription, etc.), and their modes of organization. This way, the entities, activities, and modes of organization that give rise to the phenomenon together form what we know as the mechanism underlying that phenomenon. Moreover, the boundaries of a mechanism must be fixed by its relation to its phenomenon (i.e., the mechanism is constituted only by the components, activities and organizational properties required to produce the phenomenon) (Craver 2007, 123).

Additionally, a mechanistic explanation can be related in two different ways to its explanandum phenomenon. Some mechanisms produce phenomena while others underlie phenomena (e.g., Craver 2001, 68-70; Glennan 2017, 109). When the phenomena are produced, the explanans is temporally previous to the explanandum. For example, high consumption of refined sugar in childhood and adolescence can lead to type two diabetes mellitus in an adult. In contrast, the explanandum and the explanans are simultaneous in the cases of mechanisms that underlie phenomena. The latter are the activities of the parts organized to produce the explanandum. An example mentioned above is gene expression, in which the various entities involved provide sub-processes that constitute that phenomenon.

In cognitive neuroscience, mechanistic explanations are integrated into multilevel explanations of cognitive phenomena. Boone and Piccinini (2016) describe this as follow:

> Cognitive neuroscientists study nervous systems using many techniques at many levels. They study how cortical areas and other neural systems contribute to various cognitive capacities, how the capacities of those systems are explained by the operations of the neural subsystems that compose them (columns, nuclei), how networks and circuits contribute to their containing systems, how neurons contribute to networks and circuits, and how subneuronal structures contribute to neuronal capacities.

Furthermore, the term “mechanism” can be used in either a minimal or a functional sense (Garson 2017). The minimal sense is the only sense accepted by perspectivalists. In this minimal sense, we can speak, for example, of a disease mechanism or a laptop as a mechanism for producing heat from electrical energy. As I said, it all depends on the researcher’s interests. In contrast, the functional sense of mechanism (Garson 2013) is normative and prevents speaking of diseases as the function of a mechanism. Rather, these might be understood as the breakdown of a mechanism. In the case of the laptop mentioned above, heat must be understood as a by-product, not as its function. This notion of function is narrower than the minimal one. For this reason, Garson claims that functional mechanisms are a proper subset of all minimal mechanisms.

In synthesis, the normative aspect of functional mechanisms allows us to identify malfunctions (e.g., pathology) and by-products (e.g., heat produced by a laptop). By contrast, the minimal sense of mechanism lacks the necessary tools to identify them (Garson 2013). For this reason, the causal role theory or perspectivalism, even when it is a theory about functions, cannot support the functional sense of mechanism that I will defend here.

Moreover, assuming perspectivalism would entail two undesirable consequences for a theory of computational mechanisms in cognitive neuroscience. Firstly, it would not allow us to distinguish between the physical systems to which we can objectively attribute a computational function and those to which we cannot, because this distinction would depend on the researcher's perspective[2]. Secondly, it would not allow explaining the malfunctioning of computations as a case of breakdown, since, in this approach, it would suffice that the pathological behavior is taken as a function according to the scientist's perspective.

---

[2] Bear in mind that Piccinini introduces normative function as a criterion to rule out some systems as computational. For example, why are planets not computationally solving the Kepler problem when orbiting the sun? Because planets are not functional mechanisms. We can decide that by using an objective, non-perspectival, criterion. This way, we see that certain causal transitions are insufficient to attribute computation. Therefore, we need teleological function as a basis for computational ascription under this approach. For this discussion, see Piccinini (2007, 2015, 2020).

Please note that, as Garson (2017) points out, the functional sense of mechanism is actually a family of senses that depend on the underlying theory of normative functions. The only constraint on a legitimate functional sense of mechanism is that this theory must be normative (or teleological).

In the remainder of this section, I will analyze three normative theories about functions, namely, the unified selected effects theory (Millikan 1995), the restricted selected effects theory (Garson 2016), and the goal-contribution theory (Piccinini 2015, 2020). The latter overcomes many difficulties that the previous accounts have with formulating an adequate computational theory of neural processing.

Let us begin by analyzing Millikan's unified selected effects theory. Unlike the other two theories I will present below, this theory is not framed within the new mechanistic philosophy. Even so, being a theory about the functionality of biological traits and (potentially) artifacts, it has no principled tension with mechanistic philosophy. Thus, a functional sense of mechanism based on Millikan's proposal is possible in principle.

Here is Millikan's (1995, 13-14) definition of proper function:

> To put things very roughly, for an item A to have a function F as a "proper function", it is necessary (and close to sufficient) that one of these two conditions should hold. (1) A originated as a "reproduction" (to give one example, as a copy, or a copy of a copy) of some prior item or items

> that, *due* in part to possession of the properties reproduced, have actually performed F in the past, and A exists because (causally historically because) of this or these performances. (2) A originated as the product of some prior device that, given its circumstances, had performance of F as a proper function and that, under those circumstances, normally causes F to be performed by *means* of producing an item like A.

The heart of Millikan's proposal is that function ascriptions should be made by looking at history rather than the current dispositions of an item. Of course, suppose the researcher does not know the history of an item, which is the norm in biology. In that case, she can infer from current dispositions that it is highly likely that the item has a certain reproductive history. The more complex the item, the higher the probability that the item has not arisen by chance but has undergone a process of selective reproduction. In the case of biological traits, the likelihood is even higher, given their enormous complexity. Thus, Millikan (1995, 18-19) suggests that the probability of error in the case of highly complex items is negligible. So, we can assume, with very little fear of error, the existence of a selective process and hence of a proper function even if we do not know the details of such a process.

