THE HODGKIN-HUXLEY MODEL, EXPLANATORY FORCE AND EXPLANATORY AIMS

ABSTRACT

The Hodgkin-Huxley model is a hugely influential mathematical model in the field of neuroscience and electrophysiology. The precise nature of the model and its explanatory status have been discussed by Carl Craver and Arnon Levy in their publications ( (Craver, 2006), (Craver, 2008), (Levy, 2013)). While Craver argues that the model is a mechanism sketch and hence is a deficient explanation, Levy argues that the model is an aggregative abstraction and hence is not explanatorily deficient. I compare their arguments and then explore the questions of what an adequate explanation is in neuroscience and what carries their explanatory force. Then I suggest that an answer could be found when we turn towards talk of explanatory aims.

Keywords: explanatory force, mechanism, abstraction, models

1. *Introduction*
2. *The Nature of the Model*
	1. *Craver’s argument*
	2. *Levy’s argument*
	3. *Comparison*
3. *Wherefrom Comes the Explanatory Force?*
4. *From Explanatory Aims?*
5. *Concluding Remarks*

**1. INTRODUCTION**

The Hodgkin-Huxley (HH) model is a mathematical model developed by Alan Hodgkin and Andrew Huxley (Hodgkin & Huxley, 1952) to describe and (possibly) explain the generation and propagation of action potentials along the axonal membrane of a neuron. Being a mathematical model, it uses an electrical equivalent network and differential equations to describe the action potential. The model is widely regarded as a landmark achievement in the field of neuroscience and electrophysiology because of its ability to predict accurately many phenomena associated with the action potential. The nature of the model has been a hot topic of debate among researchers in the field of history and philosophy of science as well as philosophy of neuroscience for over a decade now. In this paper, I concentrate on a particular conflict between two philosophers regarding the nature of the model, i.e., its explanatory status, because I believe that the conflict is philosophically important and that it throws up important questions about explanations in neuroscience as well as science in general. Carl Craver (Craver, 2006), (Craver, 2008) argues that the HH model is an attempt at providing a mechanistic explanation and that it has no explanatory force unless it is supplemented by an account of underlying mechanisms. Arnon Levy (Levy, 2013) argues that the model is an example of the strategy of ‘aggregative abstraction’ that scientists use to depict relations between the behaviour of the whole and lower-level events. He disagrees with Craver’s suggestion that the model has no explanatory force. While I side with Levy’s argument that the model is an example of aggregative abstraction, I go further and inquire into the source of the explanatory force of Levy’s aggregative abstraction and Craver’s mechanisms. I finally try to settle the discussion by supplementing the talk of explanatory force with talk of explanatory aims. Let us now proceed to the model.

Armed with insights from previous experimental findings, Hodgkin and Huxley conducted numerous experiments on the squid giant axon using a newly developed method called the voltage-clamp technique[[1]](#footnote-1) to determine its electrical behaviour. They used the results of the experiments to come up with the celebrated HH model of the action potential in neurons. The abstract of their paper that first introduced the HH model reads thus:

“This article concludes a series of papers concerned with the flow of electric current through the surface membrane of a giant nerve fibre.… Its general object is to discuss the results of the preceding papers (Part I), to put them into mathematical form (Part II) and to show that they will account for conduction and excitation in quantitative terms (Part III)” (Hodgkin & Huxley, 1952, p. 500).

The model uses an electrically equivalent circuit to capture the electrical behaviour of the axonal membrane and it is represented in quantitative terms in the form of mathematical equations. Those equations helped Hodgkin and Huxley to predict various features of the action potential with great accuracy. The electrical circuit and the respective equations are given below:



Fig.1. Source: (Hodgkin & Huxley, A quantitative description of membrane current and its application to conduction and excitation in nerve, 1952, p. 501)

The equation for total current passing through the membrane is the summation of capacitive current, potassium current, sodium current and leakage current, which is as follows:

$$I=C\_{M}\frac{dV}{dt}+I\_{K}+I\_{Na}+I\_{l}$$

The expressions for individual currents are arrived at using Ohm’s law:

$$I\_{K}=g\_{K}(V-V\_{K})$$

$$I\_{Na}=g\_{Na}(V-V\_{Na})$$

$$I\_{l}=\overbar{g}\_{l}(V-V\_{l})$$

From here, Hodgkin and Huxley used the voltage-clamp technique to measure sodium and potassium conductances at various membrane potential values. (The leakage current is constant and is independent of voltage.) After this, they used methods of numerical analysis and curve fitting to come up with the following equations (which they called formal assumptions) describing sodium and potassium conductances:

$$g\_{K}=\overbar{g}\_{K}n^{4}$$

$$\frac{dn}{dt}=α\_{n}\left(1-n\right)-β\_{n}n$$

$$g\_{Na}=\overbar{g}\_{Na}m^{3}h$$

$$\frac{dm}{dt}=α\_{m}\left(1-m\right)-β\_{m}m$$

$$\frac{dh}{dt}=α\_{h}\left(1-h\right)-β\_{h}h$$

The values of α and β with respect to m, n and h are all dependent on membrane potential V and the expressions relating them to V are determined through various curve-fitting procedures. (They will not be reproduced here.[[2]](#footnote-2)) After determining all the necessary equations, in the remainder of their paper, Hodgkin and Huxley show how their model succeeds in reproducing all the experimentally measured data with good accuracy as well as describing the electrical behaviour of the membrane: “The remainder of this paper will be devoted to calculations of the electrical behaviour of a model nerve whose properties are defined by the equations which were fitted in Part II to the voltage clamp records described in the earlier papers of this series” (Hodgkin & Huxley, 1952, p. 518).

**2. THE NATURE OF THE MODEL**

Even before Hodgkin and Huxley, physiologists were interested in the curious phenomenon of the action potential, which is the rapid rise and fall of the membrane potential of any excitable cell during excitation. In neurons, the phenomenon of the action potential aids in conducting nerve impulses to other neurons or target cells such as muscle cells. By the time Hodgkin and Huxley began conducting their experiments on the squid giant axon, it had already been established that “the action potential is associated with a large increase in membrane conductance” (Häusser, 2000, p. 1165). With the newly developed voltage-clamp technique, they were able to determine the presence of independent conductance mechanisms for sodium and potassium currents and measured those conductance values at different membrane potential values. From this, as we saw in the introduction section, Hodgkin and Huxley developed the quantitative model.

As to the nature of the model, Craver and Levy have conflicting views. We shall now reconstruct the arguments from each author and then proceed to compare them.

**2.1 Craver’s argument**

Carl Craver in his paper *Physical Law and Mechanistic Explanation in the Hodgkin Huxley Model of the Action Potential* (Craver, 2008) argues that the HH model is a mechanism sketch and is explanatorily deficient. The premises and conclusions involved in his argument are reconstructed here as follows:

P1: Adequate explanations in neuroscience work “by showing how the explanandum phenomenon is situated in the causal structure of the world” (Craver, 2008, p. 1024). That is, an explanandum phenomenon such as the action potential could only be explained adequately by showing how it is produced through the organised activities of the component parts and processes that are associated with it.

P2: The HH model in neuroscience consists merely of equations that reproduce the available experimental data. The model is a phenomenal model, mere “empirical description of the time course of the permeability changes in the membrane” (Craver, 2008, p. 1025).

P3: Being a phenomenal model, the HH model does not provide any details about the underlying mechanisms such as how and why sodium and potassium conductances change the way they do in response to changes in the membrane potential (cf. (Craver, 2008, pp. 1026-27)).

P4: Hodgkin and Huxley operated with some background causal knowledge regarding what underlying conductance mechanisms were possible and what were not: For example, they ruled out the ‘carrier mechanism’, according to which ions are carried in and out of the membrane using some carrier molecules, because the available evidence didn’t support it. They also found that the mechanisms of sodium and potassium conductances were independent through their experiments using the voltage-clamp technique. Thus, the HH model, complemented with the background causal knowledge, provides some information on the underlying mechanisms and leaves gaps to be filled by future research (cf. (Craver, 2008, pp. 1025-27)).

P5: Any model or explanation that is littered with explanatory gaps and that provides only limited information on the underlying mechanisms of the explanandum phenomenon is a mechanism sketch (cf. (Craver, 2008, p. 1027)).

C1: Therefore, the HH model is at best a mechanism sketch.

C2: And the model is not an adequate explanation since it has failed to elucidate the underlying causal structure behind the action potential.

**2.2 Levy’s argument**

In response to Craver’s argument that the HH model is a deficient explanation, Arnon Levy in his paper *What was Hodgkin and Huxley’s Achievement?* (Levy, 2013) tries to refute Craver’s view and provides his own argument that the model is an example of aggregative abstraction and that it is not deficient since it has its own explanatory advantage. His argument is reconstructed here as follows:

P1: Adequate explanations in neuroscience need not always be mechanistic. Explanations that use models that deliberately abstract away from mechanistic details (details concerning individual parts and processes) in order to capture the behaviour of the whole can have explanatory advantages that mechanistic explanations don’t (cf. (Levy, 2013, p. 20)).

P2: In the HH model, the equations describing the conductance changes of the membrane to sodium and potassium ions are aggregative and they represent the whole-cell conductance changes as the sum of the behaviour of the molecular parts. Levy cites the studies conducted on single channel recordings as confirming the preceding claim (cf. (Levy, 2013, pp. 15-17)).

