Functional Analyses, Mechanistic Explanations, and Explanatory Tradeoffs*

Sergio Daniel Barberis

Universidad de Buenos Aires;
Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)
sergiobarberis@gmail.com

Recently, Piccinini and Craver have stated three theses concerning the relations between functional analysis and mechanistic explanation in cognitive sciences: *No Distinctness:* functional analysis and mechanistic explanation are explanations of the same kind; *Integration:* functional analysis is a kind of mechanistic explanation; and *Subordination:* functional analyses are unsatisfactory sketches of mechanisms. In this paper, I argue, first, that functional analysis and mechanistic explanations are sub-kinds of explanation by scientific (idealized) models. From that point of view, we must take into account the tradeoff between the representational/explanatory goals of generality and precision that govern the practice of model-building. In some modeling scenarios, it is rational to maximize explanatory generality at the expense of mechanistic precision. This tradeoff allows me to put forward a problem for the mechanist position. If mechanistic modeling endorses generality as a valuable goal, then *Subordination* should be rejected. If mechanists reject generality as a goal, then *Integration* is false. I suggest that mechanists should accept that functional analysis can offer acceptable explanations of cognitive phenomena.

Key words: *functional analysis, mechanistic explanation, model explanation, generality, precision*

---

*I am extremely grateful to Liza Skidelsky for her perceptive comments and advice. I also wish to thank Sabrina Haimovici, Mariela Destéfano, and Abel Wajnerman for extensive discussions on earlier versions of this paper. This article has also been much improved due to the thoughtful questions and comments of four anonymous referees.

Journal of Cognitive Science 14: 229-251, 2013

1. Introduction

Recently, Piccinini and Craver (2011) have stated three theses concerning the relations between functional analysis (hereafter, FA) and mechanistic explanation (ME) in cognitive sciences. First, they deny the functionalist truism in philosophy of cognitive sciences according to which FA and ME are distinct kinds of explanation (*Distinctness*). This denial of *Distinctness* is equivalent to an assertion that FA and ME are explanations of the same kind. Second, these authors maintain a particular thesis concerning the direction of foundation between FA and ME: functional analyses and mechanistic explanations are explanations of the same kind because FA is a specific kind of ME (I will call this thesis: *Integration*). They advance two main arguments in favor of this second thesis. Concerning the *explanandum* of a cognitive explanation, Piccinini and Craver argue that FA is a kind of ME because both try to describe aspects of the same explanatory target, that is, the same multilevel neuronal mechanism. On the side of the *explanans*, they argue that FA is an "elliptical" mechanistic explanation, in which some (or all) of the mechanistic details might be omitted. But, what is *Integration* supposed to mean? It seems odd to assert that FA is, essentially, a kind of elliptical ME, but one in which all the explanatory features that are essential to ME are absent. I believe that the best mechanistic construal of *Integration* is one in which FA is a kind of ME because they share exactly the same explanatory ideals or the same normative commitments concerning explanation. This construal allows the authors to put forth a third thesis about the status of functional explanatory patterns. Since both functional and mechanistic modeling are committed to the same explanatory ideals, but crucial explanatory details are omitted in the former, it can be inferred that FA offers, at best, a faulty or somewhat unsatisfactory explanation for the *explanandum* phenomena it intends to explain (I will call this thesis: *Subordination*).

In this paper, I aim to make a very specific point concerning the mechanist approach to the relations between FA and ME proposed by Piccinini and Craver (2011). I will show that there is a minimalist sense in which *Distinctness* is false and FA and ME are explanations of the same

kind; namely, they are sub-kinds of model explanation (section 2). As model explanations, both FA and ME should be constrained by the representational ideals or explanatory goals that guide the practice of model-based science in general. Three of the most ubiquitously cited representational ideals in scientific modeling are generality, reality, and precision (Levins 1966). Mechanists tend to emphasize mechanistic precision or detail as the main explanatory ideal in cognitive modeling. However, there is another explanatory ideal of at least equal importance: generality. There are good reasons, acknowledged by the mechanists themselves, to believe that generality is an explanatory goal of FA. Importantly, there are tradeoffs between some of these attributes (Matthewson and Weisberg 2008). Particularly, it is not possible to maximize some types of precision without a loss in some types of generality and vice versa (section 3).

Given this context, I argue that a mechanist stance in the vein of Piccinini and Craver (2011) faces a problem (section 4). Either generality is an explanatory ideal of mechanistic explanations, or it is not. If the latter is the case, then FA is not a kind of ME. In that case, *Integration* will be false and functional modeling will not be a kind of mechanistic explanation, because they will differ in their explanatory ideals. On the other hand, if, from the standpoint of mechanistic modeling, generality is considered to be an explanatory virtue, then it is not at all clear that functional analyses are unsatisfactory explanations, despite their lack of mechanistic precision. This is because the FA modeler could be trying to maximize one legitimate explanatory ideal, generality, at the expense of other explanatory attributes, such as precision. In either case, mechanists cannot simultaneously maintain the three theses mentioned above. If generality is an explanatory ideal of mechanistic explanation, then FA will be seen as a legitimate strategy focused on that attribute, and *Subordination* will be false. If generality is not a maxim of mechanistic modeling, then *Integration* will be false and FA will not be a kind of ME, because they will differ in their explanatory ideals.

