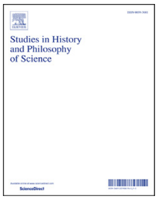
Studies in History and Philosophy of Science 91 (2022) 49–59

Contents lists available at ScienceDirect

Studies in History and Philosophy of Science

journal homepage: www.elsevier.com/locate/shpsa



Holistic idealization: An artifactual standpoint

Natalia Carrillo [*,1], Tarja Knuuttila [1]

Department of Philosophy, University of Vienna, Universitatstrase 7, 1010, Vienna, Austria

ARTICLE INFO

Keywords:
Modeling
Idealization
Artifactual account of modeling
Scientific representation
Nerve impulse
Hodgkin and Huxley model

ABSTRACT

Idealization is commonly understood as distortion: representing things differently than how they actually are. In this paper, we outline an alternative artifactual approach that does not make misrepresentation central for the analysis of idealization. We examine the contrast between the Hodgkin-Huxley (1952a, b, c) and the Heimburg-Jackson (2005, 2006) models of the nerve impulse from the artifactual perspective, and argue that, since the two models draw upon different epistemic resources and research programs, it is often difficult to tell which features of a system the central assumptions involved are supposed to distort. Many idealizations are holistic in nature. They cannot be locally undone without dismantling the model, as they occupy a central position in the entire research program. Nor is their holistic character mainly related to the use of mathematical and statistical modeling techniques as portrayed by Rice (2018, 2019). We suggest that holistic idealizations are implicit theoretical and representational assumptions that can only be understood in relation to the conceptual and representational tools exploited in modeling and experimental practices. Such holistic idealizations play a pivotal role not just in individual models, but also in defining research programs.

1. Introduction

Idealization is commonly understood as distortion in contemporary philosophy of science. This distorting character of idealization is often highlighted in contrast to the notion of abstraction. While abstractions omit features of real-world target systems, idealizations deliberately distort them (e.g. Jones, 2005; Godfrey-Smith, 2009; Levy, 2018). Examples of such idealizations are not hard to find, e.g. the assumptions of no friction, infinite number of particles, continuously increasing or decreasing populations, perfect markets, and rational expectations. The distorting nature of these assumptions is obvious, and does not even require any specialist knowledge. In concentrating on these kinds of cases of idealization, the contemporary discussion tends to approach idealization in terms of clearly identifiable false assumptions made in the modeling process. The question then becomes one of why scientists should make assumptions that strike as so obviously false. Two different accounts of idealization have emerged as a response. According to one of them, idealizations make models deficient, and many of them should, in principle, be de-idealized as modeling methods develop and the knowledge of the intended target systems accumulates (McMullin, 1985; Nowak, 1992). The other account highlights, in contrast, the epistemic benefits of idealization: it picks out causal difference makers for a particular target phenomenon, or, more generally, facilitates the study of particular causal factors in isolation of others.

Recently, Batterman and Rice (Batterman & Rice, 2014; Rice, 2018, 2019) have criticized the latter accounts of idealization on the basis that they rely on “veridical representation” of “common features” between the model and a real-world system (e.g. Batterman & Rice, 2014, p. 355). Such accounts attribute the explanatory power to the accurately represented parts of a model that identify relevant causal factors (or difference makers or shared structure). Distortion only concerns irrelevant features or non-difference makers. Batterman and Rice argue that such accounts misidentify what makes many idealized models explanatory; it is the distortion introduced by idealization, rather than the supposed common features, that is critical to the explanatory power of idealized models.[2] Consequently, “highly idealized models can play explanatory roles despite near complete representational failure” (ibid, see also Batterman, 2009).[3] Rice (2018, 2019) has further articulated this epistemically

---

* Corresponding author.
E-mail addresses: natalia.carrillo-escalera@univie.ac.at (N. Carrillo), tarja.knuuttila@univie.ac.at (T. Knuuttila).

[1] Both authors contributed equally to this work.

[2] Batterman and Rice use, as one example, the renormalization group transformation that eliminates details or degrees of freedom that are irrelevant. Their point is that the identification of the shared features of the class of systems is “a by-product” of this method (2014, p. 362–3).

[3] Batterman and Rice (2014) builds on the earlier work by Batterman on idealization (e.g. Batterman, 2009).

https://doi.org/10.1016/j.shpsa.2021.10.009

ineliminable quality of distorting idealizations in terms of holism: idealized models provide holistically distorted representations of their target systems (Rice, 2018, p. 2795). According to Rice, models themselves are idealizations, instead of being decomposable into accurate parts and distorting parts, the latter being either de-idealizable, or amenable to be "quarantined" from the accurate parts.

The holistic distortion view of idealization is based on the insight that various kinds of idealizing assumptions are required by the application of mathematical, statistical and computational modeling techniques and as such, they are crucial for particular explanations. Somewhat paradoxically, however, the holistic distortion account of idealization is still wedded to accurate representation, albeit inversely. It is the "pervasive *misrepresentations* of the features, processes and entities of the model's target system" that allow scientists to "extract the desired explanatory information that would otherwise be inaccessible" (Rice, 2018, p. 2809). Are we to conclude, then, that in the case of holistic idealization, it is misrepresentation that does the epistemic work that was earlier attributed to accurate representation? While appreciating the insight that many idealizations allow the use of mathematical, statistical and computational methods, we wish to take the focus away from distortion, which leads to questions as to whether it is the accurate representation or misrepresentation of the target system that delivers the epistemic benefits.

In this paper, we develop an artifactual view of holistic idealization that does not start from the representational assumptions inherent in idealization-as-distortion accounts, but rather focuses on the processes through which models are achieved, used and further developed. From an artifactual perspective, idealization amounts to an interrelated set of assumptions that allow, and are entailed in, the use of various kinds of epistemic resources that are coordinated in the modeling process. Such epistemic tools include mathematical, statistical and computational methods, but are not limited to them. In order to resist the temptation of analyzing idealization in terms of model-world comparisons, we study two conflicting models of the nerve signal: the Hodgkin-Huxley model (Hodgkin and Huxley, 1952c) and the more recent Heimburg-Jackson model (Heimburg & Jackson, 2005, 2006). We show that although some of the most important idealizations in these models could be interpreted as distortions—especially when each model is considered in its own terms—these central idealizations are not isolated choices concerning particular features of a representation. Instead, they result from more systematic research programs that integrate different concepts, analogies, measuring apparatus and mathematical approaches. Such idealizations are holistic in that they draw together and entail a set of interrelated assumptions that often cannot be locally undone. The artifactual approach, we submit, is able to accommodate many insights of the more traditional approaches, with a new twist: idealization can be simultaneously epistemically beneficial and detrimental.

In the following, we will first review the traditional accounts of idealization, in which we include the holistic account by Rice, although his version of holistic idealization can also be given an artifactual interpretation (Section 2). From these idealization-as-distortion accounts we turn to the artifactual account of models and idealization (Section 3). In sections 4 and 5 we compare the Hodgkin-Huxley and Heimburg-Jackson models in order to highlight the distinctive features of the artifactual approach to idealization.

2. Idealization as distortion

2.1. Deficiency and epistemic benefit accounts

In the philosophical tradition one can discern two distinct ways of understanding the functioning and justification of idealization. While one of these approaches views idealizations as deficiencies, for the other they deliver epistemic benefits. The deficiency accounts of idealization pay attention to how modelers idealize with the purpose of tackling complex real-world situations through tractable representations. Such idealizations should, in principle, be corrected. De-idealization occupies, without a doubt, a central place in many actual modeling practices, especially when it comes to applying models to real-life situations for the purposes of experimentation, prediction and policy, but de-idealization is also important for theoretical practice (Knuuttila & Morgan, 2019). The deficiency view acknowledges the epistemic access idealized models provide to the world given available representational and computational means. Yet it simultaneously considers idealizations as something to be corrected and controlled for, or eventually even replaced in the course of advancing research.

In contrast, the benefit accounts of idealization focus on the epistemic advantages of idealization. Idealization facilitates more efficient explanations and a better understanding of phenomena that would not be possible without it. While the deficiency accounts aim for correcting or controlling for idealization, the epistemic benefit accounts offer reasons for why scientists might be justified in not de-idealizing their models (or why it is not even possible, as the argument for holistic idealization goes). One crucial difference, then, between the two accounts boils down to whether de-idealization is desirable or not (irrespective of whether it in fact would be possible, see Knuuttila & Morgan, 2019).[4]

Why should scientists be justified in not de-idealizing their minimal models? Different justifications have been offered, such as difference making, isolation, and holistic distortion. We concentrate on the first two in this section and leave holistic distortion to the next section as it provides a markedly different kind of justification for minimalist idealization, challenging the difference making and isolation accounts.