Millikan's proposal has been resisted by Garson (2016). Garson objects to the underlying presupposition of unity between the selective processes of artifacts and biological traits. According to Garson, such processes possess only a distant analogy, a claim he proves by appealing to five differentiating criteria. First, changes between versions of an artifact are generally abrupt. This contrasts with the gradualness of evolutionary changes in the traits of biological organisms. Second, the species concept can identify members of a biological type with relative simplicity. In contrast, it seems arbitrary to identify a collection of artifacts as belonging to successive generations of the same artifact. Third, living organisms possess replication autonomy, whereas artifacts depend on an agent with technical expertise for their production. Fourth, artifacts have nothing remotely resembling random genetic mutation and the transmission of genetic material from organisms better suited to subsist in the environment. Finally, the first version of an artifact is already considered functional. It may even be the final version of the artifact. The first appearance of a biological trait is aleatory, and if history is the criterion for teleological adjudications, we cannot consider such a trait functional. I take Garson's arguments to be decisive in rejecting Millikan's proposal as purpoting to unify natural and artificial functions.

Because both selective processes possess no real unity, Garson argues for both a pluralism of teleofunctions (Garson 2017) and a theory of selected effects restricted to biological function (Garson 2016). This theory appeals to reproductive history when adjudicating functions but, as its name suggests, only does so in the case of biological traits and is silent on artifact functionality.

Piccinini (2015, 2020) criticizes both the causal role theory and the selected effects theory (without distinguishing the unified version from the restricted version). According to Piccinini, the causal role theory faces two main problems. First, since it is not normative, it prevents us from distinguishing objectively functional mechanisms, and therefore, it does not help to avoid undesirable consequences such as pancomputationalism. Secondly, it does not allow us to account for cases of malfunctioning since, as I have already pointed out, malfunctioning could be considered a distinct explanandum phenomenon. Thus, because the phenomenon individuates the mechanism, the causal role theory postulates two mechanisms where there is only one that, in some cases, works appropriately and in others fails. Accounting for computational failure is a necessary condition for the success of any theory of physical computation.

Against the theory of selected effects, Piccinini and Maley (2017)[3] present two arguments, one ontological and the other epistemological. On the one hand, the epistemological argument holds that the causal histories of the selection of functional features of a system are, in many cases, unknown or, in the worst case, impossible to know. Consequently, they argue that it is impossible to attribute a function to a system under the theory of selected effects in many cases. In fact, in methodological practice, biologists do not resort to the evolutionary history of a trait or a functional part of an organism to attribute current functions to traits. On the other hand, Piccinini and Maley's ontological argument is based on a thought experiment based on the logical possibility that a tiger would emerge in a

---

[3] Also in Piccinini (2015)

swamp after a lightning strike (this is an adapted version of Davidson's swampman). While this seems physically or biologically impossible, it does not seem logically impossible, at least in principle, i.e., it is not a self-contradictory possibility. This tiger's stomach would have the teleofunction of digesting, even though it has no evolutionary ancestors. Therefore, we can logically distinguish between function and selective history. Although empirically connected, these concepts are not necessarily conceptually related.

As a result of the problems Piccinini identified in the causal role theory and selected effects theories for teleofunctions, he proposed his own approach based on the contribution of the causal roles of physical systems to the goals of organisms. Thus, his definition of teleofunction (Piccinini 2020, 76) is as follows:

> Tokens of type X have function F if and only if F is a causal role and performing F by tokens of X provides a regular contribution to a goal of organisms.

It is helpful to clarify the concept of goal introduced by Piccinini. He classifies the goals of organisms into two main types, namely, biological and non-biological goals. Biological goals are survival, development, reproduction, and mutual aid (Piccinini 2020, 76). The organism must invest energy in pursuing these goals so that its species does not become extinct. In contrast, non-biological goals arise in organisms that possess an advanced degree of sentience and sapience. Examples of the latter are pleasure, avoidance of pain, knowledge, truth, beauty, justice, wealth, and the pursuit of power (Piccinini 2020, 71).

Piccinini’s proposal has the advantage of unifying the teleological functions of artifacts and biological traits without assuming the unity of selective processes. For this reason, it is not subject to Garson’s objections regarding the unity of biological and artificial functions. This unification is vital to the foundation of a computational theory of neural cognitive processing. If it turns out that the computational functions of artifacts (e.g., a laptop) and those of bio-cognitive systems (e.g., the visual system) do not belong in the same class, then this theory would turn out to be an empty hypothesis.

As said before, according to Piccinini (2020, 144), the function of all physical computing systems is to manipulate medium-independent variables in accordance with a rule defined over the variables. On the one hand, artificial computational systems will serve a user’s goals. These goals could be biological (e.g., a demand pacemaker contributing to survival) or non-biological (e.g., a video game console contributing to entertainment). On the other hand, bio-cognitive systems could also serve biological goals (e.g., the visual system contributing to survival) or non-biological goals (e.g., the same system admiring the beauty of nature) of the organism that possesses them. However, in all these cases, the contribution to these goals will be made by performing the same type of computational function, i.e., the manipulation of medium-independent variables.

The diagram in figure 1 briefly outlines the theories about the functions I have presented so far.