P3: When aggregative abstraction is involved in a model, lower-level mechanistic details are irrelevant to the behaviour of the whole. In the HH model, mechanistic details such as how and why the lower-level ion channels change their conformation in response to changes in membrane potential, leading to conductance changes, are irrelevant to the equations describing the whole-cell sodium and potassium conductance changes. Hodgkin and Huxley arrived at these equations without such structural accounts of ion channels that are available today because their model is a mathematical abstraction (cf. (Levy, 2013, p. 14)).

P4: The main explanatory advantage of the HH model “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” (Levy, 2013, p. 13). This conceptual framework (called conductance-based modelling) inspired further models of neuronal firing in computational neuroscience (cf. (Levy, 2013, p. 18)).

C1. The HH model is an example of aggregative abstraction. It is not explanatorily deficient since it has the explanatory advantage of having introduced a new conceptual framework that inspired further research.

C2. Mechanistic explanations are not the only adequate explanations in neuroscience.

**2.3 Comparison**

Both Craver’s and Levy’s arguments are concerned with the explanatory status of the original mathematical model provided by Hodgkin and Huxley. Both use their interpretations of Hodgkin and Huxley’s comments on the explanatory status of their own model to justify their arguments.[[3]](#footnote-3) The first premise (P1) in both their arguments is concerned with what kind of explanations is good and adequate in neuroscience. Craver argues that only mechanistic explanations can be adequate, and he takes the explanandum of the HH model to be the process/activity of action potential of the whole membrane. Therefore, only a “constitutive explanation”[[4]](#footnote-4) that provides an account of the combined processes/activities of the components (including underlying mechanisms of conductance changes) that result in the action potential is an adequate explanation of the phenomenon of the action potential. As a result, he deems the HH model provided by Hodgkin and Huxley incomplete and argues that future research inspired by the model into structural studies of the components of the axonal membrane such as the studies into the structure and function of ion channels alone can make the explanation complete (cf. (Craver, 2008, pp. 1024-25)).

But here, we need to note that Craver considers the model incomplete insofar as it is formulated with some background causal knowledge such as that sodium and potassium conductances change independently of each other and are only dependent on membrane potential and that certain mechanisms such as the ‘carrier mechanism’ are ruled out of the picture. This is evident in P4 of his argument. The equations that constitute the quantitative model are mere empirical descriptions and carry no causal implications for Craver. This is evident in P2. It is the background knowledge possessed by Hodgkin and Huxley that gives the original model a causal interpretation. Since such background knowledge does not fully capture the underlying causal picture, it should be considered incomplete, argues Craver. The quantitative model has no explanatory force in neuroscience unless it is supplemented by a qualitative causal picture. This is the major conclusion that Craver argues for in his paper:

“The equations, supplemented with a diagram of the electrical circuit in a membrane and supplemented with details about how membranes and ion channels work, carry considerable explanatory weight. The equations without such interpretation … do not constitute the explanation” (Craver, 2008, p. 1028).

Another important point to note is that Craver considers the body of all experimental work done prior to the HH model as attempts to uncover the underlying causal picture. Thus, he doesn’t discard the quantitative model as something irrelevant to the qualitative causal picture. He rather treats quantitative models as evidence for or against the available qualitative pictures. A quantitative model provides an argument in support of a specific qualitative causal picture. As such, quantitative models are only arguments and are not explanations (cf. (Craver, 2008, pp. 1032-33)).

Levy, on the other hand, argues that quantitative models do have explanatory force, albeit of a kind different from what qualitative causal pictures have. This explanatory force consists in the conceptual framework that the HH model put in place in the field of neuroscience (P4). Author and professor of neuroscience Michael Häusser (Häusser, 2000, p. 1165) stresses this point many times. Referring to the discovery by Hodgkin and Huxley that the net current in the cell membrane was the result of independent sodium and potassium conductance mechanisms and that the respective conductances change only with time and membrane potential, he says, “This was a stunning conceptual breakthrough, later termed the ‘ionic hypothesis,’ a unifying framework for the field that triggered the search for the underlying molecular structures” (ibid.). As to Levy’s second claim in P4 that the HH model inspired further research, Häusser concurs:

“What have computational neuroscientists learned from Hodgkin and Huxley? … [T]he H-H model introduced the power of computers for solving quantitative problems in neuroscience. … The H-H model was so elegant and unprecedented in the quantitative and complete nature of its description that it provided an intellectual framework for biophysical and modelling work that would influence the field for decades” (ibid.).

