An Artifactual Perspective on Idealization:

Constant Capacitance and the Hodgkin and Huxley Model[1]

Natalia Carrillo and Tarja Knuuttila[2]

1. Introduction

There are two traditions of thinking about idealization offering almost opposite views on their functioning and epistemic status. While one tradition views idealizations as epistemic deficiencies, the other one highlights the epistemic benefits of idealization. The deficiency account of idealization focuses on how modelers idealize for the purpose of tackling complicated real-life phenomena and achieving tractable representations. The hope is that the advancement of science and the availability of better modeling methods could eventually deliver more accurate representations of worldly target systems. In contrast, the epistemic benefit account of idealization emphasizes the fact that in scientific modeling a detailed depiction is not often sought for. Instead, idealization facilitates more efficient explanations, and better understanding of phenomena that would not be possible without it, and so the justification of idealization does not lie in its future eliminability (see Batterman 2009, 16). Indeed, the crucial difference between the two accounts boils down to whether de-idealization is desirable or not (irrespective of whether it would be possible) (see Knuuttila

[2] Natalia Carrillo and Tarja Knuuttila contributed equally to this work.

and Morgan, 2019). While the deficiency accounts aim for de-idealization, the epistemic benefit accounts offer reasons for why scientists might be justified in not de-idealizing their models.

Although the deficiency and epistemic benefit accounts appear thus diametrically opposed with regard to the status of idealization, there is still something that many, if not most of these accounts agree upon: idealizations introduce distortion into models with respect to our knowledge of worldly target systems. In other words, idealizations deliberately misrepresent. Distortions of these kinds are not difficult to find: the classic examples concern limiting concepts, e.g. when assuming that a thermodynamic system has an infinite number of particles, or treating populations of discrete individuals as continuous. In these kinds of cases, the model world undoubtedly involves features that are known not to hold in worldly target systems. But in many other cases the model is such an elaborate construct that it is difficult to tell how exactly it is supposed to differ from the worldly systems of interest – even in cases in which we had a lot of knowledge of them already. Consider, for example, assuming in an economic model that people form their beliefs of a value by drawing a value from a probability distribution (Alexandrova 2006), or modeling biochemical networks in an analogy to electric circuits (Knuuttila and Loettgers 2014). We suggest that in these and many other cases, idealization is better understood from an artifactual perspective that does not take the representational model-world relationship as a point of departure, presupposing the possibility of some straightforward comparisons between models and their supposed target systems. From the artifactual perspective, idealization can be treated holistically, as a set of interrelated assumptions emerging in, and entailed by, the model-building process. In focusing on model construction, the artifactual approach pays attention to the characteristics of actual tools of representation, and how they shape the target system.

In this article, we approach idealization from the artifactual perspective (Knuuttila 2005, 2011, 2017), comparing it to the distortion-to-reality accounts of idealization, and exemplifying it through the case of the Hodgkin and Huxley model of nerve impulse. This early modeling achievement within neurophysiology has engendered a lively discussion within mechanistic philosophy of science. In this discussion, Craver (2006, 2007) and Levy (2013) have offered opposite interpretations of the epistemic value of the Hodgkin and Huxley model. While they do not explicitly discuss the model in terms of the notion of idealization, they nevertheless address its schematic and simplified nature. In evaluating the

epistemic character of the Hodgkin and Huxley model, Craver gives it a deficiency reading, while Levy highlights the epistemic benefits its "aggregative" abstractions offer.

From the artifactual perspective, the epistemic benefits and deficiencies introduced by idealization frequently come in a package due to the way idealization draws together different resources in model construction. Accordingly, idealization tends to be holistic in that it is not often easily attributable to just some specific parts of the model (even though it might seem so at first glance). Instead, the idealizing process tightly embeds theoretical concepts and formal tools into the construction of a model. Frequently, analogies are employed to recruit epistemic resources from other areas of research. In this process, idealization enables coordination between theoretical concepts, formal tools, measuring apparatus, diagrams, and experimental preparations (Carrillo 2018, 2019). The Hodgkin and Huxley model provides a good example of such intersection of analogical reasoning and idealization.

2. Galilean and minimalist idealization

The contrast between the deficiency and epistemic benefit accounts of idealization occupies a center stage in Weisberg's seminal discussion of Galilean and minimalist idealization.[3] While Galilean idealizations make representations deficient, minimalist idealization brings epistemic benefits. In Galilean idealization, according to Weisberg, "[o]ne starts with some idea of what a nonidealized theory would look like. Then one mentally and mathematically creates a simplified model of the target." (Weisberg 2007, 640). In other words, through mental and mathematical effort scientists seek to translate the existing knowledge into a computationally tractable and in principle corrigible simplified model. The notion of Galilean idealization, in Weisberg's construal, draws attention to both the complexities of real-world situations, and the difficulties of representation that require simplification – those simplifications being subsequently alleviated by the advancement in mathematical techniques and computational power.

Minimalist idealization, in contrast, focuses on causal factors instead of the challenges of representation. It seeks to single out "only the core causal factors" that "*make a difference* to

---

[3] Frigg and Hartmann (2012) have introduced a somewhat similar distinction between Aristotelian and Galilean idealization.

the occurrence and essential character of the phenomenon in question" (Weisberg 2007, 642). As the quote shows, Weisberg characterizes minimalist idealization in line with the difference-making account of explanation put forth by Strevens (2008), although his construal of minimalist idealization covers also Cartwright's, Mäki's and Batterman's accounts. But even though the aforementioned authors have argued for minimalist idealization, their accounts differ substantially. What Weisberg glosses over is that the accounts put forth by Cartwright, Mäki and Batterman cannot easily, if at all, be characterized as variants of the difference-making account.[4]

Mäki (1992) and Cartwright (1999) rely on the notion of isolation: in idealizing, scientists isolate some causal factors in an analogy to an experimental setup. Mäki builds his account on the idea of how various unrealistic model assumptions are used to theoretically "seal off" a set of relations from the influence of others. Cartwright (1999) invokes what she calls a "Galilean experiment" that studies the effect of one cause operating on its own by eliminating *all* other possible causes. Mäki's and Cartwright's accounts seem more modest in addressing the contributions of separable causal factors instead of aiming to pick out *the* causal difference-makers. In other words, idealization may only aim at studying causal capacities of a system. Given that Cartwright refers to "Galilean experiments" in delineating her isolationist account, and McMullin's classic essay on "Galilean idealization" (1985) gives a much broader account of Galilean idealization than what Weisberg does, the term Galilean idealization seems to be a partially ambiguous label for deficiency accounts of idealization.