The difference making account by Strevens (2009) provides the most heroic benefit interpretation of idealization. It is based on a causal mechanistic understanding of explanation, and assigns idealization the role of showing which salient causal factors are explanatorily irrelevant due to their "failure to make a difference" (p. 318). A good idealizing explanation, according to Strevens, conveys explanatorily essential information and "is always better than its veridical counterpart" (p. 318), and so "cannot be further improved" (p. 300). Clearly, then, any attempt to de-idealize an idealized model would lessen its explanatory optimality. The content of an idealizing model is composed of two parts: difference-makers, and non-difference-makers. In focusing on those causal factors that are relevant for explaining the phenomenon of interest, an idealized model is simpler and allows for more straightforward derivations of the explanatory target bringing about communicative, descriptive and computational benefits (p. 322). At the same time, the idealized model delivers information on which causal influences *do not* play a role, setting them aside by misrepresenting them.

These non-difference-makers are falsified by "conspicuously distorting their properties" either by assigning them extreme or default values, or by structural simplification. As a result, there will be an overlap between an idealized model and reality that amounts to a "standalone set of difference makers for the target" (p. 318).

The isolation account of idealization shares with the difference making account the idea of a partial overlap between an idealized model and reality, and the central role that idealizing assumptions play in achieving such an overlap. However, instead of aiming to pick out *the* causal difference-makers, the isolation account is more modest in addressing the contributions of separable causal factors for some target phenomenon. Idealization is approached in an analogy to the

---

[4] Weisberg (2007) presents a somewhat similar account. He calls Galilean idealizations such distortive idealizations that simplify in order to make the model tractable (see also Frigg & Hartmann, 2012) and minimalist idealizations many of those accounts that we categorize as benefit accounts. We prefer to talk about deficiency and benefit accounts since the distinction highlights the basic difference between these two accounts. Moreover, in talking about deficiency accounts of idealization instead of Galilean idealization we wish to avoid the different interpretations of Galilean idealization, e.g. McMullin's (1985) and Cartwright's (1999) accounts differ from Weisberg's.

experimental set-up. Mäki (1992) builds his approach on the idea of how various unrealistic model assumptions are used to theoretically "seal off" a set of relations from the influence of others.[5] 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. In other words, idealization may only aim at studying causal tendencies or capacities (Cartwright, 1998).

Not all idealized models involve isolation, however. Cartwright argues that many economic models cannot be considered as isolations because they are "over-constrained". Such models employ purpose-built assumptions that provide a way of securing "deductively validated" results with the consequence that those results remain model-dependent (Cartwright, 1999, p. 18). The question is whether or not this model-dependency applies to modeling more generally, especially as model construction depends on particular mathematical, statistical, and computational methods. One way to look at this problem is to acknowledge the pervasiveness and ineliminability of model-wide idealization—with the good news that such holistic distortion can in fact be epistemically productive.

2.2. Holistic distortion

Batterman and Rice (2014) argue that minimal modeling does not need to rely on causal, mechanical, or difference-making strategies that tie the explanatory power of minimal models to features that they have in common with real systems (p. 349). In fact, it is the other way around. Using the example of the Lattice Gas Automaton model, Batterman and Rice argue that the common features that different fluids have in common—locality, conservation, and symmetry—need to be explained as well (and not just used as an explanatory resource). Such explanation is provided by an idealization: in the case of Lattice Gas Automaton by "a renormalization group-like story" that delimits the universality class (p. 374).

It shows that differences between the model system and the real systems are irrelevant as "the key connection between the model and the diverse real-world systems is that they are in the same universality class." (p. 350). While Batterman and Rice make use of the notion of universality, Rice (2018, 2019) extends the idea of epistemic ineliminability of idealization to cover, more generally, the application of mathematical and statistical modeling techniques. These formal techniques give rise to what Rice calls "pervasive system-wide distortions" that "drastically distort the kinds of entities, interactions, and basic ontology of their target systems." (Rice, 2018, p. 2799). Mathematical representation introduces distortions without which the system would not display the behaviors modelers are interested in studying.

What Rice pays attention to is that the distortions necessitated by the use of mathematical tools are not piecewise, they are holistic in that they relate to "overall mathematical techniques" and are not locatable to some isolated parts of the model. Consequently, the distortions cannot be "quarantined" such that idealization would only concern irrelevant features (or features that scientists choose not to study in the model). The virtue of such system-wide idealization is that it "*extracts (or reveals) explanatory information on which features are relevant and irrelevant for the occurrence of the target explanandum.*" (p. 2802, italics in the original). Moreover, in revealing what is relevant and irrelevant it yields modal insight, e.g. on how the phenomenon of interest might counterfactually depend on various constraints or trade-offs. In contrast to the traditional accounts that at least partially rely on accurate representation, models that distort holistically "purposefully move us away from even attempting to accurately represent some isolable part" of the target phenomenon (p. 2808).

In our view, the account by Batterman and Rice (2014) provides a new twist to traditional distortion accounts, and Rice's further articulation of holistic distortion captures the intimate relation of idealization and the use of mathematical and statistical modeling techniques. However, in treating idealization as distortion, Rice's account still seems at least partially rooted in the representationalist[6] tradition by allocating the explanatory power to distortion (instead of accurate representation). Namely, the use of the notion of distortion intimates the possibility of representation free model-world comparisons. Rice claims that the counterfactual relations revealed "will just hold for (perhaps very) different reasons in the model system [than in the real system] and perhaps only in limiting cases." (2018, p. 2808). The mathematical frameworks employed by an idealized model "represent the target system as a fundamentally different *kind* of system in which qualitatively different kinds of behaviors are expected to occur" (p. 2809, italics in the original). However, it does not seem epistemically too instructive to consider the idealized model system mainly in terms of its being *different* from real-world systems. Instead, what one would like to know is *how* the system is represented. Scientific models typically involve representation-as (Elgin, 2009; Frigg & Nguyen, 2016; Hughes, 1997; Van Fraassen, 2008). That is, they represent the modeled system as being of a particular kind, and not merely as different from the real-world systems.

While Rice criticizes the "post hoc way" the traditional difference making and other isolationist accounts treat idealizations, as if "they already knew (or ultimately discovered) that they were only distorting irrelevant features" (Rice, 2018, p. 2810),[7] one cannot avoid the suspicion that, for Rice as well, the features to be distorted were somehow available, to be shown to be relevant or irrelevant by particular modeling endeavors. This may seem to be the case when one looks at a certain model, or group of related models, already firmly rooted in a particular research tradition. Nevertheless, our supposed ability to compare the features and basic ontology of a real-world system with those of the model begins to appear more questionable if one considers conflicting models that make very different assumptions of the system under investigation, as shown by the two contrasting models of the nerve impulse. The artifactual approach to modeling provides an alternative way of approaching idealization. It does not start by assuming that we either knew, or were able to know, the entities, and interactions of real-world target systems *independently* from our means of investigating them. Consequently, the artifactual account frees philosophical theorizing of idealization from the allure of uniquely determinable model-world comparisons, thus accommodating the holistic nature of idealization without invoking distortion to deliver the epistemic benefits.

3. The artifactual perspective on idealization

The basic motivation of the artifactual approach to modeling is to provide an alternative to the traditional accounts of models that assume

---

[5] For Mäki, isolation is the central notion, and idealization is just one form of theoretical isolation, understood in terms of making use of limiting concepts. Another important form of isolation is omission.

[6] We refer with representationalism to those (philosophical and other) theories that approach knowledge in terms of representations that reproduce accurately, i.e. stand truthfully for, mind-independent real systems (see Knuuttila, 2011, p. 264). A representational approach to models does not need to be representationalist, as the pragmatist approaches to representation show.

[7] One way of defending the isolation and difference-making accounts from this charge is to claim that they provide a *criterion* for what makes an idealization good, or successful. For Strevens, "difference-making" is a criterion for explanatory relevance (2009, p. 55), while Mäki (2009) claims that isolation should be understood from the perspective of the product and not the process of modeling endeavor. One can still wonder how the modeling heuristic would look like from these perspectives. In arguing that idealization gives knowledge of non-difference makers, Strevens depicts how from the canonical explanation of Boolean behavior the non-difference makers are "discarded", "stripped away", and "removed" (pp. 312–313). What he describes, moreover, is a result of the intersection of many long lines of research.

that models give knowledge in virtue of accurately representing their target systems or their parts. Such representationalist commitment leads to many familiar epistemological worries, among them the problem of how the notion of representation should be understood. From the representationalist perspective, the question of idealization becomes translated into misrepresentation by distortion. The philosophical question then concerns how to understand *and* justify such prevalent practice of misrepresentation. While the deficiency and epistemic benefit accounts adopt opposite approaches to the epistemic status of idealization, they both subscribe to the idea of idealization-as-distortion: idealizing models represent the worldly systems differently from how they *actually* are. Of course, there is no denying that scientific practice is replete with assumptions that ascribe model systems properties that the real-world systems do not and cannot have. As the falsity of such habitual assumptions is often self-evident, the notion of idealization as distortion has seemed inescapable.