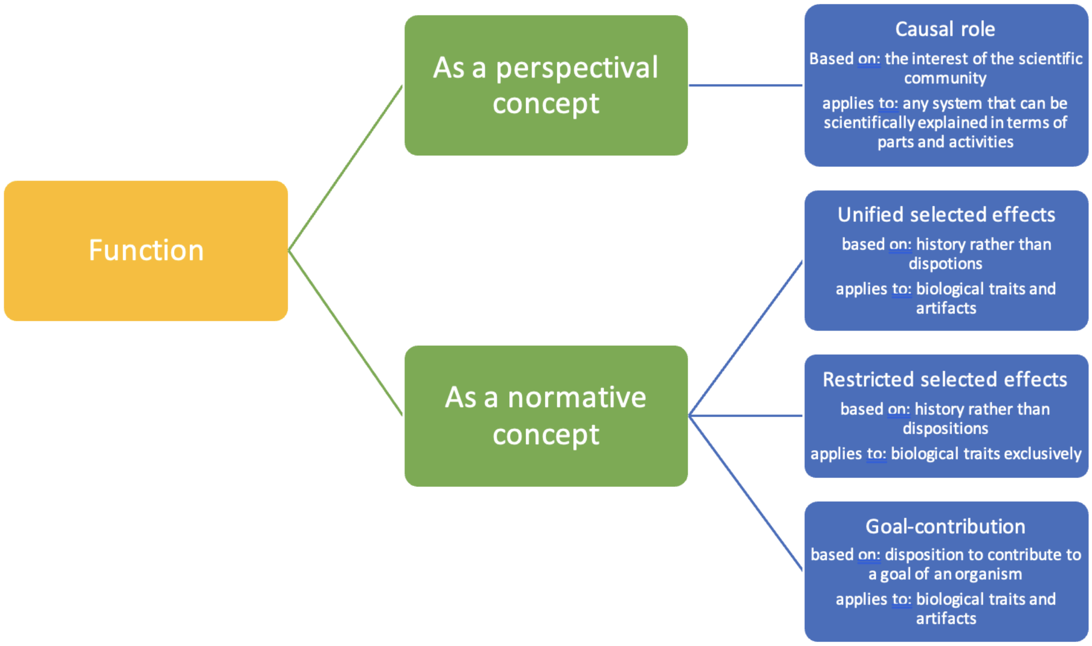

Figure 1: Theories about the functions of mechanisms

To summarize, in the first place, perspectivalism lacks the tools to identify functional mechanisms objectively. This objectivity is necessary for any mechanistic account of computation that avoids pancomputationalism. In the second place, Garson's arguments against the unified selected effects theory are conclusive as he establishes that natural and artificial selective processes are essentially different. Finally, in the third place, the restricted selected effects theory does not allow us to account for a unified computational function between artifacts and biological traits. Consequently, if we adopt such a theory of what a neurocognitive function is, we cannot state a non-vacuous computational theory of neurocognitive processing. Piccinini's goal-contribution approach addresses all these difficulties. As I will argue in the next section, it can be complemented with efficiency considerations that target specific aspects of neural processing.

3. Optimality explanations in cognitive neuroscience

Efficiency can be crucial for teleology since function attribution sometimes depends on whether a system's capability can lead to an organism reaching a goal *at an acceptable price*. Paradigmatically, optimality explanations in neuroscience assign a function to a system depending explicitly on the suitability of the strategy observed in a system to fulfill it. Many organized mechanisms could contribute to a goal of organisms. However, some of them can contribute without reaching the minimum threshold needed to fulfill that goal, so in some instances, a contribution is not enough for a mechanism to accomplish certain functions.

The goal-contribution approach has confronted this challenge by distinguishing appropriate performance rates from inappropriate ones (Piccinini and Garson 2014). When a brain mechanism applies a strategy that does not seem optimal in this respect, neuroscientists often assume that it is essential to amend the initial functional ascription. It is a methodological assumption of *optimality explanations* in neuroscience that biological mechanisms almost always employ (nearly) the most efficient of all available strategies to fulfill a particular goal. This is a heuristic principle grounded on facts about natural selection and organism development, which matches scientific practices of teleological ascriptions on efficient coding and other kinds of optimality explanations.

I am not claiming that everything we call the function or purpose of an artifact, or of a biological trait, has efficiency as a requirement. Still, it seems clear to me that this is an important requirement for at least some cognitive functions and, crucially, for some neurocomputational functions, as has been noted by Barlow, Chirimuuta, and Sterling and Laughlin, just to name a few. Therefore, my proposal in section 5 will focus on functions that require (near) the highest degree of efficiency in their execution.

My purpose in this section is to show that an important range of explanations in neuroscience, most notably including efficient coding explanations, *seems* to be elusive to the mechanistic approach to scientific explanation. In the subsequent section, I will show that these explanations can be integrated into the teleofunctional mechanistic approach. Thus, while constitutive explanations refer to how a neurocognitive system carries out a teleological function, optimality explanations show why the function is carried out by a neurocognitive system with such characteristics. In other words, there are important cases in which optimality explanations allow us to identify the systems suitable for a given functional task, i.e., the teleofunctional mechanisms.

Optimality explanations based on Barlow's efficient coding hypothesis were inserted into the philosophical debate by Chirimuuta (2014, 2018) to discuss the nature of computational explanation in cognitive neuroscience. A shortcoming that Chirimuuta (2014, 131) attributed to mechanistic models of explanation is that they do not account for the functional nature of neurocognitive systems:

> imagine if neuroscientists had only descriptive and mechanistic models of the action potential. They would know what an action potential is, and how it is generated, but that does not entail that they know anything of its function – what it's for […] A central dogma of neuroscience is that the action potential is for information transmission, and that information theory and computational principles are needed to understand why action potentials have particular patterns of generation.

Chirimuuta is aware of the connection between teleological functions (*what it is for*) and optimality explanations. However, Chirimuuta does not seem to consider efforts to combine teleology and mechanism (e.g., Coelho Mollo 2018; Garson 2013; Piccinini 2015, 2020). On the other hand, none of these positions on functional mechanisms addresses explicitly the relationship between optimality and teleology that Chirimuuta remarks as central. Her main thesis is that optimality explanations are autonomous and independent of mechanistic explanations, and, for that reason, some neurocognitive explanations are not mechanistic in nature.

As a result, the relationship between mechanistic and optimality explanations, or more specifically, how the latter could explain the normative functional character of the first, has not been studied with the degree of attention required. On the one hand, those who consider teleological function and mechanism together do not consider the connection between this pair of concepts and optimality explanations. On the other hand, Chirimuuta, who does contemplate this connection, believes that optimality explanations are not related to mechanistic explanations. My goal is to show not only that optimality and mechanistic explanations are interdependent and complementary but also that optimality explanations could play a key role in mechanistic function attribution.

Chirimuuta distinguishes three possible explanatory models in cognitive neuroscience to characterize optimality explanations, namely, the interpretive model, the phenomenal model, and the mechanistic model (2014, 128). Interpretive models would answer the question about why systems are structured the way they are, phenomenal models explain what is there, and mechanistic models explain how systems work. Thus, Chirimuuta (2014, 143) defines an interpretative minimal model (I-minimal Model) as follows

> Models which ignore biophysical specifics in order to describe the information-processing capacity of a neuron or neuronal population. They figure in computational or information-theoretic explanations of why the neurons should behave in ways described by the model.

