Brit. J. Phil. Sci. **0** (2013), 1–24

What was Hodgkin and Huxley's Achievement?

Arnon Levy

ABSTRACT

The Hodgkin–Huxley (HH) model of the action potential is a theoretical pillar of modern neurobiology. In a number of recent publications, Carl Craver ([2006], [2007], [2008]) has argued that the model is explanatorily deficient because it does not reveal enough about underlying molecular mechanisms. I offer an alternative picture of the HH model, according to which it deliberately abstracts from molecular specifics. By doing so, the model explains whole-cell behaviour as the product of a mass of underlying low-level events. The issue goes beyond cellular neurobiology, for the strategy of abstraction exhibited in the HH case is found in a range of biological contexts. I discuss why it has been largely neglected by advocates of the mechanist approach to explanation.

doi:10.1093/bjps/axs043

1 Introduction

The Hodgkin–Huxley (HH) model of the action potential is perhaps the single most important theoretical achievement in modern neurobiology. It consists of a set of differential equations that describe neuronal 'firing'. The model, and the experimental work that led up to it, earned its authors the 1963 Nobel Prize, establishing a new framework for thinking about the electrical activity of neurons.

Recently, Carl Craver has used the HH model to illustrate a certain kind of explanatory deficiency (Craver [2006], [2007], [2008]; see also Bogen [2005]; Kaplan and Craver [2011]). Hodgkin and Huxley had limited knowledge of underlying molecular structures. In particular, they knew little about ion channels—pores in the cell's membrane that allow ions to flow in and out of the cell and play a key role in the action potential. Lacking such knowledge, Craver argues, HH could supply at most an explanatory sketch, a partial account of the phenomenon. The understanding of action potentials came of age, he holds, only upon the discovery of how ion channels work. This assessment is made against the background of the mechanistic outlook on explanation, which has been very influential in recent years (Bechtel and Richardson [1993]; Machamer *et al.* [2000]; Glennan, [2002]; Craver [2007]). Mechanists hold that a good explanation decomposes a phenomenon into underlying parts and their causal interactions. Craver thinks that the HH case buttresses the mechanistic approach because it demonstrates how models that do not fully decompose a phenomenon suffer from an explanatory deficiency.

Though initially compelling, I shall argue that this is not the right way to think about Hodgkin and Huxley's work. For the model does not simply neglect the structure and functioning of ion channels. It deliberately abstracts from these molecular specifics. Hodgkin and Huxley employed a skeletal picture of underlying molecules and used it to explain whole-cell properties. Their achievement consisted in introducing a new form of explanation into neurobiology: an account that depicts cellular phenomena as the aggregate outcome of the activities of a large number of underlying constituents.

I will also suggest that it is not a coincidence that Carver, who has done much to articulate and defend mechanism about explanation, misses the key role of abstraction in the Hodgkin and Huxley story. Mechanism has a lot going for it but, at least as developed to date, it has tended to overemphasize the description of concrete parts and their spatiotemporal organization. Explanatory strategies that do not operate in this fashion have been overlooked, despite their importance in many parts of biology.

Thus, one goal of this article is to set the record straight, so to speak, with respect to the HH model. This is intended as a contribution to the philosophy of neuroscience. More broadly, the article aims to correct a bias in the

influential mechanistic view—a bias toward concreteness and against abstractness. Like Craver, I think that much can be learned from the Hodgkin and Huxley story. But rather than highlighting the vices of omitting mechanistic detail, I shall emphasize the virtues of abstracting from it.

I proceed as follows. The next two sections provide an overview of the HH model and discuss some remarks by the authors concerning its explanatory content. I then describe Craver's view and the mechanistic outlook on which it draws. Against this background I present an alternative understanding of the HH model and compare it with Craver's. I close with a broader discussion of mechanism and the role of abstraction.

Before I delve in, let me note that Marcel Weber ([2005], [2008]) has also discussed the explanatory merits of the HH model—indeed, as far as I know he was the first to draw philosophical attention to this important episode in the history of neurobiology. Weber's argument, at least as initially set out, was part of a defense of a version of the covering law account of explanation. Over time, Weber changed his view somewhat, bringing it closer to the causal view of explanation, within which the present discussion is conducted. Nevertheless, he has retained the focus on the role of physical laws in neurobiological explanation. As far as I can tell, much of what I say here is compatible with Weber's more recent thinking about the HH model. Indeed, the aggregative aspects of the action potential that my discussion highlights may help account for the role of physical laws (Schaffner [2008]). But an adequate discussion of this topic will not be possible here. So I settle for flagging my indebtedness to Weber's work.

2 A Primer on the HH Model

An action potential is a sharp transient rise—a 'spike'—in the electrical potential on the membrane of an axon, as depicted schematically in Figure 1. From a state of rest, the neuron gets excited by an external stimulus. If the stimulus exceeds a threshold, membrane potential quickly rises and then falls back, relaxing to its resting value.[1] Understanding action potentials is fundamental to neurobiology, for it is the main form of communication within the brain, as well as between it and other parts of the body, such as muscles.

Beginning in the late 1930s and through to the early 1950s—with a break during much of WWII—Hodgkin and Huxley performed a series of experiments in the giant axon of the squid (*Loligo pealii*). They demonstrated three key facts. First, they showed that an action potential arises from the

[1] Action potentials propagate along the membrane, eventually reaching the axonal terminal. The HH model deals with this aspect as well, but I set it aside here.

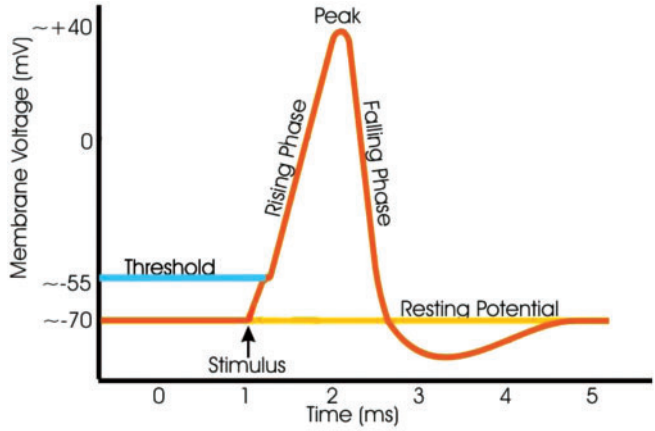

Figure 1. A schematic action potential. (Adapted from Hodgkin and Huxley [1945]).

membrane's changing its conductance to particular ions,[2] primarily sodium (Na^+) and potassium (K^+). This allows ions to move through the membrane, generating electrical currents, which change membrane potential. Second, they proved that changes in membrane conductance are themselves dependent on the membrane's potential, so that the process involves an element of feedback. Potential changes affect conductance, which affect current, which further affects potential and so on. Finally, HH demonstrated that the Na^+ and K^+ currents can be manipulated separately, suggesting that the corresponding conductances are independent. Putting these clues together (along with some basic principles of electricity), the following picture emerged.