The artifactual approach does not seek to contest the falsity of many such idealizing assumptions. Rather, it challenges, more generally, the idea of uncontextualized model-world comparisons that the idealization-as-distortion view builds on. In tackling the epistemic functioning of modeling, it does not then start from the representational relation between models and targets, but rather from the examination of models themselves and the often intertwined processes of their construction and use. The focus is on how epistemic access to the world is created through various kinds of epistemic resources – models, theoretical concepts, analogies, and experimental, mathematical and computational methods – that are integrated in model construction, in which idealization plays a central role.

The artifactual approach views scientific models as human-made or altered objects that are purposefully created to study certain problems, by making use of available representational tools in the context of specific scientific practices (Knuuttila, 2011, 2017). The epistemic value of a model is analyzed through the purposes it is constructed for, the representational tools it embodies, and its place in scientific practice. Instead of invoking the representational relation to account for their epistemic functioning, models are regarded as *erotetic vehicles*—as artificial systems of dependencies that are constrained in view of answering a pending scientific question, motivated by theoretical and/or empirical considerations.

As erotetic vehicles, models enable scientists to learn by building and manipulating models (Morrison & Morgan, 1999). This constructive and interventional side of modeling is crucial for their epistemic value. It shows how scientists can gain knowledge through articulating and working with the different relationships built into the model—instead of supposing that the only way to knowledge is through at least partially accurate reproduction of the actual state of affairs in the world. In contrast, models as variously materialized artifacts allow an epistemic access to many theoretical and empirical problems by enabling various inferences (Suárez, 2004, de Donato Rodríguez & Zamora Bonilla, 2009), providing new results, and in doing so also possessing considerable modal reach (Godfrey-Smith, 2006). Consequently, the epistemological puzzle of how the epistemic relationship between a model and the world is established is already partially answered by the construction of a model that is targeted at answering a specific question making use of, and also guided by, the symbolic, mathematical, conceptual and other available resources. Models typically draw upon many resources, as our discussion of the Hodgkin-Huxley and Heimburg-Jackson models in the next section will show.

Central among the resources that provide the epistemic access to the questions scientists are interested in are the representational tools with which models are composed. The representational tools come with their own enablements and limitations. For instance, one can do different things with diagrams, mathematics or natural language. Mathematical and computational tools have been central for the discussion of idealization, the deficiency and holistic distortion accounts of idealization giving different emphases to their enablements and limitations. While deficiency accounts have concentrated on the constraints mathematical and computational tools impose on model construction in stressing tractability considerations, the holistic distortion account has called attention to their epistemic productivity and modal character. Both perspectives are valuable, and in concentrating on the affordances of the actual tools of representation the artifactual approach is able to accommodate both. But it also goes further in paying attention to other resources—conceptual, theoretical, and empirical—and the contributions and limitations they introduce. What is important to note is that in model construction, these other resources need to be rendered by the representational tools used, and so the idealizing assumptions involved in the representing process have an important integrative function. The intersection of these different resources typically point to different lines of research that come together in particular modeling practice, each bringing their own results, methods, and norms of evaluation.

The critical question for the artifactual approach is to give an account on how the various inferences, new results and learning from models can be justified. Traditionally, the *justification* of model-based results has relied on representation—except for the deflationary pragmatists, who separate the analysis of the notion of representation from its success. As the artifactual perspective does not rely on any privileged representational relation providing a warrant for the correctness of model-based results, the justification eventually boils down to coherence with earlier theoretical and methodological commitments as well as empirical results. Within a particular modeling practice, the justification of models and the interpretations based on them is two-fold. On the one hand, is partly already *built-in* due to the independent reliability of the theoretical, empirical, mathematical, computational and other resources utilized in model construction (Boumans, 1999). On the other hand, model-based results are typically triangulated with different epistemic means: other models, experiments, observations and background theories. Such processes of triangulation are distributed in terms of epistemic labor, likely very complex and indirect, and usually inconclusive in character.

The artifactual account offers a different perspective on idealization than the idealization-as-distortion accounts. The artifactual account follows the traditional approaches in viewing idealizations as assumptions involved in model *construction*, but it focuses on how the various empirical, interventional, theoretical, mathematical and computational resources are rendered into a model through different representational tools. The representational and other resources integrated in the model construction carry their own commitments, and implicit or explicit idealizing assumptions. Not all of such idealizations are easily identifiable in terms of uncontextualized model-world comparisons, if only because many of them are not obvious falsifications—as we will show in our discussion of the Hodgkin-Huxley and Heimburg-Jackson models. The artifactual perspective is needed to appreciate the origin and evolution of commitments and idealizations originating from different artifactual contexts that provide resources for model building. Among these resources are conceptual ones that are crucial for understanding holistic idealization. Mathematical operations do not boil down to just formal techniques, usually they are coupled with particular ways of conceptualizing and representing the phenomenon of interest as being of a certain kind. The epistemic productivity of holistic idealization is, then, not due only to the application of formal methods, such as e.g. the renormalization group apparatus,[8] and the distortions they introduce. It also involves a historical dimension of representing-as (Elgin, 2009), extending to theoretical considerations, and empirical data through various intervening devices and representational tools.

In a nutshell, the artifactual account of idealization maintains the following five theses[9]:

---

[8] Knuuttila and Loettgers (2016) show in their study of the template transfer from the Ising model to spin glasses and neural networks that central for this transfer was not just the use of renormalization group methods, but also understanding the modeled system as a cooperative system.

[9] We do not claim that this list is exhaustive.

1. *Idealizations are implicit or explicit assumptions made in, or entailed by, model construction.* They embed formal, conceptual, theoretical and empirical resources into the model. Such assumptions are typically *ideal* in character in that they aim to render the phenomenon into a more pure and universal form in introducing, for example, limiting procedures, or turning some variables into constants. While many such idealizing assumptions are required by the mathematization of a phenomenon, they also are crucial for its conceptualization, with implications for empirical practice.
2. *Idealizing assumptions are frequently intertwined.* Models do not easily decompose into separable, distinct assumptions, as the assumptions of a model are typically intertwined (de Donato Rodríguez & Arroyo Santos, 2012; Rice, 2019). Many idealizations are holistic in nature, coordinating various theoretical, empirical and mathematical resources, frequently originating from different lineages of scientific work before coming together in a particular research program. Consequently, they often cannot be locally undone without dismantling the model, and undermining the research program.
3. *Idealizations are simultaneously both enabling and limiting.*[10] Idealizations facilitate the use of particular formal, representational, experimental and theoretical tools that come with their own specific affordances and commitments. In focusing on both enablements and limitations of idealization, the artifactual approach provides a unifying perspective that takes into account many insights of both epistemic benefit and deficiency accounts.
4. *Idealizations are crucial for reformulating the original problem in view of the purpose of the model.* In approaching models as erotetic devices the artifactual account is in line with most benefit accounts of idealization in that it emphasizes how idealization furthers the explanatory or exploratory goals of the model. However, it does not suppose that the main task of idealizing assumptions would be to isolate separable causal factors in a selective fashion.
5. *Holistic idealization involves representing-as.* Holistic idealization does not boil down to the application of mathematical and statistical tools (cf. Rice, 2018). Some idealizations are more central: they are *holistic* also in the sense that they depict the system as being of a certain kind, making it difficult to consider such idealizations as more or less accurate or distorting assumptions, or dissecting models into distorting and non-distorting parts. Whether or not such assumptions distort, depends on the context.

In anticipating our discussion of the Hodgkin-Huxley and Heimburg-Jackson models, it is important to notice that in the course of the advancing research, holistic idealizations tend to become reified.[11] For instance, Craver (2007) has described the discoveries made after the introduction of the Hodgkin-Huxley model—in particular the details of the mechanism of ion transport—as a transition from a “how possibly model” to a “how-actually mechanism” (p. 117). In order to avoid a direct translation of the model system into a target of interest, we contrast the Hodgkin-Huxley model with the Heimburg-Jackson model that also models the nerve impulse. The two models approach the nerve impulse from different and partially conflicting perspectives. The contrast between these two models illustrates the difficulty of making straightforward model-world comparisons, and casts doubt on depicting idealization as distortion. Holistic idealizations, in particular, should not be characterized in terms of distortion.

