THEORIA

Wiring optimization explanation in neuroscience: What is special about it?[&]

(*¿Qué tiene de especial la explicación de optimización de cableado en neurociencia?*)

Sergio Daniel BARBERIS*

Universidad de Buenos Aires, Argentina

ABSTRACT: This paper examines the explanatory distinctness of wiring optimization models in neuroscience. Wiring optimization models aim to represent the organizational features of neural and brain systems as optimal (or near-optimal) solutions to wiring optimization problems. My claim is that that wiring optimization models provide design explanations. In particular, they support ideal interventions on the decision variables of the relevant design problem and assess the impact of such interventions on the viability of the target system.

KEYWORDS: Neuroscience, Wiring Optimization, Mechanism, Darwinian Explanation, Design Explanation, Ideal Intervention.

RESUMEN: Este artículo examina la distintividad explicativa de los modelos de optimización del cableado en neurociencia. Los modelos de optimización del cableado se proponen representar los rasgos organizativos de los sistemas neuronales y cerebrales como soluciones óptimas (o casi óptimas) a problemas de optimización del cableado. Afirmo que los modelos de optimización proporcionan explicaciones de diseño. En particular, estos modelos soportan intervenciones ideales sobre las variables de decisión de un problema de diseño relevante, y evalúan el impacto de esas intervenciones sobre la viabilidad del sistema en cuestión.

PALABRAS CLAVE: Neurociencia; optimización de cableado; mecanismo, explicación darwiniana; explicación de diseño; intervención ideal

---

* *Correspondence to:* Sergio Daniel Barberis. Instituto de Filosofía «Dr. Alejandro Korn». Facultad de Filosofía y Letras. Universidad de Buenos Aires. Puán 480. Ciudad de Buenos Aires. PO:1406 Argentina – sbarberis@filo.uba.ar – http://orcid.org/0000-0002-5381-3772

1. Introduction

In a recent introduction to the philosophy of neuroscience, Craver and Kaplan (2014, 268) notice that neuroscience is especially interesting to philosophers of science, among other reasons, because it is an interfield discipline that "exemplifies a form of scientific progress in the absence of an overarching paradigm." The Society for Neuroscience (SfN) has been characterized as a "menagerie of researchers" (Craver 2007, 16) with different explanatory goals, different conceptual resources, and different experimental techniques. There seems to be no single unifying theoretical framework for the variety of explanatory approaches that populate the discipline. The contribution of these approaches to scientific knowledge depends not only on the assessment of the relevant empirical evidence but also on the assessment of frequently implicit ideas about the particular standards by which purported explanations should be evaluated.

Some philosophers think that the wild diversity of explanatory approaches that bloom in the neurosciences can be tamed by integrating them into mechanistic frameworks (Craver 2007; Boone and Piccinini 2016). A mechanistic framework is a mechanism schema that defines the space of possible mechanisms for an explanandum phenomenon. The unity of neuroscience is effected when researchers from different scientific fields collaborate to set constraints on the space of possible mechanisms for the phenomenon. More specifically, scientific models contribute to the mechanistic explanation of some explanandum phenomenon when they successfully represent some components of the underlying mechanism or some causal dependence relations among those components (Kaplan and Craver 2011). This model-to-mechanism mapping (3M) constraint on mechanistic explanation is "the mechanist's gauntlet": the default assumption that the phenomena of neuroscience have mechanistic explanations (Kaplan and Craver 2011, 602). Certainly, several areas of mainstream neuroscience are explicitly concerned with discovering, modeling, and intervening the brain complex mechanisms that are responsible for human and nonhuman cognitive capacities and other phenomena of interest.

In this paper, I am interested in the explanatory distinctness of the strategy of wiring optimization (WO) modeling that thrives in neuroanatomy and systems neuroscience (Chklovskii and Koulakov 2004; Sterling and Laughlin 2015; Kaiser and Hilgetag 2015). WO models aim to represent the organizational features of neural and brain systems as optimal (or near-optimal) solutions to WO problems. A WO problem is the problem of finding the spatial layout of a given set of components that minimizes total connection costs given a fixed connectivity. WO models are empirically adequate to the extent that the optimal layout predicted by the model is relevantly similar to the actual layout in the target system.

Since WO models aim to represent the organizational features of real mechanisms in detail, they are constrained by the 3M requirement of mechanistic explanation. However, mechanisms are the *explanandum* phenomenon of optimization explanations. The organizational features of the target mechanism typically are what we want to explain by modeling the relevant design problem, i.e. by representing the structure of design goals (the maximization or minimization of some objective function), constraints, and tradeoffs that shaped the organizational features in question.

Someone could view the explanans of a WO explanation as a description, not of a mechanism in the brain/body, but of the mechanism of natural selection that influenced the evolution of the trait toward the optimal. From this view, the explanatory relevance

of optimization modeling is built on the assumption that the design goal defined by the model is a proxy of biological fitness. Such assumption enables the inference that natural selection optimized the fitness value of the trait towards a (local) optimum, outweighing the causal contribution of other evolutionary forces (mutation, drift, migration, etc.) acting on the evolution of the target system. Many advocates of the WO approach in neuroscience endorse this "Darwinian" assumption about the role of natural selection in the evolution of optimal traits. However, other, non-Darwinian optimization mechanisms might be relevant as well for the evolution of brain wiring optimization. In any case, WO models in neuroscience *bypass* evolution and development in order to focus on the synchronic costs, constraints, and tradeoffs that a design problem impose on the relevant features of the target system.

My claim is that WO models in neuroscience provide design explanations. Design explanation is a kind of functional explanation in biology that purport to explain why specific organisms have certain traits "by showing that their actual design is better than contrasting designs" (Wouters 2007, 65). WO models contribute to design explanations in neuroscience to the extent that they support ideal interventions on the decision variables values of the relevant design problem, and measure the impact of such ideal interventions on the viability of the target system.

I will proceed as follows. In section 1, I will review the main features of the WO approach to the organization of neuronal and brain systems. In section 2, I will explore some relations between the strategy of optimization modeling and mechanistic explanation. I will argue that the 3M constraint sets a necessary condition on optimization explanations, but it fails to capture the explanatory distinctness of WO models in neuroscience. In section 3, I will explore some relations between optimization modeling and Darwinian explanation. I will argue that the WO approach is autonomous in practice from the Darwinian assumption that natural selection is the overriding causal mechanism that shaped the evolution of optimal (or near-optimal) traits. In section 4, I will explore some relations between optimization modeling and design explanation. I will suggest that an interventionist account of design explanation can shed light on the explanatory distinctness of optimization models in neuroscience.

2. *Wiring Optimization: Principles, Models, and Exemplars*

In the early days of modern neuroscience, Santiago Ramón y Cajal (1897, 1899), the primary architect of the neuron doctrine, noticed several phenomena where the positioning of neuronal components seemed to "save wire", that is, to minimize the cost of creating and maintaining neural connections. Ramón y Cajal's law of dynamic polarization (i.e. that nerve impulses take the shortest path from dendrites to axons) accounted for the anatomical positioning of the dendrites and the terminal arborizations of the axon (Ramón y Cajal 1897). Everything else about the anatomy of the nerve cells, including the positioning of the soma, appeared to be variable and accommodative. Ramón y Cajal (1899, 102) raised the following question: "Are these variations merely whims of Nature, arrangements without importance, or have they some physiologic significance?". He answered that every empirical evidence suggested that the anatomical variations in question responded to some physiologic design, and were thus the output of evolutionary (maybe Darwinian) mechanisms.