2.1 The basic qualitative picture

When at rest, the membrane of an axon is polarized; its electrical potential (V_m) is non-zero. Each ion species, given its intracellular and extracellular concentrations and its intrinsic charge, has a so-called reversal (or equilibrium) potential: a value of V_m such that, if the membrane's potential is equal to it, no net movement of that ion will occur. The difference between an ion's reversal potential and membrane potential generates a driving force, that is, a tendency of ions to flow in or out of the cell. In a normal cell at rest, the driving forces are such that sodium will tend to move inward, whereas potassium will tend to flow outwards. However, at rest the membrane is largely impermeable to ions and so the driving forces do not have an effect on V_m.

[2] The membrane is often described in terms of not of conductance but of *permeability* to ions—as in the quotes from Hodgkin and Huxley, below. Permeability is a more general concept, but here we can treat the two as equivalent.

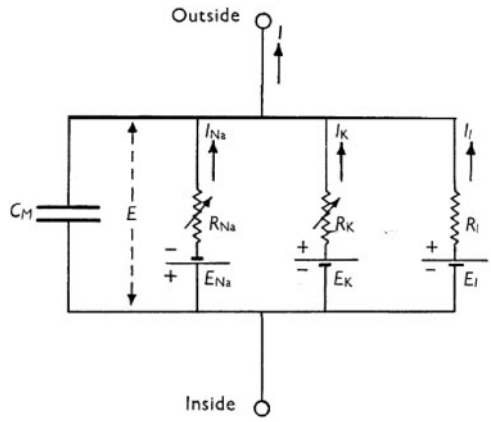

Figure 2. Equivalent circuit for a small area of an axon's membrane. (Source: Hodgkin and Huxley [1952]).

Within the membrane, there are specific channels that conduct either sodium or potassium. When V_m exceeds threshold, these channels kick into action. First, sodium channels open, raising the membrane's sodium conductance. Sodium flows inward, in accordance with its driving force, and V_m goes up. Next, potassium channels open, potassium ions flow out, and V_m drops back down. In this way a spike is generated, with the rising phase caused by sodium influx and the falling phase by potassium efflux.

2.2 The quantitative model

Hodgkin and Huxley did not settle for this qualitative causal picture. They provided a quantitative model. Its general form is represented as a circuit diagram in Figure 2, and in the corresponding Total Current Equation:

$$I_m = C_m \frac{dV}{dt} + I_K + I_{Na} + I_l \qquad (1)$$

Equation (1) is essentially a translation of the circuit diagram into mathematical symbols. I_m is the total current passing through the membrane during an action potential.[3] Each addend on the right-hand side of Equation (1) represents a separate component current. The first term is a 'capacitative' current (roughly, the membrane's ability to store potential), the second is the current due to potassium, then the current due to sodium, and finally there is a 'leakage' current, which represents a small steady flow of other ions (mainly chloride).

[3] Note that this equation tracks changes in current rather than voltage. Hodgkin and Huxley used a (then innovative) device called the voltage clamp, which measures the current needed to 'counteract' changes in V_m. Current is related to voltage via Ohm's law, $I = Vg$ (see below).

Next, Hodgkin and Huxley formulated expressions for the potassium and sodium currents (capacitative and leakage currents do not change much during the action potential). These are instances of Ohm's law, that is, a product of conductance (g) and driving force (V):

$$I_K = g_K(V_m - V_K) \tag{2}$$

$$I_{Na} = g_{Na}(V_m - V_{Na}) \tag{3}$$

Ionic conductances, in turn, are given by:

$$g_K = \overline{g_K}n^4 \tag{4}$$

$$g_{Na} = \overline{g_{Na}}m^3h \tag{5}$$

where $\overline{g_K}$ and $\overline{g_{Na}}$ are the maximal conductances for potassium and sodium, respectively. The variables h, m, and n are gating variables (a term not used by Hodgkin and Huxley). There are further differential equations, which I shall not reproduce here, that describe the dependence of the gating variables on time and voltage. Note that while $\overline{g_K}$ and $\overline{g_{Na}}$ are empirical parameters that were directly measured by Hodgkin and Huxley, gating variables are fitted expressions. I will come back to this below.

To get the full-fledged HH model, we plug in Equations (4) and (5) into Equations (2) and (3), respectively, and plug the resulting expressions into Equation (1) to obtain:

$$I_m = C_m\frac{dV}{dt} + \overline{g_K}n^4(V_m - V_K) + \overline{g_{Na}}m^3h(V_m - V_{Na}) + g_l(V_m - V_l) \tag{6}$$

Hodgkin and Huxley were able to use Equation (6) to predict the time-course, the size of the displacement in V_m, and many of the finer features of the action potential with striking accuracy. The significance of this feat is hard to overstate. It was the first major mathematical model in modern physiology, reproducing an important and complex phenomenon with unprecedented accuracy. In the words of one recent commentator: 'The Hodgkin-Huxley theory of the action potential [...] remains one of the great success stories in biology, and ranks among the most significant conceptual breakthroughs in neuroscience' (Häusser [2000], p. 1165).

3 Interlude: What Did Hodgkin and Huxley Think?

At several points in their ([1952]) article, Hodgkin and Huxley comment on the explanatory status of their model. These comments link up, albeit in different ways, with both Craver's argument and with mine. They center on the status of the gating variables, m, n, and h and some inferences made in connection with them. As noted, Hodgkin and Huxley fitted gating variables to their data concerning the time-courses of the different ionic contributions.