4. Two models of the nerve impulse

To illustrate more concretely what an artifactual approach entails, let us first consider how idealizations are frequently presented in philosophical literature. Jones (2005, p. 182) examines idealizations in a model of a cannonball trajectory, emphasizing the following:

- The model assumes that gravitational force has the same magnitude and direction at all points.
- It assumes that only the Earth exerts gravitational force.
- It assumes that the only force acting on the cannonball is gravitational, and.
- It assumes that the trajectory begins at ground level (ignoring the height of the cannon).

These idealizations are being presented as deviations from the real features of that particular system. Anyone with minimal scientific understanding knows that the elements in the list are false—they either violate the law of gravitation, assume a uniformity we never encounter in nature, or misrepresent the system of interest in some other ways. Can we expect all idealizations to be so evident? We suggest that there are idealizations that are subtler, requiring more contextualization to be identified as such. What tends to get lost in the discussion of idealization is that there are different kinds of idealizations.[12] For example, idealizations posing ideal conditions that can be approximated by experiments depend on what experimental conditions are realizable at a particular point in time, in a particular scientific practice. Such idealizations would require us to take into consideration not only the model, but also the relevant experimental practice, instrumentation and background knowledge. Likewise, such contextual factors are even more important when it comes to idealizing assumptions that seem highly plausible, but for which we cannot exclude the possibility that they turn out to be false. As an example, consider an idealization that states that organisms cannot live in temperatures of more than 1000 °C (de Donato Rodríguez & Arroyo Santos, 2012), or—as we examine in detail below—that the nerve membrane does not undergo capacitive changes during the nerve impulse. Such idealizations depend importantly on what is taken to be plausible (or implausible) at a particular moment of scientific development. For these kinds of idealizations, identifying idealization as distortion is particularly problematic, since by definition the material conditions by which we could verify their falsity are unattainable. Any philosophical account of idealization should also be able to cover these kinds of cases.

Instead of assuming that idealization boils down to choices concerning how truthfully or falsely to represent certain features of the target, the artifactual account investigates the motivations and origins of idealizations in the artifactual milieu that the scientists inhabit. The focus is on the intertwinement of idealizing assumptions—intended and unintended, explicit and implicit—and the representational tools involved in model construction and manipulation. The point is that many idealizations are not necessarily known to be false. Moreover, there is no straightforward way to individuate the idealizations of a certain model (Jones, 2005). The truthfulness or falsity of many built-in elements of models are often not addressed in scientific practice, and can even be ‘swept under the rug’ as the models they are embedded in turn out to be increasingly fruitful. For that reason, we maintain that understanding the epistemological status of many idealizations includes tracing down the model construction process with attention to the representational tools used, the scientific agenda involved, and the relevant skills of the scientists.

In the following, we (re)consider a central idealization in the Hodgkin-Huxley model that presumes that *the membrane capacitance is constant* (Hodgkin et al., 1952, p. 426; Hodgkin and Huxley, 1952c, p.

---

[10] For another view of idealization as both enabling and limiting, see Potochnik (2017).

[11] On ‘reification’ see Winther (2014).

[12] de Donato Rodríguez & Arroyo Santos, 2012 present a classification of idealizations on a modal scale. It categorizes idealizations in terms of their “degree of contingency” vis-à-vis the accepted body of knowledge. Their account comes closer to ours than the traditional idealization-as-distortion accounts as they emphasize the interrelatedness of idealizations, as well as their various epistemic, methodological or heuristic purposes.

505). In the circuit with which Hodgkin and Huxley model the membrane there are no current leaks between the conductors of the capacitor and the conductance is also otherwise constant (e.g.: there are no variations in the distance between the conductors or changes in their area). In discussing the idealizations in the Hodgkin-Huxley model, Levy (2018) focuses on its unrealistic commitments to a membrane that is "perfectly insulating" and an axon that is "perfectly symmetrical." Like the examples of idealization mentioned by Jones, the status of these idealizations *qua* idealizations is clear: to think of a natural process as *perfect* is to idealize it. Moreover, one could easily tell a standard philosophical story of why such idealizations would be justified; for instance, in the case of perfect insulation, by suggesting that any leaks in the capacitor are negligible. In the Hodgkin-Huxley model, perfect insulation is related to the assumption of constant capacitance: the membrane is assumed to be perfectly insulating *and* to have otherwise no variability in capacitance. A capacitor could be made from perfectly insulating material and still have variable capacitance if, for instance, the distance between the conductors was not fixed. So, the interesting question is why the idealization of constant capacitance is made, what is the rationale behind it, and what is its place in the scientific practice.

In anticipating our discussion of the Hodgkin-Huxley model, we suggest that the most interesting aspect of the idealization of constant capacitance lies beyond the false assumption of perfect insulation (see also Carrillo & Knuuttila, 2021). In assuming constant capacitance, Hodgkin and Huxley are also implicitly disregarding the thickness changes of the membrane that would result in capacitive currents. This latter aspect of the idealization of constant capacitance was not perceived as a distortion at the time it was introduced. As we see ahead, however, the assumption was contested a few years later and continues to be challenged to date. Yet, constant capacitance was central for Hodgkin and Huxley in permitting the coordination of previous representational tools with the novel electric circuit models. First, constant capacitance was representationally consistent with the previous models of the nerve cell. Second, capacitive changes were actually considered so implausible that they were only superficially discussed. Third, the assumption also played a role in the design of experiments and interpretation of empirical results.

The overarching presence of the assumption of constant capacitance in the research program within which the Hodgkin-Huxley model can be located justifies our characterizing it as a holistic idealization. The contrast between the Hodgkin-Huxley model and the Heimburg-Jackson model, which idealizes the membrane in a different way, shows that the notion of idealization as distortion is not the best overall framework to address holistic idealization. There does not seem to be any firm model-independent target system serving as a benchmark for evaluating whether many central idealizations in the models in question do in fact lead to a misrepresentation. Instead, the question of whether the capacitance is assumed to be constant or variable can serve as a starting point for the exploration of different possible mechanisms of nervous transmission. Both idealized models provide epistemic benefits, but these benefits also do come at the cost of narrowing down the possible explanations—without a guarantee that those left out are unfruitful.

4.1. Constant capacitance and the Hodgkin-Huxley model

The accepted explanation of the mechanism of nerve impulse before the work of Hodgkin and Huxley was Bernstein's "membrane theory". Around 1890, Ostwald and Nernst were interested in the electrochemical equilibrium of ions in galvanic cells. In 1888, W. H. Nernst expressed the equilibrium voltage as a function of the difference in concentration of the ion species that can cross the semipermeable membrane of a galvanic cell.[13] Two years later, Ostwald proposed that bioelectric potentials might arise because of the presence of semipermeable membranes in nerve fibers. Bernstein developed this insight in his theory that considered the nerve impulse to be a collapse of the nerve cell membrane. During the collapse, the membrane would become semipermeable, allowing free diffusion of (mostly potassium) ions. Published in 1902, Bernstein's "membrane theory" was still accepted when J. Z. Young discovered the squid giant axon in 1936. These axons of up to 1 mm in diameter offered an ideal experimental material for neurophysiologists, enabling them to advance these theoretical developments.

Hodgkin and Huxley focused on the design and study of the first intracellular electrical recordings in squid giant axons (Hodgkin & Huxley, 1945). One of their immediate findings was that Bernstein's explanation was partly mistaken. Hodgkin and Huxley showed that the membrane does not simply collapse as Bernstein suggested, and rather proposed independent mechanisms of variable permeability for sodium and potassium ions.

The prospect of investigating the electrical features of the giant axon provided one important motivation for the electrical rendering of the nerve cell (Tasaki, 1982). The electrical renderings soon played a role also in the development of theoretical models. In 1941, Cole and Baker described the membrane in terms of an analogical circuit composed of a capacitor, a resistor and an inductor (Cole & Baker, 1941), Hodgkin and Huxley later modeled the dynamics of ionic currents across the nerve cell membrane in terms of currents in a resistor-capacitor (RC) circuit. In this circuit (Fig. 1), the semipermeability of the membrane is divided functionally into two components: A capacitor with constant capacitance that accounts for the insulating aspect of the membrane and variable resistances connected in parallel accounting for membrane permeability (Hodgkin et al., 1952, p. 426).