2. Explanation by Scientific Models

Piccinini and Craver (2011) argue against what they call "the received view"

in philosophy of cognitive sciences regarding the relationships between functional analyses and mechanistic explanations. The received view can be stated by two different, but related, theses. The first tenet is *Distinctness*: FA and ME are distinct kinds of explanation. The second thesis is *Autonomy:* FA and ME are autonomous from one another. According to Piccinini and Craver, *Autonomy* implies *Distinctness* (but not vice versa), so the authors aim to deny *Autonomy* by arguing against *Distinctness*. My goal in this section is to argue that there is a minimal sense in which the negation of *Distinctness,* which we may call *No Distinctness,* is acceptable and in full agreement with the mechanistic stance. Nonetheless, that minimal sense does not imply the negation (or the vindication) of *Autonomy*. In particular, I hold that both FA and ME are kinds of model explanations (Bokulich 2009, 2011). As such, these explanatory patterns are constrained by many representational or explanatory ideals that will be important in the assessment of *Integration* and *Subordination*.

In the context of the cognitive sciences, FA is the analysis of some cognitive capacity, such as visual perception or episodic memory, in terms of the functional components of a system and their organization. Some functional analyses individuate the relevant components both in functional and structural terms, but, crucially, other functional analyses individuate the components only in functional terms, that is, in terms of the causal/functional profile of each of the purported components of the system.[1] What these latter analyses aim to describe is the abstract functional and dynamical organization of the system, disregarding, at least temporarily, the details of the concrete realization of that functional superstructure. The explanatory interest of a FA is partially determined by the relative complexity of the organization of the components attributed to the target system (Cummins 2010, p. 292).

[1] In this paper, I adopt the characterization of functional explanation developed by Cummins (2010). Of course, it is not the only view concerning this subject. Alternatively, one could adopt the "etiological" conception of functional analysis developed by Wright (1973) and then assess the prospects of that explanatory pattern as a kind of mechanistic explanation. I favor the discussion in terms of Cummins's proposal because it seems to be the main philosophical target of Piccinini and Craver's arguments.

In contrast, ME is the analysis of a cognitive phenomenon in terms of the constitutively relevant parts, the causally relevant activities of those parts, and the relevant organizational aspects of those parts and activities in the mechanism that produces the phenomenon (Machamer, Darden and Craver 2000, Craver 2007). The component parts in an acceptable ME must be real parts, which means that they should have a stable cluster of properties, they should be experimentally and theoretically robust, it should be possible to use them in interventions into other parts of the mechanism, and they should be physiologically plausible (Craver 2007, p. 132). Piccinini and Craver admit that, especially in the case of complex biological systems studied by the cognitive sciences, the neurobiological realization of a functional component (its corresponding "structural component") might be so distributed and diffuse as to defy the decomposition and localization heuristics essential to ME. Crucially, the mechanist conception of explanation is supposed to be compatible with the multiple realizability of functional kinds. It should be possible for the same functional component to be realized in different neurobiological kinds (Craver 2007, p. 198; Piccinini and Craver 2011).

According to *Distinctness*, FA and ME, thus characterized, are distinct types of explanations. Piccinini and Craver aim to deny this tenet. The specific way in which they articulate the negation of *Distinctness* is by sustaining a different thesis: namely, that FA is a kind of ME, an elliptical or incomplete description of a mechanism, in particular. If *Integration* is true and FA is an elliptical form of ME, it follows that *Autonomy* is false. I will discuss this assertion in section 3. But first, I would like to explore another conceptual possibility. I believe that there is a minimal sense in which FA and ME are explanations of the same kind (independently of the question concerning the status of FA as a mechanism sketch). Remarkably, this minimal sense in which *Distinctness* is false is neutral regarding the acceptability of *Autonomy*.

The idea I have in mind is that both FA and ME are explanations by scientific models or "model explanations." The key feature of model explanation is that the *explanans* must make essential reference to a scientific model (Bokulich 2011). There is no philosophical consensus concerning the characterization of scientific models or their relations

to scientific theories and empirical data (cf. Morgan and Morrison 1999, Godfrey-Smith 2006). Minimally, a model is a kind of scientific representation of some aspects of the world, but one which essentially involves some degree of idealization, abstraction, and/or fictionalization (Bokulich 2011, Weiskopf 2011a).

There is some diversity among the articulation of the additional conditions that have to be added to this analysis of model explanation in order to distinguish between non-explanatory or "phenomenological" models and potentially explanatory models. The importance of this distinction is highlighted by Craver as follows:

> Models play many roles in science beyond providing explanations […] They are used to make precise and accurate predictions. They are used to summarize data. They are used as heuristics for designing experiments. They are used to demonstrate surprising and counterintuitive consequences of particular forms of systematic organization. But some models have an additional property beyond these others: they are explanations. (Craver 2006, p. 335)

What conditions must be met by scientific models for them to be explanatory? Morrison (1999) holds that models are potentially explanatory to the extent that they exhibit certain kinds of "structural dependencies" to the represented system. Bokulich (2011) holds that a model *M* potentially explains a phenomenon *P* if "the counterfactual structure" of *M* is isomorphic in the relevant respects to the counterfactual structure of *P*. The way in which these features of structural dependency or counterfactual isomorphism must be understood is somewhat obscure. In the context of this paper, it would be wise to stay away from this weighty conceptual issue. For the interim, I will adopt the procedural account of explanatory relevance advanced by Woodward:

> We have at least the beginnings of an explanation when we have identified factors or conditions such that manipulations or changes in those factors or conditions will produce changes in the outcome being explained. Descriptive knowledge, by contrast, is knowledge that (…)

> does not provide information potentially relevant to manipulation. It is in this that the fundamental contrast between causal explanation and description consists. (2003, p. 10)

According to Woodward, then, explanatory models enable us to answer a wide range of counterfactual questions concerning how the system would behave if the factors cited in the *explanans* were different in various ways. Explanations enable us to say both how the target system behaves and how it would behave under a variety of counterfactual conditions or ideal interventions (Craver 2006).

An adequate proposal regarding scientific model explanation must be able to distinguish not only between phenomenological and potentially explanatory models, but also between possible explanations and genuinely acceptable explanations of some phenomena. The idea is that not all model-based explanations are equally acceptable. They must be evaluated relative to the representational/explanatory goals that govern and guide the practice of model-building in that domain. Generality, simplicity, precision, empirical support, and coherence with the rest of scientific knowledge are some of the most cited explanatory ideals of model-based science (Levins 1966, Weisberg 2007). I will describe some of the relations among these explanatory ideals in section 3.

What I would like to stress is that if FA and ME make essential references to scientific (idealized) models, then they are to be considered model explanations, and therefore, there is a clear (although minimal) sense in which both FA and ME belong to the same kind of explanation. There is a great deal of evidence that supports this minimal assertion. First, a mechanistic explanation of some *explanandum* phenomena is essentially linked to the exhibition of at least one mechanistic model. A model M for some phenomenon φ is a mechanistic model of φ if, and only if: (i) M offers a phenomenally adequate description of φ; (ii) the variables for component parts in M represent some of the real parts that are constitutively relevant for the mechanism that produces φ; (iii) the variables for component activities in M represent some of the causally relevant dependencies among component parts in the mechanism that produces φ; and (iv) the organizational features between variables of parts and variables of activities

in *M* represent some of the organizational features in the mechanism that produces *φ* (Craver 2007, Kaplan 2011). Some examples of mechanistic model explanations are MacKinnon's model of the structure of potassium ion channels (Doyle et al. 1998) or the textbook representation of the mechanism of chemical neurotransmission (cf. Kandel et al. 2006).

Furthermore, many paradigmatic functional analyses in cognitive sciences involve the deployment of representational or "cognitive models." A cognitive model aims to explain some psychological capacity by postulating several kinds of (usually subpersonal) mental representations, computational processes that manipulate and transform those representations, and several resources that can be accessed by those computational processes. Some exemplars of cognitive model explanations are Treisman's theory of feature integration in perception (Treisman 1983), Costello and Keane's C3 model of concept combination (Costello and Keane 2000), and the ACT-R model of declarative memory retrieval (Anderson 2007). Weiskopf summarizes the main features of cognitive models as follows:

> Thus a cognitive model can be seen as an organized set of elements that depicts how the system takes input representations into output representations in accord with its available processes and operations, as constrained by its available resources. (2011a, p. 323)

Paradigmatically, then, functional analyses of psychological capacities involve the exhibition and development of some representational/cognitive models. Since both FA and ME make reference to idealized scientific models, both of them can be seen as model explanations.

A remarkable feature of this (minimal) negation of *Distinctness* is that, *pace* Piccinini and Craver (2011), it does not imply the rejection of *Autonomy.* It could be the case that even when both FA and ME are model explanations, they are autonomous in relation to each other in a relevant sense. Analogously, mechanistic explanations and covering law model explanations are both subtypes of model explanations (Bokulich 2011), but they are *prima facie* autonomous from one another. The only general commitment of covering law model explanations is that the *explanans* must make essential use of laws of nature (roughly, non-accidental regularities).

Many of the well-known "optimality models" in evolutionary biology illustrate this explanatory pattern (Elgin and Sober 2002).[2] Even when these optimality models instantiate the general structure of models explanations, it does not necessarily follow that there should be direct constraints between these models and other kinds of model explanations, such as mechanistic ones. Similar considerations might apply to the relations between cognitive and mechanistic model explanations.

Indeed, Piccinini and Craver (2011) accept that functional modeling and mechanistic modeling are "autonomous" to the extent that each one of these practices is allowed to choose which phenomena to explain, which experimental designs to apply, which conceptual resources to adopt, and the precise way in which they are constrained by scientific evidence from adjacent fields. It seems to me that these four kinds of autonomy render functional modeling quite autonomous from mechanist modeling.[3]

Of course, one could argue that there is a more robust or stringent sense in which *Distinctness* is false. It is not only the case that FA and ME are both sub-kinds of the same general kind of explanation, namely model explanation. Furthermore, FA is a sub-kind of ME. This last thesis, which I have called *Integration,* entails the rejection of *Autonomy*. But more remains to be said about the reasons to endorse *Integration.* I will analyze those reasons in the following section.