They state explicitly that gating expressions were chosen on the basis of mathematical convenience: '[they] describe the conductances with reasonable accuracy and are sufficiently simple for theoretical calculation of the action potential' ([1952], p. 506). However, at one point, Hodgkin and Huxley indulge in some speculation on the basis of these fitted expressions. Concerning potassium they suggest:

> These equations [i.e. Equation (4) plus the differential equations for n] may be given a physical basis if we assume that potassium ions can only cross the membrane when four similar particles occupy a certain region of the membrane. n represents the proportion of the particles in a certain position (for example at the inside of the membrane) and $1-n$ represents the proportion that are somewhere else (for example, at the outside of the membrane). ([1952], p.507)

Similarly, they propose that the sodium current equation [i.e. Equation (5)]:

> [M]ay be given a physical basis if sodium conductance is assumed to be proportional to the number of sites on the inside of the membrane which are occupied simultaneously by three activating molecules but are not blocked by an inactivating molecule. m then represents the proportion of activating molecules on the inside and $1-m$ the proportion on the outside; h is the proportion of inactivating molecules on the outside and $1-h$ the proportion on the inside. ([1952], p. 512)

In these paragraphs, Hodgkin and Huxley are suggesting an inference from the form of the current equations to the nature of underlying conductance mechanisms. Specifically, they are proposing that the order of the equation is indicative of the structure of the underlying molecules. So, for instance, they suppose that the potassium gating mechanism is composed of 'particles', and that n is the probability of each particle's moving. As n^4 would then be the joint probability of four particles moving (independently) at once, they suggest that the relevant mechanism allows potassium ions to cross the membrane only when four of these elements change location. As structural studies were conducted from the 1960s onwards, some aspects of these speculations were found to be on the mark, while others were not. For instance, it is true that the potassium channel has a tetrameric structure: four subunits change conformation as it opens. But this does not involve movement from the inside to the outside of the membrane. Rather, the channel is a water-filled conduit, which twists into an open configuration upon electrical stimulation. Closing is due to a specialized 'ball-and-chain' structure: voltage causes a spherical protein to move into the mouth of the channel, physically blocking the passage of ions (Hille, [2001]).

Hodgkin and Huxley wouldn't have been surprised to learn that some of their speculations were off track. Gating expressions were experimentally fitted functions, chosen on the basis of mathematical convenience.

Their form alone was no basis for inferring the underlying structure of ionic gating. Indeed, at several points Hodgkin and Huxley make sure to highlight this. They put the point most strongly near the very end of the article:

> The agreement [with experimental data] must not be taken as evidence that our equations are anything more than an empirical description of the time-course of the changes in permeability to sodium and potassium. An equally satisfactory description of the voltage clamp data could no doubt have been achieved with equations of very different form, which would probably have been equally successful in predicting the electrical behaviour of the membrane. It was pointed out in Part II of this paper that certain features of our equations were capable of a physical interpretation, but the success of the equations is no evidence in favour of the mechanism of permeability change that we tentatively had in mind when formulating them. ([1952], p. 541)

However, this cautionary note represents only one side of Hodgkin and Huxley's attitude. For they immediately go on to say:

> The point that we do consider to be established is that fairly simple permeability changes, in response to alterations in membrane potential, of the kind deduced from the voltage clamp results, are a sufficient explanation of the wide range of phenomena that have been fitted by solutions of the equations.

Thus, Hodgkin and Huxley note a gap in their work and caution against misinterpreting their speculative remarks about it. But they also highlight the explanatory progress that has been made. As we shall see below, Craver's portrayal of the story centers on the gap. Mine will center on how it was bridged.

4 Craver's View

Craver has taken up the HH model as a challenge to the mechanistic conception of explanation that he upholds ([2006], [2007], [2008]; see also Kaplan and Craver [2011]). Partly, he has sought to show that the HH model does not support a covering law conception of explanation (cf. Weber [2005], [2008]). At the same time, he uses it as an illustration of deficiencies in mechanistic explanation—deficiencies that are due to the lack of molecular detail. I will focus on the second issue.

The content and tone of Craver's discussion are not uniform. At some points, he seems to evince a rather dismissive attitude, suggesting that the HH model has no explanatory content whatsoever.[4] But his considered view, I think, is more moderate, namely that the model is a partial

[4] Thus, he opens the concluding section of his ([2006]) by stating that '[t]he historical example of the Hodgkin-Huxley model of the action potential [...] illustrates that models are often not explanatory' ([2006], p. 373).

explanation, lacking in certain respects. In support, he offers three arguments. The first two are premised on common distinctions between the explanatory and the non-explanatory. Craver views the HH model as falling on the non-explanatory side of both distinctions. First, the model is phenomenological, that is, reproduces the phenomenon rather than depicting its causes. Second, Hodgkin and Huxley's suggestions concerning underlying mechanisms were speculative, amounting to a how-possibly model: a claim about how the phenomenon might be caused, rather than an empirically well-grounded account.

In these arguments, Craver relies heavily on the comments discussed in the previous section. These statements show, he holds, that Hodgkin and Huxley thought of their model as merely an accurate predictive device—they saw them as '[nothing] more than an empirical description' (Hodgkin and Huxley [1952], p. 52)—and not as capturing the underlying causal processes. It is therefore phenomenological rather than explanatory. In a way, this is a reformulation of the claim that the model lacks explanatory content. But it is a reformulation that makes closer contact with what Hodgkin and Huxley themselves thought and said. Similarly, with respect to the how-possibly character of the model, Craver points to the above-cited comment that 'the success of the equations is no evidence in favour of the mechanism of permeability change that we tentatively had in mind when formulating them'. This shows that Hodgkin and Huxley thought of the suggested 'physical basis' for sodium and potassium conductance as pure speculation. They did not have empirical grounds for thinking that the particles underlying conductance existed nor that they were arranged in the ways that the gating expressions suggested. Moreover, as subsequent developments revealed, the actual physical basis is not quite in conformity with the speculations. Thus, Hodgkin and Huxley's suggestions were no more than a how-possibly model, not a well-founded how-actually account.

Craver's third argument is less historical. Because the HH model did not include information about how conductance changes occur, it is best viewed as a sketch, that is, as a description that omits indispensable explanatory information.[5] I believe that the notion of a sketch and the question of whether the HH model is one are, in the end, the most important elements in Craver's discussion. For they reflect a specific requirement concerning explanation in neuroscience (and potentially beyond). Craver argues that the HH model is explanatorily deficient because of the lack of underlying molecular detail. Such detail, he thinks, is needed in order to show how conductance changes, and its absence results in an incomplete explanation. The designation 'sketch'

[5] In a sense, this is also the argument that (Craver thinks) Hodgkin and Huxley had in mind when they deemed the model phenomenological. But I take it that there are two arguments here: one from the views of Hodgkin and Huxley, the other from the significance of the lack of mechanistic information.

is therefore derogatory from an explanatory standpoint. I will argue, in contrast, that the lack of mechanistic information serves an explanatory function: it is an abstraction that allows the HH model to account for the relationship between underlying constituents and the cell's overall behaviour.