In order to arrive at a mathematical expression of the current across the membrane, Hodgkin and Huxley derived the equations for the current in the circuit from the laws of electrodynamics. The law of conservation of charge states that the total of current coming out of any node in a circuit minus the current coming in equals zero. Applied to the node where current is being injected (I_{app}), the following equation is obtained:

$$I_{app} = I_c + I_K + I_{Na} + I_L \tag{1}$$

Hodgkin and Huxley proceeded to get an expression for each of these addends. The sodium and potassium currents were calculated with Ohm's

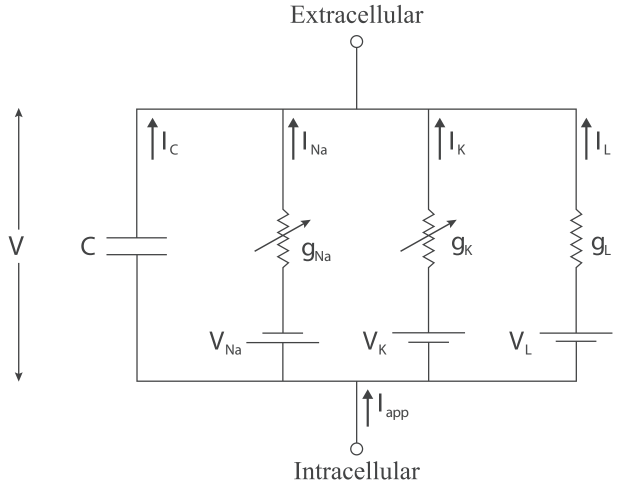


Fig. 1. The Hodgkin-Huxley circuit with modern conventions (original in Hodgkin & Huxley, 1952c, p. 501). Current is applied at I_{app}. I_{Na} and I_K correspond to sodium and potassium currents, and V_{Na} and V_K correspond to the equilibrium voltages of each of these ions, calculated with Nernst's equation. C is the capacitance of the membrane, assumed to be constant. I_L is a corrective linear factor, introduced to capture the contribution of other ions that make a very tiny contribution.

---

[13] Nernst's equation predicts the electrochemical equilibrium voltage given the concentrations of uni-univalent ions at each side of the semipermeable membrane.

law and the resistance's variation was adjusted to the experimental results of voltage-clamp experiments (we explain this experimental procedure below). The current in the capacitor is obtained via the relationship $Q = CV$, where Q is charge, C is capacitance and V is voltage. The derivative of charge with respect to time gives us the current across the capacitor (I_c):

$$I_c = \frac{dQ}{dt} = \frac{d}{dt}(CV) = C\frac{dV}{dt} + V\frac{dC}{dt} \tag{2}$$

If the capacitance is a constant function, its derivative is zero, which eliminates the term $V\frac{dC}{dt}$. In this case, the equation for the current in the circuit is much simpler; it becomes $I_c = C\frac{dV}{dt}$. Consequently Eq. (1) becomes $I_{app} = C\frac{dV}{dt} + I_K + I_{Na} + I_L$. This is the main equation that Hodgkin and Huxley used to develop their model.

What support is there for the assumption of constant capacitance? The assumption suggests that the membrane would not undergo physical changes (in thickness, for instance) such that its capacitance would vary significantly. The semipermeable membranes used in the late 19th century were macroscopic barriers that did not undergo density or volume changes, often made of ceramic treated with copper sulfate and potassium ferrocyanide solutions (Kleinzeller, 1995, p. 32). It is possible that galvanic cell models with rigid membranes influenced the idea that the neuronal membrane does not change shape. Moreover, in galvanic cells all the explanatory weight lies on the permeability of the membrane and concentrations of ions, so galvanic cell models of the nerve cell membrane implicitly discard potential contributions from capacitive currents.

Considering that previous models did not attend to capacitive changes, it was not surprising that Hodgkin and Huxley did not discuss the constant capacitance idealization to any extent in the presentation of their mathematical model. Whereas they were meticulous about whether the ionic currents could be described as Ohmic (Hodgkin & Huxley, 1952b, p. 477), they did not experimentally test whether membrane capacitance could vary between passive and excited states. Hodgkin and Huxley (1952a, 1952c) cite previous measurements that stated that the membrane capacitance is around $1\,\mu F/cm^2$ (Cole & Curtis, 1939). However, these experiments did not properly examine whether membrane capacitance is voltage-dependent (Takashima, 1979). Had variable capacitance been on the radar of the scientific community, these results would likely have struck the community as providing insufficient support for the assumption of constant capacitance.

The assumption of constant capacitance was also crucial for experimental work. In the voltage-clamp experiments on giant axons the scientist fixes a value for the membrane potential, and the voltage clamp injects the current necessary to keep it fixed. The voltage clamp setup records the time course of the current that had to be injected to this end. The underlying rationale is that the injected current is counterbalancing currents of ions across the membrane, so the recording is taken to be the inverse of the transmembrane currents. However, this presupposes that there are no capacitive currents at play during excitation. Consequently, the empirical results also assume constant capacitance (see below). In turn, the experimental results obtained in this manner influenced the design of the model. For instance, the voltage-clamp experiments supporting the conclusion that voltage sensitive permeabilities of sodium and potassium are independent from each other (Hodgkin and Huxley, 1952a, 1952b) gave justification for treating sodium and potassium resistances as appearing in parallel in the Hodgkin-Huxley analogous circuit (Hodgkin & Huxley, 1952a; 1952c, p. 500).

What would have happened if Hodgkin and Huxley had suspected that capacitance changes could also be associated with excitation? If capacitance is not constant, the currents could occur *in the near vicinity* of the membrane (due to displacements provoked by the membrane changing shape, for instance), instead of across the membrane. The assumption that capacitance is constant is precisely what allowed Hodgkin and Huxley to interpret the readings of the voltage clamp experimental setup as reflecting the *transmembranal currents.*[14] Without this idealization the experimental design would have lost its rationale. On the other hand, from the perspective of the mathematical model, the system of equations of an analogous circuit with variable capacitances *and* variable permeabilities would have become too complex due to too many degrees of freedom. It is most likely that the whole research program would have been undermined by such considerations.

In sum, there is more to the assumption of constant capacitance than meets the eye. It involves more than perfect insulation, and can be viewed as a commitment that was implicitly acquired as different representational tools were built upon each other, also influencing the design of experiments. Initially, as we have seen, it was considered implausible that this idealization would turn out to be false. Clearly, the idealization of constant capacitance was not intended as a misrepresentation. Moreover, we contend that it is unclear whether it could properly be understood as a distortion at all. The membrane was represented as a circuit *with* a capacitor, so ultimately to decide whether this idealization is a distortion would amount to evaluating whether the membrane-represented-as-a-circuit is a distortion, which is either a trivial question—of course the cell is not an electric circuit—or a misplaced one, since the circuit is not supposed to represent the membrane in the sense that it would be structurally similar to it. At most, the basic rationale is that the overall *dynamics* of both systems—the circuit and the membrane—are similar.

To recapitulate, the idealization of constant capacitance played a role in the design of experiments and the interpretation of their results, and the results influenced the design of the model. Representing the nerve impulse as an electric circuit allowed the scientists to obtain a system of equations that models the behavior of the nerve impulse (provided that the idealization of constant capacitance is in place). Consequently, the idealization of constant capacitance was wedded to the dialectics between the model and the experiments. Although Hodgkin and Huxley did not thus consider the possibility of variable capacitance, the assumption of constant capacitance being central for their research program, alternative modeling strategies that do not rely on this idealization were examined later on by a number of scientists (El Hady & Machta, 2015; Heimburg & Jackson, 2005; Tasaki, 1982).[15] We examine one of these efforts in the next section.

4.2. Adiabaticity and the Heimburg-Jackson model

The electrical approach that enabled Hodgkin and Huxley to develop their model furnishes the backbone of the current accepted explanation of nervous transmission. As a result, the idea that there are no significant contributions from capacitive currents to the dynamics of transmembrane voltage became deeply ingrained in electrophysiology (Takashima, 1979, p. 133). The reification of constant capacitance made it difficult for modeling approaches that do not commit to this idealization to flourish. According to Takashima "[t]he historic experiments by Cole and Curtis (1939) established the concept of membrane capacitance as a static and passive quantity that has no bearing on the physiology of nerve axons" (Takashima, 1979, p. 140). Still, some scientists resisted the

---

[14] Fixing the voltage in the experiments implies that the derivative of voltage in time is zero so the addend disappears from the equation. As mentioned before, because of conservation of charge, the equation for the node in which current is injected is: $I_{app} = C\frac{dV}{dt} + I_K + I_{Na} + I_L$. This equation already assumes constant capacitance, since capacitive current is taken as $C\frac{dV}{dt}$ (otherwise the capacitive current would be given by Eq. (1) above). Under the voltage clamp, the voltage is fixed at a constant value, so $\frac{dV}{dt} = 0$, which implies that the applied current is of the same magnitude as the transmembranal ionic currents (the currents in the resistances).