Furthermore, the accounts of Strevens, Mäki and Cartwright differ substantially from that of Batterman, although all four of them have addressed the benefits of minimal modeling that make de-idealization undesirable. What the three former accounts have in common is that they cast idealization in terms of singling out a few causally effective factors: idealization makes a positive epistemic contribution in identifying the contributions of these causal factors/ difference-makers. Batterman, in turn, has explicitly argued against isolationist and other representational accounts that rely on "veridical representation of difference-making features within the model", or more generally "common features" between the model and a

---

[4] Potochnik (2017) considers idealization rampant and unchecked in science. In her account idealizations help limited cognitive agents to set aside complicating factors to identify causal patterns. Consequently, Potochnik's notion of idealization contains features of both deficiency and epistemic benefit accounts.

real-world system (e.g. Batterman and Rice 2014, 355). Batterman and Rice argue that idealization can make salient how diverse real-world systems, despite the differences in their micro-causal make-up, can "exhibit the same patterns of behavior at much higher scale" (Batterman and Rice 2014, 350). Idealization thus functions as a device for coarse graining, enabling the recognition of multiply realizable patterns across various phenomena. Such idealizations may show which features are irrelevant, yet Batterman's account does not boil down to a difference-making account. The point is that idealizations themselves perform positive explanatory work, instead of assigning the explanatory task only to causal difference makers (through separating their contributions from those of non-difference makers, see Rice 2018). Although Batterman thus emphasizes the epistemic benefits of idealization, his version of minimalist idealization does not share the representationalist commitments of the other minimalist accounts.

Indeed, despite their opposite approaches to the epistemic status of idealization, the traditional deficiency and epistemic benefit accounts of idealization implicitly apply the criterion of representational accuracy in their analysis of idealization. At the bottom there is the idea of idealization as a distortion that already suggests a model-world comparison, i.e. representing the worldly systems differently from how they actually are, and ascribing to them properties they do not have (see Godfrey-Smith 2009, 47). The criterion of representational accuracy leads in the case of deficiency accounts to the quest for de-idealization. The benefit accounts, in turn, both in their difference-making and isolationist guises, suppose that some parts of the model accurately describe some causal factors and their functioning.

Yet, from the perspective of scientific practice, the criterion of representational accuracy does not seem to adequately capture the way scientists employ idealization in their modeling practices. The deficiency account presumes that successive rounds of modeling could bring the model more accurate and realistic. Yet this presumption may be unwarranted, already because of the way the model was constructed. Such gradual de-idealization is often not achievable, or feasible, firstly, due to the affordances and limitations of the representational tools employed. The specific mathematical and computational techniques chosen, for example, shape the modeled phenomenon in particular ways (just think of how heterogeneous phenomena are rendered similar through the application of network methods). Secondly, and relatedly, in scientific practice mathematical and computational techniques are often coupled

with particular theoretical concepts and perspectives and so intersected in the actual construction of a model. Consequently, any representation-independent correspondence to a target system seems an unattainable goal at the outset, and not just for practical reasons (see Knuuttila and Morgan 2019). Some discussions of idealization affirm this difficulty of de-idealization in a roundabout way. For instance, Sklar distinguishes between controllable and uncontrollable idealizations on the basis of how tractable they are and whether scientists know how to compensate for them (e.g. Sklar 2000). In other words, in the case of uncontrollable idealizations it seems misguided to talk about distortion of reality when scientists do not even know how they might be corrected.

Minimalist idealization, in its difference-making, and isolationist variants appear to make less taxing demands than the deficiency view, but a closer examination shows that this is not the case. The underlying commitments of these accounts concerning both the decomposability of models and the causal structure of the world make them more heavily dependent than the deficiency account on representational and conceptual transparency. Though the deficiency account presumes that models could be de-idealized such that they would better approximate the real situations, they at least acknowledge the challenges of representation.

The way to move forward, we suggest, is to detach the discussion of idealization from the ideas of distortion and misrepresentation. Idealization may distort but it also does something else: it keeps the model together. We will claim that two dimensions of the artifactual account, namely, its focus on the actual representational tools on one hand, and on the constrained construction of the model on the other hand, can recover many basic insights of the deficiency and epistemic benefit accounts. However, the artifactual account does not primarily offer a new account of idealization, but rather an alternative metalevel perspective to modeling. There is no need to throw many of the insights of the traditional discussions away – once they are freed from their traditional representational and realist assumptions.

3. The artifactual perspective on idealization

The artifactual approach sees models for what they are; human made, altered or engendered objects[5], whose affordances are being utilized for epistemic purposes, in the context of specific scientific practices. Although most models can be considered as human made objects, there are other kinds of entities, such as model organisms and laboratory populations that might better be characterized as human altered and/or human engendered. Knuuttila (2005, 2011, 2017) develops the artifactual account of models as an alternative to the representational approach since the latter does not duly recognize the epistemic value of modeling:

> The characteristic unit of analysis of the representational approach, the relationship of a single model and its supposed real target system, is too limiting in that it pays no attention to the models themselves as unfolding, constructed entities, or to the model-based theoretical practice that typically proceeds on the basis of many related, and also complementary, models (Knuuttila 2011, 263).

Knuuttila suggests that models, understood as particular kinds of epistemic artifacts, can be approached from two intertwined perspectives. First, scientific models are objects, whose construction is constrained in view of some epistemic goals.[6] The traditional answer to the question of how models are able to give us knowledge appeals to the notion of representation. In contrast, perceiving a model as an artifact pays heed to its constrained arrangement that renders a certain scientific problem more accessible and manageable, helping scientists to address it in a systematic manner. That idealization enables an epistemic access to complicated phenomena is recognized in the literature (though with representationalist overtones). But idealization is also inherently related to the second dimension of the artifactual approach: the focus on the actual representational tools made use of in modeling. The constrained constitution of a model is not only due to its purposeful construction, but also to the representational tools employed in the theoretical, mathematical and computational framing of the research question. These tools enter their own specific constraints into models. It is important to note that such constraints are both enabling and limiting. For example, mathematical and computational methods allow particular inferences

---

[5] Artifacts do not have to be human made objects. A rock used to open a clam may be thought of as an artifact even though it is not constructed by humans for that purpose. Artifactuality in general can rather be defined through the roles an object plays in some human (or animal) activity.