To clarify what is at stake, I think it will be helpful to say more about sketches. To this end, we first need a brief review of the mechanistic outlook on explanation.

4.1 Mechanistic explanation

Machamer, Darden, and Craver (henceforth MDC) ([2000]) provide this well-known characterization:

> Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions. ([2000], p. 3)

Other authors on mechanistic explanation offer different formulations (e.g. Bechtel and Abrahamsen, [2005]), but there is a shared basic picture. Suppose we have a system, *S*—say, an organism, or some suborganismal structure—that exhibits a characteristic behaviour, *B*. The mechanism for *B* is the set of parts of *S* and interactions among them. Together, these gives rise to (are constitutive of) *B*. Thus, a mechanism is an underlying causal structure: a set of lower-level constituents whose joint product is an overall behaviour. Examples of a mechanism include the way in which the heart pumps blood—how blood moves through the chambers, how the heart muscle contracts to expel it and so on, or the process by which neurotransmitters are released—calcium enters the cell, triggering a chain of biochemical reactions that lead to vesicles loaded with neurotransmitter releasing their contents into the synaptic cleft (Machamer *et al.* [2000], Section 4).

Corresponding to such a characterization of mechanism-hood, there is a view of explanation that privileges parts over wholes. The thought is straightforward: if the behaviour of a mechanistic system is determined by its components and their organization, then to explain that behaviour we need to describe the components, their interrelations, and how these give rise to the behaviour in question. So, in a sense, a mechanistic explanation works by decomposing the explanandum (not literally, of course, although literal experimental decomposition may play a part in justifying mechanistic explanations). Decompositional accounts are often given by graphical means, such as drawings and directed graphs, which are common in cellular and molecular biology texts. But a verbal or mathematical description, suitably interpreted, may also capture the mechanism for some phenomenon. In sum, mechanistic explanations 'look under the hood'—they seek to describe the causal

interactions among parts in order to show how some system-level behaviour is constituted.

4.2 Sketches

Sketches are deficient mechanistic models. 'A mechanism sketch', says Craver, 'is an incomplete model of a mechanism. It characterizes some parts, activities or features of a mechanism's organization, but it leaves gaps' ([2007], p. 113). A couple of contrasts will further clarify the notion.[6]

First, a sketch differs from a schema, which is a generalized mechanistic description (Machamer *et al.* [2000], Section 5). A sketch lacks detail with respect to a specific mechanism, it 'leaves gaps' in the explanation of a particular phenomenon. A schema, in contrast, describes a kind of mechanism, generalizing over its specific instances. It trades detail for mechanistic generality. Most importantly, a schema as such does not explain anything; it must be instantiated, given the specifics of the explanandum at hand. Thus, a schema serves as a template for the description of diverse systems that share a common structure. A sketch, in contrast, gets elaborated: details that were previously unknown are discovered and gaps are thereby filled.

Second, the gaps left in a sketch involve information that it would have been (explanatorily) better to include. In this sense, a sketch points to the way forward: fill in the missing details and you get a better explanation. Scientists typically construct a sketch when they cannot do any better; they lack knowledge of some of the relevant details. So sketches are typically steps along the way to a better explanation. If all goes well, the gaps are filled and the mechanism is described in full detail. Once that occurs, the sketch is transformed into a satisfactory explanation. Contrast this with an abstraction, which is a deliberate non-inclusion of detail. An abstraction does not, as such, call for filling in. Often the opposite: scientists abstract because they believe that detail is unnecessary and irrelevant and that including it would impede understanding. So whereas a sketch is an underdeveloped explanation, an abstraction often represents explanatory progress. It should be noted here that in speaking of abstraction, I am referring not only to situations in which available detail is left out, but also to situations in which detail is unavailable, yet deemed unnecessary. That, I shall argue shortly, is what we see in the HH case.

[6] Craver ([2006], [2007]) and Machamer *et al.* ([2000]) both discuss sketches, but not at length. I draw on their explicit statements, but also on how they apply the notion.

4.3 Craver's view: The HH model as a mechanism sketch

When Craver speaks of the HH model as a sketch, he has in mind primarily the explanation of conductance changes: what accounts for the effect of voltage on the ability of ions to cross the membrane? Here Hodgkin and Huxley lacked firm knowledge: they did not know about ion channels, at least not in anything like the detail we now possess. Thus, the gating variables they fitted to their data do not carry information about channel structure and the effects of voltage on it. They reproduce the phenomenon of voltage-dependent conductance changes, but do not explain it. The only thing HH could do in this regard was to offer speculations (reviewed above in Section 3). These were how-possibly stories about conductance change mechanisms, and not empirically grounded descriptions of the molecules at work. For these reasons, Craver takes the HH model to be 'a partial explanation (an explanation sketch) for how neurons generate action potentials' ([2007], p. 56).

If the lack of information about ion channels means that the HH model was a sketch, then filling in the gaps ought to turn it into a full explanation. This is indeed how Craver sees the history. He discusses at some length work that occurred subsequent to HH, aimed at uncovering the fine structure of ion channels. He highlights current knowledge of the three-dimensional structure of channels, and molecular-level information about their response to voltage (e.g. the 'ball-and-chain' structure described above). It is these kinds of facts, Craver thinks, facts that eluded HH, that were necessary for a genuine understanding of the action potential. 'Only with the discovery of these molecular mechanisms [was] the action potential not merely modelled but explained' ([2007], p. 58).

Thus, the overall picture we get from Craver is this: HH's success in explaining the action potential was partial. While they managed to isolate ionic currents and to describe the membrane's basic circuitry, they failed to uncover the molecular structures responsible for changes in conductance. So they settled for a sketch. They described the components they knew about and speculated about those they were ignorant of. It was only when detailed descriptions of ion channels became available that the mechanism of the action potential was fully elucidated. This transformed the HH model from a sketch to a full-blooded explanation.