What is at stake is not that the response properties of bio-cognitive systems must be optimal but that comparing the actual system with a hypothetical optimal system is relevant in explanatory terms (Chirimuuta 2014, 146). For example, it is known that the brain is the organ that consumes the most ATP in organisms. Therefore, as Barlow's hypothesis suggests, to perform its functions, it must adopt neuronal coding strategies that optimize the available metabolic resources so that the organism can stay alive. Thus, efficient coding explanations account for the fact that the system employs a certain computational or coding strategy and not another due to resource constraints. As Chirimuuta emphasizes, the epistemic rules of this type of explanation are very different from those of ordinary mechanistic explanation, whether constitutive or etiological.

Chirimuuta (2014) also argues that computational strategies do not depend on their biological substrate. The same strategies can be realized by very different neural mechanisms. This is illustrated through the case of so-called "canonical neural computations" (CNC). CNCs correspond to certain types of cognitive modules with which the brain operates in a variety of contexts (Carandini and Heeger 2012). Since the biophysical properties described by CNC models seem to lack the degree of detail required for mechanistic explanation, Chirimuuta claims that the mechanist is forced to qualify this kind of explanation as sketches of mechanisms. However, the explanations provided by CNC models are recognized as legitimate and complete by computational neuroscientists. Therefore, mechanism would be unable to account for key explanatory practices in cognitive neuroscience.

In response, Wajnerman Paz (2017) argues that (a) the efficient coding explanation of why a system recruited a given computational strategy to perform a given function necessarily presupposes an explanation of how the system performs that function (that is, what strategy is actually recruited); (b) pace Chirimuuta, this 'how' model should not be considered a sketch, but rather a mechanistic model in the full sense; and (c) optimality explanations can be integrated with mechanistic explanations to provide a richer perspective on neurocognitive phenomena, that is, they are not explanations of different phenomena but rather different perspectives on a single phenomenon.

To be sure, (a) is only claiming that no evaluation of an actual strategy's optimality could be made if we cannot have access to that strategy. The points of controversy are (b) and (c). To defend (b), Wajnerman Paz analyzes the case of normalization, a particular CNC, and argues that the different symbolic expressions in the model's equation (see below) stand for a series of *causal* roles. For example, the summation in the denominator describes a *suppressive effect* schematically.

$$\overline{E}_i(t) = \frac{E_i(t)}{\sigma^2 + \Sigma_i E_i(t)}$$

Moreover, Wajnerman Paz (2017, 169) adds,

> [T]he omission of information in this model is not due to an imprecise knowledge of the relevant mechanism. We have seen that mechanisms implementing normalization, such as shunting inhibition or synaptic depression, are well known. Furthermore, the model has the features

> associated with abstract mechanistic models (what Chirimuuta calls “A-minimal models”). The core idea behind these models is that they must only describe the aspects of a mechanism that are difference makers for a relevant phenomenon.

The last sentence of the above quote represents the principle that regulates the maximum possible omission of detail in mechanism types, i.e., the LB principle (after Levy and Bechtel 2013). Difference makers are features such that, when they are changed or replaced, the behavior of the mechanistic system is modified.

As a corollary of (b), Wajnerman Paz (2017, 173) states that both approaches, the efficient coding and the mechanist, are appropriate perspectives about what happens in implementing CNCs. This is what (c) states. However, he cautions that the idea of different perspectives says nothing about how these should be integrated. With that in mind, I will propose how such integration should be conceived in the next section.

According to my proposal, functional mechanisms require both types of explanation. On the one hand, they require a mechanistic explanation, either constitutive or etiological. On the other hand, they require a teleological explanation, which accounts for the assignment of a function to the mechanism. This is accomplished in some cases through an optimality explanation. These cases include neurocomputational mechanisms.

4. Efficient computational mechanisms and neurocognitive sciences

This section has two purposes. The first is to provide a definition of efficient teleofunctional mechanisms and determine how this definition interacts with the goal-contribution approach to teleofunctional mechanisms. The second goal is to show that the notion of efficient teleofunctional mechanisms can be used to characterize a class of computational mechanisms, namely, *efficient* computational mechanisms, in which neural computational mechanisms seem to be included. Characterizing neural computation as efficient computation provides a deeper insight into its operations than approaches exclusively focused on features of its vehicles, such as medium independence (e.g., Piccinini and Bahar 2013; Piccinini 2015, 2020).

I propose the following definition:

> **A mechanism M is an efficient teleofunctional mechanism iff** (1) M has the disposition to perform a particular task T with the regularity required by the functionality of the upper-level mechanism in the mechanistic hierarchy of which it is a part and (2) M performs T by using a strategy S that optimizes physical variables {q1, q2,…,qn} that define resources for doing T and/or T performance-related properties.

We say that doing T through S is the efficient teleofunction of the mechanism. The mentioned physical variables generally interact in a tradeoff. The first type is constituted by quantities that measure a resource taken from the environment of the mechanism (e.g., energy, materials), while the other type is constituted by quantities that affect the quality of the performance of the mechanism's task (e.g., response time, noise reduction in an information channel).

Note that, if optimality considerations directly imply facts about fitness, we can bypass goals, as conceived by the goal contribution account, in assigning efficient teleofunctions. The reason is that one of Piccinini's biological goals is precisely fitness. Nevertheless, the relationship between optimality and fitness is neither analytical nor purely conceptual, nor mathematically linear, but a complex contingent relationship (for discussion, see Roffé and Ginnobili 2019). We can make inferences from optimality to fitness, and it is likely that more optimality entails more fitness. However, we must discover empirically how they relate to each other in each particular situation.

I should reemphasize that I am not claiming that efficient teleofunctions could cover all the functionality cases in all teleofunctional mechanisms. I am saying that efficient teleofunctions are an important case of functionality because: (1) they seem to cover the type of functionality that neurocomputational mechanisms have, and (2) efficient functions emphasize an essential characteristic of these mechanisms, namely, that they are constrained to be highly efficient for biological reasons.