[6] Models are typically multipurpose tools: the aims of modeling may and do change, and multiple aims may co-exist as models are being reused, reconstructed and repurposed by different groups and stakeholders.

and solutions, but not others. Different representational tools are suitable for expressing different things. From the artifactual perspective idealization emerges as an ineliminable part of model-building, sitting at the intersection of various constraints – theoretical, conceptual and representational – both affording and narrowing down.

The artifactual approach lends a unifying view into idealization in that it is able to recover several basic philosophical insights motivating both the deficiency and epistemic benefit accounts, being simultaneously detached from the idea of distortion by misrepresentation. It can accommodate both the importance of tractability stressed by the deficiency account, and the epistemic benefits delivered by minimalist idealization. These insights can be elicited from the twofold character of models as unfolding objects constructed by employing already established *representational tools* in view of some *epistemic aims*.

Although the focus of the artifactual account on representational tools highlights the importance of tractability, it also proves richer in this regard. As the artifactual account is not based on misrepresentation but addresses model construction instead, idealization can be seen to fulfill many other tasks than that of rendering the model tractable. In fact, idealization makes the model possible in the first place. It enables modelers to apply different mathematical and computational techniques as well as measuring and intervening apparatus to investigate the scientific question at hand. While idealization has traditionally been understood as deliberate misrepresentation of a feature of the target system, the artifactual approach is not hung up on the accuracy / distortion of a model or its parts. In contrast, the artifactual approach views idealization as a set of assumptions that aligns different representational tools in the pursuit of constructing a model capable of answering pertinent research questions. Thus, idealization enables the application of mathematical and computational tools such that the model holds together, allows manipulations and generates useful results.

Whereas the focus on the use of representational tools addresses the tractability concerns of the deficiency account, the focus on the constrained construction of a model highlights the benefits of idealization. And it does so without the (too) heavy realist and representationalist commitments of the traditional benefit accounts. The artifactual account perceives theoretical models as highly constrained objects that are frequently built to study some pending theoretical or empirical questions. Modelers study and manipulate models in order to better understand some interdependencies lying behind actual and possible phenomena. Although

such understanding-bearing models typically are minimal, they also are often too artificial, preliminary and hypothetical to really pick any difference-making factors. And if they seem to do so, it seems rather a result of a successful modeling endeavor than a feature of the modeling heuristic itself. More can be said in favor of the isolationist accounts that are causally more cautious in not expecting the isolated causal factors to be actual difference makers. But, as already mentioned, such isolation accounts make strong (de)decomposability assumptions concerning both the modularity of worldly target systems and models themselves.[7]

While thus the artifactual approach to idealization accommodates many important insights of deficiency and epistemic benefit accounts, it does not presume that idealization boils down to deliberately misrepresenting "*what the world is actually like*" by introducing distortion to scientific representations through "known falsehoods" (cf. Levy 2018). One central problem of the idea of distortion is that it supposes too much to be known. It is as if in modeling scientists were representing systems of which they already had well-articulated and certified knowledge. Second, viewing idealization as distortion disregards the modal dimension of science. Namely, scientific modeling is an explorative endeavor that aims to probe how certain phenomena *could* be produced, instead of just studying actual phenomena (Gelfert 2016).[8] The focus is on genuine possibilities and not just some counterfactual scenarios within the range of known behavior of some actual systems. Minimal modeling, from the artifactual perspective, frequently involves narrowing down to the minimal elements and interactions that could be sufficient to produce a behavior or a pattern of interest. Yet, the possibilities that can be explored are constrained by the tools available (i.e. what mathematical methods can be used, what analogies are available, what can be measured). These constraints point towards the third important problem of distortion accounts: their neglect of the challenges of representation. The distortion view implicitly assumes that the representational tools used in science would furnish scientists malleable and transparent enough means for choosing how to misrepresent the already known world. This last criticism concerns especially the minimalist variant of the distortion accounts of idealization.

---

[7] The epistemic benefit account by Batterman and Rice (2014) seems to be in line with the artifactual account, since they focus on the epistemic work done by limiting operations.

[8] Rice has recently argued that applying mathematical methods involves holistic distortion that enables researchers to "to extract various kinds of modal information" (2018, 2802).

Finally, and perhaps paradoxically, attending to the actual model construction highlights the conceptual dimension of modeling. The strange consequence of the notion of distortion by misrepresentation is the virtual disappearance of theoretical and conceptual activity from our notion of modeling. It becomes a matter of at least partial accurate representation of the world (as it is actually like), instead of providing theoretical perspectives for understanding of actual or possible phenomena. The artifactual perspective witnesses how theoretical, conceptual, interventional and representational resources become intertwined in the modeling process. In the following sections we examine the Hodgkin and Huxley model from the artifactual vantage point, highlighting the role various kinds of idealizing assumptions have played in its construction. We pay particular attention to the assumption of constant membrane capacity, and how it enabled the analogical transfer of the toolbox of electrical engineering to nerve signal research.

4.   The Hodgkin and Huxley model

The Hodgkin and Huxley model consists of a set of differential equations reproducing the behavior of nerve cells as registered in experiments on squid giant axons. These experiments were performed with voltage-clamp techniques invented in the 1940s after the discovery that squids have a giant axon that is suitable for nerve signal experimental research. The Hodgkin and Huxley model (HH model) is based on an analogy between an electrical circuit and the nervous membrane, equating ionic currents across the membrane with electrons flowing across resistances in an electric circuit. Drawing an analogy between the nerve cell and electrical circuits, Hodgkin and Huxley were able to derive equations that establish relations between current, voltage and ionic permeability at one point in the membrane.

Before the consideration of the nerve cell membrane as an electrical circuit, the electrophysiological program in which Hodgkin and Huxley were trained had already benefitted from other analogies. The analogy between galvanic cells and nerve cells, in particular, played a vital role. Galvanic cells were used in physical chemistry at the end of the 19th century to study the relationship between chemical and electrical gradients in ionic solutions. They consist of two compartments separated by a semipermeable membrane (see Figure 1 below). Semi-permeable membranes allow some species of ions to cross but not others. The experiments in these devices involve dissolving different amounts of salt in each

compartment and measuring differences in electrical potential between the compartments with a voltmeter. The ions in the salt diffuse, and the ion species that the membrane is permeable to, can move across to the other compartment. If more salt is placed in one compartment, then there will be a diffusion gradient from one compartment to the other. Since ions are electrically charged, movement of ions from one compartment to the other breaks the electrical balance, generating an electrical potential amongst the two compartments. Eventually, the concentration of ions of the species that can travel across the semipermeable membrane is stabilized. This is called electrochemical equilibrium.