While Craver accords the explanatory force to the (incomplete and yet to be identified) descriptions of the underlying mechanisms that together constitute the phenomenon of the action potential, Levy accords the explanatory force also to the model’s aspect of aggregative abstraction. While the model’s epistemic achievement for Craver is its support for a particular causal picture and the resulting structural studies into the components of the membrane to elucidate the causal picture, for Levy, the model’s achievement is its introduction of a new conceptual framework in the field of neuroscience and the resulting studies into more complex models that incorporate various current dynamics. Levy cites the Connor-Stevens (CS) model as an example here (cf. (Levy, 2013, pp. 17-18)).

With both the arguments being valid, one way to reconcile the conflict between Craver and Levy is to argue that they both value two different explananda. David Pence (Pence, 2017) seems to be arguing for a similar conclusion. The explanandum of the original quantitative model provided by Hodgkin and Huxley is the action potential that is “tied to a specific set of results, including things like anode breaks, voltage “overshoot,” and refractory periods … (Pence, 2017, p. 1186).” But for Craver, the explanandum is the action potential that “may be better described as a complex thing-in-the-world … the kind of object molecular processes could be said to “make up”” (ibid.). Thus, it seems that researchers who argue that the HH model has explanatory force take the former to be the explanandum, whereas those who argue that the quantitative model has no explanatory force without an accompanying causal/mechanistic picture take the latter to be the explanandum. Levy considers this point, but quickly dismisses it. He says, “I doubt that this is a fruitful way to frame the issue. One can always ask further ‘whys’ and one can always reformulate an explanandum post hoc” (Levy, 2013, p. 20). Immediately, Levy adds that while he considers knowledge of mechanistic details to be explanatorily valuable, he also considers models that omit such mechanistic details to be explanatorily valuable. The question now is whether proponents of mechanistic explanations such as Craver would acknowledge that quantitative models that omit mechanistic details are explanatorily valuable. From what we have seen in Craver’s argument, he doesn’t acknowledge that quantitative models have any explanatory force. As we saw above, they are mere arguments for him. And arguments are not explanations. Without the causal interpretation, they are not even mechanism sketches.

**3. WHEREFROM COMES THE EXPLANATORY FORCE?**

There is no doubt that the HH model is a predictively successful model. Hodgkin and Huxley explicitly mention this in Part III of their paper:

“The results presented here show that the equations derived in Part II of this paper predict with fair accuracy many of the electrical properties of the squid giant axon: the form, duration and amplitude of spike, both 'membrane' and propagated; the conduction velocity; the impedance changes during the spike; the refractory period; ionic exchanges; subthreshold responses; and oscillations. In addition, they account at least qualitatively for many of the phenomena of excitation, including anode break excitation and accommodation” (Hodgkin & Huxley, 1952, pp. 540-41).

Several authors including Craver, Levy and Pence acknowledge the predictive success of the quantitative model. David Pence argues that Hodgkin and Huxley explained the phenomenon of the action potential as they understood it, and he differentiates between two explananda. Levy argues that they explained the action potential through aggregative abstraction, and he campaigns for the inclusion of aggregative abstraction as a model of explanation in neuroscience and other fields of science. Craver, however, denies the quantitative model any explanatory force. We can now try to challenge Craver’s view as follows:

Heather Douglas in her paper *Reintroducing Prediction to Explanation* (Douglas, 2009) argues for a conception of explanation as a cognitive tool that helps scientists in generating new predictions. She says, “It is the ability of an explanation to generate new predictions, which then serve as a check on the explanation, that improves the accuracy of our scientific explanations. So the relationship between explanation and prediction is a tight, functional one: explanations provide the cognitive path to predictions, which then serve to test and refine the explanations” (Douglas, 2009, p. 454). Under such a conception of explanation, a quantitative model, such as the HH model, could be given explanatory force if it leads to new predictions that help refine the model. Douglas also notes that “Predictions are valuable because they force us (when followed through) to test our theories, because they have the potential to expand our knowledge into new realms and because they hold out the possibility (if successful) of gaining some measure of control over natural processes” (Douglas, 2009, p. 455).

Let us now ask from where mechanistic explanations derive their explanatory force. In their landmark paper *Thinking about Mechanisms* (Machamer, Darden, & Craver, 2000), where their conception of mechanisms is introduced, Machamer, Darden and Craver argue that, to explain how the world works, scientists postulate entities and activities and provide conditions where those together constitute a mechanism, and, to test if that mechanism works, they draw out predictions from that hypothesised mechanism: “When a prediction made on the basis of a hypothesi[s]ed mechanism fails, then one has an anomaly and a number of responses are possible. … Reasoning in the light of failed predictions involves, first, a diagnostic process to isolate where the mechanism schema is failing, and, then, a redesign process to change one or more entities or activities or stages to improve the hypothesi[s]ed schema” (Machamer, Darden, & Craver, 2000, p. 17).