In particular, Ramón y Cajal (1899, 102) considered that the observed variations were "morphological adaptations ruled by laws of economy of time, space and material". Ramón y Cajal's conservation laws represented his ultimate "inductive effort" to glimpse into the "utilitarian principles" that governed the infinite variation of form, size, position, and direction of neurons, and nerve fibers (Ramón y Cajal 1917, 314). Why do sensitive ganglion cells in adult mammals exhibit a monopolar morphology (rather than a bipolar or multipolar morphology)? Because there is, in normal conditions, a need to save material, space, and conduction time, and moving the cell bodies to the periphery of the ganglion while arranging the sensitive conduits in rectilinear bundles in the central region represents the most advantageous disposition to satisfy that need (Ramón y Cajal 1897). Why do sensitive and motor neurons concentrate in ganglia in invertebrates? Because there is a need to economize protoplasm and space, and the successive concentration of initially separated neurons economizes conductors and allows one fiber to communicate the excitation to a high number of neurons (Ramón y Cajal 1899). Why do the nerve fibers bifurcate in the form of a Y (instead of a T) upon their arrival at the posterior cord? Because there is, in normal conditions, a need to save conduction time, and the Y-shaped bifurcation represents the shortest path between the entry point and the exit point of the nerve impulse (Ramón y Cajal 1899).

Modern neuroscientists know that the wiring cost of connections in the brain has multiple origins: it arises from volume, metabolic requirements, signal delay and attenuation, and guidance defects in development, among other factors (Chklovskii and Koulakov 2004). In any case, it is a general fact that the wiring cost must increase with the distance between connected elements. Since long-range connections in the brain are a limited resource, it is reasonable to assume that, given a fixed connectivity, nature would prefer the layout with minimum length. This principle of neural design is known as the WO principle (Sterling and Laughlin 2015; Kaiser and Hilgetag 2015). Briefly, what the WO principle states is that, given a fixed connectivity pattern between neuronal or brain components, the wiring cost determines the spatial organization of those components. From this "brain as a microchip" perspective (Cherniak 1994), we can model the different organizations of neuronal or brain components as solutions to specific wiring minimization problems constrained by a fixed connectivity.

As other guiding principles elsewhere in the sciences, such as Newton's mechanical principles, or the Hardy-Weinberg law in population genetics, if the WO principle is understood as an universal empirical generalization, then it either vacuously true, or evidently false (Cartwright 1999; Giere 1999). The general validity and empirical scope of the WO principle are tested indirectly by the construction, analysis, and assessment of particular WO models. The WO principle has been used to generate a whole family of optimization models that account for many phenomena related to neuronal and brain organization, to wit: why there are separate visual areas (Barlow 1986; Mitchison 1991); why the neocortex folds in a characteristic species-specific pattern (Essen 1997); why neurons in the mammalian visual cortex are organized into multiple maps (Chklovskii and Koulakov 2004); why ocular dominance patterns exist (Mitchison 1991); why orientation preference patterns are found in the visual cortex (Mitchison 1991); why axonal and dendritic arbors have particular branching angles (Cherniak 1992), and why axons and dendrites occupy a 3/5 fraction of gray matter (Chklovskii, Schikorski, and Stevens 2002), among many other applications (see Sterling and Laughlin 2015).

An optimization model defines an optimization problem, that is, the problem of minimizing or maximizing an objective function of a set of decision variables by systematically

choosing the values of those variables from within an allowed set. The objective function associates an element belonging to a totally ordered set of costs or utilities to each element of the allowed set. A feasible solution is any value of the decision variables that satisfies the parameters, constraints, and tradeoffs defined by the model. The goal is to find the best feasible solution by comparing different feasible solutions based on the corresponding objective function values. Any best feasible solution is called an optimal solution. Many important optimization problems have been proved to be non-deterministic polynomial-time (NP)-complete, thus heuristic optimization methods are used to approximate optimal solutions.

Adopting the WO approach, Cherniak and colleagues (Cherniak 1994; Cherniak et al. 2004) applied combinatorial network optimization theory to construct an optimization model which represented and solved the saving wire problem for the placement of ganglia in *Caenorhabditis elegans*' nervous system. A problem of component placement optimization can be stated as follows: "Given connections among a set of components, find the spatial layout of the components that minimizes total connection costs" (Cherniak et al. 2004, 1081). The problem of ganglion-level optimization is treated as involving 11 movable components, with 11! possible layouts (39916800 alternative solutions). All these possible placements have been searched for following a genetic algorithm. Cherniak and colleagues found, surprisingly, that the actual placement of ganglia in *C.Elegans* is very close to the optimal layout for those components, i.e. the layout that preserves the desired connectivity between elements using the smallest amount of wire length. From this analysis, we can infer that the actual placement of ganglia in the nematode is mainly driven by the need to minimize the wiring cost, given a fixed connectivity.
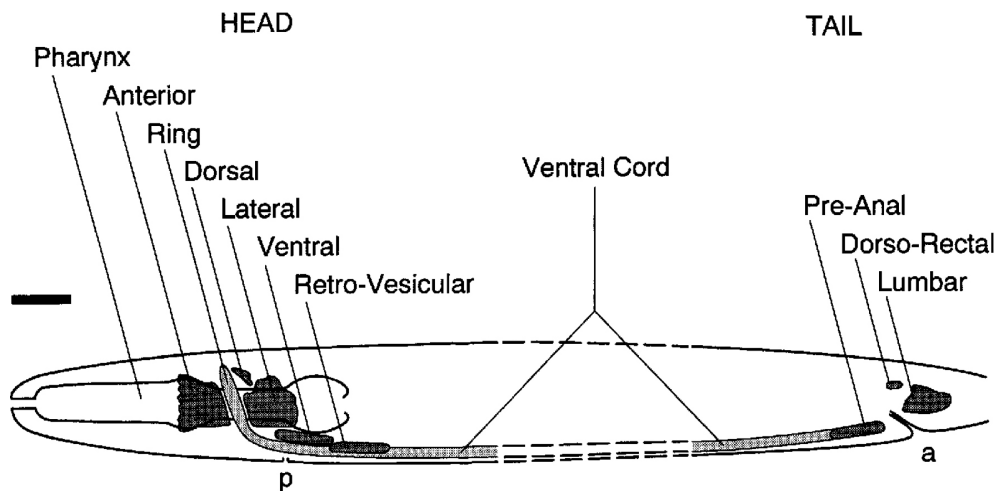

Figure 1

C. Elegans ganglion components. Taken from Cherniak (1994, 2420)

Cherniak and colleagues (2004) have also applied the WO approach to the placement of core functional areas of the cat cortex and the macaque cerebral cortex. The application of

this approach to the mammalian cortex demands for some simplifications and approximations in the model. Given the impracticality of measuring long-range wire length in the mammalian brain, they have used a network wire-minimizing heuristic —the adjacency rule— as an optimization measure. According to the adjacency rule, if two components *a* and *b* are connected, then *a* and *b* are adjacent, i.e. they are immediately contiguous topologically. Furthermore, given that the information on cortical connection and adjacency in the mammalian cortex is not complete, these researchers introduced a size law as a working hypothesis: "If a set of connected components is optimally placed, then the smaller a subset of the total layour, the less optimal it will tend to be" (Cherniak et al. 2004, 1083). The rationale behind the size law is that local subsystem optimization is usually sacrificed to obtain global optimization. The most striking finding of this model is that optimization improves exponentially with the size of subset visual areas. When a subset of 20 areas was considered, only three layouts of a billion sampled were cheaper than the actual one. For a 25-area subset, "a billion-layout random sample yielded no placements cheaper than the actual one," which suggests that mammals may have evolved "the best of possible brains" considering the wiring cost (Cherniak et al. 2004, 1084).
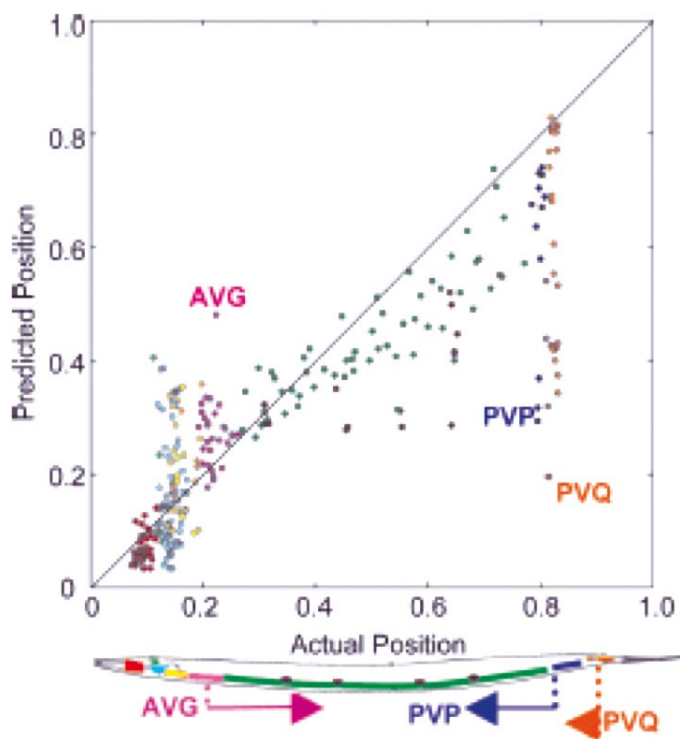