As this physicochemical balance of forces came to be understood better, the idea that a similar mechanism could lie behind nervous transmission was proposed (Ostwald 1890). This approach defined the research agenda for many physiologists in the next decades. An important resource for this research was the Nernst equation. In his famous work "The electromotive action of ions" (1889), Nernst developed a mathematical expression for the electromotive force of ions in a galvanic cell in terms of their concentrations and voltage across membranes. With this equation it is possible to calculate the electric potential difference between two solutions of a uni-univalent electrolyte[9] at different concentrations separated by a semi-permeable membrane. Based on Nernst's work, the excitatory process of nerves began to be associated to the "nerve membrane" on physicochemical grounds. A particularly influential development of this sort was Bernstein's membrane theory (1902).

Bernstein had detected an "action current" in nerves, associated to muscle contraction. To explain the rise and fall of the current, Bernstein assumed that a difference in concentration exists between the inside and outside of the membrane and suggested that the observed current was due to a "collapse" of the membrane. Thus, in the excited state the cell becomes permeable to all ions such that the ions are momentarily free to cross according to the diffusion and electrical gradients. That was the main theory at the time when Hodgkin and Huxley started their research.

---

[9] A solution in which each ion has a valence of 1, and produces two ions when dissociated.
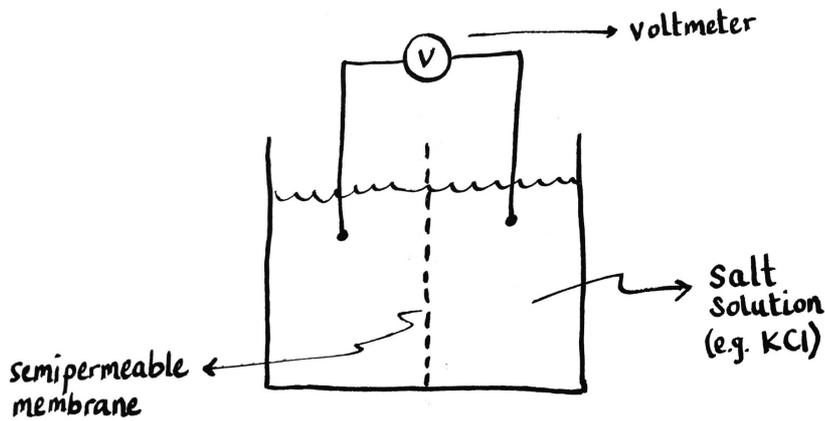
Figure 1. Diagram of a galvanic cell. Source: Natalia Carrillo.

Giant axons of squid had been discovered in 1936 by J. Z. Young, and were immediately recruited by neurophysiologists for their wonderfully advantageous proportions (they can be up to 1mm in diameter). This experimental material was susceptible to measurements that were impossible in other tissues with the technology of the time. The prospect of intervening electrically in the giant axon of squid motivated researchers to investigate the electrical features of the nerve cell membrane in order to understand the interactions between the electrical devices and this newly discovered material. This allowed the development of apparatus such as the voltage clamp that injected and recorded electrical currents in the nerve cell.[10] With this equipment, new empirical discoveries were made. After performing experiments on squid giant axons, Hodgkin and Huxley realized that Bernstein's account was not entirely correct, since the membrane did not "collapse" but briefly became first permeable to potassium and later permeable to sodium (Hodgkin and Huxley 1952a, Huxley 1999).

As these measuring artifacts were developed and the nerve cell was rendered in electrical terms, various theoretical and representational tools traditionally associated to electrical engineering and electromagnetism became available for theorizing about nerve impulse generation and transmission. Electric circuit diagrams were used to depict the dynamics of

---

[10] The voltage clamp fixes the transmembrane voltage of the membrane to a value set by the experimenter, and records the current that had to be injected in order to maintain voltage steady at that value. That recording is the inverse of the actions that took place in the membrane, and was interpreted by Hodgkin and Huxley and others as the inverse of the transmembrane currents.

charge distribution around and across the membrane. Cole and Baker were the first to model the membrane in terms of what they described as an "approximate equivalent" electrical circuit (Cole and Baker 1941). Hodgkin and Huxley later on exploited this idea successfully, obtaining a system of equations that can simulate the electrical recordings in giant axons.

Hodgkin and Huxley contributed to the aforementioned line of research that was trying to bring together the previous physicochemical model with resources from the field of electrical engineering. Whereas Cole and Baker focused on a circuit with a resistor and an inductor, Hodgkin and Huxley analyzed the dynamics of ionic currents around the membrane in terms of currents in a resistor-capacitor circuit with a constant capacitance and ohmic variable resistances. The HH equivalent circuit interprets the insulating features of the membrane as the behavior of a capacitor,[11] the difference in ionic concentration across the membrane as an electric potential source,[12] and the mechanism of permeability as variable resistances. In the equivalent circuit the voltages of the different batteries were set to the value of the electrochemical equilibrium of the corresponding ionic species (sodium and potassium), calculated with Nernst's equation. The resistances[13] were the most difficult to characterize, and were thought of as variable resistances with first order dynamics that were fine-tuned to empirical data.

Hodgkin and Huxley consider the capacitor of their circuit as a one with constant capacity. A capacitor would change its capacity to store charge if the distance between its plates changes, or if the area of the plates changes. In the devices used in electrical circuits this is seldom the case, that is, the capacitance of capacitors is usually fixed. Although the membrane is materially quite different from a capacitor like the one used in circuits, Hodgkin and Huxley assumed that the capacitance of the capacitor in the "equivalent" circuit is

---

[11] A capacitor is usually constructed of two plates of conducting material separated at a distance small enough so that charges on one side will feel electric repulsion or attraction from charges on the other plate. If connected to a battery, one plate becomes negatively charged with respect to the other.

[12] This is the equivalent of a battery. When a cable connects one pole of the battery to the other, the charges tend to move from the pole with excess negative charge to the one with less negative charge.

[13] Hodgkin and Huxley used the term "conductance," which is the multiplicative inverse of the resistance, but for simplicity we describe them as resistances since both terms retain the same underlying concept.

constant (Hodgkin and Huxley 1952b, 505; Hodgkin, Huxley and Katz 1952, 426). This was an important assumption for Hodgkin and Huxley for a number of interrelated reasons that include but also go beyond the formulation of the mathematical model.

