Models, Mechanisms, and Coherence

Matteo Colombo, Stephan Hartmann, and Robert van Iersel

Abstract Life-science phenomena are often explained by specifying the mechanisms that bring them about. The new mechanistic philosophers have done much to substantiate this claim, and to provide us with a better understanding of what mechanisms are and how they explain. While there is disagreement among current mechanists on various issues, they share a common core position and a seeming commitment to some form of scientific realism. But is such a commitment necessary? Is it the best way to go about mechanistic explanation? In this paper, we propose an alternative antirealist account that also fits explanatory practice in the life sciences. We pay special attention to mechanistic models, i.e. scientific models that involve a mechanism, and to the role of coherence considerations in building such models. To illustrate our points, we consider the mechanism for the action potential.

1. Introduction

According to the mechanistic account of scientific explanation, to explain some phenomenon is to describe the mechanism that produces that phenomenon, where a mechanism consists of components and their associated activities organized such that they produce the phenomenon to be explained (Bechtel and Richardson [2010]; Bechtel and Abrahamsen [2005]; Craver [2007]; Glennan [1996], [2002]; Machamer, Darden and Craver [2000]).

Although not every explanation is mechanistic, mechanistic explanation has proved to be especially fit for advancing our understanding of the structure of several actual explanations in science—particularly in the life-sciences—and of the properties that a good explanation should have. As currently developed, however, the mechanistic account seems to have also an important and problematic feature, which has been largely ignored in the literature (see McKay Illari and Williamson [2011], for an exception). The account appears to be construed so as to presuppose some form of scientific realism.

As noted by Stuart Glennan ([2005], p. 1), mechanistic philosophers share 'realist tendencies,' which emerge in the development of their proposals about the nature and scope of mechanistic explanation. William Bechtel and Adele Abrahamsen ([2005], pp. 424-5), for instance, claim that 'mechanisms are real systems in nature' and that correct mechanistic explanations involve accurate descriptions of mechanisms themselves 'operative in nature'. Carl Craver ([2006], p. 370) holds that '[t]o distinguish good mechanistic explanations from bad, one must distinguish real components from fictional posits'. McKay Illari and Williamson ([2011], p. 820) are yet more explicit. They argue that mechanistic explanation requires mechanisms to be real and point out that 'the mechanisms literature is implicitly or explicitly committed to mechanisms being real. The claim—they go on—is seldom made baldly: perhaps it has seemed too obvious to comment on'. Although the claim might be obvious to mechanist philosophers, scientific realism is in fact problematic and the case for it is clearly not settled.[1]

Here we are not going to provide any novel argument against scientific realism. Our novel contribution is to show that the scientific realism debate is orthogonal to the debate over the nature and scope of mechanistic explanation—despite widespread presumption to the contrary. Accordingly, our aim is to show that good mechanistic explanation can be construed without presupposing scientific realism. Put differently, an antirealist version of mechanistic explanation can be developed such that the distinctive features of mechanistic explanation are preserved. In pursuing this aim, we share Hitchcock's ([1992], p. 175) hope that 'divorcing issues in the debate about scientific explanation from 'prejudices' imported from the orthogonal debate over scientific realism' will facilitate a rapprochement between different accounts of explanation.

If the mechanistic account could steer clear from realist commitments while retaining its distinctive attractive features, it would be an advantage for those who wish to subscribe to it, but do not have realist tendencies. Furthermore, if good mechanistic explanations can be construed without presupposing scientific realism, it would be clear that the distinctive attractive features of mechanistic explanations do not depend on a commitment to scientific realism. Hence, it will be the realist mechanists' burden to show us what exactly their extra-commitments contribute to scientific explanation.

The main claims defended in this paper are twofold: First, scientific realism is unnecessary for good mechanistic explanation; and hence one can subscribe to the mechanistic account of explanation without thereby committing herself to scientific realism. Second, there is prima facie reason to think that a mechanistic account with a realist bent faces problems that an antirealist mechanist can avoid; hence, there is at least

[1] There are at least four types of considerations that make scientific realism problematic. First, the epistemic status of unobservable aspects of the world, to which scientific explanations often appeal, is controversial. Second, arguments from the underdetermination of theory by data put pressure on the idea that an explanation should be chosen among its rivals based solely on evidence that may be relevant to its truth. Third, the history of science suggests that a realist understanding of science is implausible. Fourth, some patterns in the history of science give us reason to believe that there are, at present unconceived by us, fundamentally distinct, equally promising and well-confirmed alternatives to our best contemporary scientific theories (Chakravartty [2011] for a recent survey of arguments for and against scientific realism).

prima facie reason to think that it is preferable for mechanists to adopt an antirealist stance instead of a realist one. Most of this paper will be devoted to the first claim.

The paper will proceed as follows. Section 2 outlines some core, distinctive aspects of mechanistic explanation, and highlights important distinctions bearing on the explanatory status of mechanistic models. Section 3 explicates the realist stance, which seems to be widely presupposed by current mechanisms literature. Section 4 begins by introducing the case-study of the mechanism for the neural action potential. This example will help us to argue that presupposing realism is unnecessary, as an antirealist stance does at least equally well in preserving the core features of the mechanistic account. Our argument hinges on the role of modelling procedures, and coherence considerations in the process of specifying a mechanism. Section 5 clarifies further the proposal developed in the previous section by addressing three outstanding questions for the antirealist mechanist. Section 6 provides grounds for the second claim advanced by the paper. It argues that the antirealist stance might be preferable because some of the problems that the new mechanists face can be avoided *more easily* from this stance, rather than from the realist stance. Section 7 concludes.

2. Some Core Features of Mechanistic Explanation

According to the mechanistic account, scientific explanation often involves a description (or a model) of the mechanism that produces a phenomenon of interest.[2] Mechanistic explanation involves constructing a model that represents the spatio-temporally organized parts of a mechanism and their causal interactions (i.e. the activities of the mechanism) such that they enable the mechanism to produce the phenomenon of interest (see Craver and Bechtel [2006] for a concise overview). Organelles, cells, hearts, brains, and whole organisms are examples of mechanisms studied and modelled in the life sciences. Mechanisms studied in the life sciences produce phenomena like blood circulation, the action potential, long-term potentiation (LTP) photosynthesis, and protein synthesis.

Adequate mechanistic models need not only include information about the parts, causal activities, and spatial-temporal organization of the mechanism producing the phenomenon of interest. According to Craver ([2006], [2007]), what is required of an adequate mechanistic model is that it is representationally accurate and “how-actually.” Craver lays out three distinctions, which are useful to assess the explanatory adequacy of a given model. He distinguishes between (1) how-possibly and how-actually models, (2) sketches and ideally complete models, and (3) merely phenomenal models and genuinely explanatory models.

The first distinction, introduced in the literature by Machamer, Darden and Craver ([2000], pp. 21-2), is elaborated by Craver ([2006], [2007]). *How-possibly* models are ‘only loosely constrained conjectures about the mechanism that produces the explanandum phenomenon’ (Craver [2006], p. 361). Although how-possibly models describe organized parts and activities such that they can produce the phenomenon to be

[2] For present purposes, we assume that models can be physical objects, fictional objects, set-theoretic structures, equations, or linguistic descriptions, and that one of the primary functions of models is to represent some aspect of the world (Frigg and Hartmann [2012]).

explained, these are mere posits, which might not pick out any entity or activity in the world. In Craver's words 'one can have no idea if the conjectured parts exist and, if they do, whether they can engage in the activities attributed to them in the model' (Ibid.). *How-actually* models, instead, 'describe real components, activities, and organizational features of the mechanism that in fact produces the phenomenon' (Ibid.). How-actually models are not 'loose conjectures.' Unlike how-possibly models, in providing a how-actually model, one should believe that the organized parts and causal activities described in the model pick out corresponding entities and activities in the world. How-actually models are explanatory; how-possibly models are not.

Craver ([2006], [2007]) does not spell out how we should determine whether a given model is nearer to the how-possibly or how-actually end. We agree with Weiskopf ([2011]) that, although terminologically it might sound more natural to say that a how-actually model is just the true or accurate model of its target mechanism, it is plausible that the distinction between how-possibly and how-actually models is in fact an epistemic one that turns on degrees of evidential support. To begin with, Machamer et al. ([2000], p. 21) frame the distinction in epistemic terms, in reference to the intelligibility bestowed by a mechanistic explanation upon the phenomenon it purports to explain. Second, Craver's ([2006], [2007]) own reconstruction of case-studies that illustrate the distinction between how-possibly and how-actually models is best understood in terms of confirmation. For example, Craver's argument that the Hodgkin-Huxley model of the action potential is a how-possibly model relies on the fact that Hodgkin and Huxely 'had no evidence favoring their model over other possible models' (Craver [2007], p. 115; see also Craver [2007], pp. 117ff on how Bertil Hille's work on the action potential fits this characterisation of how-possibly/how-actually models). Accordingly, even if a model is *in fact* a true or accurate representation of its target mechanism, if we lack evidence in its favour, then it is a how-possibly model. Third, as argued by Weiskopf ([2011]), if the distinction were in terms of truth, then it would make little sense to say that a how-possibly model can turn out to be a how-actually model. But, in fact, 'any one of a set of how-possibly models might turn out to accurately model the system, so the difference in how they are placed along this dimension cannot just be in terms of accuracy. So it seems that this is fundamentally an epistemic dimension. It represents something like the degree of confirmation of the claim that the model corresponds to the mechanism' (Ibid., p. 316).