---

[2] Of course, there are ongoing debates in philosophy of biology concerning the very existence and status of laws of nature. It is not my intention to advance any bold claim concerning this topic here. If one is unsympathetic to the idea that there are natural laws governing the biological realm, it is perfectly acceptable to interpret the "laws" that appear in optimality models and other covering law model explanations as "principles" that govern, in any case, the modeled world, in the vein of the semantic conception of scientific theories (cf. van Fraassen 1989; Giere 1999).

[3] Since the "methodological" varieties of autonomy mentioned in this paragraph are explicitly acknowledged by Piccinini and Craver (2011), it is evident that these comments do not constitute an argument against their position. I mention these kinds of autonomy in order to put them aside and concentrate on the key to *Integration* being possible according to Piccinini and Craver; namely, that FAs are supposedly (bad) mechanistic explanations.

3. Sketchiness, Generality, and Explanatory Tradeoffs

Piccinini and Craver (2011) maintain that FA in cognitive sciences (typically, the specification of representational models in cognitive psychology) is a kind of mechanistic explanation. This thesis is controversial, since we have seen that functional analyses usually characterize the components of a system only in terms of their functional/causal roles in that system, while mechanistic analyses demand not only specification of the functional profiles of the purported components, but also a detailed description of the concrete structures in which those functional properties are realized. In consequence, these authors develop a sophisticated version of *Integration* according to which FA is an "elliptical" mechanistic explanation, one in which the details or structural aspects of the mechanistic explanation are omitted.

What are the features common to FA and ME that justify this sophisticated version of *Integration*? As we have seen, FA and ME diverge in the structure of their *explanantia*. For a model to be mechanistically explanatory, it must necessarily identify in its *explanans* those structural components that can be considered real parts of the mechanism producing a certain phenomenon. If a cognitive model does not meet this necessary condition, then it does not have the structure of a mechanistic explanation. Indeed, the phrase "elliptical mechanistic explanation" seems to be a euphemism for a scientific representation in which all explanatorily relevant factors are absent or implicit; in other words, a model that is not explanatory at all. But that is not the tenet of Piccinini and Craver. What they set out to argue is that purely functional cognitive models are potentially explanatory models. In fact, these models intend or purport to identify the relevant constituents of the mechanism that produces the *explanandum* phenomenon. The point is simply that purely functional cognitive models fail to yield acceptable mechanistic explanations, or, crudely put, that these models offer bad mechanistic explanations.

The argument behind the sophisticated version of *Integration* involves two main tenets. In the first place, Piccinini and Craver (2011) stress that the defenders of FA are committed to precisely the same norms of explanation

or explanatory ideals that mechanists embrace. Two of these explanatory ideals are discussed in the mechanist literature: plausibility, which provides the continuum of how-possibly, how-plausibly, and how-actually mechanistic models; and accuracy/precision, which provides the continuum of sketches, schemas, and ideally complete models of mechanisms (Craver 2007).

Let us consider the first dimension: plausibility. How-possibly models are not phenomenological models, but "loosely constrained conjectures" about the structure and function of the target system. They may exhibit some kind of dynamical organization of parts and activities, but the modeler cannot be sure if those components are real or if they are organized as the model describes. How-actually models, on the other hand, describe all and only real parts, activities, and organizational features of the mechanism that are relevant to the production of the *explanandum* phenomenon (Craver 2006, 2007). In between how-possible and how-actually models are those models that vary in their degree of mechanistic plausibility. Following Weiskopf (2011a), it seems accurate to characterize plausibility as an epistemic dimension of model assessment. In particular, the placement of a given model on this continuum seems to be determined by the degree of evidential support that exists in regard to it. A how-actually model for a domain would fit the majority of empirical evidence that has been gathered for that domain in the relevant scientific fields.

We can now turn to the second dimension of assessment: completeness. The mechanists seem to rely on a pre-analytical or intuitive notion of completeness. A mechanism sketch is a model that may specify some parts and activities of the target system, but that leaves various representational gaps for components whose functional or structural properties are unknown. On the other extreme of the continuum, an ideally complete model does not incorporate any "filler terms" and describes all the features that are relevant for production of the *explanandum* phenomenon. While sketches suppress many mechanistic details concerning the target system, more "complete" models or schemas exhibit greater precision in their description of the system (Weiskopf 2011a). Since we are discussing the assessment of the products of model-based science, and because scientific models always involve some degree of idealization and/or abstraction, it follows that this

account of ideally complete models is more akin to the specification of a regulative ideal than to any description of actual scientific models:

> Few if any mechanistic models provide ideally complete descriptions of a mechanism. In fact, such descriptions would include so many potential factors that they would be unwieldy for the purposes of prediction and control and utterly unilluminating to human beings (…) Ideally complete mechanistic models are the causal/mechanical analogue to Peter Railton's notion of an "ideally explanatory text," which includes all of the information relevant to the explanandum. (Craver 2006, p. 360)

I will expand upon this characterization of completeness as a regulative ideal in the following section. Here, I would simply like to stress that the mechanists tend to limit the discussion of the explanatory goals of modeling to the ideals of plausibility and completeness/precision.

The second tenet of Piccinini and Craver's argument for *Integration* is that purely functional cognitive models are sketchy or imprecise descriptions of the target system, given that they ignore or overlook the specification of values for the relevant parameters that represent structural aspects of the target mechanism. Functionalist modelers aim to maximize both plausibility and precision in the description of a mechanism, but they fall short of that goal. Therefore, cognitive models are unsatisfactory mechanistic models, or so the mechanists conclude.