Efficient functions are normative in two ways. They regulate not only what a mechanism is for (performing T) but also the strategy S to achieve it. Efficient teleofunctions, as defined, are epistemically transparent in the conditions they propose for teleofunctional attribution as they depend on whether a mechanism employs an optimal strategy in executing a task. If not, then no efficient teleofunction could be attributed to the mechanism. Crucially, as Chirimuuta remarks, this is the type of functionality at stake in many computational explanations in neuroscience of why the mechanisms have the structure they actually possess.

Furthermore, if we combine the definition of efficient functional mechanisms with an adjusted version of Piccinini's (2020, 144) definition of computational function, we can define efficient computational mechanisms:

> **A mechanism M is an efficient computational mechanism iff** (1) M is an efficient teleofunctional mechanism, (2) M's task T is to manipulate medium-independent variables in accordance with a rule defined over those variables, and (3) M performs T through a strategy S that optimizes design variables {q1,q2,…, qn} some of which characterize signal processing parameters (e.g., redundancy, transmission rate, noise, etc.).

The brain’s efficiency in processing information is apparently determined by long-standing evolutionary and individual developmental processes of organisms situated in an environment. For the current account, it is sufficient to take its efficiency as a given fact without delving into how it emerged. Also, the efficiency of neural computation is a key theoretical assumption of current widely endorsed models, such as those that belong to the predictive coding framework (see Ali et al. 2022), which are supposed to characterize the mechanisms underlying a significant variety of cognitive domains (e.g. Friston 2010; Clark 2013; Hohwy 2014; Clark 2015; Buzsáki 2019). I will briefly review these approaches and how they presuppose the efficiency of brain processes.

During the second half of the twentieth century, many computer engineers worked on the problem of compressing the size of images or sounds. The practical challenge was to fit as many images or sounds as possible into a computer storage device or a fraction of it. With this goal in sight, they developed a series of strategies that exploited redundancies in natural images or sounds to code only the essential features that allow the computer to reconstruct them (cf. Shi and Sun 1999). For instance, in the case of black and white natural images, it is not necessary to encode the brightness of every single pixel since most of them determine a narrow range of possibilities for the pixels in their neighborhood. For this reason, these coding regimes are labeled as predictive since they exploit these possibilities for predicting complete patterns from incomplete data and only encode prediction *errors* that occur during this process. This type of processing has inspired hierarchical predictive processing approaches to cognition (Clark 2013). In these, generative models attempt to predict signals from lower processing levels by sending top-down signals to the earliest perceptual levels. The bottom-up signals that are sent back from these lower levels to a higher level encode the divergence between the expected and the actual signal, i.e., the prediction error. What is important about this story for the current discussion is that the optimal coding principles predicted by computational engineering for improving artificial systems coincide with those that some leading cognitive neuroscientists (e.g., Buzsáki 2019) ascribe to neural processing. That is, the predicted optimal strategy seems to match the strategy currently employed by the brain.

But predictive approaches to cognition can explain psychological phenomena beyond visual or auditory perceptual processing. One prominent example is the sense of agency. According to Hohwy (2015a), one of the features that explain that I can feel my body as my own is that my internal priors are much more precise and provide highly accurate predictions about my future states than those made about the temporal evolution of the bodies of other agents. Another important example is emotion (e.g., Seth 2013). Here the key is an internal model that predicts not future states of movement and action, as in the agency case, but the internal states of our body. This approach captures James' (1884) insight that emotions arise from perceptual awareness of changes in our bodies. Other mental phenomena that predictive processing aims to explain, among many others, include conditions such as schizophrenia or autism (e.g., Hohwy 2004; Palmer et al. 2017), consciousness (e.g., Hohwy 2015b), and top-down cognitive modulation (e.g., Vetter and Newen 2014). All these predictive models of the mind or, equivalently, of the brain's functional processing, share the feature of ascribing highly efficient computational strategies to neural processing.

This is not a general definition of computation. For instance, logic gates made with vacuum tubes are computational mechanisms in the usual sense. However, they cannot be considered efficient computational mechanisms because they expend more energy and take longer to perform their operations than a logic gate made with transistors. As a second example, we can mention hard disk drives (HDD) which are digital data storage mechanisms whose response time is significantly longer than solid-state drives (SSD). Thus, a laptop equipped with an HDD is not an efficient digital computing mechanism, though it is a computing one. One equipped with an SSD could be efficient if its other components are optimal and if it turns out that there is no better strategy to optimize response time and energy resources for data storage. Nevertheless, this subclass of computational mechanisms is key because, as I have argued, neural mechanisms appear to be computational in this specific sense.

Based on the examples, it is clear that we often do not know whether the strategy used is the optimal one simpliciter or the optimal one between those available, but this does not always seem to be the case. Theoretically, this could be seen in cases such as the Shannon limit (Shannon 1948), which refers to the maximum rate of error-free data that can be transmitted over a channel if it is subject to random transmission errors. The Shannon limit and other limits postulated by the mathematical communication theory entail objective optimization criteria for some significant informational quantities in computational mechanisms. Similarly, Landauer's principle entails an objective optimization criterion for the theoretical lower limit of energy consumption in a computational system (see Maroney 2009).

In neurobiology, Sterling and Laughlin (2015, 441) conclude their analysis with an important hypothesis regarding what they call the Principles of Neural Design:

> Many neural computations already operate near the lower bounds on space and energy costs defined by thermodynamics.

And they continue below (444)

> So it seems that with regard to brain design, natural selection has followed different routes but reached the same destination. This suggests that the core principles of neural design were established already in their shared ancestors. If we evolve for another 450 Myr, we will certainly look different; we will think and behave differently because we will have continued to be shaped by natural selection. But our brains will probably be no more efficient.

To sum up, it seems to be the case that neurocomputational mechanisms operate, if not in the most efficient way possible, at least very close to it.

Additionally, my definition of efficient computational mechanism fits with the explanatory model of optimality explanations put forward by both Chirimuuta (2014, 2017) and Wajnerman Paz (2017), to which I referred at the end of the previous section. Since the last part of this type of explanation is to check whether the optimal strategy fits the mechanism's actual implementation, we note that this check is an integral part of the assignment of an efficient teleofunction to the mechanism. All of this seems to suggest that the efficiency of brain mechanisms is not just another feature of their processing but one that determines their teleofunctional structure and, consequently, the type of computations they perform, but one that refers to the particularities of their computational processing.