Thus, we see here that it is in the light of new predictions that mechanistic explanations get refined, and they derive their explanatory force from their ability to generate testable predictions. We have also seen previously that the quantitative model provided by Hodgkin and Huxley generate numerous testable predictions. And Häusser notes, “Like any good theory, the H-H model inspired many new experiments. … Naturally, experiments over the last few decades also revealed phenomena incompatible with the original H-H model, such as the dependence of inactivation on activation …” (Häusser, 2000, p. 1165). Hodgkin and Huxley also note the shortcomings in their model: “The other important respect in which the model results disagreed with the experimental was that the calculated exchange of internal and external potassium ions per impulse was too large. … We have no satisfactory explanation for this discrepancy, but it is probably connected with the fact that the value of the potassium potential was less strongly affected by changes in external potassium concentration than is required by the Nernst equation” (Hodgkin & Huxley, 1952, p. 543). This proves that they left their model open for further refinement.

We see that Craver’s mechanistic explanation has explanatory force because it fits the conception of explanation put forward by Douglas. We see that the quantitative model provided by Hodgkin and Huxley also fits the above conception of explanation. Using the same rationale, we need to grant the HH model explanatory force. If Craver disagrees with this conclusion, then he must show from what source his mechanistic explanation derives its explanatory force. Until he shows that, he must give up his argument and admit that the quantitative HH model has explanatory force even without any accompanying causal picture.

**4. FROM EXPLANATORY AIMS?**

Before we seek out Craver’s possible response to the challenge posed above, let us first elucidate the causal picture that Craver has in mind when talking about the phenomenon of the action potential. The following causal picture can be found in any textbook (although the details may vary slightly in each book).

Initially the membrane potential is at rest (around -70 mV). When the membrane potential reaches
-55 mV, through excitatory signals, voltage-gated sodium channels open in the axon hillock. This results in an influx of sodium ions following its electrochemical gradient, which results in further depolarisation of the membrane. Now the membrane potential overshoots to around +30 mV. At this point, the sodium channels begin to close and voltage-gated potassium channels open. This results in an efflux of potassium ions and the membrane is repolarised. Again, the membrane potential falls below the threshold and even undershoots to around -85 mV. Now, the potassium channels begin to close, and the membrane potential slowly reaches the value of the resting potential owing to leak channels and sodium-potassium pumps. This is how an action potential is generated in an axonal membrane. It gets propagated along the membrane to the axon terminal through the same mechanism that gets repeated in successive membrane segments. That is, after the influx of sodium ions in the first segment (axon hillock), a fraction of those ions moves towards the next segment. This again makes the voltage in the next segment reach the threshold and causes another influx of sodium ions through the voltage-gated sodium channels present in that segment. Then the same mechanism of the action potential follows. This is how action potentials are propagated along axonal membranes.

Craver thinks that it is these details (details about the parts and processes forming the underlying mechanisms that constitute the action potential, i.e., the opening and closing of various ion channels, the influx and efflux of ions, the actions of sodium-potassium pumps, etc.,) that carry the explanatory force in the causal picture provided for the generation and propagation of action potentials in neurons. Why are they the carriers of explanatory force? Because they are the actual things-in-the-world, the actual flesh-and-blood parts and processes of a neuron, that make happen the thing-in-the-world called the action potential in a neuron. Craver is of the idea that knowledge about these parts, processes and their interrelations will help us manipulate the world to a greater extent than mere mathematical models can. He says, “… an explanation containing more relevant detail about the components, properties, and activities that compose a given mechanism is more likely, all things equal, to be able to answer more questions about how it will behave in a variety of circumstances …” (Craver & Kaplan, 2011, p. 613). Thus, a mechanistic explanation leads us to more prediction and hence more control of the external world. Craver goes on: “This connection between explanation and control might help to explain why scientific explanation remains prized as a distinct scientific virtue” (ibid.). This line of thinking from Craver is in accordance with Douglas’s conception of explanation as a cognitive tool to arrive at more prediction and control, which we saw earlier.

Now, again, as we saw earlier, both Craver’s mechanisms and Levy’s aggregative abstractions seem to have explanatory force as per Douglas’s conception of explanation. So why should Craver say that the mathematical model (the set of equations and the equivalent electrical network) provided by Hodgkin and Huxley does not have explanatory force without any accompanying causal picture?