Figure 2

Neurons positions predicted by the Dedicated-Wire Model versus actual neuron positions.
Perfect predictions fall on the diagonal.
Taken from Chen, Hall, and Chklovskii (2006, 4725)

In their influential Dedicated-Wire Model of *C. Elegans*'s nervous system, Chklovskii and his colleagues (Chen, Hall, and Chklovskii 2006) extended Cherniak's approach to the neuronal level. They considered the placement of 279 non-pharyngeal neurons as decision variables and estimated the total cost of connecting neurons to each other (i.e. the internal cost) and of attaching neurons to sensory endings and muscles (i.e. the external cost). They solved for the optimal neuronal layout using powerful placement algorithms and found that most actual neuron positions and predicted neuron positions coincided with each other. There were some deviations, however, particularly when it came to internal (neuron-to-neuron) connectivity (see Fig. 2). To account for these deviations, these researchers suggest that design constraints other than connectivity (e.g. volume) may be incorporated into the model.

Revisiting the structural connectivity and neuronal layout of *C. Elegans* and macaque cerebral cortexes, Kaiser and colleagues(Kaiser and Hilgetag 2006; Chen et al. 2013) have argued that high processing efficiency, characterized as reduction of path lengths, is at least as important as low wiring cost, characterized as short wiring length. Long distance connections are metabolically expensive, but they have the benefit of reducing the number of intermediate transmission steps in neural pathways. In fact, processing efficiency may arise from the synchronization of near and distant regions, the reduction of intermediate transmission steps in neural pathways, the reduction of cross-modal interference and/or the increase of reliability of transmission (Kaiser and Hilgetag 2015). The benefits of processing efficiency obtained by adding long-distance projections might outweigh the wiring costs of establishing those additional connections. Recently, Chen and colleagues (2013) have explored a model combining the wiring cost and processing efficiency constraints and found that both reconstructed and real networks had a similar organization of network modules (to save wiring) and hubs (to reduce path lengths). They concluded that these findings provide clear evidence to support the hypothesis that the small-world topology of neural networks results from a trade-off between the two constraints of wiring cost and processing efficiency (although further constraints might be relevant as well).

An elegant application of the wiring minimization approach is the explanation of retinotopy in visual cortical areas of the mammalian brain. A visual cortical area is said to contain a retinotopic map if neighboring neurons in that area receive inputs from neighboring retinal neurons, in such a way that the cortical neurons form a 2D representation of the visual image formed on the retina. Retinotopic maps are present in many visual areas of the cerebral cortex beyond V1 (Swindale 2008). Why does retinotopic maps exist? The classical explanation for the existence of retinotopic maps rely on the WO principle and the fact that wiring cost is minimized if neurons with adjacent receptive fields are placed as close to each other as possible:

> Because of spatial correlations in the external world, the early stages of the visual system combine information coming from the adjacent points in the visual field. For example, recognition of a face on a portrait relies on exchanging information about facial features represented by neurons with adjacent fields. Such processing requires connecting such neurons by costly wiring, i.e., axons and dendrites. To minimize the wiring length, neurons with adjacent receptive fields should be placed as close to each other as possible. This is exactly what an ordered retinotopic map accomplishes. Thus, the reason retinotopic maps exist is to minimize the total length of intracortical connections that are required for processing local features of the visual space. (Chklovskii and Koulakov 2004, 371)

From similar findings concerning ocular dominance patterns, ocular dominance stripes, orientation preference maps, and direction of motion preference maps, Chklovskii and Koulakov (2004) conclude that, given a fixed connectivity pattern, the wiring cost decides the neuronal layout in the mammalian cortex. Furthermore, they claim that the existence of an animal with a suboptimal layout with 10% more wiring that the actual one is "close to impossible" and, thus, that the mammalian cortex is on the verge of a "wiring catastrophe" (Chklovskii and Koulakov 2004, 385; see also Cowey 1979; Kaas 1997; Nelson and Bower 1990). The hope is that the "wiring optimization approach may provide a general unifying framework, which will help to organize the multitude of experimental facts about cortical architecture" (Chklovskii and Koulakov 2004, 288).

To sum up, a WO model in neuroscience relates the spatial organization of a target neuronal or brain system with the wiring costs, connectivity constraints, and tradeoffs defined by the relevant optimization problem. What kind of scientific explanation, if any, is conveyed by WO models in neuroscience? This is a thorny question. In the following section, I explore some relations between WO modeling and mechanistic explanation.

3. Wiring Optimization and Mechanistic Explanation

In this section, I acknowledge that the 3M constraint (see section 1) sets a necessary condition on WO explanations in neuroscience. However, I claim that the 3M constraint is not sufficient to capture the explanatory distinctness of the WO approach. In the practice of WO modeling, the actual organizational features of some target mechanism constitute the phenomenon that demands for an explanation. An optimization model aims to explain that phenomenon by representing the structure of costs, constraints, and tradeoffs that shaped its design. This structure is not the kind of thing that can participate in componential or productive relations. It is not a mechanism in the brain/body.

The influential idea behind the 3M constraint is that scientific models are explanatory in neuroscience to the extent, and only to the extent, that they contribute to mechanistic frameworks (see section 1). This involves representing some of the component parts, activities or organizational features of the target system. The 3M constraint make this representational requirement on mechanistic explanation explicit.

> (3M) In neuroscience, a model of a target phenomenon [mechanistically] explains that phenomenon to the extent that (a) the variables in the model correspond to identifiable components, activities, and organizational features of the target mechanism that produces, maintains, or underlies the phenomenon, and (b) the (perhaps mathematical) dependencies posited among these (perhaps) mathematical variables in the model correspond to causal relations among the components of the target mechanism (Kaplan and Craver 2011, 611).

The 3M constraint sets necessary and sufficient conditions on mechanistic explanation (Kaplan 2011; Craver and Kaplan 2014). The 3M constraint does not require a fully detailed representation of every component or every causal relation among the components of the target mechanism. It does not imply that more mechanistic detail is always better (Kaplan 2011; Chirimuuta 2014; Weiskopf 2011), and it is compatible with many kinds of idealization and abstraction in the model (Boone and Piccinini 2016; Levy and Bechtel 2013).

The 3M requirement set necessary conditions on WO explanation, in the sense that the practice of WO modeling in neuroscience is actually constrained by mechanistic adequacy considerations. A good WO model must adequately represent the actual organizational features of the target mechanism (*as* the optimal solution of an optimization problem). If the optimal organizational features deviate too much from the actual, then we can infer that the model is mechanistically inadequate, and non-explanatory. A necessary step in the practice of optimization modeling is the assessment of the fit between the optimal organizational features predicted by the model and the actual organizational features of the target mechanism.