5 An Alternative Take on the HH Model

Craver's story fits nicely with the mechanistic outlook, and it has some basis in HH's work. But I think that, ultimately, it is not the right way to think about HH's achievement. By focussing on the absence of molecular detail, Craver misses what many neuroscientists would regard as the key feature of

the HH model: it describes whole-cell properties as arising from the aggregate activities of a large number of lower-level constitutes, primarily ions and ion channels. This type of modelling involves minimal commitments with respect to underlying constituents. It proceeds by abstracting from the concrete structural aspects of parts in order to describe the overall properties of a collective. In particular, HH viewed conductance mechanisms as discrete, independent, diffusion-permitting devices, and no more. The overall electrical behaviour of the axon can then be seen as the aggregate effect of a mass of lower-level events.

This minimal picture was not purely speculative: HH had some empirical justification for it. A fuller justification emerged in subsequent decades, but it did not require structural studies into ion channels. Most importantly, the enduring significance of HH's work consists in the conceptual framework it put in place, the key to which is abstracting from channel structure in order to capture whole-cell behaviour. I will now flesh out these ideas in several steps.

5.1 Another look at the equations

Let us start with a closer look at Equations (4) and (5):

$$I_K = \overbrace{n^4}^{\text{a}}\,\overbrace{\overline{g_K}}^{\text{b}}\,\overbrace{(V_m - V_K)}^{\text{c}} \tag{4}$$

$$I_{Na} = \overbrace{m^3 h}^{\text{a}}\,\overbrace{\overline{g_{Na}}}^{\text{b}}\,\overbrace{(V_m - V_{Na})}^{\text{c}} \tag{5}$$

These equations describe the time-courses of the main ionic contributions—those of potassium and sodium—to total membrane current during the spike. Recall that n, m, and h are gating variables. They range between 0 and 1, and are functions of voltage and time (the equations for which I have not reproduced here). Meanwhile $\overline{g_K}$ and $\overline{g_{Na}}$ are the maximal conductances for potassium and sodium, respectively. The term in brackets in each equation represents the driving force: the difference between the membrane's actual potential (V_m) and the reversal potential for each ion (V_K for potassium; V_{Na} for sodium). Thus, the three elements of each current are a variable (dimensonless) fraction, a parameter for maximal conductance and a variable force term.

To this division in Equations (4) and (5), there corresponds a causal picture as follows. Consider, first, the terms designated *a* and *b* in both equations. This is the conductance part of the equation. It is a product of a gating expression that captures the proportion of 'active' or available conductance (*a* terms) and maximal conductance (*b* terms). The product, therefore, gives the magnitude of the available conductance. That magnitude is then combined with the

driving forces (*c* terms) for each ion. This quantity, too, changes over the time course of the action potential. Together, conductance changes and driving forces define a transmembrane current—one for each type of ion and a total current resulting from their sum. This current alters membrane voltage, thereby constituting a spike.

So far, as can be seen, this is all at the macro-, whole-cell level. The magnitudes involved pertain to the axon as a whole and not to channels or ions. This macro model is a mechanistic explanation: it decomposes the total current into distinct ionic currents, accounts for them in terms of an interaction between conductance and driving force, and describes how they add up to form the total current. But if that were all there was to the HH model, then, in one sense, it would be shallow or lacking. For what occurs at the whole-cell level is obviously dependent on molecular events and understanding that dependency relationship looks to be an important element of explaining the action potential. However, we need to distinguish two explanatory questions here. First, one may ask for an explanation of lower-level events, i.e. what are the processes that change conductance? Second, one can ask how lower-level events relate to macro-level changes, i.e., how do the changing states of individual conducting molecules give rise to a spike? The model provides little by way of answering the first question. Indeed, it is implausible to regard it as an attempt to do so. But, I want to argue, it is aimed at answering the second question and it succeeds in this.

The key to success in this latter task lies in viewing whole-cell behaviour as an aggregate product of events at the molecular level. Such an account does rely on assumptions regarding the molecular goings on. But they amount to an abstract, skeletal picture, far more minimal than the detailed, structural account of channels that is nowadays available.

5.2 The discrete-gating picture

The skeletal picture I have in mind has it that the molecules involved in ionic conductance are discrete, selective, independently acting gates: each one can be either open, in which case ions of a particular type may diffuse through it, or else closed. The probability of a particular gate's opening and closing is related to voltage and can be specified independently of the behaviour of other gates. The membrane as a whole is equipped with many gates, operating in parallel. We may call this the discrete-gating picture.

The discrete-gating picture does not explain gating, at least not in any depth. But what it says suffices to account for the relationship between molecular-level conductance mechanisms and whole-cell spikes. For it implies that whole-cell conductance changes with the sum of the number of channels that are open at any given moment. At the molecular level, each channel is a

stochastic device, switching between open and closed states in a random fashion. But the membrane contains many channels, each responding to voltage independently. Thus (via a law of large numbers) the overall behaviour of the membrane is very nearly a smooth and deterministic sum of the conductances of the channels that are open at any given moment. Thus, the discrete-gating picture relates whole-cell behaviour to events at a lower level via aggregation: the system's total behaviour is the sum of the behaviours of its parts.

The discrete-gating picture implies a certain interpretation for gating expressions: they represent the proportion of channels that are open. I believe that this is how HH interpreted them too, and there are several indications for this. I get to these shortly. First, it is important to note that what I am attributing to Hodgkin and Huxley is consistent with holding, as Craver rightly does, that Hodgkin and Huxley's inferences about conductance mechanisms (reviewed above in Section 3) were largely speculative. Those inferences, recall, took the structure of gating expressions—which variables they contain, the power to which they are raised—as indicative of the detailed molecular structure of the gating molecules—how many subunits they have, what motions these undergo, and so on. The inference I am alluding to here has to do with relating gating expressions to the skeletal conception I referred to as the discrete-gating picture. That picture is (substantially) weaker than a detailed structural account, and Hodgkin and Huxley had (substantially) more evidence for it.

There were at least two pieces of such evidence. First, Hodgkin and Huxley were able to rule out an alternative to the discrete-gating picture—what they refer to as a 'carrier' mechanism. In this model, ions cross the membrane bound to a specialized molecule, rather than diffusing through it. Hodgkin and Huxley observe that such a mechanism would leave a distinct electrical signature, which they were unable to detect. On the positive side, they were able to show that while the rate of conductance change is affected by temperature, its maximal value is not, a result consistent with discrete-gating.[7]

Let me be clear that Hodgkin and Huxley did not, as far as I know, provide an explicit and organized statement of the discrete-gating picture. However, it is evident in many points in their 1952 article. Indeed, it even underlies the structural speculations discussed in Section 3: the leading idea there was that 'activation particles' can be in one of the two distinct states, and that total membrane conductance is a function of the proportion of particles that occupy each of the two states. The discrete-gating picture is apparent elsewhere too, for example, when Hodgkin and Huxley use the two-state

[7] The rate of conductance rise changes because channels open or close faster with temperature. Maximal conductance is not affected because temperature does not change the overall number of channels. This observation does not rule out all possible forms of non-discrete gating, but it narrows down the options considerably.