While a how-actually model is both *well*-confirmed by the available evidence *and* the *best*-confirmed model among the available models of a given explanandum phenomenon, how-possibly models are not well-confirmed (but neither disconfirmed). Given a set of models of a given phenomenon to be explained, the model that is both *well*-confirmed and the *best*-confirmed by the evidence available at the time is a how-actually model. In light of this consideration, and from the way Craver ([2006], [2007]) characterizes a how-possibly model in terms of 'loosely constrained conjecture,' it follows that in providing a how-actually model, one is justified in believing that the claims made by the model correspond to how things stand in the world because that is the model that is currently best confirmed, above a certain evidential threshold, by the available evidence. This distinction is crucial for the argument articulated in this paper, and we shall return on it in the next section.

Let us consider the second type of distinction, the one between sketches and ideally complete models. A *sketch* "is an incomplete model of a mechanism. It characterizes some parts, activities, and features of the mechanism's organization, but it has gaps" (Craver [2006], p. 360). Instead, *ideally complete descriptions of a mechanism* 'include all of the entities, properties, activities, and organizational features that are relevant to every aspect of the phenomenon to be explained. Few if any mechanistic models provide ideally complete description of a mechanism' (Ibid.).

The distinction between sketches and ideally complete models turns on two dimensions: the model's detailedness and relevance with respect to the explanandum phenomenon displayed by its target system. If a model takes account of more features of the system modelled than another, then it is more detailed. If a model takes account of more features of the system that are relevant to the aspects of the phenomenon to be explained than another, then it is more pertinent. So, given a set of models to account for an explanandum phenomenon, the most representationally complete model is the one that takes account of more of the features of the system modelled that are relevant to the aspects of the phenomenon.

The third distinction is between merely phenomenological and genuinely explanatory models. Merely phenomenological models provide only descriptions of the phenomena to be explained. They are consistent with observations about the system modelled, since they are often constructed via ad hoc fitting of the model to the empirical data. And they often allow one to make predictions about the system. Hence, phenomenological models are descriptively and often predictively adequate. Yet, they do not give us genuine explanatory insight, because they do not reveal the causal structure underlying the explanandum phenomenon, and, more generally, do not show why the dependencies captured by the models are as the models describe them. Ptolemaic models of the solar system are prominent examples of merely phenomenological models.

Genuinely explanatory models can be distinguished from merely phenomenological models on the basis of the control and manipulation they afford, and of their ability to answer "what-if-things-had-been-different" questions (Craver [2006], p. 358). Accordingly, the more a model affords us with possibilities for manipulations and control over the target system's behaviour, and the more it allows us to answer counterfactual questions, the more it is genuinely explanatory as opposed to merely phenomenological.

Before concluding this section, we should note another distinction between current accounts of mechanistic explanation, viz.: the distinction between the epistemic and the ontic sense of explanation (Salmon [1984]). According to the epistemic sense, mechanistic explanation is an epistemic activity aimed at increasing understanding of the world (Bechtel and Abrahamsen [2005]). 'Thus, since explanation is itself an epistemic activity, what figures in it are not the mechanisms in the world, but representations of them' (Ibid, p. 425). According to the ontic sense, mechanistic explanation consists in a portion of the world producing the explanandum phenomenon. It is the mechanism *itself*, instead of some representation of the mechanism, that does the explaining independent of scientists' epistemic goals, interests, and abilities at identification (Craver [2007]). It might be expected that mechanists' realist presuppositions depend on the sense of explanation to which they subscribe. For it seems more natural for the ontic mechanistic account of explanation than for the epistemic account to be developed against a realist background. However, this is not the case.

Current mechanistic accounts of explanation, in either the epistemic or the ontic sense, equally share realist presuppositions. The realist background of mechanistic explanation is perhaps clearer on the ontic view, since for mechanisms themselves to explain they must be real entities in the world that actually produce the phenomena they explain. But even on the epistemic view championed by Bechtel and Abrahamsen, a realist stance is explicitly maintained: good mechanistic explanations provide descriptions of 'real systems in nature' (Becthel and Abrahamsen [2005], pp. 424-5). As McKay Illari and Williamson ([2011], p. 820) argue: 'If the epistemic sense of explanation is to succeed in increasing understanding of the *world*, rather than merely making up interesting stories about it, the stories had better be describing the mechanisms in the world'. According to this view, explanatory power is a feature of a model (or some other representation); and a model has this feature only if it has the "right" relationship to the world (e.g. the model is true, correct, accurate, or isomorphic to worldly facts).

In sum, in spite of two distinct senses of mechanistic explanation, viz. the epistemic and the ontic sense, the shared core of the mechanistic position can be summarized in light of the distinctions laid out above. Mechanistic explanation involves the description (or a model) of the mechanism producing the explanandum phenomenon. A mechanism is a set of entities (i.e. the component parts of the mechanism) and their associated causal activities spatio-temporally organized in such a way as to produce the explanandum phenomenon. How-possibly and merely phenomenological models are not genuinely explanatory. A model is genuinely explanatory to the extent that:

- it describes 'real components, activities, and organizational features of the mechanism that in fact produces the phenomenon' (Craver [2007], p. 112). That is, to the extent that it is a how-actually model.
- The model is representationally complete in that it is detailed and relevant to the explanandum phenomenon displayed by its target system.
- The model affords several possibilities for manipulation and control over the behaviour of the target system, and it answers several counterfactual questions about the behaviour of the target system.

With these claims in the background, the next section will specify in what consists the seeming realist presuppositions of current accounts of mechanistic explanation.

3. Scientific Realism and Mechanistic Explanation

At the core of scientific realism there is an optimistic epistemic attitude towards the outputs of scientific investigation (e.g. Chakravartty [2011]; Psillos [1999]). According to this epistemic attitude, our best scientific explanations, theories and models are (at least approximately and/or probably) true descriptions of how things stand in the world, and we can know them to be so. Scientific realism holds that we are justified in believing that the claims made by our best scientific explanations, theories and models about those parts of the world that are extremely small, remote in space or time, or otherwise inaccessible are literally (or at least approximately and/or probably) true. The core epistemic attitude underlying scientific realism can involve a metaphysical thesis, and a semantic thesis. The metaphysical thesis is that the world has a mind-independent natural-kind structure; when engaging in scientific investigation we discover features of the world that exist independently of any observer. The semantic thesis is that our best scientific theories give us literal descriptions of their intended domain and that theoretical assertions have truth values.

The realist presuppositions of mechanist philosophers include epistemic as well as metaphysical claims directed at both the components and the causal activities of which mechanisms consist. At the metaphysical level, mechanists have claimed that the components and activities, of which mechanisms consist, are mind-independent features of the world (e.g. Bechtel and Abrahamsen [2005], pp. 423-5; Glennan [2011], p. 12; Machamer, Darden and Craver [2000], Section 3). At the epistemic level, mechanists have claimed that we can have epistemic access to mechanisms. In particular, we are justified in believing that the descriptions of otherwise inaccessible parts and activities of the mechanisms involved in mechanistic explanations are literally (or at least approximately and/or probably) true (e.g. Bechtel [2008], Section 1.5; Glennan [2005], p. 19; Machamer, Darden and Craver [2000], Section 7). According to the optimistic epistemic stance that is shared by current mechanists, adequate mechanistic explanations yield knowledge of all the aspects of the world that they describe. This kind of epistemic stance, along with the claim that mechanistic models are explanatory only if they are how-actually models, describing real mechanistic components and activities in the world, implies that we are justified in believing the claims made by mechanistic explanations to be literally (or at least approximately and/or probably) true (cf. Craver [2007], Chapter 4; Glennan [1996], p. 52).[3]

We shall be mostly concerned with this optimistic epistemic stance. We put forward an antirealist[4] variant of mechanistic explanation, which shows that mechanistic explanation can do without a commitment to scientific realism while still fitting scientific practice. The antirealism we advocate denies two claims. First, it denies that adequate mechanistic explanations need be (at least approximately and/or probably) true descriptions of mechanisms in nature. This is because the notion of a how-actually model, which—as we saw in the previous section—provides a necessary condition on the adequacy of a given mechanistic explanation, is most plausibly understood epistemically as a model that is best confirmed, above a certain threshold, by the available evidence. From this notion, however, it does not follow that how-actually models should be regarded as true descriptions. To the extent that mechanistic explanations involve how-actually models of the explananda phenomena, they need not be true descriptions of pieces of the world.

[3] This conclusion is underwritten by claims in the mechanism literature such as: 'How-possibly models […] are not adequate explanations. In saying this, I am saying not merely that the description must be true (or true enough) but further, that the model must correctly characterize the details of the mechanism [producing the explanandum-phenomenon]' (Craver [2006], p. 361). 'Descriptions of mechanisms are good descriptions insofar as they describe what is "really" there' (Glennan [1996], p. 52).