Furthermore, the evolution of the WO approach exhibits a pattern of scientific progress in the direction of mechanistic accuracy. Cherniak's (1994) model of the *C. Elegans*' nervous system represented the target system as constituted by 11 ganglion-level components. The Dedicated-Wire Model used a more detailed wiring diagram of 279 non-pharyngeal neurons (Chen, Hall, and Chklovskii 2006). Using new electron micrographs, Varshney et al. (2011) assembled whole, self-consistent gap junction and chemical synapse networks of *C. elegans*. On that basis, Chen et al. (2013) optimization model with multiple constraints used an updated connectivity diagram of 2902 (rather than 2105) connections. The best explanation of this pattern from mechanism sketches to more complete mechanism schemata is that the strategy of WO modeling is restricted by the 3M requirement.

Someone could reply that, even if the optimization approach draws upon up-to-date mechanistic information about the organizational features of the target system, the approach does not provide any new mechanistic information about the target. Therefore, WO models do not genuinely contribute to mechanistic frameworks. I contend that, if not about their target systems, mechanistically accurate WO models can provide valuable information about the unknown mechanisms that underlie related phenomena. In particular, mechanistically adequate optimization models of normal brain organization may become relevant for the explanation of abnormal brain organization syndromes, whose underlying mechanisms remain unknown, by changing the probability distribution over the space of possible mechanisms for those syndromes. In fact, the design constraints on normal brain organization identified by some optimization models in systems neuroscience are becoming increasingly relevant for the mechanistic explanation of many abnormal brain organization syndromes like schizophrenia, autism, and Alzheimer's disease.

Let us consider in some detail the mechanistic explanation for schizophrenia. This disease is a heritable psychiatric disorder characterized by the disintegration of psychological function resulting in the loss of unity of mind and consciousness (Sporns 2011). The actual biological mechanism(s) for schizophrenia is (are) unknown, although several authors suggest a "disconnection hypothesis" according to which the behavioral and cognitive symptoms are related to a disconnection of cognitive networks in the brain (Friston and Frith 1995; Friston 1998). Recently, Bullmore and Sporns (2012) have presented an optimization approach on the neuropsychiatric disorder, grounded in the fact that brain network organization optimizes an economic trade-off between minimizing wiring cost and maximizing the behavioral value of network function. In particular, they argue that the loss of cognitive functions with disease may be accounted for by a shift in the cost-efficiency trade-off in the direction of lowering metabolic connection costs at the expense of integrative capacity.

Vertés et al. (2012) adopt this perspective and propose an economical clustering model in which the spatially embedded network of the brain emerges from two competing factors: a distance penalty based on the metabolic cost of maintaining long-range connections and a topological term that favors the formation of highly connected hubs. They estimated economical clustering model parameters from resting-state functional fMRI measures in a group with childhood-onset schizophrenia and found that the abnormal profile of brain network topology extracted from the group could be well matched by the economical clustering model with rather different model parameters compared with those estimated from a group of healthy volunteers. The abnormal organization of brain functional networks in schizophrenia is seen as the outcome of an abnormally biased tradeoff between the generative factors of distance penalization (wiring cost) and topological clustering (efficiency).

The economical clustering model contributes to the mechanistic explanation of schizophrenia to the extent that it provides new evidence for the disconnection hypothesis. The abnormal trade-off identified by the economical clustering model results in disturbances of large-scale structural and functional brain connectivity, supporting Friston's hypothesis. The model exemplifies a kind of mechanistic progress in the explanation of schizophrenia to the extent that it changes the probability distribution over the space of possible mechanisms for the disease. In this way, WO models can provide mechanistically relevant information about the target system.

Even if the 3M constraint sets necessary conditions on WO explanation, and even if WO models contribute to mechanistic frameworks about target phenomena, the 3M constraint is not sufficient to capture the explanatory distinctness of WO models in neuroscience. The target mechanism constitutes the explanandum phenomenon of an optimization explanation. Once the fit between the model and the target system is established, the WO approach proceed to explain the organizational features of the target system by referring to the structure of the relevant design problem (design goals, constraints, and tradeoffs) that shaped that trait.

Do the design problem describe a mechanism? From an metaphysical point of view, costs, constraints, and tradeoffs are not themselves the kind of entities that can enter into causal or componential relations (Rice 2012, 2015). They are not events nor are they causal properties. Tradeoffs have no productive relation whatsoever with their purported effects. They do not compress, or pull, or unwind anything in the system (Rice 2015).While causal relations are essentially diachronic, the mathematical representation of costs, constraints, and tradeoffs does not refer to any processes that unfold over time or any events that take place prior to the explanandum (Rice 2012). Consider, for example, the Dedicated-Wire Model of *C. Elegans*' nervous system (Chen, Hall, and Chklovskii 2006). The model represents the total wiring cost of the nervous system as the sum of an internal cost to connect neurons to each other and an external cost to attach neurons to fixed structures. The wiring cost does not constitute, however, a part, an activity, or an organizational feature of *C. Elegans*' nervous system. In general, the elements of the design problem are not component parts or causal relations within the brain/body.

One mechanism-friendly way to flesh out the explanans of an optimization explanation is to say that it is not a mechanism in the brain/body at all, but an evolutionary mechanism that occurred over epochs of time and countless generations, a hugely messy process of ever-present natural selection towards cheaper wiring, one involving relative rates of reproduc-

tion for those animals with more versus less optimized wiring in the nervous system.[1] Thus, the optimization hypothesis would concern a mechanism, i.e. the mechanism of natural selection. The idea that natural selection is literally a mechanism is not undisputed (Skipper and Millstein 2005). It is not evident what aspects of the evolutionary process would count as parts, activities, or organizational features of such a ubiquitous mechanism. Anyway, the explanatory claim of WO models would be the following: "Had the minimal wiring morphology not have been of such high import, and thus not selected for, we would not end up with the peculiar morphologies we find." I examine some relations between optimization modeling and Darwinian explanation in the next section.

4. *Wiring Optimization and Darwinian Explanation*

Do WO models in neuroscience provide Darwinian explanations? A Darwinian explanation accounts for efficient design in phenotypic traits in terms of its contribution to the biological fitness of the organism. If a WO model fits the structure of a design problem in nature, and the wiring cost identified by the model (and optimized by the target phenotype) is a proxy of biological fitness, then it is reasonable to conclude that natural selection has been the superseding evolutionary mechanism that moved the phenotype toward the optimum. From this view, what is special about WO models is that they contribute to Darwinian explanations of some target feature to the extent that they abstract away from most of the causal history that produced the outcome, in order to focus on the fitness-enhancing character of the trait in question.

The idea that optimization models provide Darwinian explanations pervades the practice of optimization modeling in evolutionary biology (Maynard Smith 1982) and thus, most philosophers of evolutionary biology endorse it as well. This mainstream opinion is expressed by Potochnik (2007, 681), who claims that "the long-term importance of the optimality approach is assured by its role in evolutionary explanation." According to Potochnik, in optimization modeling, one determines the range of possible values for some phenotype and the fitness function relating these phenotypes to the environment. On this basis, the model predicts which phenotypic values will predominate in the population. From her view, optimization models provide causal explanations that highlight a modular part of the causal process that produce an evolutionary outcome. In particular, optimization models emphasize how natural selection has shaped the trait in question. Rice (2012, 2015) call this view a product of "a censored causal approach" to optimization explanation.

Rejecting the censored causal approach, Rice considers that optimization models do not represent a modular part of the causal evolutionary process that leads to an adaptation. He endorses the idea that optimization models in evolutionary biology typically provide highly idealized equilibrium explanations (Sober 1983). Causal explanations aim to describe the particular causal path that leads to the target outcome state. In contrast, equilibrium explanations aim to show that a very large number of initial states will evolve in such a way that the system ends up in the outcome state we want to explain, "regardless of which of a vari-

---

[1] I thank to an anonymous reviewer for suggesting this mechanistic interpretation of optimization explanation.

ety of causal scenarios actually transpired" (Sober 1983, 202). Core to equilibrium explanation is the idea that the phenomenon is explained as a stable point in a field of unstable options.[2] The reason why the stable point happens is that the other points in the landscape for the relevant phenomenon are unstable and "tip" toward the stable point. Why does a phenotype (e.g. a particular neural structure) exist? Because the other structures represent unstable strategies from an evolutionary perspective: if by chance a system moves into one of those structures, it will be outcompeted by organisms with better structures.

