Scientific Constitutive Abduction

In 1952, Alan Hodgkin and Andrew Huxley argued that fluxes of sodium and potassium ions generate the action potential. Part of their case rested on the idea that such fluxes would explain various features of the currents associated with action potentials. AUTHOR 1 proposed that Hodgkin and Huxley used abductive reasoning, that is, reasoning that infers some hypothesis because of what that hypothesis explains. Magnani, 2001, Aliseda, 2005, Bird, 2005, and Douven, 2017, understand abduction this way. AUTHOR 1 also mentioned features of abduction that support this interpretation of Hodgkin and Huxley's reasoning.

This paper expands on AUTHOR 1, on three fronts. First, we provide a more detailed account of constitutive abduction. More specifically we describe four of its features. Scientific constitutive abduction 1) is sometimes meant to offer some degree of confirmation for the hypothesis in the explanans, 2) may be used to postulate entities that are qualitatively distinct from the entities cited in the supporting evidence, 3) may be used to postulate entities that are not directly empirically detected, and 4) may rely on background beliefs.

Second, we present two additional scientific examples of this of constitutive abduction. We discuss two experiments Edward Tolman proposed to support the hypothesis that rats navigate mazes using cognitive maps rather than stimulus-response strategies. We also review Baumgartner's 1960 argument for the biological basis of the Hermann grid illusion. Baumgartner's example provides a striking illustration of the role of background beliefs in supporting abductive inferences, as Baumgartner's argument rests entirely on background beliefs. He used no new experiments. The point of these additional examples is to show that the account of constitutive abduction that applies to the Hodgkin-Huxley examples is not limited to that example. The account is, one might say, more robust than that.

The first and second fronts help us on the third. New Mechanists have recently invited philosophers of science to give an account of the confirmation of constitutive hypotheses that might serve as an alternative or complement to the "manipulationist" idea that such hypotheses are confirmed by paired combinations of top-down and bottom-up experiments. See Craver, Glennan, & Povich, 2021. Insofar as scientists use constitutive abduction to try to offer some degree of confirmation for the hypothesis in the explanans, they are using it to do the work manipulationist approaches might be supposed to do. Further, since constitutive abduction allows a scientist to postulate entities in the explanans that are qualitatively distinct from those in the explanandum, it follows that constitutive abduction can support hypotheses in which some entities are constitutively explained by qualitatively distinct entities. Thus, the products of both constitutive abduction and manipulationism are, to this degree, the same. Third, because constitutive abduction allows the postulation of entities that are not directly empirically detected, it can be used when experimental techniques do not allow the manipulation or measurement of lower level entities. In other words, constitutive abduction can be used when manipulability approaches cannot. This suggests the conditions under which constitutive abduction provides an alternative to manipulability approaches. Fourth, constitutive abduction captures the view, endorsed by some New Mechanists, that scientific confirmation may rely on background beliefs. Thus, constitutive abduction—as found in the Hodgkin-Huxley, Tolman, and Baumgartner cases and characterized in 1)-4)—answers the New Mechanist invitation for an alternative to manipulability approaches.

1.0 Constitutive Abduction

1.1 A Target Case.

We begin by reviewing an example of the scientific reasoning that is a "target" of our account, namely, the first two of Hodgkin and Huxley's 1952 arguments for the role of sodium in the initial inward current of the action potential. Just to be clear, it should be noted that we are not proposing an account of Hodgkin and Huxley's reasoning in the last of their 1952 papers, the "Quantitative Description" paper.

In their first experiment, Hodgkin and Huxley depolarize an axon by 65mV and measure the change in current across the membrane, first in a sodium-containing medium, then in a sodium-free medium, then again in a sodium-containing medium. (See Figure 1.) In the case where the axon is placed in the sodium-free medium, the initial inward current (represented by the initial hill at the left of the curve) disappears. In reporting on this result, Hodgkin and Huxley do not state the obvious, namely, that what would explain this feature is sodium ions being responsible for the transient inward current seen in the top and bottom panels.

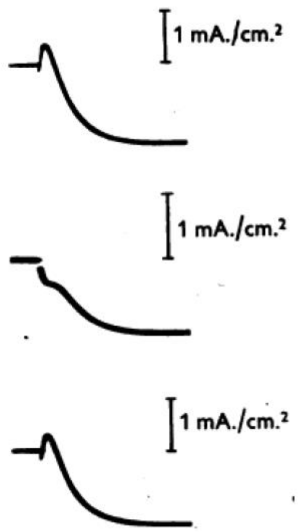

Figure 1. Measurements of current in an axon depolarized by 65mV in a sodium-containing medium (top panel), sodium-free medium (middle panel), and sodium-containing medium (bottom panel).

Consider a philosophical interpretation of Hodgkin and Huxley's experiment. On the abductive approach, the application of the voltage clamp is the manipulation, and the current is what is measured. Each curve in Figure 1 corresponds to a manipulation-measurement pair. Each manipulation-measurement pair invites a mechanistic constitutive why-question, such as, "Why did the axon voltage clamped to -65 mV in the sodium-containing medium produce an initial inward current?" or "Why did the axon voltage clamped to -65 mV in the sodium-free medium produce an initial outward current?" The answers to these questions are given by mechanistic constitutive explanations. Further, each of the answers, each of the constitutive explanations, can be part of a constitutive abduction.