 A simple answer to this question can be found in his argument itself (P2 and P3). Craver does not accord explanatory force to phenomenal models, since they are just compact descriptions of empirical data. This is reflected in (Craver, 2006), (Craver, 2008) as well as (Craver & Kaplan, 2011). In (Craver & Kaplan, 2020), Craver and Kaplan say the following:

“[T]he HH model, suitably interpreted, describes explanatory causal relationships among conductance, current, and voltage. However, the model does not provide a constitutive, mechanistic explanation for conductance changes.”[[5]](#footnote-5)

What we can infer from the above passage is that causal relationships among variables in equations can also be carriers of explanatory force for Craver. This is also reflected in Craver’s commitment to Woodward’s interventionist criterion of causality for his theory of mechanistic explanations.[[6]](#footnote-6) However, such causal relationships among variables need to be based on actual relations that exist between constituents of the mechanism and those constituents and their relations should also be contributing towards the happening of the explanandum phenomenon. This is summarised as the condition 3M (the Model-to-Mechanism-Mapping requirement) in (Craver & Kaplan, 2011)[[7]](#footnote-7) and further refined as 3M\* in (Craver & Kaplan, 2020)[[8]](#footnote-8). This is equivalent to saying that a model stops being a phenomenal model and becomes a mechanistic model for Craver if the variables, parameters and other such abstract components of the model are shown to correspond with the actual things in the world that constitute the mechanism underlying the phenomenon in question. So, one way or another, we see that Craver brings in the causal picture when talking about explanatory force.

Here, we can try to give a ‘positive answer’ to the question from where mechanistic explanations derive their explanatory force, apart from the somewhat ‘neutral answer’ that it derives from the ability to predict and control various aspects of the explanandum phenomenon. The positive answer from Craver is that mechanistic explanations depict the relevant causal picture. And this causal picture is nothing but the parts and processes that together constitute the mechanism underlying the explanandum phenomenon. If that is the case, the original HH model is a deficient explanation because the details about how the conductances change in response to membrane potential and through what mechanisms they change are not captured by the model. The model simply captures the electrical behaviour of the axonal membrane and not the structural behaviour of the axonal membrane and its components during the phenomenon of the action potential.

Have we finally shown that Craver is right by revealing the source of explanatory force for mechanistic explanations? Not exactly. Notice what we have done here. We have just come full circle. Look at the premises P1 to P5 in Craver’s argument from sub-section 2.1. Instead of providing justification for those premises, we ended up assuming their truth. We ended up begging the question.

Perhaps asking for the source of explanatory force is not the right way to go about this problem. If we were to pass the buck to Levy and ask him for a ‘positive answer’ to the question from where his aggregative abstractions derive their explanatory force, he would say, in the case of the HH model, it is by abstracting from molecular specifics and “depicting whole-cell behaviour as an aggregate of discrete, independent voltage gates at the molecular level” (Levy, 2013, p. 20) that the model achieves explanatory payoff. In (Bechtel & Levy, 2013), Levy and Bechtel argue that phenomena such as arabinose regulation system, galactose-utilisation system, etc., in bacteria can be understood not by going into specific molecular details, but by representing those systems as abstract network structures with particular organisational features (such as coherent and incoherent Feed-Forward-Loops). So, we could say that aggregative abstractions derive their explanatory force from the depiction of only those skeletal (causal and organisational) features that matter to the explanandum phenomenon as equivalent networks (as electrical, mechanical, logical systems, etc.). This could be taken as a positive answer apart from the neutral answer that aggregative abstractions such as the HH model derive their explanatory force from their ability to predict and control various aspects of the explanandum phenomenon.

However, there is a problem with the aforesaid positive answer for the question of explanatory force of aggregative abstractions. Craver in his (Craver, 2016) argues that network models are explanatory only to the extent that they depict the actual causal and constitutive relations in our world. Models that don’t depict such relations are only depicting correlations. Such models may or may not help us find causal and constitutive relations and therefore they are not explanatory in themselves. We can see that Craver is again bringing in his talk of causal picture and the conditions 3M and 3M\* in the discussion of explanatory force. While Craver’s arguments for this position are laid out in the series of papers that have been cited here so far, we can see a pattern in all those papers. Wherever proponents of aggregative abstractions and other such non-mechanistic models argue that such models have explanatory force even though they don’t capture the mechanistic details, Craver is quick to provide a counterargument that whatever explanatory force can be found in those purported non-mechanistic models is due to the mapping between the components of the model and the constituents of the mechanism (3M, 3M\*). Of course, Craver is not saying that adding more details are always better for our explanations. This is the chief point of conflict between mechanists and non-mechanists, says Craver. His refined thesis is rather that adding more relevant details are always better for our explanations and finding out which details are relevant is dependent on how the explanandum is contrastively formulated (for example, the question why the membrane potential undershoots to -85 mV after repolarisation and does not stop at the resting potential is a contrastive formulation of the form ‘why P and not P'’).[[9]](#footnote-9) But, we need to remember that the chief point of conflict between Craver and Levy is not about whether adding more details is always better, but about whether abstracting from even certain relevant details could be better from an explanatory point of view. We can see that this has been Levy’s thesis from the beginning: Look at the premises P1 to P4 in his argument in sub-section 2.2. We have come full circle again and the conflict is still not settled. Let’s put the question of explanatory force aside for now and see if there is another way to address the conflict between Craver and Levy.