By conceiving the nerve cell as an electric circuit, it was possible to obtain a mathematical expression for the ionic currents across the membrane by applying Ohm's Law and Kirchhoff's laws to describe the movement of electric currents in the equivalent circuit. In view of the discussion of idealization, it is important to note that Hodgkin and Huxley's mathematical model retains the already imported set of theoretical and representational resources from physical chemistry, through setting the electromotive force (i.e. the voltage of the batteries in the circuit) to the value given by the Nernst's equation. Thus, the analogy to an equivalent circuit is not independent of the analogy to galvanic cells, they rather become integrated in the Hodgkin and Huxley model. In other words, the recruitment of epistemic resources from electrical engineering to nerve signal research was related to the previous idealized physico-chemical rendering of the workings of the nerve cell as a rigid barrier with changes in permeability (Carrillo 2018).

One might argue that the assumption of constant capacitance is an idealization in the sense of distortion because there are no perfect capacitors in nature. Or one could wonder whether it amounts to omission instead, since some minor variations in the membrane's width are discarded when conceiving the membrane's capacity as constant. But the artifactual perspective underlines how the idealization of constant capacitance allows for coordination of an electrical interpretation of the nerve cell with the previous physicochemical elements that were already playing a theoretical role. This role is neither entirely detrimental nor entirely beneficial, highlighting that such idealization is less understandable through model-world comparison than through its integrative role in drawing together particular representational tools and their already established uses.

Let us now comment on the role that this idealized approach to the nerve signal played in the experimental part of Hodgkin and Huxley's research. One important empirical result Hodgkin and Huxley obtained concerned the separate contributions of sodium and potassium to the overall current (Hodgkin and Huxley 1952a, fig. 5). To perform this experiment, they changed the composition of the extracellular solution such that sodium ions would be in electrochemical equilibrium. This amounts to the experimental equivalent of taking away the sodium battery from the equivalent circuit. They then stimulated the axon, generating a signal

that would, in their interpretation, only reflect potassium currents, since sodium ions were not subject to any electromotive force. The interpretation of this experiment as the isolation of potassium current relies crucially on the assumption that once the capacitor is charged, the currents in the membrane are all transmembrane currents. This reading of the experiment by Hodgkin and Huxley makes important assumptions regarding the nature of the membrane. From the viewpoint of the equivalent circuit, the idea is that through the process of nervous transmission the capacitor does not change shape or otherwise change its capacity to store charge. If the membrane would change shape, then the neighboring charges would be displaced, generating capacitive currents that could contribute to the recordings. This would imply that the recorded currents could be due to potassium ions crossing the membrane *or* because of the displacement of charges neighboring the membrane. Assuming that there are no variations in the membrane's capacitance, Hodgkin and Huxley were able to conclude that the current they registered was due to potassium ions crossing the membrane. They then subtracted this curve to the total current in normal conditions, and obtained the sodium current.[14] For this reason, the assumption of constant capacitance plays a role in the interpretation of experimental results.

The assumption of constant capacitance was also mathematically convenient. In order to derive the equations, it is necessary to describe the currents in each of the electronic devices of the equivalent circuit (Kirchoff's Law). The equation that describes the charge in the capacitor is $Q = CV$, where C is capacitance, V is voltage and Q is charge. In order to get the current in the capacitor (I_c) one derives the equation, obtaining

$$I_c = \frac{dQ}{dt} = \frac{d}{dt}(CV) = C\frac{dV}{dt} + V\frac{dC}{dt}$$

The right hand term corresponds to the capacitive current. If the capacity is a constant function the term $V\frac{dC}{dt}$ disappears (since the derivative of a constant function is zero). If capacitance is not constant, this term would of course remain, and the equations describing the system would be different than those obtained by Hodgkin and Huxley (1952b). In this manner, the assumption of constant capacitance is not only present in the interpretation of the

---

[14] Hodgkin and Huxley also considered the contribution of a "leak" current, which accounted for minor errors in prediction and measurement. We are not discussing this for simplicity since it does not contribute substantially to the excitable behavior of the axon.

experiments but also in the derivation of the equations.

Ultimately, the idea that there are no significant contributions from capacitive currents to the dynamics of transmembrane voltage during nervous transmission became deeply ingrained into the electrophysiological tradition (Takashima 1979, 133). But some researchers have stressed that this assumption does not have the empirical support it would require to be considered as unproblematic (Iwasa and Tasaki 1980, Heimburg and Jackson 2005). As we will see in more detail in the following sections, the issue goes beyond the question of whether constant capacitance is an eventually de-idealizable falsity or a beneficial lie, since it is not clear what it would mean for the membrane to *truly* be a constant capacitor or not to be one. The point is that the assumption of constant membrane capacity is foundational for the research program in question in a way that eludes any idealization-free method of finding out whether the membrane is a constant capacitor. The evaluation of the epistemic value or status of constant capacitance cannot rely on model-world comparisons in any straightforward way. We expect this finding to apply to many other idealizations as well, at least on closer inspection, and taking into account the role they play in model construction.

5. Discussion of idealizations in the Hodgkin and Huxley model

The lively philosophical discussion of the Hodgkin and Huxley model furnishes a fruitful vantage point for delineating the differences between the artifactual and idealization-as-distortion accounts of idealization. Philosophers participating in this discussion have presented widely different interpretations of the epistemic contributions of the Hodgkin and Huxley model to neuroscience. The first to address the Hodgkin and Huxley model from a philosophical perspective was Weber, who claimed that the HH-model explains in the same way as many physical explanations do: it entwines the experimental regularities and general physical laws (that are invariant under some interventions) (Weber 2005, 2008). In contrast to Weber, Craver (2006, 2007) argued that the HH model should not be understood as being derived from the laws of physics, but rather as a how-possibly sketch of a mechanism that sustains nerve impulses.[15] For Craver, the HH model is only of a how-possibly character,

---

[15] Craver's criticism of Weber is a part of his more comprehensive mechanistic account of explanation, whose main target of criticism is the covering law account of explanation.

because it does not give an account of the "nuts and bolts" of the mechanism by which ions cross the nervous membrane. Based on this fact – recognized by Hodgkin and Huxley themselves – Craver claimed that the explanation of the nerve impulse was not truly given until the proteins that form ionic channels across the membrane were discovered, thereby completing the explanatory sketch.[16] According to Craver, it was only at this stage that a *complete mechanistic explanation of the nerve impulse* was delivered.