[15] The Engelbrecht model presents a different kind of alternative to the Hodgkin-Huxley model that attempts to unify the different modeling strategies. For a discussion see (Holland et al., 2019).

idealized view of the membrane inherent to the Hodgkin-Huxley model, also on an experimental basis. The assumption of constant capacitance was contested in the eighties by showing that nerve axons shorten (Tasaki & Iwasa, 1980) and swell (Tasaki & Iwasa, 1982) when transmitting a nerve impulse. Such mechanical effects are difficult to explain from the received electrical perspective, since it does not address pressure, volume or density changes in the membrane. Additionally, evidence of heat emission and reuptake in phase with the nerve impulse suggests that it is not a dissipative phenomenon (Abbott et al., 1958). As recognized by Hodgkin himself, this result is not accommodated by the Hodgkin-Huxley model (Hodgkin, 1964, p. 70).

The Heimburg-Jackson model presents an alternative to the traditional approach with the potential of addressing the mechanical and thermodynamical features of the nerve impulse. The model is based on findings that isolated lipids of biomembranes display order-disorder (gel-fluid) phase transitions. The gel state is associated with excitation: the lipids of the membrane have a more ordered organization with a thicker and denser presentation than in the liquid phase. These observations suggest that the mechanical effects of nervous transmission could be due to the phase transitions in the lipids of the membrane. In view of these developments, interest in such mechanical effects has been rekindled, and researchers have found novel ways to investigate the swelling of the membrane during transmission using an atomic force microscope (Gonzalez-Perez et al., 2016).

Phase transitions would have to form localized signals in order to support nerve impulses. Heimburg and Jackson consider that phase transitions form solitons—waves with peculiar properties that make them ideal transmitters of information. Solitons are solitary waves that maintain their shape and velocity, and do not annihilate or change shape when colliding with other waves. Solitonical waves can propagate without a loss of energy that is in line with the previously mentioned evidence of heat emission and reuptake during nerve impulse transmission. The soliton phenomenon was discovered in water channels by Scott Russell already in the early 19th century, and was mathematically characterized by Boussinesq in 1872. Solitons are studied nowadays in as diverse areas as telecommunication, magnets, nuclear physics and molecular biology. In the case of nerve impulse propagation, Heimburg and Jackson used hydrodynamic equations as a point of departure to model the nerve impulse, obtaining an equation formally equivalent to Boussinesq's. Heimburg and Jackson explain nerve impulse generation and transmission as the result of these solitonical pulses sustained by phase transitions that travel along the axon.

Instead of making assumptions about the capacitance of the membrane, the set of representational tools that Heimburg and Jackson are exploiting keeps the possibility open that the nervous membrane undergoes capacitance changes during excitation. Heimburg and Jackson juxtapose their model with the Hodgkin-Huxley model in questioning the feasibility of constant capacitance idealization:

> They [Hodgkin and Huxley] assumed, as it is now customary in the field, that the lipid membrane acts as a simple constant capacitor and that the observed nonlinear currents are due to protein ion-channels which are embedded in an otherwise inert membrane. However, it has been shown that the membrane rather behaves as a nonlinear capacitor. (Mosgaard, Zecchi, Heimburg and Budvytye, 2015, p. 495).

Via their novel thermodynamic approach, the Membrane Biophysics group led by Heimburg focuses on investigating electrostriction, i.e.: how changes in the electric field influence the transition temperature of the membrane (Heimburg, 2012; Mosgaard et al., 2015). If electrostriction is at play, the value of the capacitance changes as a nonlinear function of the applied voltage.

The Heimburg-Jackson model proposes an alternative rendering of the nerve impulse as a localized piezoelectric sound pulse and of the nervous membrane as an elastic material that undergoes phase transitions between gel and liquid phases (see Fig. 2). The model relies on a different set of idealizations than the Hodgkin-Huxley model—for instance, it views the nerve impulse as a thermodynamically and mechanically reversible phenomenon:

> The finding of zero net heat exchange implies a conservation of entropy of the system (no entropy is dissipated). […] The consequence is that the physical process underlying the action potential must be of reversible nature. The Hodgkin-Huxley model, which is the textbook picture of the nerve impulse, however, is exclusively based on irreversible processes […] we described here an alternative approach that assumes that the nerve impulse is an entropic pulse (with zero entropy change after completion of the pulse). Such a pulse would show some features that have been described for nerves: reversible heat release, reversible changes in thickness, and reversible changes in membrane state. (Heimburg & Jackson, 2008, p. 336).

The Heimburg-Jackson model relies importantly on a few empirical findings of no net heat exchange, suggesting the conservation of entropy and the reversibility of the physical processes underlying the nerve impulse. The argument is that *if* one takes the reversible heat of the nerve pulse “seriously,” “one is inclined to state that the [Hodgkin-Huxley] model in itself cannot be correct” (Appali et al. 201, 295). However, as in the case of constant capacitance, the pieces of evidence supporting no net heat exchange are controversial (Drukarch et al., 2018). Among other things, experimental difficulties involved in measuring heat have resulted in very few experimental results regarding heat production and absorption during nervous transmission. This may be why this is described as an “assumption” of the Heimburg-Jackson model (see the quote above). The assumption of no net heat exchange during nerve impulse transmission (henceforth adiabaticity) integrates with the hypothesis of phase transitions as the basis of the signal, since such phase transitions are reversible phenomena. Moreover, solitonical waves are compatible with the notion of reversible (isentropic) pulses. In other words, adiabaticity is not a separable assumption that could be isolated since it plays a role in coordinating the use of certain theoretical and representational tools.

The adiabaticity assumption depicts an idealized phenomenon: “In real systems no process is perfectly adiabatic (completely reversible) and always some of the energy is dissipated. If the dissipation is small, however, sound propagation can occur over long distances.” (Heimburg, 2007, pp. 311–312). However, in this case as well as with constant capacitance, the crux of the issue does not lie in whether it is a distortion or not. The benefits of the adiabaticity idealization are due to the fact that it coordinates a number of theoretical and representational tools that enable scientists to approach thermal as well as mechanical aspects of the nerve impulse. The struggle between the Hodgkin-Huxley and Heimburg-Jackson models is not one of how one could make the models more realistic, or of which idealization would succeed to isolate relevant causal factors, but that of which idealized framework offers more epistemic potential to address specific questions concerning nervous transmission.

When one considers the Hodgkin-Huxley model and its target in isolation, the representation of the membrane as having constant capacitance does not seem like a distortion (beyond the capacitor being perfectly insulating). Similarly, if the Heimburg-Jackson model were the only available model of the nerve impulse, it would be more feasible to claim that the nerve impulse is being accurately approximated as an adiabatic phenomenon. However, the tensions between viewing the nervous impulse as a solitonical pulse vis-à-vis the earlier electrodynamical approach, makes such a claim difficult to maintain. The adiabaticity idealization is related to the thermodynamical approach that Heimburg and Jackson are committed to. It enables them to relate the nerve impulse with phase transitions, and is crucial for applying the concept of a soliton, as well as obtaining equations that simulate the behavior of nerve impulses.