We already saw that settling for different explananda for Craver’s mechanisms and Levy’s aggregative abstractions, à la (Pence, 2017), is also not a fruitful way to go about it. We are against the different explananda thesis because, as Levy argued, one could always ask further questions about various aspects of the explanandum phenomenon and further details could always be provided as explanations. While the practice of asking questions about every minute detail isn’t inherently bad, scientists don’t always keep that as their primary aim in their research. After saying, “[T]he HH model was not intended to explain the mechanisms underlying conductance changes, nor should it be faulted for its failure to do so” (Levy, 2013, p. 19), Levy adds the following in a footnote on the same page: “Hodgkin [ (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.”

Perhaps we are onto something here when we talk about explanatory aims. Even though Hodgkin and Huxley initially intended to uncover the molecular basis for the conductance changes against changes in membrane potential, they soon shifted their aim towards building a mathematical framework to capture the changes in sodium and potassium conductances values (gK and gNa) against the values of membrane potential V. So, it appears that Hodgkin and Huxley could have engaged in a different line of research altogether. They could have tried to find out the underlying mechanisms behind the conductance changes. But instead, they chose to come up with a “unifying conceptual framework”, as Häusser put it. Perhaps, as we saw in sub-section 2.3, the kind of explanatory force carried by a conceptual framework introduced by aggregative abstractions such as the HH model is qualitatively different from the kind of explanatory force carried by the relevant mechanistic details that such models omit. This suggestion is implicit in the premises P1 and P4 of Levy’s argument in sub-section 2.2.

What could account for this qualitative difference between the kind of explanatory force carried by aggregative abstractions and that carried by mechanistic explanations? One answer that we may venture to give here is ‘explanatory aims’. Combining all the results they obtained from their experiments using the voltage-clamp technique with the already available theories concerning the electrical behaviour of excitable membranes, Hodgkin and Huxley created a framework to talk about action potentials in general. They chose to do this instead of diving deep into the structural behaviour of specific axonal membranes because such conceptual frameworks are valued among the scientific community. As we noted earlier, Häusser puts it succinctly: “Finally, perhaps their most important and intangible influence was the style of their discovery. The H-H model was so elegant and unprecedented in the quantitative and complete nature of its description that it provided an intellectual framework for biophysical and modelling work that would influence the field for decades” (Häusser, 2000, p. 1165).

Perhaps talk of explanatory aims could help us make sense of the conflict between Craver and Levy. Researchers in neuroscience who value conceptual frameworks in their theories engage in the strategy of aggregative abstraction and hence they are guided by that specific explanatory aim of building frameworks to understand the phenomena in question. Researchers in neuroscience who value specific molecular details in their theories engage in structural studies and hence they are guided by the explanatory aim of providing constitutive (mechanistic) explanations to understand the phenomena in question. One should note that the above distinction of two explanatory aims is not to say that researchers in neuroscience only have these two aims or that research works guided by these two aims are mutually exclusive. I have begun this talk of explanatory aims with the hope of shedding some light on the recent discussion among philosophers of neuroscience regarding abstraction and mechanistic details. I believe that, at least, the conflict between Craver and Levy could be resolved with the talk of explanatory aims. As far as research works are concerned, the HH model is a great example because it has inspired further research in both structural studies into axonal membranes as well as computational neuroscience. Work in one line of research guided by one explanatory aim could contribute to another line of research guided by another explanatory aim. Also, we should note that Hodgkin and Huxley had both explanatory aims (the aim of building a mathematical framework as well the aim of uncovering the mechanisms underlying the conductance changes), but they prioritised one aim over the other, probably because of available time, technological resources, etc.

**5. CONCLUDING REMARKS**

The primary focus of this paper has been to discuss the arguments provided by Craver and Levy in their papers on the explanatory status of the Hodgkin-Huxley model. But I also focussed on the question of what an adequate explanation in neuroscience is. Craver argues that only those explanations that capture causal-mechanistic details are adequate because only such details carry explanatory force. Levy argues that models that abstract from certain relevant mechanistic details can also have explanatory force. In this discussion, I have tried to show that talk of explanatory force supplemented with talk of explanatory aims could resolve the conflict between Craver and Levy. Neuroscientists and philosophers of neuroscience who value the explanatory aim of providing conceptual frameworks to understand phenomena would see adequate explanatory force in the equations and the equivalent electrical network provided by Hodgkin and Huxley. However, those who value the explanatory aim of providing mechanistic details to understand phenomena would not see adequate explanatory force in the equations and the electrical network and would probably consider the HH model to be either a phenomenal model or at least a deficient explanation in neuroscience.