I would like to argue that, even if one accepts that cognitive models intend to represent the same multilevel mechanisms that mechanistic models do, and even if one concedes that cognitive models represent their target mechanisms without maximizing precision, it does not follow that they are unsatisfactory explanations of the phenomena they intend to explain. My argument in favor of the acceptability of "sketchy" cognitive models relies in the philosophical work of Weisberg and colleagues concerning the structure of tradeoffs in model building (Weisberg 2006, 2007, Matthewson and Weisberg 2008, *inter alia*).

The ideas of this research tradition in philosophy of science were advanced by Levins in the sixties. According to Levins (1966), when

confronted with the task of theoretically representing the structure and internal dynamics of complex systems, the modeler has two main options or approaches. First, she can adopt a "brute-force approach" in which the aim is to build as much of the target system's complexity into the model as possible; that is, to build a model which is "a faithful, one-to-one reflection of this complexity" (Levins 1966, p. 421). The representational ideal associated with this brute-force approach is *Completeness.* According to this ideal, the best representation is one that represents "all aspects of the target phenomenon with an arbitrarily high degree of precision and accuracy" and one in which the "causal connections within the target phenomenon must be reflected in the structure of the representation" (Weisberg 2006, p. 626).

Levins (1966) mentions three main problems with the brute-force approach to complex systems: there would be far too many parameters to measure, the dynamical equations would be insoluble analytically, and, even if they were soluble, the results of those equations would have no meaning for us. Considering these obstacles, a modeler may disregard the ideal of *Completeness* and the brute-force approach and accept from the outset that some aspects of the *explanandum* phenomena will not be incorporated into the model. Weisberg (2006) calls this the "idealization approach." This approach is constrained by many different representational ideals, *Completeness* being only one of them. The philosophers of this tradition usually concentrate on another three *desiderata* of modeling.

The first ideal is generality. It is a *desideratum* of most models (Weisberg 2007) and refers, roughly, to the number of target systems that a particular model or set of models applies to. This notion is ambiguous, containing two different "components" of generality: A-generality and P-generality. A-generality corresponds to the number of target systems the model actually captures. P-generality is the number of possible, but not necessarily actual, target systems it applies to. According to Weisberg (2007), P-generality is often thought to be associated with explanatory power, as we will soon see.

The second ideal is realism. The term "realism" is used, though not clearly explained, by Levins (1996). Weisberg (2006) construes this ideal as being related to the dynamical fidelity or accuracy of the *output* of the model to some aspects of the target phenomenon (predictive accuracy) and/or the fidelity or accuracy in the description of the target system's causal structure.

The particular assessments of fidelity depend on the criteria the modeler adopts when determining whether the model applies to a target. The fidelity criteria that the modeler adopts to assess a particular model will affect its generality, as more permissive criteria will tend to increase the generality of that model, *ceteris paribus*.

Finally, the third representational ideal is precision. It corresponds to the fineness of specification of the parameters, variables, and other parts of model's descriptions (Weisberg 2006). Matthewson and Weisberg (2008) represent a parameter value as the central value for the parameter plus or minus the uncertainty associated with it. The idea is that precision increases as uncertainty decreases. This ideal of precision, in conjunction with realism, seems to be presupposed in the diatribe of mechanists against the use of "black boxes", such as those that are common in purely functional cognitive models.

A crucial feature of the idealization approach to scientific modeling is that there are several tradeoffs among the representational/explanatory ideals mentioned above. It would be perfect to maximize the three *desiderata* of generality, realism, and precision, but it seems that this is not possible. When modeling complex systems, the perfect is the enemy of the good. Tradeoffs are relationships of attenuation that hold between two or more *desiderata* or attributes of model building (Matthewson and Weisberg 2008). Two modeling attributes exhibit attenuation when increasing the magnitude of one attribute makes achievement of the other more difficult. Two *desiderata*, A and B, exhibit a *strict tradeoff* if, and only if, an increase in the magnitude of A results in a decrease in the magnitude of B and vice versa. When two attributes exhibit a strict tradeoff, the modeler must make strategic decisions concerning which attribute ought to be maximized, because when the magnitude of one of these two attributes goes up, the magnitude of the other must go down (Matthewson and Weisberg 2008).

Relevant to my present interests is the well-established fact that there exists a strict tradeoff between precision and P-generality. It is impossible to increase the magnitude of these attributes at the same time; if there is an increase in precision, there follows a decrease in P-generality and vice versa. To present the argument these authors advance in favor of this

thesis, it is indispensable that we introduce the distinction between model descriptions, models, and the target of models (Matthewson and Weisberg 2008, p. 178). Any model description selects a set of models considered as mathematical or abstract structures. A single model description may pick out several models, and one single model may be selected by several descriptions. It is important to bear in mind that precision is an attribute of model descriptions. Suppose that a model description *d* selects a set of models *M1*. If model description *d'* is more precise than model description *d,* then *d'* selects a proper subset *M2* of the models of *d*. Since *M2* is a proper subset of *M1*, the models in *M2* apply to a proper subset of the possible target systems that *M1* applies to. Then, *M2* is less P-general than *M1* and, therefore, increasing the precision of a model description decreases the P-generality of the corresponding model set. The reverse of this argument proves that the attenuation is symmetrical: increasing the P-generality of a set of models decreases the precision of the model description (Matthewson and Weisberg 2008).