[4] One referee suggested that our position could be best labelled "non-realist" rather than "antirealist" to emphasise that it does not presuppose scientific realism as all the existing accounts of mechanistic explanation do. However, we prefer to use the term "antirealism" because, as it will be clearer, we hold not only that mechanistic accounts of explanation need not presuppose scientific realism (i.e. a non-realist position), but also that there is reason to prefer a mechanistic account with an antirealist bent.

Our second denial is that mechanistic explanations require us to believe all the claims they make about nature itself. The first denial is motivated by pragmatic considerations: ineliminable aspects of scientific practice suggest that at least some mechanistic explanantia can simply be seen ‘as powerful and reliable instruments for mediating [our] engagement with a variety of mechanical phenomena’ without also their needing to be (approximately and/or probably) true descriptions of nature itself (Stanford [2009], p. 388). The second denial is motivated by the sort of epistemic prudence endorsed by Stanford ([2006], p. 211): the historical record of scientific inquiry makes it likely that even our best-confirmed and pragmatically successful current explanations, theories and models will ultimately be found wanting and ‘replaced by more powerful conceptual tools offering fundamentally different conceptions of nature that have presently not yet even been conceived.’

Informed by such motivations, our antirealist variant of mechanistic explanation has two aspects, on which we rely to show that good mechanistic explanation need not presuppose realist commitments but only a pragmatist, epistemically prudent attitude. The first aspect consists in a certain understanding of the process of modelling a mechanism. The second aspect of our proposal focuses on the role of coherence in mechanistic modelling. The first aspect is concerned with how mechanisms can be established in the course of modelling phenomena, while the second one is concerned with the justification of a postulated mechanism. Mechanistic modelling and coherence play a prominent role in scientific practice. Hence, our antirealist account will fit at least these two prominent features of scientific practice. After having specified our realist target, and qualified our own antirealist proposal, it is now time to introduce these two basic aspects of our proposal, which will be articulated in Section 4.

According to the antirealist position on offer here, mechanistic modelling is understood as the process by which a mechanistic model is *defined*. A mechanistic model is a model that describes a target system *as* a mechanism. The model defines organized entities and activities such that its target mechanism displays certain phenomena of interest. The model does not require us to believe that all the features of the target system it defines *are* real features of a mechanism existing in nature. Nonetheless, this way of understanding modelling does not block us from making useful distinctions of the sort made in Section 2.

A how-actually model at a given historical time will be the one that is best-confirmed, above a certain evidential threshold, by the evidence available at that time. A genuinely explanatory model will be one that answers several counterfactual questions as well as affords opportunities for controlling and manipulating the behaviour of the target system. A representationally complete model will be one that is both detailed and relevant to its target system.

These distinctions, however, are not sufficient to address the issues of (i) under which circumstances we should believe that a mechanistic model is a true description of some aspect of the world, and (ii) which of the particular claims made by a mechanistic model we are most entitled to believe. Even if a model is considerably better confirmed, above a certain evidential threshold, by the available evidence than any rival that we have considered—being thereby a how-actually model—two questions relevant to address issues (i) and (ii) remain. The first question is whether we have reason to believe that there are equally well-confirmed alternatives to that model that are presently unconceived by us; the second question is whether for some of the claims made by the model we have epistemic access *independent* of the underlying theory's own descriptive apparatus (cf. Roush [2005]; Stanford [2006]). The fact that we have a how-actually model does not justify per se a realist attitude towards the model.

Good mechanistic models will serve as reliable instruments for engaging with the phenomena displayed by the mechanism without the models being necessarily true or without the need to commit ourselves to believe all claims made by the model. This will allow us to maintain a relationship of epistemic access towards those features of a target mechanism that e.g. can be understood independently of the underlying theory, while remaining epistemically agnostic about the truth of at least some of the current how-actually models, or about the truth of some claims derived from a model about its target. The case study in the next section will further clarify the type of modelling involved in our antirealist proposal.

For now, let us introduce the second aspect of our proposal: *coherence*-based justification. While realists might attempt to justify the success of a proposed mechanistic explanation by appealing to the truth of the mechanistic explanans, arguing that the description of the mechanism featuring in the explanation picks out certain entities, activities, and their properties existing in nature, this avenue is not open to the antirealist. The main constraint of antirealist mechanistic modelling is that the model is a reliable instrument that allows us to guide our pragmatic engagement with the world, satisfy our interests, and achieve our practical ends. This constraint opens the door to two concerns. First, on which grounds should a mechanistic explanation be accepted if not on its truth? Second, how can it be warranted that mechanistic explanations are improved over time? After all, there could be many, mutually inconsistent, mechanistic models that serve our interests and practical equally well, and also save the phenomena. These types of concerns can be defused pragmatically, and in agreement with many episodes in the history of science (cf. Kuhn [1962]), by appealing to *coherence* considerations, and by highlighting that, at least under a plausible explication of this notion, coherence is in general *not* truth-conducive.

A mechanistic model is never proposed in isolation, but is always embedded in a network of other mechanistic models, explanatory theories, and knowledge about the world. A mechanistic model is always embedded in a background belief system. Different research programs might incorporate different beliefs in their respective background belief system at a given time in history. Within a research program, once certain beliefs are provisionally assimilated in the program's background belief system, new beliefs and posits typically cohere with such a background belief system.

'Coherence' can be understood here with BonJour's ([1985], p. 93), intuitively, as 'how well a body of belief "hangs together".' Following this line of thought we would then accept an explanation delivered by a proposed mechanistic model only if it makes a given system of beliefs more coherent, or at least only if it does not make it less coherent. Coherence considerations play an important role in many examples in the history of science,[5] and—as shown below—in current practice in cognitive neuroscience, where, for

[5] Kosso ([1992], Chapter 8) and Thagard ([1992]) emphasise the role of coherence in scientific practice as driving force in the development and justification of explanatory theoretical systems. They both illustrate this role of coherence with several cases-studies

example, an explanation of a certain neural phenomenon at a given level of structural description—say at the level of neural systems—is accepted in so far as it is coherent with accepted explanations at other levels of organization (cf. Churchland [1986]). Drawing on coherence considerations, the extent to which a proposed mechanistic explanation is a good one, and constitutes scientific progress will depend on whether it coheres well with the best available, accepted background knowledge (Hartmann [2001]).

One way to precisely explicate 'coherence' quantitatively is within a probabilistic framework, where coherence is taken to be "a confidence boosting property" of an information set (e.g. a set of beliefs, or a model), and the coherence of a given information set exhibits itself in the probabilistic relations that hold between the propositions in the set (Bovens and Hartmann [2003]). Remarkably, a number of results within this area of research have shown that under certain plausible conditions coherence cannot fully succeed in being truth-conducive (Bovens and Hartmann [2003], Chapter 2; Olsson [2005]). Coherence considerations could thus guide us in choosing between alternative mechanistic explanations within a certain belief system, thereby helping us to make progress in science; and yet a high degree of coherence of a belief system will not generally secure a high probability of truth.

Realists might be happy to concede that coherence plays an important role in scientific practice and mechanistic modelling. And yet they may complain that this does not underwrite an antirealist attitude towards mechanistic explanation. The problem with this complaint is that 'coherence' is a vague notion. As we pointed out, one way to make it precise is within formal epistemology. But, once 'coherence' is made precise in this way, the link between coherence and truth becomes problematic. Realists' rejoinders to our proposal will then have to meet three challenges. First, realists will have to provide us with a precise characterisation of 'coherence'; second, they will have to show us that coherence is generally truth-conducive, and, therefore, does not obviously underwrite an antirealist attitude; third, they will have to construct new arguments that the conditions required for coherence to be truth-conducive are generally satisfied in scientific cases.

By pursuing coherence-based mechanistic modelling in the process of explaining a certain phenomenon from an antirealist stance, it can be demonstrated that 1) the distinctive features of the mechanistic account remain unaffected, 2) the resulting mechanistic explanation is genuinely explanatory, and 3) a realist stance is an unnecessary baggage. The next three sections will establish these points, illustrating them with the case of the action potential.

4. Antirealist Mechanistic Explanation. The Case of the Action Potential

We focus on the mechanism for the action potential for two reasons. First, the action potential is a fundamental neural mechanism, whose modelling has been of considerable importance for the formation of neuroscience as a discipline. Second, this mechanism is

form the history of science. More recently, Thagard ([2007]) explores the relationship between scientific theories, truth, and coherence, and argues that explanatory coherence leads to approximate truth. *However*, his argument overlooks the impossibility results proved by Bovens and Hartmann ([2003], Chapter 2) and Olsson ([2005]) that coherence does not lead to higher likelihood of truth.

now fairly well-understood and has already been dealt with in current mechanisms literature (Bogen [2005]; Craver [2006], [2007]; Machamer et al. [2000]; Weber [2008]). We can therefore refer to this literature and build on it. Although we shall refer in passing to the well-known Hodgkin-Huxley model, our focus is not the Hodgkin-Huxley model, which, as we'll note, continues to be actively debated among current mechanists. Rather, our focus is the action potential mechanism, which we take to be representative of the mechanistic literature in general.