Boltzmann equation, in an attempt to estimate the charge carried by gating particles. Without getting into technical details, we can note that this technique presupposes that a gating particle can be in one of the two distinct and randomly fluctuating states, and that particles behave independently. (It is noteworthy that the estimate Hodgkin and Huxley reached via this technique has turned out to be accurate.)

Historically speaking, it appears that Hodgkin and Huxley held the discrete-gating picture, albeit tentatively. Moreover, they had some empirical justification for it. Having said that, I want to emphasize that my main aim is not to elucidate Hodgkin and Huxley's conception of the action potential. The goal is, first and foremost, to clarify what Hodgkin and Huxley actually contributed to the study of action potentials. I claim that a chief contribution of theirs was in the understanding of how lower-level events generate the electrical behaviour of whole axons. That this is so can also be seen by attending to work that followed Hodgkin and Huxley's seminal article, to which I now turn.

5.3 The road paved by Hodgkin and Huxley

As discussed above, one post-HH development in neuroscience has been the immense growth in structural knowledge pertaining to ion channels. This is a development that Craver draws attention to, as it has special significance from his point of view. But I believe there are at least two further developments that are pertinent to understanding the significance of the HH model. The first is the empirical confirmation of the discrete-gating hypothesis, which did not require structural studies into ion channels. The second is the development of HH-like equations that account for action potentials not covered by Hodgkin and Huxley's original model. Let us take a quick look at each in turn.

Confirmation of discrete-gating was made possible by the advent of single channel recordings, which became possible in the late 1970s with the development of 'patch clamping'. This technique was available prior to substantial progress on the structure of ion channels, and did not depend on it (though it did enable subsequent structural work). By the time patch-clamp studies were conducted, the discrete-gating picture had received clear articulation, and single channel records were readily interpreted as confirming it. Figure 3 is drawn from one of the first patch-clamp articles, in which recordings from single sodium channels were reported (Sigworth and Neher [1980]). It can be seen—and can be verified quantitatively—that single-channel currents exhibit stochastic, step-like behaviour. Furthermore, summing over a large number of single-channel currents gives a current similar to one recoded at a whole-cell level. Thus, recording from single channels, irrespective of structural studies, suffices to show that action potentials conform to Hodgkin and Huxley's basic

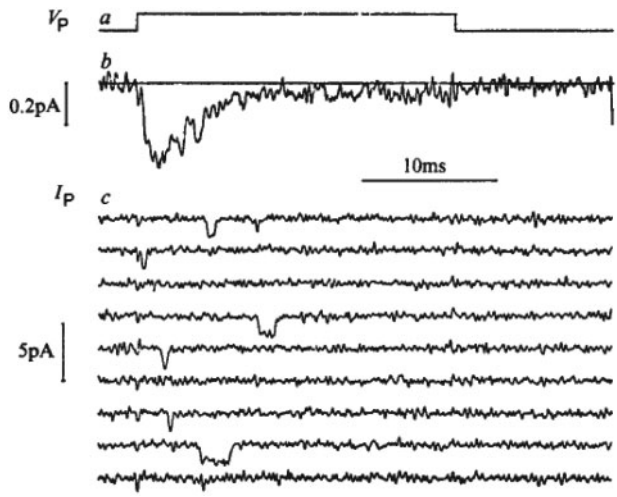

Figure 3. Currents recorded from single Na^+ channels. The bottom panel (c) shows stochastic, step-like behaviour of single channels. Panel (b) shows the result of summing several hundred channel currents (membrane patches typically have thousands of channels per square micrometer, or more). At the top (panel (a)) is the current that was applied to the membrane. (Source: Sigworth and Neher [1980]). Reprinted by permission from Macmillan Publishers Ltd: *Nature*, 287, pp. 447–9, copyright 1980.

picture: channels are discrete and stochastic voltage gates, and whole-cell behaviour changes with the sum of the behaviour of a large number of them. The authors of this article note these facts and relate them explicitly to the HH model.[8]

A second important post-HH development is theoretical. Recall that Hodgkin and Huxley worked on the squid's giant axon. Their model applies precisely only to that system. However, the basic form of Hodgkin and Huxley's equations, and the discrete-gating picture on which it is founded, has been applied to a wide range of neurons. This framework is nowadays often referred to as 'conductance-based modelling'. To give its flavour let us take a brief look at one well-known HH-like model—the Connor-Stevens (CS) model (Connor and Stevens [1971]). The CS model is based on observations in gastropod neurons. Its general form, which is all we will cover here, is the same as the HH model, albeit with an extra potassium current (the 'A current'):

$$I_m = \overline{g_K} n^4 (V_m - V_K) + \overline{g_{Na}} m^3 h (V_m - V_{Na}) + \overline{g_A} a^3 b (V_m - V_A) + g_l (V_m - V_l) \tag{7}$$

[8] Theoretical simulations carried out in the 1970s and 1980s provided further support for the discrete-gating picture (see Koch [1999], Section 8.3.1 for a review and references.)

Each addend represents a separate component current, and gating expressions are used to capture the kinetics of available conductance. In the CS model, the extra A current is carried by potassium, but it behaves somewhat like the sodium current in the HH model. One important result of this additional current is that CS neurons do not display a discontinuous 'jump' in firing rate near their threshold potential. Another consequence is that spikes are shorter.

Other conductance-based models involve different gating expressions and may include currents carried by different ions, such as calcium. But the overall framework follows the HH model. These models are regarded by most theoretical neuroscientists as the most realistic models of neuronal firing. Although they are not typically derived from a lower-level model of channel behaviour, they are readily interpreted in these terms, as we saw in the HH case. Their great advantage is that they operate at a whole-cell level, aggregating the behaviour of many micro-level constituents into a compact set of macro variables. Pioneering this type of modelling, and the picture of the relation between whole-cell and molecular-level events that underlies it, was the key contribution made by Hodgkin and Huxley.