The fact that there is a strict tradeoff between precision and P-generality leaves two available strategies for the idealization approach to model building. A modeler can either sacrifice generality to gain precision and realism, or she can sacrifice precision to gain generality and realism. The first strategy is very similar to the brute-force approach. Indeed, according to Weisberg (2006), they are indistinguishable. The sacrifice of generality amounts to the search for a complete and detailed representation of particular phenomena. The second strategy (maximizing generality and realism in detriment of precision) is the one favored by Levins and Weisberg. The reason is that, as I have mentioned, P-generality seems to be directly linked to the explanatory strength of the model. A general characterization of the causal structure of a target system allows us to capture similar but distinct phenomena under the principles or equations of the model (Weisberg 2006). Thus,

> Increasing the generality of a set of models, perhaps by lowering precision, lets theorists treat these systems in a common framework. In so doing, theorists may have a greater ability to determine the underlying features common to these systems, features which may be

> responsible for understanding patterns of interest. (Matthewson and Weisberg 2008, p. 188)

Now, we must remember that mechanists such as Piccinini and Craver (2011) emphasize the representational ideals of realism (accuracy) and precision in the assessment of potentially explanatory models. From that point of view, mechanists criticize purely functional cognitive models as inadequate or faulty explanations, since the modelers who defend cognitive models tend to concentrate on the causal/functional superstructure of the target system while omitting the structural details of the realization or implementation of the functional aspects they identify. It seems evident that modelers of cognitive models are adopting the strategy favored by Levins and Weisberg; that is, those modelers make the strategic choice to maximize the attribute of generality in detriment of the precise details concerning the neurobiological implementation of the abstract system they describe.

I believe that the main reason or rationale behind the strategic choice of cognitive model modelers in favor of generality is related to certain ideas concerning the status of functional kinds. Particularly, the maximization of generality is linked with the (perhaps implicit) acceptance of the multiple realizability of functional kinds. The mechanists themselves acknowledge that the same psychological capacity is fulfilled at different times by entirely different configurations of neural structures (Piccinini and Craver 2011). But more relevant for our purposes is the purported fact that one psychological capacity can be realized in multiple neurobiological *substrata* across species. Weiskopf (2011b) exemplifies the thesis of multiple realizability with the case of multiple realization of lateral inhibition in arthropod compound eyes and vertebrate eyes.

Considered from the mechanist point of view, the compound eye of the horseshoe crab and the camera eye of some vertebrates are as different as two kinds of neurobiological mechanisms can be. The lateral eyes of the horseshoe crab are composed of simple structures known as ommatidia. Each of these ommatidia contains photoreceptive cells that can activate a central eccentric cell. This central cell is connected to adjacent ommatidia, constituting the "lateral plexus". These ommatidia are organized in such a way that the activity of one ommatidium can be inhibited by the

depolarization of adjacent ones. In contrast with this relatively simple structure of the lateral plexus of the crab, the retina of the vertebrate eye is extremely complex. It is organized into several layers, and there is a greater range of cell types with highly specific connectivity patterns. Since these mechanisms differ in the number and complexity of their parts, in the nature of their activities, and in the dynamical organization of those parts and activities, it is not bold to infer that they are two distinct mechanisms. Despite having different neurobiological properties, however, both mechanisms can produce the same phenomenon of lateral inhibition. In this phenomenon, the activity in one kind of photoreceptor inhibits activity in other receptors. This pattern of activation may produce a particular experience known as Mach bands, the appearance of light or dark stripes after the end of a brightness gradient (Weiskopf 2011b).

What this example illustrates is that the same functional property (lateral inhibition between receptors) that accounts for a phenomenon (the perception of Mach bands) is realized in significantly distinct neurobiological mechanisms across different species (the relatively simple compound eye of the horseshoe crab and the complex mammalian camera eye). A very detailed and/or precise model of the horseshoe crab's compound eyes would fail to capture the causal superstructure that those eyes share with vertebrate eyes, namely lateral inhibition, a functional property that accounts for some relevant phenomena, such as the formation of Mach bands. Consequently, it would be rational for a modeler interested in capturing that causal superstructure common to different neurobiological structures to choose a modeling strategy that increases the model's generality. Such a strategy would trade neurobiological precision for explanatory scope.

Therefore, even if a functional model aims to represent the same inter-level mechanism as other mechanistic models, and even if the functional model sacrifices neurobiological precision, it could be the case that the modeler's choice of maximizing generality results in an increase of explanatory scope. The mechanists' choice to maximize precision is not justified *per se* in every explanatory context, and the same normative commitments of the idealization approach to model building legitimate the maximization of generality when we are trying to model complex biological

systems, such as those usually studied by the cognitive sciences.