We begin with some science. Neurons mainly communicate using action potentials (or spikes). The action potential is a brief (roughly 1 ms) event, consisting of a roughly 100 mV fluctuation in the electrical potential across the cell membrane. In its resting state the neuron maintains a potential inside its membrane of about – 70 mV relative to that outside the cell (which is conventionally defined to be 0 mV). The cell membrane is thus said to be polarized. This polarization depends on the difference in ion concentrations inside and outside the cell, predominantly sodium (Na^+), potassium (K^+), calcium (Ca^{2+}) and chloride (Cl^-). The flow of ions across the cell membrane is controlled by ion channels, which are selectively permeable proteins in the cell membrane. They open and close stochastically, letting ions flowing into and out of a cell. Whether a channel is open or close depends on the voltage and ion concentration gradients across the membrane as well as on a variety of internal and external signals. The process whereby the membrane potential becomes more negative is called hyperpolarization. Conversely, the process whereby the membrane potential becomes less negative or even positive is called depolarization. These two processes depend on the flow of positively charged or negatively charged ions into and out of the cell. If the neuron is depolarized sufficiently to raise the membrane potential above a certain threshold level (about – 50mV), a positive feedback process triggers the action potential: the membrane potential abruptly shoots upward (to about + 40mV) and then abruptly shoots back downward, generally below the resting level. It is noteworthy that the generation of action potentials also depends on the recent history of the cell electric activity. It may be impossible to evoke another action potential for a few milliseconds after an action potential has been generated. This period is called the absolute refractory period, and is largely responsible for the unidirectional propagation of the action potential along the axon. For a longer interval, known as relative refractory period, which lasts up to tens of milliseconds after an action potential, it is more unlikely that the neuron will fire another action potential.

In 1963 Alan Hodgkin and Andrew Huxley were awarded the Nobel Prize in Physiology or Medicine 'for their discoveries concerning the ionic mechanisms involved in excitation and inhibition in the peripheral and central portions of the nerve cell membrane.' Their model of the action potential, formulated as a set of nonlinear ordinary differential equations, is one of the landmarks in the history of neuroscience. Yet, in order to satisfactorily explain the phenomenon described above, Hodgkin and Huxley's model may not suffice. Some mechanist philosophers have argued that one reason why this model is not explanatory is that it doesn't describe '*how* the membrane changes its permeability' (Craver [2006], p. 364, emphasis in original; see also Bogen [2005]). In other words, Hodgkin and Huxley's model provides a description of how action potentials are initiated and propagated, but it "embodies no commitments as to the mechanisms that change the membrane conductance, allow the ionic currents to flow, and coordinate them so that the action potential has its characteristic shape" (Craver [2006], p. 364). Hence, it is not explanatory.[6]

According to Bogen ([2005]) and Craver ([2006]), there are two reasons why Hodgkin and Huxley's model is not explanatory. First, it is incomplete (or 'sketchy'), as it only characterizes a small sub-set of the entities and activities that are relevant to explain the action potential. Second, and more importantly, it is a how-possibly model instead of a how-actually model, since it is was put forward as a conjecture, uncommitted to the existence of the entities and activities that it includes. Sketchy models and how-possibly models would not adequately explain because they do not show how all the relevant components, properties, and activities of a real mechanism produce the phenomenon to be explained. 'At most [the Hodgkin and Huxley's model] provides a 'how-possibly' sketch of the action potential' (Craver [2006], p. 366). Hence, it does not adequately explain.

Hodgkin and Huxley did acknowledge that more needed to be known about the biophysical features of ion channels in order to reach firmer conclusions about the mechanism of permeability change that they 'tentatively had in mind when formulating' their model (Hodgkin and Huxley [1952], p. 541). Nonetheless, after more than half a century, neuroscientists recognize that, in spite of several limitations, the 'Hodgkin–Huxley explanation of the action potential—regenerative depolarization due to a steeply voltage-dependent sodium-selective conductance followed by repolarization due to inactivation of this conductance along with activation of a slower potassium conductance—has been confirmed beyond doubt' (Bean [2007], p. 462, Box 3). The conceptual framework that the Hodgkin–Huxley model put in place is in fact still widely used to quantitatively understand how lower-level events such as voltage-dependent membrane conductances relate to such macro-level changes as spike generation (Dayan and Abbott [2001], Chapter 5). In general, it is not obvious that a mechanistic model is explanatory only in so far as it provides us with a non-gappy and (approximately and/or probably) true description of its target mechanism (cf. Bokulich [2011]; Levy [forthcoming]). We deny this claim, and by drawing on the case of the action potential, we now lay out our antirealist mechanistic proposal.[7]

In order to provide such a mechanistic model, the first step is to clearly identify the phenomenon to be explained. Phenomena can be plausibly understood as 'features of the world that in principle could recur under different contexts or conditions' (Woodward [2011], p. 166; see also Bogen and Woodward [1988]). Both realists and antirealists can subscribe to this characterization—at least if we do not restrict it in question-begging ways.[8] The disagreement between realists and antirealists is in fact not about the notion of 'phenomenon,' but about epistemic commitments.

[6] Against Craver's and Bogen's views, Weber ([2008]) argues that the Hodgkin-Huxley model is explanatory, as it successfully identifies causal relations underlying the action potential.

[7] The step-by-step reconstruction of mechanistic explanation we articulate should not suggest that scientists follow some "recipe" in providing a mechanistic model, nor that, in practice, the steps we describe are pursued in an orderly manner.

[8] Phenomena should be distinguished from data. 'Data are public records produced by measurement and experiment that serve as evidence for the existence or features of

One way to identify the explanandum phenomenon is in two steps: First, we posit some mechanism producing the phenomenon, treating this mechanism as a black-box at this initial stage. Second, we describe the outputs of the posited underlying mechanism. These consist of what can be measured, or data, about the phenomenon of interest, given available methods and technologies. For instance, some important outputs of the mechanism of the action potential are the rate at which the action potential raises, its peak magnitude, its subsequent rate of decline, and its refractory period. At least some of the measurements of these outputs may only require a pragmatist attitude from us. Besides accepting these results, on the grounds that they cohere with the rest of our background knowledge, and we can use them to fulfil our pragmatic goals and epistemic interests in the course of the modelling of the mechanistic phenomenon, we need not believe in the truth of any claim about specific features of the posited mechanism. This step is informed by a pragmatic attitude, and respects the epistemic prudence involved in our antirealist proposal. Such a pragmatic attitude towards the relationship between at least some of the measured outputs of a putative mechanism of interest and specific features of the mechanism itself also fits much of current work directed at understanding how the generation of action potentials depend on complex lower-level channels' interactions. For example, although the somatic voltage-clamp method is currently one of the most widespread approaches to investigate synaptic physiology and neural excitability, this method suffers from "space-clamp" limitations,[9] which can significantly affect measurements of several features of dendritic synapses and ion channels activity (Bar-Yehuda and Korngreen [2008]; Williams and Mitchell [2008]).

It is also important to note that insofar as phenomena are relatively stable and general features of the world that are interesting from a scientific point of view, realists and antirealists will agree on the identification of the phenomenon. Both realist and antirealist mechanist accounts will then fit one core aspect of mechanistic explanation, viz. mechanisms produce phenomena.

After the explanandum phenomenon has been sufficiently characterized, the underlying black-box mechanism needs to be "opened-up." For the realist mechanist this is a process of discovery, called decomposition (Bechtel and Richardson [2010]). Decomposition is about taking the mechanism (physically or conceptually) apart with the aim of identifying its constitutive parts and the related activities. It involves both structural and functional decomposition of the putative mechanism producing the phenomenon of interest. Although some parts and operations can be directly observed,

phenomena' (Woodward [2011], p. 166). While data, unlike phenomena, seem to be necessarily observable, observability is not crucial to the data/phenomena distinction. In Woodward's words: 'in many cases standard discussions of the role of observation in science shed little light on how phenomena are detected or on the considerations that make data to phenomena reasoning reliable—asking whether phenomena are observable is often not the right question to ask if one wishes to understand how such reasoning works. This is because the reliability of such reasoning often has little to do with how human perception works' (Woodward [2011], p. 171).

[9] The so-called "space-clamp" problem consists in reliably inferring properties of the dendritic synapses and ion channels from electrophysiological recordings made from the soma of the neuron.

most of them can only be indirectly inferred. For many mechanisms, including the action potential, there are parts and activities behind the phenomena that can only be inferred from patterns of measurements obtained from sophisticated instruments. The identification process typically involves the usage of various experimental tools and methods, and the reliance on ampliative-abductive inference, which could justify knowledge claims about the unobservable parts and activities featuring in mechanistic explanation (Bechtel and Abrahamsen [2009]; Glennan [2005]; Machamer, Darden and Craver [2000]).

From an antirealist stance, this step is not a discovery process, but a *modelling process*, as understood in the previous section. The decomposition[10] of a mechanism can be considered as an exercise, whereby the details of the mechanism are *defined*, instead of discovered, in terms of organized parts and activities. This modelling process relies on relevant background knowledge, well-established empirical results about the target mechanism, inferential methods connecting background knowledge to such results, and on practical ends and epistemic interests. Being an exercise in finding appropriate definitions, driven by pragmatic considerations, such modelling process will allow us to remain agnostic about the claims made by the resulting model about nature itself. Hence, even if the resulting model is empirically successful, makes successful predictions, and enables us to achieve other practical ends, we can still be justified in maintaining an attitude of epistemic prudence towards at least some of its claims.