Ultimately, Craver claims that in order for mechanistic models to explain, they would need to "account for all aspects of the phenomenon by describing how the component entities and activities are organized such that the phenomenon occurs" (Craver 2006, 374). Levy (2013) picked up this requirement of completeness arguing, contra Craver, that the explanatory achievement of the HH model is in fact due to its *abstract* character. Because the HH model abstracts from the individual movement of ions, it is able to more generally account for the ionic currents—without having to open the "black box" of the mechanism of ion transport. For Levy, the contribution of the model is due to its characterization of regularities at an *aggregative level:* "the discrete-gating picture relates whole-cell behavior to events at a lower level via aggregation: the system's total behavior is the sum of the behaviors of its parts." (Levy 2013, 15). He goes on to explain that such "aggregative abstraction" could be "truer to the mechanistic ideal, because it explains the relationship between lower-level mechanisms and higher-level ones" (20).[17]

Although this philosophical discussion of the epistemic contribution of the Hodgkin and Huxley model is not framed in terms of idealization, the way it focuses on the

---

[16] Discrete ion fluxes were detected by Neher and Sackmann in the 1970s, supporting the idea of a passive mechanism of ionic transport. Later, in the late 1990s, evidence of the existence of the potassium ion channel was obtained with x-ray crystallography. These results were considered as sound evidence for the hypothesis that it is voltage-sensitive protein ion channels that change the permeability of the membrane during a nerve impulse.

[17] We consider aggregative abstraction an idealization, since we take that omissions and distortions cannot sharply be distinguished, in contrast to Levy (2018), who holds that we ought – and can – differentiate between them. Another justification for considering aggregative abstraction as an idealization is due to the distinction between Galilean and minimalist idealization. If minimalist idealization is understood as removing non-difference-making factors (or their contributions), it would amount to the conventional abstraction-by-omission account. So, by the standards of already established discussion of idealization, it is legitimate to regard Levy's "aggregative abstraction" as idealization.

simplified/schematic nature of the HH model certainly allows for such an interpretation. Indeed, the contrast between Levy and Craver revolves around the question of whether the HH model should be de-idealized in order to be explanatory, or if the model was explanatory *because* it ignored such detail. Consequently, Craver gives a deficiency reading of the HH model, as it does not make explicit which are the nuts and bolts of the mechanism of ionic transport. Levy, on the contrary, views the simplified nature of the HH model as an epistemic benefit, since in abstracting from the mechanism of ion transport the HH model is able to account for the ionic currents. Note how both authors, despite their differences, nevertheless ascribe to the representational approach in line with what we claimed above concerning the common supposition shared by both deficiency and epistemic benefit accounts of idealization. Both authors evaluate idealizations in terms of how they are able to pick out what is relevant for the explanation of the target phenomenon (and leave out what is not), and how such choices could be detrimental should they ignore *relevant* parts (or levels of abstraction). In other words, both authors implicitly agree that the assessment of idealizations should be done on the basis of model-world comparisons.

The analogical bedrock of the HH model casts doubt on the shared reliance of both Levy and Craver on the ability of scientists to hand-pick relevant factors or levels, or omit them in some representation-neutral manner. Interestingly, other philosophical discussants have paid attention to the role different formal and theoretical tools play in the derivation of the HH equations. Bogen (2008) suggested that the laws, such as Kirchhoff's Laws and Ohm's Law, are used in the HH models as calculation tools. That is, they function as formal relations used to derive the equations, and to hypothesize about electrical quantities of interest (Bogen 2008, see also Schaffner 2008). We take Bogen to be pointing at the artifactual role of the equivalent circuit and the laws used to describe the currents in it. Under this interpretation, rather than considering the electric circuit as a (mis)representation of the membrane, it is better viewed as an analogy that is associated with representational tools allowing the study of the nervous membrane's excitable behavior. Indeed, Hodgkin and Huxley themselves underlined the fact that the equations do not point towards specific mechanisms but provide a way of mathematically describing the overall dynamics (1952b p. 541, also discussed in Craver 2008, 1023; and in Levy 2013, 8).

Regarding the constant capacitance idealization, it is interesting to note that the previous philosophical discussion of the HH model has not addressed it at all. This may be due to the

fact that this particular idealization is not easily rendered representationally. If considered in terms of the deficiency account, one would expect that further improvements of the HH model might be able to correct this idealization. But correcting this idealization would mean changing the interpretation of the empirical results, and ultimately debunking the whole research program (something that is attempted by some neuroscientists, see Tasaki 1982, Lowenhaupt 1996, Heimburg and Jackson 2005, El Hady et al 2015).

On the other hand, from the epistemic benefit perspective, this idealization could be viewed as a distortion that ignores details that are not difference makers. This interpretation would be more suitable than the deficiency view with regard to the assumption of constant capacitance. It exhibits the fact that, to the best of their knowledge, Hodgkin and Huxley thought that capacitance was constant, and therefore did not consider it as a difference maker. However, this is not the whole story of why they made this assumption. The evidence for constant capacitance at the time was insufficient for it to be considered along the lines of the difference-making account, i.e. there was no conclusive evidence that capacitive currents would not contribute to the overall measured current (see Takashima 1979). So why did Hodgkin and Huxley make this assumption? The artifactual account, we claim, can give an answer. The assumption of constant capacitance did not emerge just for tractability reasons, and neither can it be cast solely in terms of difference-making. The isolation of a causal factor story by Cartwright and Mäki might fare better, yet it relies too much on decomposability. The point is that the previous renderings of the nerve signal were operating on assumptions that, when forced to be thought of in electrical terms, would translate into constant capacitance. Consequently, the assumption was already shaped by previous modeling attempts and conceptualizations of the nerve signal that both enabled and bounded the way the HH model was achieved.

Last but not least, the research program developed by Hodgkin and Huxley (and many others) would not have made sense if the capacitance were allowed to vary. Such acknowledgement would have obliged scientists to consider that the currents could be due to either capacitance changes *or* permeability changes (in regard to both the experiments and the equations!). As a result, the two accomplishments of Hodgkin and Huxley, the experiments and the model, would have been nullified. And it is highly likely that the whole research program would have been led into an entirely different direction. The artifactual perspective on Hodgkin and Huxley's achievement attests, then, to the role of idealizing assumptions in

the intertwinement of different theoretical, mathematical and empirical considerations that cannot be related to misrepresentation alone.