5.4 Summary and comparison to Craver

As we saw, Craver takes the HH model to be a mechanism sketch—a gappy and therefore deficient explanation, omitting key parts of the relevant molecular mechanism. For this reason, he deems the model phenomenological and regards Hodgkin and Huxley's speculations as a how-possibly account, not grounded in molecular facts. In contrast, I have sought to show that the structure of the underlying molecules is largely beside the HH model's point: the model abstracts from these details in order to capture the relationship between whole-cell circuitry and lower-level gating.

On the picture I presented, the HH model embodies a coarse-grained conception of the molecular machinery underlying conductance—a conception of gating molecules as discrete, independent, on/off molecular conduits. The HH equations can then be read as aggregating the behaviour of many such gates and the diffusion of ions through them. Importantly, such a picture does not rely on nor suggest structural specifics concerning the sorts of molecules involved, the manner in which gating is achieved, and so on. Therefore, and contra the impression given by Craver, research into such details, while independently interesting and important, was not necessary (nor was it sufficient, for that matter) to fill a lacuna in Hodgkin and Huxley's work. What was necessary was confirmation of the discrete-gating picture. That was achieved by single channel recordings and not via structural studies. In my view, then,

the HH model was not intended to explain the mechanisms underlying conductance changes, nor should it be faulted for its failure to do so.[9]

The issue, as I see it, isn't merely the virtues of Hodgkin and Huxley's original work. It is equally important to understand how Hodgkin and Huxley influenced subsequent progress in neuroscience, as well as the current significance of their work. Craver takes it that Hodgkin and Huxley provided inspiration for, and guidance to, structural studies into ion channels. In contrast, I have argued for the equal, if not greater importance the mathematical framework Hodgkin and Huxley put in place, conductance-based modelling, and a concomitant understanding of the relationship between events at the molecular level and whole-cell neural excitability. That was Hodgkin and Huxley's achievement.

6 Conclusion: The HH Model and Mechanistic Explanation

A proper understanding of the explanatory status of the HH model is important within a philosophy of neuroscience context. But I think it has broader significance. As I noted above, Craver's view of the HH model is guided by, and presented as support for, the mechanistic approach to explanation. In this closing section, I want to situate the HH case within the wider context of mechanistic explanation. I begin with sketches.

6.1 Sketches and abstractions

Recall that a mechanism sketch is a model that leaves certain parts and/or operations underspecified, thereby falling short of a full explanation. Craver deems the HH model a sketch, because he thinks that explaining the action potential requires that one specifies how conductance changes occur, where this is seen as a matter of uncovering the structure of ion channels, and the process by which voltage affects their opening and closing. Now, as I have acknowledged, the HH model did indeed fail to explain the mechanism of conductance changes. So, clearly, there is a sense in which the HH model falls short of a full explanation of the process underlying action potentials. However, in another sense, the model is a satisfactory explanation. It answers a well-defined question concerning the whole-cell currents involved and how they arise from underlying molecular events. It is possible, therefore, to construe this as a difference over what counts as a satisfactory explanation of a given phenomenon. Or, alternatively, over what the precise explanadum of the HH model is.

[9] Hodgkin ([1992]) attests that he and Huxley set out thinking they might be able to uncover the molecular mechanisms involved, but abandoned this goal over time. This is not in conflict with my claim—I argue that the model Hodgkin and Huxley ended up producing had a different explanatory aim.

I doubt that this is a fruitful way to frame the issue. One can always ask further 'whys' and one can always reformulate an explanadum *post hoc*.

The deeper issue, I think, is not whether knowledge of gating mechanisms is explanatorily valuable. Surely it is. The question is whether there is a distinctive explanatory payoff to a model that omits such mechanistic information. I have argued that there is: abstracting from channel structure allowed HH to depict whole-cell behaviour as an aggregate of discrete, independent voltage gates at the molecular level. Here we see that, at least in some respects, an explanation that ignores underlying molecules isn't mechanistically deficient. Indeed it might be thought truer to the mechanist ideal, because it explains the relationship between lower-level mechanisms and higher-level ones. Abstracting from molecular mechanisms is a way of achieving inter-level integration.

From this perspective, what seems to be missing from the mechanist outlook is an analytical category: a notion that would cover cases in which a model is deliberately 'sketchy', that is, where gaps aren't the product of ignorance or theoretical limitations, but of an intentional strategy. The notion of a sketch does not capture this explanatory practice. In other words, the judgment that the HH model is a sketch stems, I think, from a gap in the mechanistic outlook itself, in which room has not been made for the explanatory fruits of abstracting away from structural detail.

This is not the place to develop an analysis of abstraction and its explanatory role (Strevens [2008] provides a comprehensive treatment. See also, Levy and Bechtel [forthcoming]). Let me note, however, that the type of abstraction we have seen in the HH case plays a role in many other areas of biology—perhaps most prominently in populational sciences such as ecology, epidemiology, population genetics, and some parts of evolutionary biology. In these disciplines, one typically seeks to understand the dynamics of a collective-level property—the spread of an allele, fluctuations in population sizes, and the like—as arising from the behaviour of a large number of lower-level individuals, such as genes or organisms. This is typically done by positing a skeletal description of the lower-level entities, then tracking how a collection of such entities changes over time. It is an interesting observation that this explanatory strategy—call it aggregative abstraction—exists in both cellular neuroscience and in these macro-sciences. There is surely much to learn by comparing its uses in different areas. That is a task for another day.

6.2 Why has aggregative abstraction been overlooked?

There are two interrelated reasons, I think, why aggregative abstraction has not received attention from advocates of mechanistic explanation. First, mechanistic accounts of explanation are often developed under the guidance

of a machine image of explanation. What I have in mind is the sort of understanding that one has when one decomposes a machine-like structure: a system with localized parts, in which operations often occur in sequence, and where geometrical–mechanical properties and processes play a key role. While typically not an official part of the mechanist picture, I think the machine image plays a background role. Thus, characterizations of the notion of mechanism are often couched in terms which best fit a machine-like system with spatially localized components. In particular, formulations typically refer to 'entities' (Machamer *et al.* [2000]) and 'parts' (Bechtel and Abrahamsen [2005]). Moreover, accounts of mechanistic explanation are often illustrated with examples from the world of machines or machine-like systems. For instance, Glennan ([2002], [2005]) illustrates his views with examples such as a coke machine, a clock, and blood-pumping by the heart; Craver and Bechtel ([2007]) appeal to a mouse trap. Moreover, the role of spatial contiguity and temporal sequences, and that of geometrical properties such as shape, size, and orientation, are often emphasized in discussions of mechanistic organization (e.g. Darden [2006], [2008, Section 3]; Machamer *et al.* [2000]). Overall, I think these are indications that machine-like structures are paradigms for mechanists.