For example, with respect to the parts of the action potential, while realist mechanists will identify the membrane of the axon, the sodium and potassium channels, sodium and potassium ions, ion pumps and so on, antirealist mechanists will define those parts over a modelling process. While the realist's process of identification necessarily involves a reference to the extra-theoretical relation between the descriptions of parts and the world, the antirealist's process of definition need not involve any direct reference to anything in the world. Nonetheless, antirealists can claim that their modelling process tallies, just like the realists' process of discovery, with another core aspect of mechanistic explanation, viz.: mechanisms consist of parts.

Realist mechanists, to counter this antirealist reinterpretation, can say that the experiments undertaken by the scientists during the process of "decomposition" are actively intervening in the world and, thus, identifying via manipulation real causal relationships between variables and outcomes. These experimental interventions drive the realist's mechanism-decomposition perspective, since the interventions are doing more than simply "defining" parts and operations.

Antirealists can rebut this objection by appealing to at least three types of considerations. First, active intervention and experimental manipulation alone are not sufficient to justify a realist attitude towards parts and operations of a mechanism. Manipulation does not occur in a theoretical vacuum. Even when manipulation occurs in a context such as Hodgkin and Huxley's, where there are no firmly held beliefs about the parts and operations being manipulated in a mechanism, some kind of theory must be presupposed in claiming confidence in the correct functioning of the experimental apparatus and in characterising the results obtained. Manipulation may be 'merely a step in the process of theorizing and experimentation that [can] eventually produce some

[10] We shall stick to this and related terms despite their realistic connotations.

degree of conviction. Hence, *even within the context of scientific practice* manipulation seems insufficient to bear the burden of a commitment to realism' (Morrison [1990], p. 13).

Second, scientists' evaluation of whether an intervention has been successful or not often involves coherence considerations. Especially in cognitive neuroscience, "the degree to which the evidence produced through the intervention supports or coheres with what are regarded as plausible theories or models of the phenomenon" as well as with results produced by different instruments and methods (Bechtel [2008], p. 36-7; see also Bechtel [in press]; Hartmann [2001]). If coherence is generally not truth conducive, and coherence considerations loom large in scientists' evaluation of the evidence produced by interventions, then there is reason to maintain that even successful interventions do not generally warrant a realist attitude.

Third, the perspective of Woodward's ([2003]) account of manipulation and experimentation that informs much of the literature on mechanistic explanation, in particular Craver's ([2007]) theory, can be described 'as that of a modeler: pragmatic, piece-meal, and anti-foundational' (Woodward [2008], p. 195). This is congenial to antirealist mechanists. Woodward's ([2003]) primary focus is in fact not the metaphysics of causation, nor scientific realism. Rather, it is '*methodological*: how we think about, learn about, and reason with various causal notions and about their role in causal explanation' (Ibid.). Woodward's manipulationalism 'requires only that there be facts of the matter […] about which counterfactual claims about the outcome of hypothetical experiments are true or false and about whether a correlation between *C* and *E* reflects a causal relationship between *C* and *E* or not' (Woodward [2003], p. 121). This modest form of realism about causation, however, does not entail scientific realism. For it remains neutral about the conditions under which particular causal beliefs, formed through some experimental manipulation of a certain target mechanism, rest justified.

Realist mechanists may insist that merely defining the parts of a mechanism does not suffice for mechanistic explanation. The mechanistic account cannot do with just definitional posits towards which we cannot claim knowledge, because in this case the mechanistic account would provide us with no guide for individuating the conditions under which a mechanistic model explains. This argument is too quick, however.

According to the antirealist position on offer, parts can be understood as model entities so that the question about their ontic status is in fact irrelevant. Definitional posits are to be judged by their usefulness for various purposes, for example, in informing further experimentation, in clarifying notions that were previously ambiguous or obscure, in establishing links with other concepts, and so on. This position is informed by epistemic prudence and pragmatism so that the modelling process will be driven by knowledge of mechanisms' phenomena and by the modeller's pragmatic ends and interests. This does not imply that all mechanistic parts can be posited only if they are useful. Neither does it imply that we do not have any means besides predictive success to assess whether a mechanistic model delivers a good explanation. Both coherence and evidential considerations can in fact be brought to bear on our assessment of a mechanistic model.

The parts comprised by the mechanism of the action potential will always be defined against a background of relevant accepted knowledge with which they must cohere, as well as of relevant evidence, with which they must fit. This background knowledge will be reflected in various modelling rules, which should be taken into consideration both in the definition of mechanistic parts and in the assessment of the resulting mechanistic model. A model that explains how ‘ions move across the membrane by […] a mechanism made of Swiss cheese’ will obviously not do (Craver [2006], p. 370). But to use this type of Swiss cheese argument against the antirealist is misguided. For also the antirealist’s mechanistic model of the action potential will, as it should be, fit into the broader picture of the world. The justification of a model of the action potential should be in light of the degree of coherence of the model with established theories, other relevant models and accepted background knowledge and evidence (cf. Bechtel [in press]). This is why some of the posits in Hodgkin and Huxley’s model, such as their “activation particle,” were legitimate although they did not pick out any real feature of the target mechanism. While Hodgkin and Huxley were modelling the action potential, there were indeed a host of accepted results and theories about the build-up of the cell membrane, the behaviours of various ions, electrical currents, and so on, which informed and constrained their model. This case illustrates that, as mechanistic models are never defined in isolation, Swiss cheese-like mechanisms of the action potential will never see the light of the day, while Hodgkin and Huxley’s model of the mechanism of the action potential did.

After the parts are defined, the next step is functional decomposition, in which the activities of the mechanism are defined. There are a number of modelling frameworks for defining the causal relationships comprised by a mechanism, including graphical models (causal diagrams) and structural equations models (see e.g. Jordan and Sejnowski [2001]; Pearl [2000]). In the case of the action potential, the operations to be modelled comprise the movement of ions through the axons membrane and the movement of an electrical charge down the axon. The antirealist will understand operations as model entities or events, constrained by certain modelling rules based on coherence-considerations, and informed by pragmatic ends. Antirealists can then claim that their modelling process tallies, just like the realists’ process of discovery, the fact that mechanisms consist of activities.

In sum, with respect to both structural and functional decomposition, mechanistic models can be provided by both realist and antirealist mechanists. Mechanistic parts and activities can be defined against a background of available knowledge. Pragmatic and coherence considerations will guide both the modelling process and the assessment of the resulting model.

Once the constitutive parts and activities of the mechanism are defined, localization is the next step. Localization is about ‘mapping […] operations onto […] parts’ (Wright and Bechtel [2007], p. 63). This should be interpreted as the attempt to synthesize the structural and functional decompositions of a mechanism. Structural and functional decompositions are partial and complementary in the sense that each draws on distinct properties of the other. Yet, in order to get to a comprehensive mechanism, the activities must be localized in specific parts. During the localization process, working parts are established.

In the action potential mechanism, ion channels (parts) open and close (activity) resulting in inward and outward fluxes (activity) of ions (parts) through the cell membrane at a specific location (part). As we described at the beginning of this section, an electrical spike is thus generated (activity), and in an ongoing process, the spike travels swiftly down to the termination point of the axon (parts and activities).

Localization involves, according to current mechanisms literature, a realist commitment to believing that there exist organized components and activities that produce the behaviour of the mechanism in the process of being discovered (see e.g. Bechtel and Richardson [2010]). The antirealist, instead, sees both parts and activities as model entities defined over the course of the modelling process. No belief in the model's claims about any particular feature of nature itself is required from this standpoint. We cannot map any activity onto just any part. But the specific defined build-up of the parts constrains the activities possibly performed by those parts, and in turn, as discussed above, the build-up of the parts and activities are constrained by coherence considerations and background knowledge. The antirealist has therefore a convincing story to tell also about localization.

Decomposition and localization frequently represent the most successful steps of mechanistic discovery, or mechanistic modelling. Nevertheless, current mechanists—particularly Bill Bechtel—insist that those steps by themselves are not sufficient to warrant a successful mechanistic explanation (Bechtel [2011]; Bechtel and Abrahamsen [2009]). A full mechanistic explanation requires a fourth step regarding the organization of the mechanism, called recomposition. This process underwrites that a mechanism is not just any aggregate of working parts lumped together, but an organized structure. In fact, it is only in virtue of the specifically organized interaction of working parts that the phenomenon comes about. Properly understood, organization is a precondition for the very possibility of interaction among the parts. Causality obviously plays an important role in organization, as it is the causal interaction between working parts that furnishes the foundation of organization.

The idea behind recomposition is intuitive from a realist point of view; we can only be sure of having fully explained and understood a mechanism once we are able to reconstruct it from its real parts and operations. But there is an equally convincing antirealist account of recomposition. On such an account, spatial and temporal relationships between model parts and activities should be defined such that the required output is generated.[11] Causation, defined within an appropriate modelling framework that will provide us with rules of interaction between working parts, is crucial here. These rules of interaction underlying causation will bind the defined working parts together in such a way that they jointly and in an organized fashion produce what is relevant about the phenomenon.