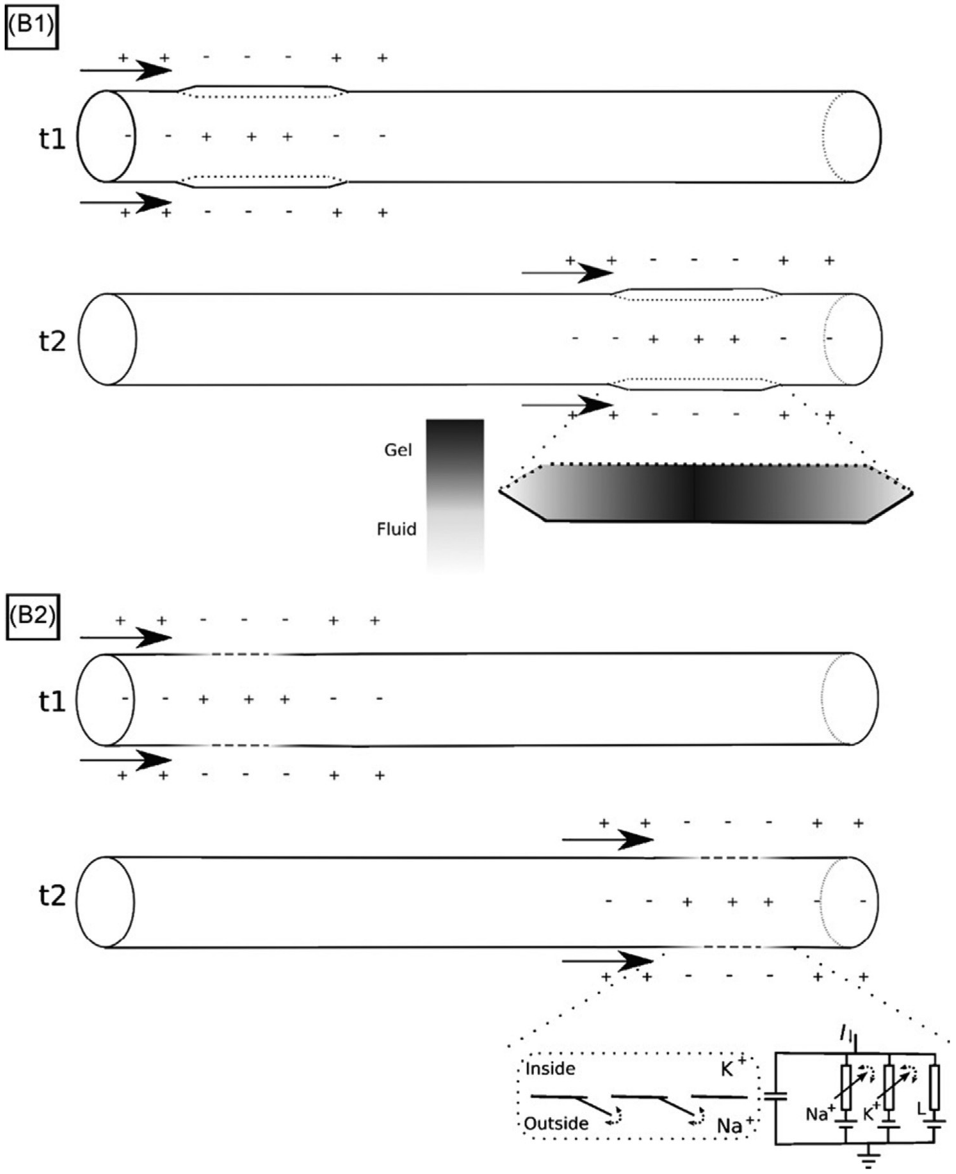


Fig. 2. A comparison of Heimburg-Jackson (B1) and Hodgkin-Huxley (B2) models. Image from (Andersen et al., 2009). The Hodgkin-Huxley model focuses on the electric effect of permeability changes, while the Heimburg-Jackson model stresses the thermal and mechanical effects of propagating phase transitions.

5. Holistic idealization

In our assessment of the contemporary discussion of idealization in the philosophy of science, we identified two main currents. According to one, idealizations make models deficient and should be de-idealized if possible; according to the other, idealizations are beneficial and not to be eliminated lest the model lose much of its epistemic power. Both approaches have shortcomings. On the one hand, the deficiency accounts cannot adequately explain the pervasiveness of minimal modeling and the epistemic usefulness of idealization. On the other hand, the benefit accounts of idealization tend to rely on our ability to decompose models (and isolate separable causal factors) that does not apply to many instances of modeling—as argued by Rice in his discussion of holistic distortion (Batterman & Rice, 2014; Rice 2018, 2019). Despite their differences, the three approaches to idealization – deficiency, benefit and Rice's holistic view– assume that idealization amounts to *distortion*.

We claim that fixating on distortion is misleading in that it cannot give a comprehensive account of idealization. From the perspective of model construction, many idealizations appear holistic, and not separable into assumptions whose distorting nature would be self-evident. The idealization of constant capacitance is a case in point. As we have shown in our analysis of the Hodgkin-Huxley model, the assumption of constant capacitance is a reinterpretation of the earlier tradition of representing the membrane as a galvanic cell, via the integration of representational tools from electrodynamics. It is thus an assumption both related to the previous models and constrained by the choice to represent the cell as a Resistor-Capacitor circuit. Moreover, this idealization was intertwined with the experimental procedures. The idealization of constant capacitance is an instance of a central, holistic idealization in the sense that the model, and even the whole research program, would collapse without it. We also showed that whether or not this idealization could be viewed as a distortion depends on historical, and other contextual factors. Hodgkin and Huxley certainly did not conceive of it as a distortion, yet some empirical results appear to contest it, and some contemporary models do not in fact make such an assumption.

In contrast, we examined the Heimburg-Jackson model that does not commit to the constant capacitance assumption. However, this model does not simply aim to correct for this idealization, but draws inspiration from entirely different research traditions. The model proposes an alternative idealized framework for studying the nerve impulse, as it too engages in what can be described as holistic idealization. The rendering of the nerve impulse as a thermodynamically reversible phenomenon is crucial for the development of the program, coordinating a number of theoretical and representational tools. If idealizations were considered mainly as distortions, the question of whether the idealizations of constant capacitance or adiabaticity are idealizations, would depend on who considers them, when, and from which perspective. Perhaps this is a conclusion that idealization-as-distortion theorists would be willing to embrace, but it serves to show that it is often difficult to judge which modeling assumptions are distortions, and which are not.

Viewing idealizations as distortions assumes that we could have ways of assessing them independently of the artifactual contexts that give us epistemic access to phenomena. We do not think that this is often the case, although many standard examples of idealizations clearly misrepresent what is known to be the case. There are other cases, however, that are not so easily rendered as distortions and that nevertheless bear the mark of idealizing in that they e.g. assume a ‘perfect’ or ‘constant’ behavior of a system or parts of it, reflected in the values that certain variables and parameters can take. If we base our notion of idealization on the obvious cases of misrepresentation, we lose sight of many other assumptions that have similar functions in model construction ending up, as a result, with an arbitrary account of idealization.

We have offered an alternative artifactual account of idealization that considers idealizations as implicit or explicit assumptions that draw formal, conceptual, theoretical, and even empirical resources together in model construction. These assumptions are crucial for reformulating the original problem, are often intertwined, and holistic in that they represent the system of interest as being of a certain kind. The two models we studied approach the membrane from fundamentally different perspectives. The Hodgkin-Huxley model represents the membrane as a semipermeable barrier and gives it an electrical reinterpretation, whereas the Heimburg-Jackson model renders the membrane as an elastic material undergoing phase transitions. What is at stake does not reduce to the correctness of separable assumptions, and to the consequent question of whether or not idealizing assumptions are beneficial or detrimental. The artifactual account regards idealization as *both enabling and limiting,* thus accommodating insights of both benefit and deficiency accounts. For example, though the Hodgkin-Huxley model offered a convincing explanation of the nerve impulse, as the research program matured it could not address some recalcitrant empirical evidence that the alternative Heimburg-Jackson model was able to explain.

In conclusion, the distortion notion of idealization tends to set aside the modeling processes, focusing instead on model-world comparisons, adopting ‘a view from nowhere’ for judging which assumptions misrepresent and how. In a context where there are no alternative models, it is tempting to reify holistic idealizations, taking the success of the model for the truthfulness of its idealizing assumptions—as was the case of the idealization of constant capacitance for most electrophysiologists in the 20th century. The idea of misrepresentation is not particularly helpful for capturing the complexities of such holistic idealizations, and to portray them as distortions would often require more than what the epistemic situation in question allows for.

Acknowledgements

This project received funding from the European Research Council under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 818772). The authors would like to thank the audience of the 11th Auburn Philosophy Conference, as well as the anonymous reviewers for their comments. We also thank Thomas Heimburg, Andrew Jackson and Edgar Villagrán for discussing the thermodynamical features of nerve impulses with us.

References

Abbott, B. C., Hill, A. V., & Howarth, J. V. (1958). The positive and negative heat production associated with a nerve impulse. *Proceedings of the Royal Society B: Biological Sciences, 148*(931), 149–187.

Andersen, S. S. L., Jackson, A. D., & Heimburg, T. (2009). Towards a thermodynamic theory of nerve pulse propagation. *Progress in Neurobiology, 88*(2), 104–113.

Batterman, R. W. (2009). Idealization and modeling. *Synthese, 169*(3), 427–446.

Batterman, R. W., & Rice, C. C. (2014). Minimal model explanations. *Philosophy of Science, 81*(3), 349–376.

Boumans, M. (1999). Built-in justification. In M. S. Morgan, & M. Morrison (Eds.), *Models as mediators. Perspectives on natural and social science* (pp. 66–96). Cambridge University Press.