The classical example of equilibrium explanation is Fisher's ([1931] 1999) explanation of the fact that the sex ratio is 1:1 in many species. The insight of Fisher's argument, according to Sober (1983, 201), is that "if a population departs from equal numbers of males and females, there will be a reproductive advantage favoring parental pairs that overproduce the minority sex", thus a "a 1:1 ratio will be the resulting equilibrium point." In modern terminology, the target structure would be an "evolutionary stable strategy" (Smith and Price 1973). Once the strategy is fixed in a population, the mechanism of natural selection alone is sufficient to prevent mutant strategies from invading successfully. Rice (2012, 687) believes that optimization models explain—just like Fisher's model of sex ratios does—when they represent the evolution of a particular phenotype under the optimization assumption that "the strategy that optimizes the criterion of the model is the equilibrium point of the evolving population."

I acknowledge that, to the extent that the design goal of an adequate optimization model is a proxy of biological fitness, there is no harm in interpreting the model as conveying a Darwinian explanation, be it a censored causal explanation or an equilibrium one. Many optimization models in evolutionary biology, and elsewhere in the sciences, corroborate this interpretation. Sometimes, the advocates of the WO approach in neuroscience formulate the WO principle with a Darwinian flavor, e.g., "among various functionally equivalent arrangements of neurons, the one having connected neurons as close as possible is most evolutionarily fit and, therefore, likely to be selected" (Chklovskii and Koulakov 2004, 369). In practice, however, the minimization of total wiring cost is not automatically conceived as a proxy of biological fitness.

According to Cherniak (2006), it is an open question whether the near-optimal component placement in *C. Elegans'* nervous system is due to phylogenetic mechanisms. Cherniak (2009) suggests a "non-genomic" nativist interpretation of many WO findings, in which complex biological structure can originate without need of DNA involvement. I have reviewed in section 2 that a very simple genetic algorithm, with total wirelength as fitness measure, will robustly and reliably find that optimal layout (Cherniak 1994). Yet so will a force-directed placement algorithm or vector-mechanical "mesh of springs" (Cherniak 2006). This suggests the possibility that the mechanism of WO in the nervous system may involve basic physical processes only. From this view, the genome seems to get WO automatically and directly from energy-minimization phenomena involving classical mechanics, i.e., vector-mechanical energy minimization.

A similar case has been made for local optimization of neuron arbors. Cherniak (1992) modeled the neural optimization of dendritic and axonic arborizations as a Steiner tree op-

---

2   I thank an anonymous reviewer for suggesting this framing of the idea that optimization models provide equilibrium explanations.

timization problem and found that local branch-junction geometry of neuronal connecting structures fits a volume minimization model well. Neuron arbor junctions arrive at angles that, for their branch diameters, minimize local volume. Yet ensembles of non-neuronal junctions appear to approximate volume-minimizing behavior comparably well. Non-neuronal junctions include living cases: astrocytes, capillary networks, elm tree roots, and non-living structures: Lichtenberg electric discharges, river deltas, and converging erosion gulleys. Since the evolution of river junctions and electric discharge patterns is not mediated by genetic information, the possibility arises that a wide range of local volume-minimizing tree-forming phenomena may all arise by simply energy-minimization mechanisms. It might well be that WO constitutes a "high level" constraint on brain organization, that is, a constraint to which nearly any physically possible kind of brain will also be subject (Changizi 2003).

When Chen et al. (2006) discuss the optimization mechanisms that are responsible for near-optimal wiring in *C. Elegans,* they suggest that neuronal placement may be driven by genetics as well as by forces generated during embryonic and postembrionyc development. They hope that future research in evo-devo may shed light on the nature of the relevant optimization mechanisms.

Finally, Chklovskii and Koulakov (2004) mention the "smoking gun" that shows the WO principle in action. When neurons are grown in low-density culture, they have regular shape with straight neurite branches (Bray 1979) (Figure 3). Because a straight line is the shortest path connecting points, this is exactly what the WO principle would predict. A likely biophysical mechanism for straight segments and, thus, for WO is tension along neurites. The growth cone of cultured neurons can exert mechanical tension, and the direction of advance of the growth cone is determined by the tension between it and the rest of the neuron (Bray 1979).
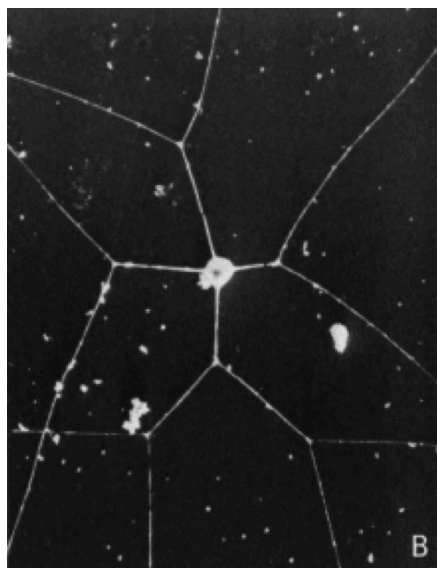

Figure 3

Scanning electron micrograph of a single sensory neuron in culture.
Taken from Bray (1979)

In sum, non-Darwinian optimization mechanisms might account for WO phenomena in the nervous system. In evolutionary biology, the optimization criterion is usually conceived of as a proxy of biological fitness. However, there is a reverse-engineering sense of optimization in which a design is said to be optimal "if it complies with its functional requirements as well as possible" (Vilarroya 2002, 251). Optimization models in neuroscience can be seen as embodying this reverse-engineering sense of optimization, that is, as attempting to analyze "an already existing intelligent artifact or system in terms of the design considerations that must have governed its creation" (*ibid*).

Green, Levy, and Bechtel (2015) remark that optimization modeling, and other design-related methods, play a specific methodological role that is tangential to the study of adaptation. For these philosophers, the cognitive strategy of reverse engineering can be detached from the study of natural selection. They introduce a valuable distinction between a system's *having a design* and a system's *being designed*. Optimization modeling explains organizational features of some target system by appealing to the system's having a design, that is, to the hypothesis that the organizational features in question approximate the optimal solution to some optimization problem relevant for the target system. In this same vein, I claim that WO explanation need not be grounded in adaptationist presuppositions. In other words, the concept of "design" in WO modeling has a present-looking, thin character (Green, Levy, and Bechtel 2015), and it does not itself entail any historical assumptions about the trait in question.

Modelers' assumptions about the optimization mechanisms underlying WO are always speculative and programmatic, and they are not purported to carry any explanatory weight. Most optimization studies in neuroscience are not concerned with the nature and variety of optimization mechanisms that may have produced the target system. Advocates of the WO approach admit that our current level of knowledge about the mechanisms underlying WO is insufficient to formulate testable hypotheses (Chklovskii and Koulakov 2004). Addressing the explanatory distinctness of the WO approach to the organization of cortical maps, Chklovskii and Koulakov (2004, 385) emphasize that "the wiring optimization theory *bypasses* development and provides a link between map structure and intracortical connectivity, both experimentally measurable quantities." This supports the idea that the optimization approach in neuroscience is autonomous, in practice, from any particular assumption about the relevant optimization mechanisms. In particular, the strategy of WO modeling is autonomous from the Darwinian assumption that natural selection is the relevant mechanism that shaped the evolution of WO.

In the next section, I explore some relations between optimization modeling and a different kind of explanation in functional biology, to wit: design explanation.