In the case of the action potential, after relevant parts and activities have been defined over the course of mechanistic modelling, we should temporally and spatially relate them so that they display the specific organization of opening and closing ion channels that plays a key role in the firing of the neurons. As noted above, one

[11] Bechtel's emphasis in his work on recomposition is on how computational modelling is integral to this process. His description of the computer modelling enterprise fits an antirealist stance nicely (cf. e.g. Bechtel and Abrahamsen [2010]; Bechtel [2008], pp. 246-8 on the relationship between modelling, computer simulations, and the action potential mechanism). We are grateful to an anonymous referee to draw our attention to this point.

framework whereby we can specify such types of organizational relationships is Pearl's ([2000]) framework. Within this framework, we can define spatial and temporal organizational of a mechanism by using a causal network. This is a graph comprising nodes, which may correspond to components of the mechanism under considerations, and edges, which may correspond to organizational features of the mechanism, whose behaviour is governed by a set of structural equations. The point that bears emphasis here is that such type of causal modelling allows us to define the organizational features that are partly responsible for the phenomena produced by the mechanism (see also Halpern and Pearl [2005]). Such modelling tools are congenial to an antirealist account of the recomposition process. By drawing upon them, the antirealist mechanist can show how the phenomenon produced by some mechanism depends on its organized parts and operations, which are defined over a modelling process. Thereby, the antirealist can successfully carry out mechanistic recomposition without adopting any realist stance.

5. Some Outstanding Issues for the Antirealist Mechanist

The above discussion suggests that an antirealist approach to mechanistic explanations is tenable since all core aspects of the mechanistic account are underwritten also from an antirealist's standpoint. It should be clear that our aim in the sections above was to provide a plausible antirealist reconstruction, which could fit scientific practice, so as to show that the mechanistic account of scientific explanation is independent of scientific realism. Our aim was *not* to argue for a full-fledged account of explanation, or to argue against scientific realism. We take it that our contribution will challenge realist mechanists to tell us what exactly the payoff of realism is with respect to mechanistic explanation; it will also serve antirealist mechanists as a starting point to develop a more detailed account of mechanistic explanation.

This section provides additional considerations and details to articulate the proposal that we have been advancing in the previous sections, by beginning to address three outstanding issues. The first question concerns the relationship between modelling and explanation. It asks how is it that a mechanistic model that results from an antirealist modelling process of the sort described above successfully explains the phenomenon in question?

As pointed out in Section 2, an important difference between current mechanistic accounts is the one between epistemic and ontic views of explanation. Although the ontic view is arguably in tension with an antirealist approach to mechanistic explanation, such a view is not essential to the mechanistic account (Wright [2012]). On most accounts, to successfully explain a phenomenon is to provide some description of how a certain mechanism produces that phenomenon. In Machamer, Darden and Craver's ([2000], p. 3) words: 'To give a *description* of a mechanism for a phenomenon is to explain that phenomenon, i.e., to explain how it was produced' (emphasis added). As a description, the explanation will be a representation of the mechanism, rather than the mechanism itself. So, insofar as models provide descriptions of their targets, given an explanandum phenomenon, to provide a model of the mechanism for that phenomenon is thereby to explain that phenomenon. To provide a mechanistic model, it should now be clear, does not force upon us any realist attitude. We need not believe that an explanatory mechanistic model is a true description of nature itself.

The second question concerns our view of coherence and asks: How is the internal consistency of a proposed mechanism guaranteed? Under what conditions is coherence with background knowledge warranted?

Internal consistency and coherence with background knowledge emerge during the iterative process of defining (i.e. modelling) the mechanism. To begin with, the mechanistic parts and operations are defined, independently of each other, in such a way that the parts structurally, and the activities functionally, build up the mechanism; but they are also defined in such a way that these parts and activities comply with relevant background knowledge. Then the parts and activities are wedded into working parts, respecting the constraints that the build-up and characteristics of the parts put on the possibility to perform certain activities. In this wedding process it might turn out that the fit between parts and activities is not sufficiently good. In this case, the first step of the modelling process is reiterated.

Next, the working parts are organized into an interconnected and interacting structure, respecting the constraints put on this structure by causality rules, so that the explanandum phenomenon is produced. It might be that, given the available working parts and activities in the model, we are not able to spelling out the right organizational features of the mechanism under consideration. In this case, the previous steps of the modelling process should be repeated.

This iterative process guarantees that a consistent structure will emerge and that the resulting mechanistic model is consistent. When the development process is iterated, the model is constantly being adjusted and, possibly, detailed. This process of adjusting is basically a process of putting ever more severe constraints on the model. It is a process of putting constraints on the build-up of the parts, activities, on the way certain parts can perform certain activities, and on the way working parts interact, and hence on the causal organization of the mechanism as a whole.

Hodgkin and Huxley's model of the action potential illustrates this point well. This model is constrained by knowledge about (i) the ways in which an action potential can be initiated (e.g. through depolarizations in the axon hillock), (ii) the ways in which it can be stopped (e.g. by blocking the flow of sodium ions through ion channels by applying tetrodotoxin to the process), and (iii) the build-up of the parts (e.g. the maximum numbers of channels in the membrane or the maximum and minimum concentration of ions). The modelling process was constrained by the available data, and embedded the resulting mechanism into a larger body of accepted background knowledge. Although incomplete and non-committal on the reality of at least some of the components and activities that it included, Hodgkin-Huxley types of models are at least still useful and popular for understanding ion channels interactions.

Finally, in which sense is the claim that an antirealist account of mechanistic explanation fits scientific practice warranted?

Recall that one attractive feature of the mechanistic account is that it fits practice in the special sciences in a way that the covering law or unificationist models fail to do. Life scientists, for example, do not typically explain a phenomenon by subsuming it under some law. Rather, their explanations typically involve the specification of a mechanism. Such mechanistic explanations are developed by starting from a characterization of the phenomenon to be explained; the explanatory process proceeds by functionally and structurally identifying components and activities comprised by the mechanism producing the phenomenon; ultimately, a successful explanation provided by the scientist specifies how the organized behaviour of such components and activities produces the explanandum phenomenon.

The previous section showed that an antirealist reconstruction of mechanistic explanation seats well with such practice. Moreover, as pragmatic and coherence considerations are the driving forces of the antirealist's modelling process, our proposal is congenial to understanding many historical cases and current scientific practice, where pragmatic ends and coherence are essential to theoretical advancement and model choice. Therefore, salient features of scientific practice do not force upon the mechanistic account scientific realism. Antirealism is at least an equally viable option.

6. Two Problems for the Realist Mechanist

Thus far we have argued that mechanistic explanation does not require a commitment with scientific realism. Specifically, it does not require an "optimistic epistemic" stance towards mechanistic explanations. This section provides some prima facie grounds for why it can be preferable for the mechanist to adopt an antirealist stance, instead of a realist one. We shall discuss two problems that the realist mechanist faces, which can be *more easily* avoided by the antirealist. These are the identification problem and the type-token problem.

Let us begin with the identification problem. On the mechanistic account, both in its realist and antirealist guise, explaining a phenomenon depends on identifying the mechanism for that phenomenon. Apart from telling us that mechanisms are identified by the parts and activities that constitute them, most of the current philosophical accounts of mechanisms do not provide precise answers to the questions of where one mechanism ends and another begins and how a mechanism is to be set apart from its environment (acknowledging this fact, Craver [2009] discusses this issue while he provides an assessment of the relationship between mechanisms and natural kinds). For instance, some of the ions constituting the mechanism of the action potential are inside the membrane, others are outside. Action potentials, moreover, have long-lasting effects, for instance on blood oxygenation, which are typically not included in the mechanism. How do we tell, then, where and when the mechanism of the action potential begins and another, different mechanism begins?

For a realist, the spatial and temporal boundaries between a mechanism and its environment are real, extra-theoretical, boundaries. For a realist, there is a fact of the matter about where and when the mechanism of the action potential begins and where and when it ends. Thus, from this perspective, it is not enough to apply certain pragmatic criteria to a description of a part of the world and carve out a certain mechanism according to our explanatory interests and needs.

This approach suits our antirealism; but it would be problematic if one's mechanistic account presupposes some form of realism. The realist might address this difficulty by emphasizing the causal-organizational aspect of a mechanism. Mechanisms are causal structures, and a proper causal characterization of them can enable us to reliably carve out distinct mechanisms from the total causal structure of the world. By drawing on Simon ([1962]) and Wimsatt ([1972]), for example, the realist mechanist can identify the boundaries of mechanisms based on the strength of the causal relations within a putative mechanism, and between the putative mechanism and the outside environment. The idea is that the strength of the causal relations between the mechanism and the environment are not as abundant and tenacious as the relations among mechanistic components.

This proposal has two problems, however. First, it does not fit ‘our scientifically informed intuitions about the boundaries of many mechanisms’ (Craver [2009], p. 590). As Craver ([2009], p. 590) acknowledges, the causal relations between the mechanism of the action potential and some of the factors that are typically treated as outside its boundaries ‘are as strong as any of the causal ties among mechanism’s components.’ Second, and more importantly, identifying a mechanism’s boundaries based on the strength of its causal relations makes the identification problem shift from the identification of the mechanism o the identification of the (strength of the) causal relations. So the question now becomes how causal relations are identified.