4. A Problem for Mechanicism

We have seen that even when both FA and ME intend to offer model explanations (section 2), they seem to adopt different strategies in the context of the idealization approach to complex system modeling (section 3). While FA tends to maximize generality at the expense of structural precision or detail, ME tends to maximize precision at the expense of generality. The same point can be formulated as a problem for the mechanist image of the relationships between FA and ME. The relevant question is: does ME endorse generality as a representational or explanatory goal? I have already mentioned that generality is a common ideal for most models in the idealization approach, but the mechanist emphasis on complete descriptions of mechanisms suggests that perhaps the mechanist conception does not adopt the idealization approach, but rather the brute-force approach. If the latter were the case, then the mechanist stance would drop generality as a *desideratum* for model building. This is the interpretation that Bokulich (2011) suggests of Craver's construal of the sketch/schema/ideally complete model *continuum*:

> In my view, this requirement [of completeness] for a model to be explanatory is far too strong. If one has a complete and accurate description of the phenomenon, it is not clear to me that one has a model at all. Indeed this sounds much closer to a theoretical description of the system, than a model. I think that RIG Hughes was absolutely right to say that 'To have a model… is *not* to have a literally true account of the process or entity in question' (Hughes 1990, p. 71). Hence Craver's account succeeds in defending the view that models can explain, only by reducing the notion of a model to a complete and accurate description of the system. (Bokulich 2011, p. 35)

I believe that Bokulich is correct in her diagnosis of Craver's proposal. However, the argument I want to formulate against the mechanist conception of FA does not require considering the mechanist conception

as an instance of the brute-force approach. The objection runs as follows: either the mechanistic pattern of explanation presupposes P-generality as a legitimate representational/explanatory ideal for model building, or it does not. If P-generality is a valuable attribute for mechanist modelers, then they ought to admit that there are relevant modeling scenarios in which adopting a FA can be genuinely explanatory in spite of its lack of structural detail. In that case, it would be false of functional analyses that they are faulty models of mechanisms. It would likewise be false that "explanations that capture these mechanistic details are deeper that those that do not" and that "full-blown mechanistic models are to be preferred" (Piccinini and Craver 2011, p. 307). If P-generality is not a valuable attribute for mechanist modelers, then there is no reason to believe that functional analyses are elliptical kinds of mechanistic explanations. It is important to remember that the best construal of *Integration* is that the defenders of FA and the defenders of ME share the same normative commitments concerning model assessment. However, if FA defenders have an additional commitment to generality as a *desideratum,* and if generality is in tension with other requirements for ME, then it is not the case that FA belongs to the kind of mechanistic explanations. Either way, Piccinini and Craver's philosophical image of the relationships between FA and ME needs to be amended.

There are many promissory notes concerning the direction that such an amendment should take. Despite the emphasis that mechanists put on the precision *desideratum,* it is my position that mechanists do accept P-generality as a valuable goal of model building and that an increase in generality (even at the cost of precision) may enhance, in some contexts, the explanatory power of scientific models. Craver introduces this matter explicitly:

> What, then, is the appropriate degree of abstraction to use in characterizing a kind of mechanism? Characterizing the mechanism very abstractly potentially glosses over sub-kinds of mechanism. Characterizing the mechanism in maximal detail threatens to make each particular mechanism a kind unto itself. (Craver 2009, p. 587)

Craver (2009) exemplifies this issue with different schemata of the

hippocampus. First, we have Ramon y Cajal's schemas of particular hippocampal specimens. These schemas exhibit the precise and detailed locations, shapes, and orientations of the constituent neurons. Next, we have the textbook diagrams of the hippocampal trisynaptic circuit, which are relatively more abstract. These schemata capture the spatial organization of excitatory synapses, but they omit other details concerning inhibitory neurons, support cells, etc. Finally, we have the diagrams in computational models of the hippocampus. These computational models abstract away from most structural details and represent the abstract functional organization among sub-regions of the hippocampus. Each region is represented as performing different functions. This last schema need not be applied only to biological organisms; it could apply to any system that shares its abstract functional superstructure (Craver 2009). In this context, Craver happily accepts that the computational model of the hippocampus, even if it is abstract in regard to almost every neurobiological detail, can offer a genuine explanatory step relative to other, more concrete, scientific representations.

> [E]ach of these schemata makes a nonredundant contribution to our ability to predict, *explain*, and control what the hippocampus does. Where more precise schemata capture relevant distinctions among the mechanisms classed together by the abstract schemata, *the abstract schemata reveal regularities in the behavior of the hippocampus that are invisible in the more precise schemata* (…) For some purposes (surgery, for example) precision is of the utmost importance. For other purposes (such as building an abstract computational model) generality is more important. (Craver 2009, p. 588; emphasis added)

It is evident that mechanists should accept that generality constitutes an important and desirable feature of mechanistic models. The consequence of such an endorsement is that functional models (the kind of models that are legion in cognitive psychology) should not be considered to be faulty or inadequate models of mechanisms. They may be maximizing a legitimate *desideratum* of scientific modeling other than precision: generality. This conclusion is a natural consequence of conceiving both FA and ME as

genuine kinds of model explanation and taking into account the tradeoff between generality and precision.[4]

5. Conclusion

Piccinini and Craver (2011) maintain three different theses. First, functional analyses and mechanistic explanations are explanations of the same kind. Second, FA is a kind of ME because they share the same representational/explanatory ideals. Third, FA offers, at best, a faulty or inadequate explanation of a given phenomenon, because it fails to satisfy the ideal of precision. In this paper, I have argued that these three theses cannot be maintained simultaneously. In section 2, I argued that FA and ME are sub-kinds of model explanation. This minimal assertion denies *Distinctness*, but in a way that does not imply the rejection of *Autonomy*. In section 3, I sustained that if a modeler adopts the idealization approach, there are multiple representational/explanatory ideals that might influence the acceptance of a certain modeling strategy. The three most important of these ideals are generality, realism, and precision. Following the work of Weisberg, I have argued that there are tradeoffs among some of those ideals and that, particularly, there is a strict tradeoff between generality and precision. A modeler cannot increase both generality and precision at the same time, and any increase in one of those magnitudes would imply a decrease in the other.