5. *Wiring Optimization and Design Explanation*

A telling feature of optimization modeling in biology it that the strategy demands a functional perspective on the target system. Wouters (2005) identifies this functional perspective as a particular way of understanding biological organization by depicting the parts and behaviors of an organism as solutions to certain design problems. Adopting this perspective enables us to provide design explanations of target phenomena (Wouters 1999; 2007). Design explanations are usually brought up in answer to explanatory demands in functional

biology, that is, that part of biology that is concerned with the way individual organisms are built (e.g., anatomy, morphology), the way they work (e.g., physiology), and the way they behave (e.g., ethology).

The core idea is that design explanations purport to explain why specific organisms have certain traits by showing that their actual design is better than contrasting designs. The hallmark of design explanations is their concern with the utility of a particular trait, often in comparison with merely possible alternatives. The hypothetical organisms are similar to the real organisms but the trait in question is absent or present to a lesser degree. The results of such a comparison are expressed by means of a functional counterfactual of the form: "if s-organisms had one of the alternative traits instead of the trait in question they would have problem P" (Wouters 1999, 232).

In this section, I show, first, that the WO approach in neuroscience fits the basic structure of design explanation. Second, I argue that WO models contribute to design explanations in neuroscience to the extent that they support ideal interventions on the relevant decision variables of the problem and evaluate the impact of such interventions on the viability of the target system.

Concerning the explanandum phenomenon, Wouters (1999, 222) suggests that design explanations are answers to questions of the following form:

> Design Explanandum
>
> (Q) Why do s-organisms have/perform t_1 rather than t_2, t_3, t_n?

Where s is a set of organisms, t_1 is the trait in question, and t_2, t_3, t_n are the alternative traits. The trait in question is the presence or character of a certain item or behavior of s-organisms. Q is neither a question about causes at the individual level nor a question about evolutionary causes at the population level. Design explanations are explicitly or implicitly contrastive: they compare real organisms to hypothetical organisms that may have never existed. Q ask about the utility of a trait in terms of what is needed or useful to stay alive, i.e. to maintain the organism, to grow, to develop, and to produce offspring.

The answer to a Q-question has the following structure (Wouters 1999, 223):

> Design Explanans
>
> (C) s-organisms live in condition c_u.
> (U) In condition c_u trait t_1 is more useful than trait t_2, t_3, t_n.

Where c_u is a conjunction of one or more conditions of organisms or environments in which organisms live. Statements like C specify conditions that apply to the relevant organisms, and statements like U claim that, due to those conditions, the trait in question is more useful to the s-organisms than the alternative traits.

An example of design explanation in morphology is Schwenk's (1994) explanation of why snakes have a forked tongue, as presented in Wouters (1999, 229):

(Q) Why do snakes have a forked tongue rather than a blunt one?
(C_1) The tongue of snakes has a causal role in trail following.
(C_2) Snakes follow trails by comparing chemical stimuli simultaneously sampled on two sides.

(*U*) In order to sample chemical stimuli simultaneously on two sides, it is more useful to have a forked tongue than a blunt one.

Claims about utilities may vary in strength (Wouters 1999). Simple design explanations can be classified either as optimization or as requirement explanations depending on whether the utility claim is an optimization claim or a requirement claim. A requirement claim has the following form: "In condition c_u, trait t_1 is the only useful one among the following traits: t_1, t_2, t_3, t_n." Requirement claims assert that the trait in question is the only one in the reference class that works. Many design explanations derive requirements directly from the laws of physics and chemistry. In contrast, an optimization claim is relatively weaker and has the following form: "In condition c_u, trait t_1 is more useful than each of the following traits: t_2, t_3, t_n." Optimization claims assert that the trait in question is the best one in the reference class.

The schema of design explanation smoothly applies to core exemplars of the WO approach reviewed in section 2. In the first place, consider Cherniak's (1994) component placement optimization model of *C. Elegans*' nervous system. The design explanation empowered by the model has the following structure:

(*Q*) Why are ganglia located in *C. Elegans* nervous system as they are?
(C_1) Neural wiring is metabolically expensive.
(C_2) The wiring cost increases with wire length.
(C_3) The actual interconnections between ganglia are [such and such].
(*U*) In conditions C_1-C_3, the worm's actual layout requires the least total wire length of any of the millions of possible layouts.

Cherniak's placement optimization model of *C. Elegans*' nervous system explains why the ganglia are placed as they are by representing the particular design problem the target feature solves par excellence, i.e. by representing the structure of functional, counterfactual dependence relations among the actual layout in the worm, the connectivity of the system, and the wiring cost.

In the second place, consider the Dedicated Wire model of *C. Elegans*'s nervous system (Chen, Hall, and Chklovskii 2006). The design explanation the model provides is the following:

(*Q*) Why are non-pharyngeal neurons of *C. Elegans* positioned as they are?
(C_1) Neural wiring is metabolically expensive.
(C_2) The wiring cost increases with wire length.
(C_3) The actual interconnections between non-pharyngeal neurons are [such and such].
(*U*) In conditions C_1-C_3, the actual neuronal layout requires less total wire length than most members of the set of possible neuronal layouts.

The Dedicated Wire model explains the neuronal layout of the nematode's nervous system by representing the structure of the WO problem the actual neuronal layout solves better than most alternative placements.

Other optimization explanations are more complex and include tradeoffs between several design constraints. Consider Chen et al. (2013) optimization model of small-world network organization (i.e. modules and hubs) in the macaque cortical areas. The design explanation empowered by the model runs as follows:

(Q) Why is the Macaque cortex organized as a small-word network?
(C_1) Neural wiring is metabolically expensive.
(C_2) The wiring cost increases with wire length.
(C_3) Processing efficiency is has several functional advantages (reduction of noise, synchronous information processing, etc.).
(C_4) Processing efficiency increases with long-distance projections.
(C_5) There is a tradeoff between wiring cost and processing efficiency.
(U) In conditions C_1-C_5, the small-world topology of the Macaque cortex represents a better compromise between wiring cost and processing efficiency than most alternative topologies.

Finally, consider Chklovskii's and Koulakov's (2004) claim that WO explains retinotopic organization in visual cortical areas of the brain. The structure of that explanation is the following:

(Q) Why do visual cortical areas exhibit retinotopy?
(C_1) Neural wiring is metabolically expensive.
(C_2) The metabolic cost of neural wiring is minimized if neurons with adjacent receptive fields are placed as close to each other as possible.
(C_3) The actual interconnections of neurons in visual cortical areas are [such and such].
(U) In conditions C_1-C_3, retinotopic organization of visual cortical areas requires less total length of connecting fiber than any other possible organization.

Chklovskii and Koulakov (2004) explains why visual cortical areas are organized as maps by representing the design problem that cortical maps solve par excellence, i.e., by representing the functional dependencies among the organization of a given set of cortical neurons, the interconnections of those neurons with other visual cortical and retinal neurons, and the wiring cost.

In sum, Wouters's (1999) schema for design explanation correctly characterizes the strategy of WO modeling in neuroscience. Now someone may insist that this does not answer the relevant question: what is the specific contribution of WO models to design explanation?

I recommend to analyze the contribution of WO models in terms of ideal interventions (Woodward 2003). I suggest that the explanatory distinctness of an adequate optimization model is that it enables us to identify "what sort of difference it would have made for the explanandum if the factors cited in the explanans had been different in various possible ways", and thus it allows us to address a range of what-if-things-had-been-different-questions, or w-questions (Woodward 2003, 11). The model defines an optimization problem that helps us to visualize how the target system would behave or be like in a set of counterfactual scenarios in which the values of the relevant decision variables are different. Crucially, such a model would allow us to evaluate the impact on the system's viability of ideal interventions on the decision variables defined by the design problem, even when those interventions are mere conceptual possibilities (Woodward 2003; Schindler 2014).