The realist might reply that our best theories of causation can help us to answer this question. The realist might appeal to Woodward’s ([2003]) manipulationist counterfactual approach, to which—as we saw above—accounts of mechanistic explanation such as Craver’s ([2007]) appeal. Here causal relations are seen as relations between two variable-types (c and e)—where, roughly, manipulations of the value of c result in a change in the value of e, and if the value of c is not manipulated, then the value of e does not undergo any change. Still the identification problem remains. It is not clear how variables (c and e), which should refer to real parts and operations in the world, should be identified *independently* of our interests and epistemic goals (Strevens [2007], [2008] touch upon this sort of identification problem in relation to Woodward’s manipulationist theory of causation). Thus, we are back at the initial problem of how to identify a mechanism’s parts and activities independently of our interests and epistemic goals.

Hence, there is reason to conclude that realist mechanists, who try to identify mechanisms’ joints in nature, face an important challenge. They should be able to carve up the world into specific entities, viz. mechanisms, in order to identify mechanisms with real organized parts and activities. But there seems to be no purely ontic arbitrator that could serve them to cut the world at its mechanistic joints. Consequently, the foundations of the mechanistic account developed from a realist perspective seem to rest on problematic grounds.[12]

It may well be that the mechanistic structure of the world depends, at least in part, on our explanatory interests and pragmatic considerations. Also Craver acknowledges that ‘the boundaries of a mechanism depend on our choice of the explanandum phenomenon […] and on the way that we choose to describe that phenomenon’ (Craver [2009], p. 591; see also Machamer, Darden and Craver [2000], pp. 11-2 on the role of ‘idealization’ in specifying a mechanism’s ‘start and finish conditions’). If this is so, then

[12] It is possible that a realist could be a conventionalist about mechanism boundaries while still believing that terms standing for parts and activities must refer, picking out real parts and activities in the world independently of our pragmatic and epistemic interests. We do not deny this possibility. Our point is that, in the face of a more natural solution to the problem of identification put forward by the antirealist, the realist should tell us what her extra-commitments would buy us.

the identification problem will not arise for the antirealist mechanist. For her, it will be sufficient to rely on methodological and pragmatic selection criteria, which provide goods grounds for the identification of the mechanism as well as its parts and operations.

Let us now turn to the type-token problem. From a realist perspective, mechanisms should be understood as *things*, which occupy unique space-time points (cf. Glennan [2002], p. S345). A mechanism, that is, is a concrete entity, a particular or a token. If this is so, then an important feature of our world is that it often contains many tokens of a single type of mechanism; it contains, for example, many tokens of the mechanism for the action potential. Scientists are generally interested in *types* of mechanisms, and it is types of mechanisms that feature prominently in scientific explanation. Machamer, Darden and Craver note this fact by quoting Wimsatt: 'At least in biology, most scientists see their work as explaining types of phenomena by discovering mechanisms' (Wimsatt [1972], p. 67; cited by Machamer, Darden and Craver [2000], p. 34).

Now, from the realist's standpoint, if the world contains many tokens of the same type of mechanism, then there must be some sort of regularity (or fact of the matter) in the world in virtue of which those tokens are of the same type of mechanism. Different regularities would thus warrant the inference that if a particular phenomenon has already been produced many times and reliably by a particular (token) mechanism, then a similar phenomenon will be produced by a similar (token) mechanism. Furthermore, these regularities in the structure of the world would make intelligible the appeal to types of mechanisms in scientific explanation. For such an appeal would exactly pick out some natural regularity.

However, if the realist mechanist has to rely on natural regularities to account for the fact that there are many mechanisms of the same type in the world, then two troubles would arise for her. First, many particular mechanisms behave too stochastically, unreliably or are too fragile for underwriting the claim that their operations instantiate natural regularities. As noted by Bogen ([2005], p. 400), 'the mechanisms which initiate electrical activity in post-synaptic neurons by releasing neurotransmitters are a case in point: […] each one fails more often than it succeeds, and so far, no one has found differences among background conditions which account for this.' To insist that in spite of such unreliability, stochasticity and fragility displayed by many mechanisms, which scientists would surely consider of the same type, there must be underlying regularities, and that tokens of the same type of mechanism 'must operate in accordance with them […] is an article of faith," which does not have enough empirical support' (Ibid.).

Second, even if we assume that the evidence justifies the generic claim that the operations of many token mechanisms instantiate *some* natural regularities corresponding to true generalizations, the realist mechanist still faces an important challenge. She should be able to tell us more about *which* natural regularities among certain token-mechanisms enable us to justifiably induce that those particular mechanisms are of the same type. This is not a trivial challenge.

The same neuron, for instance, can produce action potentials with different shapes, varying heights and widths, whose underlying particular mechanisms comprise ionic conductances with varying biophysical properties. Furthermore, different neurons may have different voltage sensitivity and different temporal dynamics. In short, "all spikes are not alike" (Bean [2007]; Nusser [2009]). Even if there is some regularity among token mechanisms for the action potential, there also is a lot of variance. Thus, picking out the "right" regularities in order to warrant the induction from tokens to types will involve a trade-off between generality and precision considerations. If many varying details are considered so as to take account of potentially important distinctions among particular mechanisms, then the induced type might fail to reveal some potentially significant regularity. Vice versa, if too many varying details among token mechanisms are discarded, then the induced type might fail to reveal potentially significant differences. Hence, 'whether two mechanisms are mechanisms of the same kind depends upon which grain of abstraction one chooses to describe them. If there is no objectively appropriate degree of abstraction for typing mechanisms, then judgments about whether two mechanisms are mechanisms of the same kind rely ineliminably on judgments by people (in concert) about the appropriate degree of abstraction required for the problem at hand' (Craver [2009], p. 589). Therefore, as a mechanist antirealist would claim, there is probably no fact of the matter about natural regularities, independent of our background knowledge, interests and methodological decisions, that warrants the induction from token-mechanisms to types.[13]

In sum, if all or even most token mechanisms do not operate regularly, then the justification of the inductive inferences from properties of token-mechanisms to properties of a type cannot rely on some fact of the matter in nature as a realist may claim. Furthermore, in lack of such natural regularities, scientists' appeal to types of mechanisms in their explanations would be hard to understand from a realist standpoint.

7. Conclusions

The growing literature on mechanistic explanation represents one of the major success stories in recent philosophy of science. In this paper, we showed that the realist background that seems to be generally presupposed within such literature can be replaced by an antirealist background, preserving the core features of the mechanistic account. In particular, the antirealist mechanist as well as the realist can make sense of the explanatory practice in special sciences such as neuroscience. What is more, we have provided some grounds to maintain that an antirealist background might be preferable to the realist one, as it avoids more easily some of the methodological and epistemological problems the realist faces.

Our project invites several follow-ups. First, it would be interesting to examine other case studies, in light of which to evaluate the different merits of the realist and antirealist backgrounds. For example, Lindley Darden's extensive work on protein synthesis (Darden [2006]), which has generated little debate in comparison to the action potential, may be another good test-case to assess our claim that good mechanistic explanations do not require a commitment to scientific realism. Second, an antirealist mechanistic account would need further development. For the aims of this paper, our

[13] Realists may have answers also to this problem. For example, they may claim that one could be a realist about token-mechanisms, while allowing that classification of mechanisms into types involves various pragmatic considerations. The challenge, however, remains the same: Why should we make extra commitments, in the face of a plausible antirealist solution that fits scientific practice?

approach has been minimal: it has been informed just by pragmatism and some type of epistemic prudence motivated by general considerations about the history of science and scientific practice. But one may want to build on it to add more details. One may want to know more about the relationship between how-actually models, mechanistic explanation and confirmation, which seems to be well set in an antirealist context, and, more generally, about the prospects of an antirealist account of explanation.

Acknowledgements

We are sincerely grateful to Liz Irvine, Alex Reutlinger, Mark Sprevak, Jon Williamson, Carlos Zednik, and especially to one of the editors and three anonymous referees for this journal for their encouragement, constructive comments and helpful suggestions. The work on this project was supported by the Deutsche Forschungsgemeinschaft (DFG) as part of the priority program "New Frameworks of Rationality" ([SPP 1516]) [to M.C.] and [S.H.], and the Alexander von Humboldt Foundation [to S.H.].

Matteo Colombo
Tilburg Center for Logic and Philosophy of Science
Tilburg University
P.O. Box 90153
5000 LE Tilburg, The Netherlands
m.colombo@uvt.nl

Stephan Hartmann
Munich Center for Mathematical Philosophy
LMU Munich
Ludwigstr. 31
80539 Munich, Germany
s.hartmann@lmu.de

Robert van Iersel
Tilburg Center for Logic and Philosophy of Science
Tilburg University,
P.O. Box 90153
5000 LE Tilburg, The Netherlands
robertvaniersel55@gmail.com

References

Bar-Yehuda D. and Korngreen, A. [2008]: 'Space-clamp problems when voltage clamping neurons expressing voltage-gated conductances', *Journal of Neurophysiology*, **99**, pp. 1127-36.

Bean, B. P. [2007]: 'The action potential in mammalian central neurons', *Nature Review Neuroscience*, **8**, pp. 451-65.

Bechtel, W. [2007]: 'Biological Mechanisms: Organized to Maintain Autonomy', in F. C. Boogerd, F. J. Bruggeman, J. S. Hofmeyr and H. V. Westerhoff (eds), 2007, *Systems Biology: Philosophical Foundations*. New York: Elsevier, pp. 269-302.