From the artifactual viewpoint, the idealization of constant capacitance (or of the membrane as a rigid semipermeable membrane) emerged from the effort of aligning and integrating diverse epistemic resources that had previously been exploited. These resources include the galvanic cells, Nernst's equation, electric circuit diagrams transferred from electrodynamics (and the laws applied to them), and the actual electrical devices implemented in the recording and intervening apparatus. If capacitive currents had not been assumed to disappear after the initial rearrangement of charges, the scientific problem itself would have changed dramatically, both empirically and also from the perspective of the construction of the target to be explained. For Hodgkin and Huxley this "simplification" (Hodgkin and Huxley 1952b, 505) seemed harmless, and without it the research program would probably have been paralyzed. It seems, then, that Hodgkin and Huxley, as well as many electrophysiologists before and after them, were exploring the explanatory potential of a series of assumptions that could not be de-idealized without corrupting the research program as a whole. The study of the nerve signal as a phenomenon in which there are no capacitive currents contributing to the global measured current in voltage clamp[18] delivered important explanatory benefits.

From the artifactual perspective, idealization-as-distortion view results in a naïve account of model-building, if only because the contact with the world is through various kinds of epistemic artifacts that are recruited and integrated in model construction. The upshot is that there are holistic idealizations for the evaluation of which there are no theory or representation free model-world comparisons (cf. Teller 2008). This is clearly the case with the HH model, where the model simulates experimental results that are interpreted under the assumption of constant capacitance – committing to the same idealization as required for the derivation of the equations. This means that the experimental results cannot arbitrate between the model and the world regarding that idealization.

6. Conclusions

---

[18] Except for the initial charging of the capacitor at the instant when the voltage is fixed to a particular value with the voltage-clamp.

In this paper we have argued for an artifactual approach to idealization by showing how it makes salient some important features of modeling that we analyzed through the case of nerve signal modeling. Most philosophical accounts approach idealization as distortion and consequently presume, either explicitly or implicitly, the possibility of establishing determinable representation free model-world comparisons for evaluating idealizations and their epistemic roles. In contrast, we have focused on how models are achieved by using actual representational tools and other epistemic resources.

The artifactual perspective emphasizes that even those idealizing assumptions that would traditionally be rendered as misrepresentations may have intricate relationships to various epistemic resources exploited in model construction. Accordingly, we have examined idealization by focusing on the different renderings of the nerve signal in the unfolding modeling process. In this process, modeling of the nerve cell with representational tools from physical chemistry led to the idealization of the nerve cell membrane as a semipermeable membrane like those in galvanic cells. Then we examined how this assumption was further developed into the assumption of constant capacitance in the equivalent circuit that Hodgkin and Huxley used to model the nerve cell membrane. The resulting set of assumptions is an example of holistic idealization that aligns and integrates different empirical, theoretical and representational resources. It is our claim that the relationship between these assumptions, and their role in coordinating diverse representational tools, only becomes salient when we adopt an artifactual approach instead of viewing idealizations as distorting misrepresentations.

As a result, the artifactual account recognizes the often holistic nature of idealization: idealizing assumptions hold the model together in configuring different epistemic artifacts. Such holistic nature of idealizing assumptions means that they are both enabling and limiting in a manner that is not dissectible into either one of them, as the deficiency and epistemic benefit accounts imply. The artifactual account both occupies a middle ground between the traditional deficiency and epistemic benefit accounts, and goes beyond them in attending to features of idealization that have largely been passed by in the contemporary discussion. In particular, the engagement of the artifactual approach with actual representational tools highlights the role of idealization in aligning and integrating mathematical, statistical, or

computational methods with theoretical notions, concepts, and results from recording and intervening apparatus.

Another important dimension of the artifactual account, the focus on the constrained construction of a model, is able to account for the epistemic benefits of idealization without too heavy realist baggage or too demanding decompositionality requirements that riddle the difference-making and isolation accounts. In regard to the decomposability issue, the artifactual approach is in agreement with Rice (2019) in that scientific models are not modular arrays, whose parts could, in some straightforward way, be compared to matters of fact about a target system. Scientific models are commonly de-idealized, of course, but the challenges of such processes are many, and partly overwhelming, as would only be expected should one take notice of the artifactual dimension of modeling (see Knuuttila and Morgan 2019).

Finally, our analysis of the Hodgkin and Huxley model reveals an intimate link between analogical reasoning and idealization that has been overlooked in the present discussion of idealization. Once the nerve impulse was modeled in terms of an electric circuit, it became possible to establish relations between formal laws and theoretical concepts, calculation methods and measuring techniques. The nerve signal research shows how central such an idealization as constant capacitance can be in drawing together different resources, and in establishing links between different research fields. In view of the distortion-to-reality accounts of idealization it is important to notice that if a membrane is cast as a capacitor, what is being assumed is that the whole of the circuit and the membrane behave in the same way. Since the artifactual account does away with the representationalist commitments that require us to decide whether models or their parts accurately describe their targets or not, it enables us to better understand how analogical reasoning routinely exploits idealization.

References

Alexandrova, Anna. 2006. Connecting Economic Models to the Real World: Game Theory and the FCC Spectrum Auctions. *Philosophy of the Social Sciences* 36 (2):173-192.

Batterman, Robert. 2009. Idealization and Modeling. *Synthese* 169 (3): 427–446.

Batterman, Robert and Rice, Collin. 2014. Minimal Model Explanations. *Philosophy of Science* 81 (3): 349–376.

Bernstein, Julius. 1902. Untersuchungen Zur Thermodynamik Der Bioelektrischen Ströme. *Pflüger Archiv Für Die Gesammte Physiologie Des Menschen Und Der Thiere* 92: 521–562.

Bogen, Jim. 2008. The Hodgkin-Huxley Equations and the Concrete Model: Comments on Craver, Schaffner, and Weber. *Philosophy of Science* 75 (5): 1034–1046.

Carrillo, Natalia. 2018. Integración de analogías en la modelación científica. *Revista Colombiana de Filosofía de la Ciencia* 18 (37): 317–335.

Carrillo, Natalia. 2019. *Modelación y abstracción en la investigación de la señal nerviosa.* PhD Thesis. http://132.248.9.195/ptd2019/mayo/0789695/Index.html

Cartwright, Nancy. 1999. The Vanity of Rigour in Economics: Theoretical Models and Galilean Experiments. *Centre for Philosophy of Natural and Social Science.* Discussion paper series 43/99.