Piccinini and Craver (2011) exacerbate the centrality of precision in detriment of generality in model building, but there are reasons to believe that there are contexts in which the alternative strategy is preferable in order to obtain an acceptable explanation of the *explanandum* phenomenon. These facts concerning model building allow me to put forward a problem for Piccinini and Craver's position. Either generality is considered a

[4] It could be maintained that the argument about generality and precision tradeoff (in section 4) could have been made without the *No Distinctness* component (in section 2). However, as far as I can see, it is a non-trivial premise of the tradeoff argument that FA and ME belong to the same general class of model explanations. Otherwise, it would not be mandatory for FA and ME to be constrained by (at least some of) the representational ideals that govern model-based science in general.

valuable goal of mechanist modeling, or it is not. In the first case, a mechanist should admit that the maximization of generality in functional modeling is a legitimate strategy. In the second case, FA would not be a kind of mechanistic explanation, since FA and ME would not share the same explanatory ideals. There are good mechanistic reasons to accept that generality is a valuable attribute of mechanistic modeling. Therefore, even in light of their own arguments, mechanists should reject Piccinini and Craver's (2011) position and accept that FA can offer genuine explanations in spite of its lack of mechanistic detail.

References

Anderson, J. 2007. *How Can The Human Mind Occur in the Physical Universe?* Oxford: Oxford University Press.

Bokulich, A. 2009. Explanatory Fictions. In M. Suárez (Ed.) Fictions *in Science. Philosophical Essays on Modelling and Idealisation*. London: Routledge, 91-109.

Bokulich, A. 2011. How Scientific Models Can Explain. *Synthese* 180(1), 33-45.

Costello, F. & Keane, M. 2000. Efficient Creativity: Constraint-Guided Conceptual Combination. *Cognitive Science* 24(2), 299-349.

Craver, C. 2006. When Mechanistic Models Explain. *Synthese* 155, 355-376.

Craver, C. 2007. *Explaining the Brain.* Oxford: Clarendon Press.

Craver, C. 2009. Mechanisms and Natural Kinds. *Philosophical Psychology* 22, 575-594.

Cummins, R. 2010. *The World in the Head*. Oxford: Clarendon Press.

Doyle, D., Morais Cabral, J., Pfuetzner, R., Kuo, A., Gulbis, J., Cohen, S., Chait, B. & MacKinnon, R. (1998). The structure of the potassium channel: molecular basis of K+ conduction and selectivity. *Science* 280, 69-77.

Elgin, M. & Sober, E. 2002. Cartwright of Explanation and Idealization. *Erkenntnis* 57(3), 441-450.

Giere, R. 1999. *Science without Laws*. Chicago: University of Chicago Press.

Godfrey-Smith, P. 2006. The Strategy of Model-Based Science. *Biology and Philosophy* 21, 725-740.

Kandel, E., Schwartz, J. & Jessell, T. 2006. *Principles of Neural Science.* New York: McGraw-Hill.

Kaplan, D. 2011. Explanation and Description in Computational Neuroscience. *Synthese* 183, 339-373.

Levins, R. 1966. The Strategy of Model Building in Population Biology. *American*

Scientist 54(4), 421-431.
Machamer, P., Darden, L. & Craver, C. 2000. Thinking About Mechanisms. *Philosophy of Science* 67, 1-25.
Matthewson, J. & Weisberg, M. 2008. The Structure of Tradeoffs in Model Building. *Synthese* 170, 169-190.
Morgan, M. & Morrison, M. 1999. *Models as Mediators.* Cambridge: Cambridge University Press.
Piccinini, G. & Craver, C. 2011. Integrating Psychology and Neuroscience: Functional Analyses as Mechanisms Sketches. *Synthese* 183, 283-311.
Treisman, A. 1986. Features and Objects in Visual Processing. *Scientific American* 255, 106-115.
Van Fraassen, B. 1989. *Laws and Symmetry*. Oxford: Clarendon Press.
Weisberg, M. 2006. Forty Years of 'The Strategy'. *Biology and Philosophy* 21(5), 623-645.
Weisberg, M. 2007. Three Kinds of Idealization. *Journal of Philosophy* 104(12), 639-659.
Weiskopf, D. 2011a. Models and Mechanisms in Psychological Explanation. *Synthese* 183, 313-338.
Weiskopf, D. 2011b. The Functional Unity of Special Science Kinds. *British Journal for the Philosophy of Science* 62, 233-258.
Woodward, J. 2003. *Making Things Happen: a Theory of Causal Explanation.* Oxford: Oxford University Press.
Wright, L. 1973. Functions. *The Philosophical Review* 82, 2, 139-168.