This is as far as we get to understanding why some researchers in neuroscience accord explanatory force to models that abstract away from certain details that seem relevant to the phenomena in question and why some researchers don’t. Now, if we want to extend our talk of explanatory aims to the whole of science, we need to take scientific cases from various fields and see whether the conflict between abstractions and mechanisms can be found in those cases and whether our talk of aims applies to various scientific communities.

**Acknowledgements**

I sincerely thank my supervisor Prof. Lalit Saraswat for giving me suggestions on which papers and books to read for my research. I also thank my friends, professors and acquaintances for all the philosophical discussions they had with me.

# Works Cited

Bechtel, W., & Levy, A. (2013, April). Abstraction and the Organization of Mechanisms. *Philosophy of Science, Vol. 80, No. 2*, 241-261.

Craver, C. F. (2006). When Mechanistic Models Explain. *Synthese*, 355-376.

Craver, C. F. (2008, December). Physical Law and Mechanistic Explanation in the Hodgkin Huxley Model of the Action Potential. *Philosophy of Science, 75*, 1022-1033.

Craver, C. F. (2016, December). The Explanatory Power of Network Models. *Philosophy of Science Volume 83, Number 5, 83*, 698-709.

Craver, C. F., & Kaplan, D. M. (2011, October). The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective. *Philosophy of Science, 78*(4), 601-627.

Craver, C. F., & Kaplan, D. M. (2020). Are More Details Better? On the Norms of Completeness for Mechanistic Explanations. *The British Journal for the Philosophy of Science*, 287-319.

Douglas, H. E. (2009, October). Reintroducing Prediction to Explanation. *Philosophy of Science, 76*, 444-463.

Häusser, M. (2000, November). The Hodgkin-Huxley Theory of the Action Potential. *Nature Neuroscience, 3*, 1165.

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

Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*(117), 500-544.

Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., & Hudspeth, A. J. (Eds.). (2013). *Principles of Neural Science* (5th ed.). The McGraw-Hill Companies, Inc.

Levy, A. (2013). What was Hodgkin and Huxley's Achievement? *The British Journal for the Philosophy of Science*, 1-24.

Machamer, P., Darden, L., & Craver, C. F. (2000, March). Thinking about Mechanisms. *Philosophy of Science, 67*, 1-25.

Nelson, M. E. (2005). Electrophysiological Models. In S. H. Koslow, & S. Subramaniam, *Databasing the Brain: From Data to Knowledge* (pp. 285-301). New York: Wiley.

Pence, D. E. (2017). Potential Controversies: Causation and the Hodgkin and Huxley Equations. *Philosophy of Science*, 1177-1188.

Woodward, J. (2013). Mechanistic Explanation: Its Scope and Limits. *The Proceedings of the Aristotelian Society Supplementary Volume LXXXVII*, 39-65.

1. Refer to (Kandel, Schwartz, Jessell, Siegelbaum, & Hudspeth, 2013, pp. 150-151) to know more about the voltage-clamp technique. [↑](#footnote-ref-1)
2. For more information on these, see (Nelson, 2005) [↑](#footnote-ref-2)
3. See (Craver, 2006) as well. [↑](#footnote-ref-3)
4. Craver uses this term in his (Craver, 2008, p. 1024). Even though Craver has his own reasons for this terminology, for our purposes, we can take both the terms “constitutive” and “mechanistic” to mean the same. [↑](#footnote-ref-4)
5. See footnote 7 of (Craver & Kaplan, 2020). Craver here admits that he made an injudicious remark that the HH model does not explain the action potential in (Craver, 2006) and takes it back. However, he still holds the view that the model (with the accompanying causal picture) is a mechanism sketch and hence is only a partial explanation. [↑](#footnote-ref-5)
6. For more information, see (Woodward, 2013). [↑](#footnote-ref-6)
7. 3M: In successful explanatory models in cognitive and systems neuroscience (*a*) the variables in the model correspond to components, activities, properties, and organizational features of the target mechanism that produces, maintains, or underlies the phenomenon, and (*b*) the (perhaps mathematical) dependencies posited among these variables in the model correspond to the (perhaps quantifiable) causal relations among the components of the target mechanism. [↑](#footnote-ref-7)
8. 3M\*: A constitutive mechanistic model has explanatory force for phenomenon P versus P' if and only if (*a*) at least some of its variables refer to internal details relevant to P versus P', and (*b*) the dependencies posited among the variables refer causal dependencies among those variables (and between them and the inputs and outputs definitive of the phenomenon) relevant to P versus P'. [↑](#footnote-ref-8)
9. 3M\* makes more sense in this light. For more details, see (Craver & Kaplan, 2020). [↑](#footnote-ref-9)