Bechtel, W. [2008]: *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*, London: Routledge.

Bechtel, W. [2011]: 'Mechanism and Biological Explanation', *Philosophy of Science,* **78**, pp. 533-57.

Bechtel, W. [in press]: 'The epistemology of evidence in cognitive neuroscience', in R. Skipper Jr., C. Allen, R. A. Ankeny, C. F. Craver, L. Darden, G. Mikkelson and R. Richardson (eds), in press, *Philosophy and the Life Sciences: A Reader*, Cambridge*,* MA: MIT Press.

Bechtel, W. and Abrahamsen, A. [2005]: 'Explanation: A Mechanist Alternative', *Studies in History and Philosophy of Biological and Biomedical Sciences*, **36**, pp. 421-41.

Bechtel, W. and Abrahamsen, A. [2009]: 'Decomposing, Recomposing, and Situating Circadian Mechanisms: Three Tasks in Developing Mechanistic Explanations', in H. Leitgeb and A. Hieke (eds), 2009, *Reduction and Elimination in Philosophy of Mind and Philosophy of Neuroscience*, Frankfurt: Ontos Verlag, pp. 173-86.

Bechtel, W. and Abrahamsen, A. [2010]: 'Dynamic mechanistic explanation: Computational modelling of circadian rhythms as an exemplar for cognitive science', *Studies in History and Philosophy of Science Part A,* **41**, pp. 321-33.

Bechtel, W. and Richardson, R. C. [2010]: *Discovering complexity: Decomposition and localization as strategies in scientific research.* Second Edition. Cambridge, MA: MIT Press/Bradford Books.

Bogen J. [2005]: 'Regularities and Causality: Generalizations and Causal Explanations', *Studies in History and Philosophy of Biology and Biomedical Science*, **26**, pp. 397-420.

Bogen, J. and Woodward, J. [1988]: 'Saving the phenomena', *The Philosophical Review,* **97**, pp. 303-52.

Bokulich, A. [2011]: 'How Scientific Models Can Explain', *Synthese*, **180**, pp. 33-45.

BonJour, L. [1985]: *The Structure of Empirical Knowledge*, Cambridge, MA.: Harvard University Press.

Bovens, L. and Hartmann, S. [2003]: *Bayesian Epistemology*. Oxford: Oxford University Press.

Chakravartty, A. [2011]: 'Scientific Realism', in E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Summer 2011 Edition)*, <plato.stanford.edu/archives/sum2011/entries/scientific-realism>

Churchland, P. S. [1986]: *Neurophilosophy: Toward a Unified Science of the Mind-Brain*, Cambridge, MA: MIT Press.

Craver, C. F. and Bechtel W. [2006]: 'Mechanism', in S. Sarkar and J. Pfeifer (eds.), 2006, *Philosophy of Science: an Encyclopedia*, New York: Routledge, pp. 469-78.

Craver, C. F. [2006]: 'When Mechanistic Models Explain', *Synthese*, **153**, pp. 355-76.

Craver, C. F. [2007]: *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*, Oxford: Clarendon Press.

Craver, C. F. [2009]: 'Mechanisms and Natural Kinds', *Philosophical Psychology*, **22**, pp. 575-94.

Darden, L. [2006]: *Reasoning in Biological Discoveries: Essays on Mechanisms, Interfield Relations, and Anomaly Resolution*, Cambridge, MA: Cambridge University Press.

Frigg, R. and Hartmann, S. [2012]: ‘Models in Science’, in E. Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Fall 2012 Edition)*, <plato.stanford.edu/archives/fall2012/entries/models-science/>.

Glennan, S. [1996]: ‘Mechanisms and the Nature of Causation’, *Erkenntnis*, **44**, pp. 49-71.

Glennan, S. [2002]: ‘Rethinking Mechanistic Explanation’, *Philosophy of Science*, **69**, pp. 342-53.

Glennan, S. [2005]: ‘Modelling Mechanism’, *Studies in History and Philosophy of Science* C, **3**, pp. 375-88.

Halpern, J. Y. and Pearl, J. [2005]: ‘Causes and explanations: A structural-model approach -- Part I: Causes’, *British Journal of Philosophy of Science*, **56**, pp. 843-87.

Hartmann, S. [2001]: ‘Mechanisms, Coherence, and Theory Choice in the Cognitive Neurosciences’, in P. Machamer, R. Grush and P. McLaughlin (eds), 2001, *Theory and Method in the Neurosciences*, Pittsburgh: Pittsburgh University Press, 70-80.

Hitchcock, C. R. [1992]: ‘Causal Explanation and Scientific Realism’, *Erkenntnis,* **37**, pp. 151-72.

Hodgkin, A. L. and Huxley, A. F. [1952]: ‘A Quantitative Description of Membrane Current and its Application to Conduction and Excitation in Nerve’, *Journal of Physiology*, **117**, pp. 500-44.

Jordan, M. I. and Sejnowski, T. J. (eds) [2001]: *Graphical Models: Foundations of Neural Computation*, Computational Neuroscience, Cambridge, MA: MIT Press.

Kosso, P. [1992]: *Reading the Book of Nature: An Introduction to the Philosophy of Science*, Cambridge: Cambridge University Press.

Kuhn, T. [1962]: *The Structure of Scientific Revolution*, Chicago: Chicago University Press.

Levy, A. [Forthcoming]: ‘What was Hodgkin and Huxley’s Achievement?’, *British Journal for the Philosophy of Science.*

Machamer, P., Darden, L. and Craver, C. [2000]: ‘Thinking about Mechanisms’, *Philosophy of Science*, **67**, pp. 1-25.

McKay Illari, P. and Williamson, J. [2011]: ‘Mechanisms are Real and Local’, in P. McKay Illari, F. Russo and J. Williamson (eds), 2011, *Causality in the Sciences*, Oxford: Oxford University Press, pp. 818-44.

Morrison, M. [1990]: ‘Theory, intervention and realism’, *Synthese*, **82**, pp. 1-22.

Nusser, Z. [2009]: ‘Variability in the subcellular distribution of ion channels increases neuronal diversity’, *Trends in Neuroscience*, **32**, pp. 267-74.

Olsson, E. [2005]: *Against Coherence: Truth, Probability, and Justification*. Oxford: Oxford University Press.

Pearl, J. [2000]: *Causality: Models, Reasoning and Inference*, Cambridge: Cambridge University.

Psillos, S. [1999]: *Scientific Realism: How Science Tracks Truth*, New York: Routledge.

Roush, S. [2005]: *Tracking Truth*, Oxford: Oxford University Press.

Salmon, W. [1984]: *Scientific Explanation and the Causal Structure of the World*, Princeton: Princeton University Press.

Simon, H. A. [1962]: ‘The Architecture of Complexity: Hierarchic Systems’, *Proceedings of the American Philosophical Society*, **106**, pp. 467-82.

Stanford, P. K. [2006]: *Exceeding Our Grasp*, New York: Oxford University Press.

Stanford, P. K. [2009]: ‘Grasping at Realist Straws: Author’s Response, from Syposium Review of *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives* (New York, Oxford University Press, 2006), *Metascience*, **18**, pp. 355-90.

Strevens, M. [2008]: ‘Comments on Woodward, Making things happen’, *Philosophy and Phenomenological Research*, **77**, pp. 171-92.

Strevens, M. [2007]: ‘Essay review of Woodward, Making things happen’, *Philosophy and Phenomenological Research*, **74**, pp. 233-49.

Thagard, P. [1992]: *Conceptual Revolutions*. Princeton, NJ: Princeton University Press.

Thagard, P. [2007]: ‘Coherence, truth, and the development of scientific knowledge’, *Philosophy of Science*, **74**, pp. 28-47.

Weber, M. [2008]: ‘Causes without Mechanisms: Experimental Regularities, Physical Laws, and Neuroscientific Explanation’, *Philosophy of Science*, **75**, pp. 995-1007.

Weiskopf, D. A. [2011]: ‘Models and mechanisms in psychological explanation’, *Synthese*, **183**, pp. 313-38.

Williams, S. R. and Mitchell, S. J. [2008]: ‘Direct measurement of somatic voltage clamp errors in central neurons’, *Nature Neuroscience*, **11**, pp. 790-8.

Wimsatt, W. C. [1972]: ‘Complexity and Organization’, in: K. F. Schafner and R. S. Cohen (eds.), *PSA* 1972, Dordrecht: Reidel, pp. 67-86.

Woodward, J. [2003]: *Making Things Happen: A Theory of Causal Explanation*, Oxford: Oxford University Press.

Woodward, J. [2008]: ‘Response to Strevens’, *Philosophy and Phenomenological Research*, **75**, pp. 193-212.

Woodward, J. [2011]: ‘Data and phenomena: a restatement and defence’, *Synthese*, **182**, pp. 165-79.

Wright, C. D. [2012]: ‘Mechanistic explanation without the ontic conception’, *European journal for Philosophy of Science*, **2**, pp. 375-94.

Wright, C. D. and Bechtel W. [2007]: ‘Mechanisms and Psychological Explanation’, in P. Thagard (ed.), 2007, *Philosophy of Psychology and Cognitive Science*, New York: Elsevier, pp. 31-79.