Woodward's (2003) influential version of the counterfactual view was originally tied to the analysis of causal explanation. In his account, causal explanation has to do with the exhibition of patterns of counterfactual dependence of a special sort, namely, those involv-

ing active counterfactuals. Active counterfactuals are counterfactuals whose antecedents are made true by ideal interventions. An ideal intervention on some variable X with respect to some second variable Y is a causal process that changes X in such a way that, if a change in Y occurs, it is entirely as the result of the change in X. A counterfactual explanation, in this sense, is an answer to a question about what would happen under an ideal intervention.

Several authors have suggested that we should decouple the core idea of Woodward's counterfactualist theory—explanation involves providing counterfactual information about the world—from the interventionist part—explanation involves providing counterfactual information about ideal interventions—(Reutlinger 2016; Chirimuuta 2018). Furthermore, many of these philosophers think that the counterfactual part of the account is the more fundamental (Saatsi and Pexton 2013).[3] In my opinion, one of the comparative advantages of the interventionist approach is the elegant solution it provides to the problem of asymmetry that doomed the covering-law model of explanation, and that is particularly difficult to manage from a counterfactualist perspective. Consider the flagpole example. According to the interventionist approach, the height of the flagpole explains the length of the shadow cast by the flagpole, but not viceversa, because there is an ideal intervention on the height of the flagpole that would change the length of the shadow, making the corresponding active counterfactual conditional true. Ideal interventions have an asymmetry-individuating function, which allows us to discern the asymmetry between explanans and explanandum (Schindler 2014).

Rather than developing a non-interventionist account of counterfactual explanation, a better strategy would be to embrace the idea that ideal interventions need not be physically possible. WO models contribute to design explanations in neuroscience to the extent that they give us epistemic access to (conceptually possible) ideal interventions on the decision variables of the relevant design problem, and evaluate the impact of those ideal interventions on the viability of the target system.

There is an aspect of Woodward's analysis of ideal interventions that encourages this interpretation. In (2003), Woodward asks, "In what sense must interventions be possible?" More specifically, Must interventions on X be physically possible for X to cause Y? He claims that there appear to be cases in which X causes Y but ideal interventions on X in relation to Y are not physically possible. For example, we know that changes in the position of the moon with respect to the earth cause changes in the motions of the tides. However, it is not physically possible to change the position of the moon (e.g. by changing the position of some other massive body) without directly changing the motion of the tides. The ideal intervention on the position of the moon in relation to the motion of the

---

3   Recently, Baron, Colyvan, and Ripley (2017) provide a counterfactual account of extra mathematical explanation of empirical facts that might be helpful to understand the relations between optimization modeling and counterfactual explanation. They use a structural-equation framework to model impossible perturbations to mathematics and the resulting differences made to physical explananda. This proposal is similar to the idea I defend, that wiring optimization models support ideal interventions and assess the impact of those interventions on the viability of the target system. Impossible perturbations involves twiddling mathematical facts. Ideal interventions involves twiddling empirical facts about the organizational features of target systems. A deeper analysis of the relations between wiring optimization and mathematical explanation is very much needed, but it is beyond the scope of this paper. I thank an anonymous reviewer for bringing this work to my attention.

tides is not physically possible, but it is conceptually possible given the relevant mechanical model. Woodward concludes that ideal interventions need only to be logically or conceptually possible.

Similarly, I claim that WO models give us epistemic access to the conceptually possible ideal interventions on the placement of components in neuronal and brain systems, and they enable us to assess the impact of those interventions on the wiring cost, the behavioral value, or other viability conditions of the target systems. The comparison among alternative placements may result in a functional counterfactual of the form: "if the target nervous systems had one of the alternative layouts instead of the layout in question they would face a catastrophic wiring problem." This is the counterfactual claim at the heart of the WO approach. Thus, optimization models give us epistemic access to the relevant ideal interventions required to evaluate the explanatory traction of the WO principle in some intended domains of neuroscience.

6. Conclusions

In this paper, I have reviewed the main features of the WO approach in neuroscience. Then, I have explored some relations between WO modeling and mechanistic explanation. I consider that the 3M constraint sets a necessary condition on WO explanation, but it does not capture the explanatory distinctness of WO models. I have analyzed some relations between WO modeling and Darwinian explanation. I consider that the optimization approach in neuroscience is autonomous in practice from adaptationist assumptions. Finally, I have explored some relations between optimization modeling and design explanation. My suggestion is that an interventionist account of design explanation can shed light on the explanatory distinctness of optimization models in neuroscience.

REFERENCES

Barlow, Horace. 1986. Why have multiple cortical areas? *Vision Research* 26/1: 81-90.
Baron, Sam, Mark Colyvan, and David Ripley. 2017. How mathematics can make a difference. *Philosophers' Imprint* 17.
Boone, Worth, and Gualtiero Piccinini. 2016. The cognitive neuroscience revolution. *Synthese* 193/5: 1509-34. https://doi.org/10.1007/s11229-015-0783-4
Bray, Dennis. 1979. Mechanical tension produced by nerve cells in tissue culture. *Journal of Cell Science* 37 (1): 391.
Bullmore, Ed, and Olaf Sporns. 2012. The economy of brain network organization. *Nature Reviews Neuroscience* 13 (5): 336-49. https://doi.org/10.1038/nrn3214
Cartwright, Nancy. 1999. *The dappled world: A study of the boundaries of science*. New York: Cambridge University Press.
Changizi, Mark. 2003. *The brain from 25,000 feet*. Dordrecht: Springer Netherlands.
Chen, Beth L., David H. Hall, and Dmitri B. Chklovskii. 2006. Wiring optimization can relate neuronal structure and function. *PNAS* 103 (12) 4723-4728. https://doi.org/10.1073/pnas.0506806103
Chen, Yuhan, Shengjun Wang, Claus C. Hilgetag, and Changsong Zhou. 2013. Trade-off between multiple constraints enables simultaneous formation of modules and hubs in neural systems. Edited by Olaf Sporns. *PLoS Computational Biology* 9/3: e1002937. https://doi.org/10.1371/journal.pcbi.1002937

Cherniak, Christopher. 2009. Neuroanatomy and cosmology. In John Bickle, ed., *The Oxford handbook of philosophy and neuroscience*, 370-378. Oxford: Oxford University Press.

Cherniak, Christopher. 1994. Component placement optimization in the brain. *The Journal of Neuroscience* 14 (4): 2418-27. https://doi.org/10.1523/JNEUROSCI.14-04-02418.1994

Cherniak, Christopher. 1992. Local optimization of neuron arbors. *Biological Cybernetics* 66/6: 503-510.

Cherniak, Christopher. 2006. Innateness and brain-wiring optimization: Non-genomic nativism. In Antonio Zilhao, ed., *Evolution, rationality and cognition*, 103-112. London: Routledge. https://doi.org/10.4324/9780203012918-14

Cherniak, Christopher, Zekeria Mokhtarzada, Raul Rodriguez-Esteban, and Kelly Changizi. 2004. Global optimization of cerebral cortex layout. *Proceedings of the National Academy of Sciences* 101 (4): 1081-86. https://doi.org/10.1073/pnas.0305212101

Chirimuuta, Mazviita. 2014. Minimal models and canonical neural computations: The distinctness of computational explanation in neuroscience. *Synthese* 191 (2): 127-53. https://doi.org/10.1007/s11229-013-0369-y

Chirimuuta, Mazviita. 2018. Explanation in computational neuroscience: Causal and non-causal. *The British Journal for the Philosophy of Science* 69 (3): 849-80. https://doi.org/10.1093/bjps/axw034

Chklovskii, Dmitri B., and Alexei A. Koulakov. 2004. Maps in the brain: What can we learn from them? *Annu. Rev. Neurosci.* 27: 369-392.

Chklovskii, Dmitri B., Thomas Schikorski, and Charles F. Stevens. 2002. Wiring optimization in cortical circuits. *Neuron* 34/3: 341-347.