Carrillo, N., & Knuuttila, T. (2021). An artefactual perspective on idealization: Galvanic cells and electric circuits in nerve signal research. In A. Cassini, & J. Redmond (Eds.), *Models and idealizations in science: Fictional and artefactual approaches*. Springer.

Cartwright, N. (1998). Capacities. In J. B. Davis, D. W. Hands, & U. Mäki (Eds.), *The handbook of economic methodology* (pp. 45–48). Edgar Elgar.

Cartwright, N. (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, K. S., & Baker, R. F. (1941). Longitudinal impedance of the squid giant axon. *The Journal of General Physiology, 24*(6), 771–788.

Cole, K. S., & Curtis, H. J. (1939). Electric impedance of the squid giant axon during activity. *The Journal of General Physiology, 22*(5), 649–670.

Craver, C. F. (2007). *Explaining the brain - mechanisms and the mosaic unity of science*. Oxford University Press.

de Donato Rodríguez, X., & Arroyo Santos, A. (2012). The structure of idealization in biological theories: The case of the wright-Fisher model. *Journal for General Philosophy of Science, 43*(1), 11–27.

de Donato Rodríguez, X., & Zamora Bonilla, J. (2009). Credibility, idealisation, and model building: An inferential approach. *Erkenntnis, 70*(1), 101–118.

Drukarch, B., Holland, H. A., Velichkov, M., Geurts, J. J. G., Vroon, P., Glas, G., & de Regt, H. W. (2018). Thinking about the nerve impulse: A critical analysis of the electricity-centered conception of nerve excitability. *Progress in Neurobiology, 169*, 172–185.

El Hady, A., & Machta, B. B. (2015). Mechanical surface waves accompany Action potential propagation. *Nature Communications, 6*, 1–7.

Elgin, C. Z. (2009). Exemplification, idealization and scientific understanding. In M. Suárez (Ed.), *Fictions in science: Essays on idealization and modeling* (pp. 77–90). Routledge.

Frigg, R., & Hartmann, S. (2012). *Models in science, the stanford Encyclopedia of philosophy*. https://plato.stanford.edu/entries/models-science/. (Accessed 26 January 2020).

Frigg, R., & Nguyen, J. (2016). The fiction view of models reloaded. *The Monist, 99*, 225–242.

Godfrey-Smith, P. (2006). The strategy of model-based science. *Biology and Philosophy, 21*, 725–740.

Godfrey-Smith, P. (2009). Abstractions, idealizations, and evolutionary biology. In A. Barberousse, M. Morange, & T. Pradeu (Eds.), *Mapping the future of biology: Evolving concepts and theories* (pp. 47–56). Springer Netherlands.

Gonzalez-Perez, A., Mosgaard, L. D., Budvytyte, R., Villagran-Vargas, E., Jackson, A. D., & Heimburg, T. (2016). Solitary electromechanical pulses in lobster neurons. *Biophysical Chemistry, 216*, 51–59.

Heimburg, T. (2007). *Thermal biophysics of membranes*. Wiley - VCH.

Heimburg, T. (2012). The capacitance and electromechanical coupling of lipid membranes close to transitions: The effect of electrostriction. *Biophysical Journal, 103*(5), 918–929.

Heimburg, T., & Jackson, A. D. (2005). On soliton propagation in biomembranes and nerves. *Proceedings of the National Academy of Sciences, 102*(28), 9790–9795.

Heimburg, T., & Jackson, A. D. (2006). On the action potential as a propagating density pulse and the role of anesthetics. *Biophysical Reviews and Letters, 2*(1), 57–78.

Heimburg, T., & Jackson, A. D. (2008). Thermodynamics of the nervous impulse. In K. Nag (Ed.), *Structure and dynamics of membranous interfaces* (pp. 317–339). Wiley.

Hodgkin, A. L. (1964). *The conduction of the nervous impulse (Sherrington lecture)*. Charles C. Thomas.

Hodgkin, A. L., & Huxley, A. F. (1945). Resting and action potentials in single nerve fibres. *The Journal of Physiology, 104*(2), 176–195.

Hodgkin, A. L., & Huxley, A. F. (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, A. L., & Huxley, A. F. (1952b). The components of membrane conductance in the giant axon of Loligo. *The Journal of Physiology, 116*, 473–496.

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

Hodgkin, A. L., Huxley, A. F., & Katz, B. (1952). Measurement of current-voltage relations in the membrane of the giant axon of Loligo. *Journal of Physiology, 116*, 424–448.

Holland, L., de Regt, H. W., & Drukarch, B. (2019). Thinking about the nerve impulse: The prospects for the development of a comprehensive account of nerve impulse propagation. *Frontiers in Cellular Neuroscience, 13*(May).

Hughes, R. I. G. (1997). Models and representation. *Philosophy of Science, 64*, 325–448.

Jones, M. R. (2005). Idealization and abstraction: A framework. In M. Thomson-Jones, & N. Cartwright (Eds.), *Idealization XII: Correcting the model* (pp. 173–217). Rodopi.

Kleinzeller, A. (1995). Exploring the cell membrane. In A. Kleinzeller (Ed.), *Comprehensive biochemistry* (Vol. 39, pp. 1–26). Elsevier Science B. V.

Knuuttila, T. (2011). Modelling and representing: An artefactual approach to model-based representation. *Studies in History and Philosophy of Science Part A, 42*(2), 262–271.

Knuuttila, T. (2017). Imagination extended and embedded: Artifactual versus fictional accounts of models. *Synthese*, 1–21. https://doi.org/10.1007/s11229-017-1545-2

Knuuttila, T., & Loettgers, A. (2016). Model templates within and between disciplines; from magnets to gases—and socio-economic systems. *European Journal for Philosophy of Science, 6*(3), 377–400.

Knuuttila, T., & Morgan, M. S. (2019). De-idealization – No easy reversals. *Philosophy of Science, 86*(4), 641–661.

Levy, A. (2018). Idealization and abstraction: Refining the distinction. *Synthese*, 1–18. https://doi.org/10.1007/s11229-018-1721-z, 1999.

Mäki, U. (1992). *On the method of isolation in economics*. Poznan Studies in the Philosophy of the Sciences and the Humanities.

Mäki, U. (2009). MISSing the world: Models as isolations, representations, and credible worlds. *Erkenntnis, 70*, 29–43.

McMullin, E. (1985). Galilean idealization. *Studies in History and Philosophy of Science Part A, 16*, 247–273.

Morrison, M., & Morgan, M. S. (1999). Models as mediating instruments. In M. S. Morgan, & M. Morrison (Eds.), *Models as mediators. Perspectives on natural and social science* (pp. 10–37). Cambridge University Press.

Mosgaard, L. D., Zecchi, K. A., Heimburg, T., & Budvytyte, R. (2015). The effect of the nonlinearity of the response of lipid membranes to voltage perturbations on the interpretation of their electrical properties. A new theoretical description. *Membranes, 5*(4), 495–512.

Nowak, L. (1992). The idealizational approach to science: A survey. In J. Brzeziński, & L. Nowak (Eds.), *Idealization III: Approximation and truth* (pp. 9–66). Rodopi.

Potochnik, A. (2017). *Idealization and the aims of science*. University of Chicago Press.

Rice, C. (2018). Idealized models, holistic distortions, and universality. *Synthese, 195*(6), 2795–2819.

Rice, C. (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.

Strevens, M. (2009). *Depth: An account of scientific explanation*. Harvard University Press.

Suárez, M. (2004). An inferential conception of scientific representation. *Philosophy of Science, 71*, 767–779.

Takashima, S. (1979). Admittance change of squid axon during action potentials: Change in capacitive component due to sodium currents. *Biophysical Journal, 26*(1), 133–142.

Tasaki, I. (1982). *Physiology and electrochemistry nerve fibers*. Academic Press.

Tasaki, I., & Iwasa, K. (1980). Shortening of nerve fibers associated with propagated nerve pulse. *Biochemical and Biophysical Research Communications, 94*(2), 716–720.

Tasaki, I., & Iwasa, K. (1982). Rapid pressure changes and surface displacements in the squid giant axon associated with production of action potentials. *The Japanese Journal of Physiology, 32*, 69–81.

Van Fraassen, B. C. (2008). *Scientific representation: Paradoxes of perspective*. Oxford: Clarendon Press.

Weisberg, M. (2007). Who is a modeler? *The British Journal for the Philosophy of Science, 58*(2), 207–233.

Winther, R. G. (2014). James and dewey on abstraction. *The Pluralist, 9*(2), 1–28.