Finally, as we have seen and pace Chirimuuta, the notion of efficient teleofunctional mechanism integrates within a single approach both the mechanistic view of explanation and optimality considerations. While mechanistic explanations explain *how* a computational capacity is carried out, efficient teleofunctional explanations also show *why* that capacity is carried out in that specific manner. Thus, both are complementary when understanding mental capacities in cognitive neuroscience. Moreover, efficient computational mechanisms set an ideal to reach in the design of artificial computers.

Many computer design decisions are aimed at improving their efficiency. So, efficient functionality is vital in cognitive neuroscience and artificial computation.

5. Some objections to teleofunctional approaches to computation

Since I have embraced Piccinini's claim that teleofunctions allow us to separate computing from non-computing physical systems, I must meet recent objections to this claim. Specifically, in this section, I am going to address three objections made respectively by Dewhurst (2016), Coelho Mollo (2019), and Shagrir (2022). In doing so, further advantages of considering efficient teleofunctional mechanisms in cognitive neuroscience will become clear.

5.1 Dewhurst's irrelevance objection

According to Dewhurst (2016, 3),"[t]he fact that the mechanistic structure of an organism contributes to its survival or inclusive fitness does not by itself contribute anything to our understanding of those mechanisms". This view is inconsistent with scientific practice, as Chirimuuta's and Wajnerman Paz's work has shown. As we saw, Chirimuuta's interpretative models aim to explain global structural facts based on optimality considerations. These models are considered to be of explanatory value since they provide the *why* 'part' of the scientific story. As I have argued, this teleological aspect of scientific explanation is an important aspect that must be integrated within the mechanist framework to account for explanations in cognitive neuroscience. Furthermore, Dewhurst argues immediately before as follows:

> Once we have objective teleological functions, it is relatively simple to derive a teleosemantic account of representations, for example by positing mechanisms whose funtion is to represent […] Piccinini's argument becomes self-defeating, for by invoking objective teleology he removes one of the primary motivations for giving a non-representational account of computation in the first place.

I believe that this interpretation is misguided. What the mechanistic account of computation is trying to state is not that semantic properties are not important for cognitive agents, nor that cognition is not essentially representational. The point is rather that some computational mechanisms need not have the function to represent anything external in a relevant sense. To prove the point, it is sufficient to show one mechanism of that sort to prove that semantic properties are not essential to computation, although they can be essential for cognition. Imagine a computer that iterates a recursive numerical succession over an initial state internal to the machine. This device is a case of a computer where no representation of the external world is posed. So the point is made.

5.2 Coelho Mollo's medium-dependence of teleology objection

Coelho Mollo (2019) states that there is a problem with the idea that there are medium-independent teleofunctions. More specifically, while computation is individuated in medium-independent terms, teleology depends on specifics about constitutive physical properties. Further, he claims that to overcome this difficulty, the mechanist philosopher has to accept a version of functionalism, specifically in admitting that medium-independent explanations are not mechanistic.

According to Coelho Mollo's (2018) view on computational mechanisms, there are two types of computational explanation. Thus, he denies that explanations of the CNCs sort are mechanistic, as Wajnerman Paz and I claim. This is because they pose extremely weak structural constraints, and that is not enough to be considered mechanisms properly so called. Instead, computational explanations of that sort are to be treated as functional individuations of the computational strategies being employed. In contrast, explanations of how different specific structures carry these CNCs in the brain are mechanistic. These mechanistic explanations provide the implementational part of the story. Thus, different mechanistic explanations provide different implementational stories about how different mechanisms implement the same computation.

Since Coelho Mollo's remarks coincide with some claims Chirimuuta makes, they have already been responded to by Levy and Bechtel (2013), Wajnerman Paz (2017), and Piccinini (2020), just to give some examples. As we saw, they all claim that computational explanation poses structural constraints that are relevant for the teleofunction of a mechanism. Since the teleofunction of computational mechanisms is defined in medium-independent terms, Coelho Mollo is right that it poses weak structural constraints. Nevertheless, this is no threat to the mechanistic view since all the relevant details for the computation to occur are included. Moreover, in the case of efficient neurocognitive computation, the structural organization is precisely one that is the best for the survival of the organism. So, I agree with Coelho Mollo that computational individuation is made in medium-independent terms and that this is a distinctive characteristic of computational teleofunctions, as opposed to other kinds of teleofunctions. However, since I do not accept the distinction he makes between medium-independent and mechanistic explanations as different types, the tension he tries to establish between teleology and medium-independence does not affect my present proposal.

One thing that remains to be solved is how what Coelho Mollo calls computational individuations and mechanistic implementational stories are to be integrated into a single explanatory frame. Nevertheless, that problem is independent of the one addressed here.

5.3 Shagrir's insufficiency of details objection

According to Shagrir (2022), mechanistic approaches to computation owe us a detailed account of how they correctly classify computing and non-computing physical systems. Specifically, he asks for a specification (ch. 6, sec. 4) of the goals of computing systems and how these goals constrain the individuation of medium-independent processes. Additionally, he claims that mechanistic models downplay the role of informational aspects in cognitive neuroscience:

> an important chunk of computational theory in cognitive science of computational theory in cognitive science and neuroscience is devoted to addressing certain *why* questions whose explanations do not seem to involve causal mechanisms […] they do not aim to track (only) causal

> relationships, but rather aspects related to the fact that the described system is information-processing.

While I am unsure whether the goal-contribution approach without efficiency considerations can address all these issues, I claim that efficient teleofunctions can manage them. By doing this, the current account will show its advantages.

As I argued in section 4, efficient computational mechanisms are constructed to include the why aspect that Shagrir mentions. I agree with him that an essential part of cognitive neuroscience is devoted to these models. They are not an aspect that could be thrown out once we know how the mechanism operates, contra Dewhurst. Also, although Piccinini's mechanistic account conceptually separates computation from information processing, he adds that semantic aspects play a crucial role in biological cognition (Piccinini 2020, 4). The *why* part of efficient computational mechanisms does include information processing aspects as the strategy S is evaluated considering signal processing parameters (sec. 4). However, as I warned there, these aspects cannot be included in a general notion of computation since neither all computation is efficient nor all signal processing is efficient. Moreover, although artificial computer design could set efficiency goals as a regulative ideal, most crafted computers are far from being absolutely efficient in a relevant sense.