Cowey, Alan. 1979. Cortical maps and visual perception the grindley memorial lecture. *Quarterly Journal of Experimental Psychology* 31 (1): 1-17. https://doi.org/10.1080/14640747908400703

Craver, Carl. 2007. *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford: Oxford University Press.

Craver, Carl and David M. Kaplan. 2014. Towards a mechanistic philosophy of neuroscience. *The Bloomsbury Companion to the Philosophy of Science*, 268-292. London: Bloomsbury.

Essen, David C. Van. 1997. A tension-based theory of morphogenesis and compact wiring in the central nervous system. *Nature* 385 (6614): 313-18. https://doi.org/10.1038/385313a0.

Fisher, Ronald. 1999. *The genetical theory of natural selection: By R.A. Fisher; edited with a foreword and notes by J.H. Bennett*. A complete variorum ed. Oxford: Oxford University Press.

Friston, Karl J., and Chris Frith. 1995. Schizophrenia: A disconnection syndrome? *Clinical Neuroscience* 3/2: 89-97.

Friston, Karl J. 1998. The disconnection hypothesis. *Schizophrenia Research* 30 (2): 115-25. https://doi.org/10.1016/S0920-9964(97)00140-0

Giere, Ronald N. 1999. *Science without laws. Science and its conceptual foundations*. Chicago: University of Chicago Press.

Green, Sara, Arnon Levy, and William Bechtel. 2015. Design sans adaptation. *European Journal for Philosophy of Science* 5/1: 15-29. https://doi.org/10.1007/s13194-014-0096-3

Kaas, Jon H. 1997. Topographic maps are fundamental to sensory processing. *Brain Research Bulletin* 44 (2): 107-12. https://doi.org/10.1016/S0361-9230(97)00094-4

Kaiser, Marcus, and Claus Hilgetag. 2015. Wiring principles, optimization. *Encyclopedia of Computational Neuroscience*, 3172-3177.

Kaiser, Marcus, and Claus Hilgetag. 2006. Nonoptimal component placement, but short processing paths, due to long-distance projections in neural systems. *PLoS Computational Biology* 2/7: e95. https://doi.org/10.1371/journal.pcbi.0020095

Kaplan, David Michael. 2011. Explanation and description in computational neuroscience. *Synthese* 183/3: 339-73. https://doi.org/10.1007/s11229-011-9970-0

Kaplan, David Michael, and Carl F. Craver. 2011. The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of Science* 78/4: 601-627.

Levy, Arnon, and William Bechtel. 2013. Abstraction and the organization of mechanisms. *Philosophy of Science* 80/2: 241-261.

Maynard Smith, John. 1982. *Evolution and the theory of games*. Cambridge ; New York: Cambridge University Press.

Mitchison, Graeme. 1991. Neuronal branching patterns and the economy of cortical wiring. *Proc. R. Soc. Lond. B*, 245:151-158.

Nelson, Mark E, and James M Bower. 1990. Brain maps and parallel computers. *Trends in Neurosciences* 13/10: 403-8. https://doi.org/10.1016/0166-2236(90)90119-U

Potochnik, Angela. 2007. Optimality modeling and explanatory generality. *Philosophy of Science* 74/5: 680-91. https://doi.org/10.1086/525613

Ramón y Cajal, Santiago. 1899. *Textura del sistema nervioso del hombre y de los vertebrados*. Madrid: Librería de Nicolás de Moya.

Ramón y Cajal, Santiago. 1897. *Leyes de la morfología y dinamismo de las células nerviosas*. Madrid: Librería de Nicolás Moya.

Ramón y Cajal, Santiago. 1917. *Recuerdos de mi vida*. Madrid: Librería de Nicolás Moya.

Reutlinger, Alexander. 2016. Is there a monist theory of causal and noncausal explanations? The counterfactual theory of scientific explanation. *Philosophy of Science* 83/5: 733-45. https://doi.org/10.1086/687859

Rice, Collin. 2012. Optimality explanations: a plea for an alternative approach. *Biology & Philosophy* 27/5: 685-703. https://doi.org/10.1007/s10539-012-9322-6

Rice, Collin. 2015. Moving beyond causes: Optimality models and scientific explanation: Moving beyond causes. *Noûs* 49/3: 589-615. https://doi.org/10.1111/nous.12042.

Saatsi, Juha, and Mark Pexton. 2013. Reassessing woodward's account of explanation: Regularities, counterfactuals, and noncausal explanations. *Philosophy of Science* 80/5: 613-24. https://doi.org/10.1086/673899

Schindler, Samuel. 2014. Explanatory fictions—for real? *Synthese* 191/8: 1741-55.

Schwenk, K. 1994. Why snakes have forked tongues. *Science* 263/5153: 1573-77. https://doi.org/10.1126/science.263.5153.1573

Skipper, Robert A., and Roberta L. Millstein. 2005. Thinking about evolutionary mechanisms: Natural selection. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 36/2: 327-47. https://doi.org/10.1016/j.shpsc.2005.03.006

Smith, J. Maynard, and G. R. Price. 1973. The logic of animal conflict. *Nature* 246/5427: 15-18. https://doi.org/10.1038/246015a0

Sober, Elliott. 1983. Equilibrium explanation. *Philosophical Studies* 43/2: 201-10. https://doi.org/10.1007/BF00372383

Sporns, Olaf. 2011. *Networks of the Brain*. Cambridge, Mass: MIT Press.

Sterling, Peter, and Simon Laughlin. 2015. *Principles of neural design*. MIT Press.

Swindale, Nicholas. 2008. Visual map. *Scholarpedia* 3/6: 4607. https://doi.org/10.4249/scholarpedia.4607.

Varshney, Lav R., Beth L. Chen, Eric Paniagua, David H. Hall, and Dmitri B. Chklovskii. 2011. Structural properties of the caenorhabditis elegans neuronal network. *PLoS Computational Biology* 7/2: e1001066. https://doi.org/10.1371/journal.pcbi.1001066

Petra E. Vértes, Aaron F. Alexander-Bloch, Nitin Gogtay, Jay N. Giedd, Judith L. Rapoport, and Edward T. Bullmore. 2012. Simple models of human brain functional networks. *Proceedings of the National Academy of Sciences* 109/15: 5868-73. https://doi.org/10.1073/pnas.1111738109

Vilarroya, Oscar. 2002. Two many optimalities. *Biology & Philosophy* 17/2: 251-70. https://doi.org/10.1023/A:1015264829309

Weiskopf, Daniel A. 2011. Models and mechanisms in psychological explanation. *Synthese* 183/3: 313-38. https://doi.org/10.1007/s11229-011-9958-9

Woodward, James F. 2003. *Making things happen: A theory of causal explanation*. New York: Oxford University Press.

Wouters, Arno. 2007. Design explanation: Determining the constraints on what can be alive. *Erkenntnis* 67/1: 65-80.

Wouters, Arno. 2005. The functional perspective of organismal biology. In Thomas A.C. Reydon and Lia Hemerik, eds., *Current themes in theoretical biology*, 33-69. Berlin/Heidelberg: Springer-Verlag. https://doi.org/10.1007/1-4020-2904-7_3

Wouters, Arno. 1999. *Explanation without a cause*. Zeno, The Leiden-Utrecht Research Institute of Philosophy.

SERGIO DANIEL BARBERIS received his PhD in Philosophy from UBA in 2013. He has been a Fulbright Visiting Scholar at Georgia State University (Spring 2017). He is an Assistant Professor of Philosophy of Science, History of Science, and Metaphysics at the Department of Philosophy, UBA. His interests focus on the philosophy of neuroscience, the history of neuroscience, and the metaphysics of mind.

ADDRESS: Instituto de Filosofía "Dr. Alejandro Korn", Facultad de Filosofía y Letras, Universidad de Buenos Aires, Puán 480, Ciudad de Buenos Aires, Argentina, PO:1406. E-mail: sbarberis@filo.uba.ar