Cole, Kenneth and Baker, Richard. 1941. Longitudinal Impedance of the Squid Giant Axon. *The Journal of General Physiology* 24 (6): 771–788.

Craver, Carl. 2006. When Mechanistic Models Explain. *Synthese* 153 (3): 355–376.

Craver, Carl. 2007. *Explaining the Brain: Mechanisms and the Mosaic Unity of Science.* New York: Oxford University Press.

Craver, Carl. 2008. Physical Law and Mechanistic Explanation in the Hodgkin and Huxley Model of the Action Potential. *Philosophy of Science* 75 (5): 1022–1033.

El Hady, Ahmed and Machta, Benjamin. 2015. Mechanical Surface Waves Accompany Action Potential Propagation. *Nature Communications* 6 (6697): 1–7.

Gelfert, Axel. 2016. *How to Do Science with Models.* Cham: Springer.

Godfrey-Smith, Peter. 2009. Abstractions, Idealizations, and Evolutionary Biology. In Anouk Barberousse, Michel Morange, & Thomas Pradeu (eds.)*, Mapping the Future of Biology: Evolving Concepts and Theories*. 47–56. Dordrecht: Springer.

Frigg, Roman and Stephan Hartmann. 2012. Models in Science, *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/models-science/. Accessed 26.01.2020.

Heimburg, Thomas, and Jackson, Andrew. 2005. On Soliton Propagation in Biomembranes and Nerves. *Proceedings of the National Academy of Sciences* 102 (28): 9790–9795.

Hodgkin, Alan, Huxley, Andrew, and Katz, Bernard. 1952. Measurement of Current-Voltage Relations in the Membrane of the Giant Axon of Loligo. *Journal of Physiology* 116: 424–448.

Hodgkin, Alan and Huxley, Andrew. 1952a. Currents Carried by Sodium and Potassium Ions through the Membrane of the Giant Axon of Loligo. *The Journal of Physiology* 116: 449–472.

Hodgkin, Alan, and Huxley, Andrew. 1952b. A Quantitative Description of Membrane Current and Its Application to Conduction and Excitation in Nerve. *The Journal of Physiology* 117: 500–544.

Huxley, Andrew 1999. Overton on the Indispensability of Sodium Ions. *Brain Research Bulletin* 50 (5–6): 307–8.

Iwasa, Kunihiko, and Tasaki, Ichiji 1980. Mechanical Changes in Squid Giant Axons Associated with Production of Action Potentials. *Topics in Catalysis* 95 (3): 1328–1331.

Knuuttila, Tarja. 2005. Models, Representation, and Mediation. *Philosophy of Science* 72 (5): 1260–1271.

Knuuttila, Tarja. 2011. Modeling and Representing: An Artifactual Approach to Model-Based Representation. *Studies in History and Philosophy of Science Part A* 42 (2): 262–271.

Knuuttila, Tarja. 2017. Imagination Extended and Embedded: Artifactual versus Fictional Accounts of Models. *Synthese*.

Knuuttila, Tarja and Morgan, Mary. 2019. De-Idealization: No Easy Reversals. *Philosophy of Science* 86 (4): 641-661.

Knuuttila, Tarja and Loettgers, Andrea. 2014. Varieties of Noise: Analogical Reasoning in Synthetic Biology. *Studies in History and Philosophy of Science* 48: 76-88.

Levy, Arnon. 2013. What Was Hodgkin and Huxley's Achievement? *British Journal for the Philosophy of Science* 65 (3): 469–492.

Levy, Arnon. 2018. Idealization and Abstraction: Refining the Distinction. *Synthese*. https://doi.org/10.1007/s11229-018-1721-z

Lowenhaupt, Benjamin. 1996. Gated, Ion-Selective Sodium and Potassium Channels of the Giant Axon: Do They Have a Role in Bioelectric Excitation? *Electro-and Magnetobiology* 15 (2): 151–157.

McMullin, Ernan. 1985. Galilean Idealization. *Studies in History and Philosophy of Science Part A* 16 (3) 247-273.

Mäki, Uskali. 1992. On the Method of Isolation in Economics. *Poznan Studies in the Philosophy of the Sciences and the Humanities* (special issue *Idealization IV: Intelligibility in Science, edited by Craig Dilworth)*. 26: 319-354.

Nernst, Walther. 1889. Die elektromotorische Wirksamkeit der Jonen. *Zeitschrift Für Physikalische Chemie* 4 (1): 129–81.

Ostwald, Wilhelm. 1890. Elektrische Eigenschaften halbdurchlässiger Scheidewände. *Zeitschrift für Physikalische Chemie* . 6 (1): 71-82.

Potochnik, Angela. 2017. *Idealization and the Aims of Science*. Chicago: The University of Chicago Press

Rice, Collin. 2018. Idealized Models, Holistic Distortions, and Universality. *Synthese* 195 (6): 2795-2819.

Rice, Collin. 2019. Models Don't Decompose That Way: A Holistic View of Idealized Models. *The British Journal for the Philosophy of Science* 70 (1): 179–208.

Schaffner, Kenneth. 2008. Theories, Models, and Equations in Biology: The Heuristic Search for Emergent Simplifications in Neurobiology. *Philosophy of Science* 75 (5): 1008–1021.

Sklar, Lawrence. 2000. *Theory and Truth: Philosophical Critique within Foundational Science*. Oxford: Oxford University Press.

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

Takashima, Shiro. 1979. Admittance Change of Squid Axon during Action Potentials. Change in Capacitive Component Due to Sodium Currents. *Biophysical Journal* 26 (1): 133–142.

Tasaki, Ichiji. 1982. *Physiology and Electrochemistry of Nerve Fibers*. New York: Academic Press.

Teller, Paul. 2008. Of Course Idealizations Are Incommensurable! In *Rethinking Scientific Change and Theory Comparison*, Léna Soler, Howard Sankey, and Paul Hoyningen-Huene (eds.), 247–264. Dordrecht: Springer Netherlands.

Weber, Marcel. 2005. *Philosophy of Experimental Biology*. New York: Cambridge University Press.

Weber, Marcel. 2008. Causes without Mechanisms: Experimental Regularities, Physical Laws, and Neuroscientific Explanation. *Philosophy of Science* 75 (5): 995–1007.

Weisberg, Michael. 2007. Three Kinds of Idealization. *Journal of Philosophy* 104 (12): 639-659.