While the goals of computing mechanisms could be too diverse to respond to Shagrir's requirement, something could be said about efficient computational mechanisms. Their contribution to a biological goal G of organisms is always made through an efficient information processing strategy. For that reason, they are essentially efficient. That is the reason why CNCs are repeated across different neurocognitive systems, as Chirimuuta correctly points out. Consequently, at least in the case of efficient computational mechanisms, such as CNCs, their efficiency is directly connected to their ubiquity. Also, we can identify normalization mechanisms as performing the same efficient strategy across different neurocognitive systems, as Carandini and Heeger remark. Thus, despite the physical implementational differences, we individuate them as the same computational mechanism.

Another type of *why* explanation that Shagrir and Bechtel (2017) address also targets the strategy used by a computational mechanism but focuses on environmental constraints rather than efficiency directly. The example they use is the following. In studying vision, Marr was concerned about how the brain pairs images from each eye. To solve this problem, Marr saw that the brain exploits two facts about typical environments in our world: uniqueness and continuity. As a result, the brain employs a mathematical function that Marr named UC-pairing as its strategy. Uniqueness is the characteristic such that every point that the right eye sees must correspond exactly to one point in the left eye. Continuity is about smoothness. Discounting edges, our world appears smooth and continuous between them. If we had lived in a completely different world (201):

> [T]hen computing this function would not lead to matching, but, if anything, to something else. Computing UC-pairing function is appropriate for matching in our case due to certain contingent facts about the physical environment in which we are embedded.

Now, these are the kind of contingent facts that are previous to optimization. Remember that according to the predictive coding approach to cognition, cognitive organisms must exploit facts about typical natural percepts to optimize their use of resources. As we saw, for example, every pixel in a natural image constrains the possibilities of the pixels in its neighborhood. The specific kind of constraint is a contingent fact about our world. Since neurocognitive computation is essentially embedded, these environmental conditions provide a landscape over which optimization must be done.

Hence, this second kind of *why* explanation is previous to optimization models, and it is included in them, as a precondition. Optimization always has some sort of environmental restriction. For example, we may want to use as little fuel as possible to put a satellite in orbit. To do that, we need to take into account facts about our specific environment such as gravity, friction, and the size, mass and geometrical shape of the satellite. All these facts constrain our ability to optimize.

Given that, as argued, (a) this second kind of *why* explanation is a precondition to optimality explanations, and (b) optimality explanations can be integrated into efficient teleofunctional mechanistic explanations, this second kind of *why* explanation can be integrated into the mechanistic frame too, pace Shagrir.

6. Conclusion

We have seen that efficiency determines a distinctive sense of functionality, in which optimality, on the tradeoff between the use of resources and performance, is key to attributing an efficient teleofunction. Thus, the concept of efficient teleofunction helped us characterize a distinctive class of computational mechanisms, namely, efficient computational mechanisms.

Widely endorsed hypotheses have proposed that the brains employ highly efficient coding strategies to save energy resources critical to the organism's survival. This suggests that neural computation is efficient in the sense I proposed here. Moreover, the notion of efficient function I have provided allows us to integrate harmonically computational models that explain why a particular processing strategy is implemented by the brain (by appealing to an optimal theoretical strategy that fits the actual one) and mechanistic models of how that strategy is implemented.

I have argued that *why* and *how* are two different questions that efficient teleofunctional mechanisms should scientifically address. Despite Chirimuuta's claims that they are two separate kinds of explanations in cognitive neuroscience, I have argued that they can be integrated into a single model in which both play specific explanatory roles.

REFERENCES

Ali, A., Ahmad, N., de Groot, E., Johannes van Gerven, M. A., & Kietzmann, T. C. (2022). Predictive coding is a consequence of energy efficiency in recurrent neural networks. *Patterns*, *3*(12), 100639. https://doi.org/10.1016/j.patter.2022.100639

Barlow, H. (1959). Sensory Mechanisms, The Reduction of Redundancy, and Intelligence. *Symposium on the Mechanization of Thought Process*, *10*, 535-539.

Barlow, H. (1961/2012). Possible Principles Underlying the Transformations of Sensory Messages. In W. A. Rosenblith (Ed.), *Sensory Communication* (pp. 216–234). The MIT Press. https://doi.org/10.7551/mitpress/9780262518420.003.0013

Boone, W., & Piccinini, G. (2016). The cognitive neuroscience revolution. *Synthese*, *193*(5), 1509-1534. https://doi.org/10.1007/s11229-015-0783-4

Buzsáki, G. (2019). *The Brain from Inside Out* (1th ed.). Oxford University Press. https://doi.org/10.1093/oso/9780190905385.001.0001

Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, *13*(1), 51-62. https://doi.org/10.1038/nrn3136

Chirimuuta, M. (2014). Minimal models and canonical neural computations: The distinctness of computational explanation in neuroscience. *Synthese*, *191*(2), 127-153. https://doi.org/10.1007/s11229-013-0369-y

Chirimuuta, M. (2018). Explanation in Computational Neuroscience: Causal and Non-causal. *The British Journal for the Philosophy of Science*, *69*(3), 849-880. https://doi.org/10.1093/bjps/axw034

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181-204. Cambridge Core. https://doi.org/10.1017/S0140525X12000477

Clark, A. (2016). *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190217013.001.0001

Coelho Mollo, D. (2019). Are There Teleological Functions to Compute? *Philosophy of Science*, *86*(3), 431–452. https://doi.org/10.1086/703554

Coelho Mollo, D. (2018). Functional individuation, mechanistic implementation: The proper way of seeing the mechanistic view of concrete computation. *Synthese*, *195*(8), 3477-3497. https://doi.org/10.1007/s11229-017-1380-5

Craver, C. F. (2001). Role Functions, Mechanisms, and Hierarchy. *Philosophy of Science*, *68*(1), 53-74. Cambridge Core. https://doi.org/10.1086/392866

Craver, C. F. (2007). *Explaining the Brain*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199299317.001.0001

Craver, C. F. (2013). Functions and mechanisms: A perspectivalist view. In P. Huneman (Ed.), *Functions: Selection and Mechanisms* (pp. 133-158). Springer.