The machine image is a powerful one and points to a genuine mode of understanding. It has been especially influential in twentieth-century cellular and molecular biology (including cellular and molecular studies of the brain), areas on which mechanists have focussed. Directing attention to these sorts of explanations is certainly an important innovation of the mechanist outlook. But it sometimes looks like the machine image has become over-dominant. Some types of causal systems do not fit this image—populational and other aggregative processes are an important case. It is usually a stretch to speak of such systems in terms of parts, because their constituents are dispersed in both time and space. Geometrical properties such as shape and size and mechanical interactions such as pushing, blocking, and conformation-changing are typically of lesser importance. There is a spectrum here: some systems are more machine-like than others. The action potential, which we have looked at in depth, is a mixed case. It has machine-like aspects such as the location and role of the membrane, and, of course, the molecular structure of ion channels. It also has aggregative aspects such as ionic fluxes and overall conductance changes. It is these latter aspects that Craver's account has overlooked, and one reason for this, I suggest, is the influence of the machine image.[10]

[10] I should clarify that I am not claiming that there is no room for abstraction in modelling and explaining the behaviour of machine-like systems. My claim is merely that aggregative abstraction has less of role in that context. For more on the distinction between machines and other types of mechanisms, and on different forms of abstraction.

A second reason has to do with mechanists' attitudes toward formal reasoning. Aggregative models are invariably mathematical: it is by using the tools of statistics and probability, by looking at averages, distributions, variance, and the like, that one characterizes a collective and tracks its behaviour. These concepts are inherently mathematical. Mechanistic models can, but need not, be mathematical. They often rest on qualitative causal reasoning. Moreover, the mechanistic conception has developed in no small part out of a resistance to accounts of explanation which emphasize deductive reasoning—especially Hempel's deductive-nomological account. This, I think, has generated a tendency to focus on non-quantitative causal reasoning, and a concomitant backgrounding of the significance of mathematical theorizing, including aggregative models. Like many, I think that there are well-motivated arguments against logic-centred views of explanation and the associated notion that understanding consists in subsumption under laws. But the move away from the deductive-nomological account may have over-generated, in the sense that formal tools that contribute to causal understanding have also been cast aside. One way to read the neglect of aggregative abstraction is along these lines.

A central aim of the mechanist movement in the philosophy of science has been the development of a theory of explanation that takes into account explanatory practices in biology, especially cellular and molecular biology. Much emphasis was placed in this literature on explanations that decompose a phenomenon into concrete parts and operations, and on associated forms of qualitative causal reasoning. This represents a much needed corrective to earlier conceptions of explanation. But it is important not to let the pendulum swing too far in the other direction. Some explanatory strategies are quite far removed from the paradigms on which mechanists have focused. Abstraction of the sort seen in the HH model is one important example.

The Van Leer Jerusalem Institute
POB 4070
Jerusalem, 914070, Israel
arnondor@gmail.com

References

Bechtel, W. [2006]: *Discovering Cell Mechanisms: The Creation of Modern Cell Biology*, New York: Cambridge University Press.

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

Bechtel, W. and Richardson, R. C. [1993]: *Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research*, Princeton, NJ: University Press.

Bogen, J. [2005]: 'Regularities and Causality; Generalizations and Causal Explanations', *Studies in History and Philosophy of Biological and Biomedical Sciences*, **36**, pp. 397–420.

Connor, J. A. and Stevens, C. F. [1971]: 'Prediction of Repetitive Firing Behaviour from Voltage Clamp Data on an Isolated Neuron Soma', *Journal of Physiology*, **213**, pp. 31–53.

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*, New York: Oxford University Press.

Craver, C. F. [2008]: 'Physical Law and Mechanistic Explanation in the Hodgkin and Huxley Model of the Action Potential', *Philosophy of Science*, **75**, pp. 1022–33.

Craver, C. F. and Bechtel, W. [2007]: 'Mechanism', in W. Dubitzky, O. Wolkenhauer, H. Yokota and K.-H. Cho (*eds*), *Encyclopedia of Systems Biology*, New York: Springer.

Darden, L. [2006]: *Reasoning in Biological Discovery*, Cambridge: Cambridge University Press.

Darden, L. [2008]: 'Thinking Again about Mechanisms', *Philosophy of Science*, **75**, pp. 958–69.

Glennan, S. [2002]: 'Rethinking Mechanistic Explanation', *Philosophy of Science*, **69**, pp. S342–53.

Glennan, S. [2005]: 'Modelling Mechanisms', *Studies in History and Philosophy of Biological and Biomedical Sciences*, **36**, pp. 443–46.

Häusser, M. [2000]: 'The Hodgkin-Huxley Theory of the Action Potential', *Nature Neuroscience*, **3**, p. 1165.

Hille, B. [2001]: *Ion Channels of Excitable Membranes*, Sunderland, MA: Sinauer Associates.

Hodgkin, A. L. and Huxley, A. F. [1945]: 'Resting and Action Potentials in Single Nerve Fibers', *Journal of Physiology*, **104**, pp. 176–95.

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.

Hodgkin, A. L. [1992]: *Chance and Design: Reminiscences of Science in Peace and War*, Cambridge: Cambridge University Press.

Kaplan, D. M. and Craver, C. F. [2011]: 'The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective', *Philosophy of Science*, **78**, pp. 601–27.

Koch, K. [1999]: *Biophysics of Computation*, New York: Oxford University Press.

Levy, A. and Bechtel, W. [forthcoming]: 'Abstraction and the Organization of Mechanisms', *Philosophy of Science*.

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

Schaffner, K. F. [2008]: 'Theories, Models, and Equations in Biology: The Heuristic Search for Emergent Simplifications in Neurobiology', *Philosophy of Science*, **75**, pp. 1008–21.

Sigworth, F. J. and Neher, E. [1980]: 'Single Na^+ Channels Observed in Cultured Rat Muscle Cells', *Nature*, **287**, pp. 447–9.

Strevens, M. [2008]: *Depth: An Account of Scientific Explanation*, Cambridge, MA: Harvard University Press.

Weber, M. [2005]: *Philosophy of Experimental Biology*, Cambridge: Cambridge University Press.

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