Cummins, R. (s. f.). Analysis and Subsumption in the Behaviorism of Hull. *Philosophy of Science*, *50*, 96-111.

Dewhurst, J. (2016). Physical computation: A mechanistic account. *Philosophical Psychology*, *29*(5), 795–797. https://doi.org/10.1080/09515089.2016.1150450

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127-138. https://doi.org/10.1038/nrn2787

Garson, J. (2013). The Functional Sense of Mechanism. *Philosophy of Science*, *80*, 317-333. https://doi.org/10.1086/671173

Garson, J. (2016). *A critical overview of biological functions*. Springer Berlin Heidelberg.

Garson, J. (2017). Mechanisms, phenomena, and functions. In *The Routledge Handbook of Mechanisms and Mechanical Philosophy* (1st Edition, pp. 104-115). Routledge.

Garson, J. (2022). *Madness: A philosophical exploration* (1th ed.). Oxford University Press.

Glennan, S. (2017). *The new mechanical philosophy* (First edition). Oxford University Press.

Glennan, S. S. (1996). Mechanisms and the nature of causation. *Erkenntnis*, *44*(1), 49-71. https://doi.org/10.1007/BF00172853

Hohwy, J. (2004). Top-Down and Bottom-Up in Delusion Formation. *Philosophy, Psychiatry, & Psychology*, *11*(1), 65-70. https://doi.org/10.1353/ppp.2004.0043

Hohwy, J. (2013). *The Predictive Mind*. Oxford University Press UK.

Hohwy, J. (2015a). Prediction, agency, and body ownership. In A. Engel, K. J. Friston, & D. Kragic (Eds.), *The Pragmatic Turn: Toward Action-Oriented Views in Cognitive Science* (pp. 109-120). (Strüngman Forum Reports). The MIT Press.

Hohwy, J. (2015b). Prediction Error Minimization, Mental and Developmental Disorder, and Statistical Theories of Consciousness. In R. J. Gennaro (Ed.), *Disturbed Consciousness* (pp. 293-324). The MIT Press. https://doi.org/10.7551/mitpress/9780262029346.003.0012

James, W. (1884). What is an Emotion? *Mind*, *9*(34), 188-205. JSTOR.

Kaplan, D. M. (2011). Explanation and description in computational neuroscience. *Synthese*, *183*(3), 339-373. https://doi.org/10.1007/s11229-011-9970-0

Levy, A., & Bechtel, W. (2013). Abstraction and the Organization of Mechanisms. *Philosophy of Science*, *80*(2), 241-261. https://doi.org/10.1086/670300

Maley, C., & Piccinini, G. (2017). A unified mechanistic account of teleological functions for psychology and neuroscience. In D. Kaplan (Ed.), *Explanation and Integration in Mind and Brain Science* (1st ed.). Oxford University Press.

Maroney, Owen, "Information Processing and Thermodynamic Entropy", *The Stanford Encyclopedia of Philosophy* (Fall 2009 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/fall2009/entries/information-entropy/>.

Millikan, R. G. (1995). *White queen psychology and other essays for Alice* (1. paperback ed., 2. print). MIT Press.

Palmer, C. J., Lawson, R. P., & Hohwy, J. (2017). Bayesian approaches to autism: Towards volatility, action, and behavior. *Psychological Bulletin*, *143*(5), 521-542. https://doi.org/10.1037/bul0000097

Piccinini, G. (2007). Computational modelling vs. Computational explanation: Is everything a Turing Machine, and does it matter to the philosophy of mind?1. *Australasian Journal of Philosophy*, *85*(1), 93-115. https://doi.org/10.1080/00048400601176494

Piccinini, G. (2015). *Physical Computation: A Mechanistic Account*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199658855.001.0001

Piccinini, G. (2020). *Neurocognitive Mechanisms: Explaining Biological Cognition* (1th ed.). Oxford University Press. https://doi.org/10.1093/oso/9780198866282.001.0001

Piccinini, G., & Bahar, S. (2013). Neural Computation and the Computational Theory of Cognition: Cognitive Science. *Cognitive Science*, *37*(3), 453-488. https://doi.org/10.1111/cogs.12012

Piccinini, G., & Garson, J. (2014). Functions Must Be Performed at Appropriate Rates in Appropriate Situations. *British Journal for the Philosophy of Science*, *65*(1), 1-20.

Plato, Cooper, J. M., & Hutchinson, D. S. (1997). *Complete works*. Hackett Pub.

Roffé, A. J., & Ginnobili, S. (2020). Optimality Models and the Propensity Interpretation of Fitness. *Acta Biotheoretica*, *68*(3), 367–385. https://doi.org/10.1007/s10441-019-09369-5

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences*, *17*(11), 565-573. https://doi.org/10.1016/j.tics.2013.09.007

Shagrir, O. (2022). *The nature of physical computation*. Oxford University Press.

Shagrir, O., & Bechtel, W. (2017). Marr's computational level and delineating phenomena. In D. Kaplan (Ed.), *Explanation and Integration in Mind and Brain Science* (1st ed.). Oxford University Press

Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379-423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shi, Y.-Q., & Sun, H. (1999). *Image and Video Compression for Multimedia Engineering: Fundamentals, Algorithms, and Standards*. CRC Press.

Sterling, P., & Laughlin, S. (2015). *Principles of Neural Design*. The MIT Press. https://doi.org/10.7551/mitpress/9780262028707.001.0001

Vetter, P., & Newen, A. (2014). Varieties of cognitive penetration in visual perception. *Consciousness and Cognition*, *27*, 62-75. https://doi.org/10.1016/j.concog.2014.04.007

Wajnerman Paz, A. (2017). Pluralistic Mechanism. *THEORIA*, *32*(2), 161-175. https://doi.org/10.1387/theoria.16877