As an aside, note that the axon was "running down" or dying over the course of the experiment, and Hodgkin and Huxley knew it. Thus, strictly speaking, Hodgkin and Huxley believed that there were three distinct explanations corresponding to the three curves of Figure 1:

1. S Ψ-ing generated by x_1 ϕ_1-ing, x_2 ϕ_2-ing, … x_p ϕ_p-ing (sodium),
2. S* Ψ*-ing generated by x_1 ϕ^*_1-ing, x_2 ϕ^*_2-ing, … x_q ϕ^*_q-ing (no sodium),
3. S** Ψ**-ing generated by x_1 ϕ_1-ing, x_2 ϕ_2-ing, … x_r ϕ_r-ing (sodium).

In this notation, S, S*, and S** represents the changes in the axon over the course of the experiment. It is a different individual over time.[1] Ψ-ing represents the initial inward current, Ψ*-ing represents an initial outward current, and Ψ**-ing represents another inward current, differing slightly from Ψ-ing. Further, the notation captures the idea that S Ψ-ing is generated by p sodium ions moving into the axon, whereas S* Ψ*-ing is an outward current generated by q sodium ions moving out of the axon. Finally, the notation captures the idea that S** Ψ**-ing is generated by r (rather than p) sodium ions moving into the axon. In this notation, it is assumed that, as the axon runs down, the sodium ions are constant, differing only in number over the course of the experiment. Each explanation in 1.-3. supports a mechanistic constitutive abduction. Of course, Hodgkin and Huxley "idealized" somewhat and treated 1. and 3. as the same. The intended overarching structure of the first experiment was a controlled experiment.

Consider, now, Hodgkin and Huxley's second experiment. On Hodgkin and Huxley's account, the sodium component of the membrane current is proportional to the driving force for sodium ions across the membrane. This force is determined by the difference between the membrane potential and the sodium equilibrium potential. By moving the membrane potential closer to the sodium equilibrium potential, Hodgkin and Huxley offset

[1] One might propose that over time the same axon is multiply constituted, so that in each case there is just a single S, rather than S, S*, and S**. For present purposes, there is no need to resist this view.

the inward driving force of the sodium ion concentration gradient. This reduces the transient inward flux of sodium ions even to the point of eliminating the influx in favor of an efflux. Hodgkin and Huxley's second experiment bears all this out. They retain the sodium-containing, sodium-free, sodium-containing protocol, but vary the strength of the depolarization. (See Figure 2.) The first and third columns of Figure 2 show that, at voltage steps beyond -15mV, the putative sodium current weakens as the membrane voltage approaches the sodium reversal (-84 in Fig. 2), and then changes sign as it goes past it (-98 and -112 in Fig. 2). Hodgkin and Huxley's implicit assumption is that the influx of sodium ions driven by the sodium ion concentration gradient explains the pattern in the data. As with the first experiment, each curve of Figure 2 implicates a mechanistic constitutive explanation that may serve as the basis for a mechanistic constitutive abduction. And, here again, a full accounting of this experiment must include an account of controlled experiments.

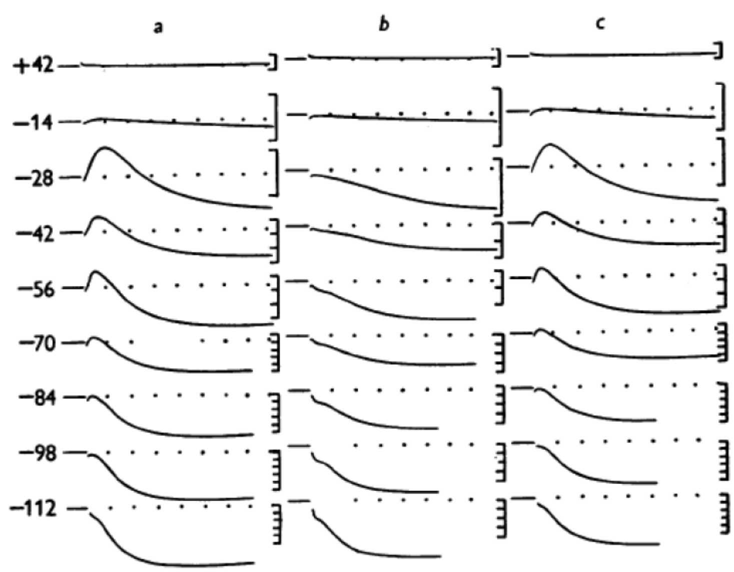

Figure 2. Measurements of current in an axon depolarized to different values in a sodium-containing medium (left panel), sodium-free medium (center panel), and sodium-containing medium (right panel).

Hodgkin and Huxley make the following suggestive comment regarding the middle column of Figure 2:

> When the axon is placed in a sodium-free medium, such as the "choline sea water", there can be no inward flux of sodium, and the sodium current must always be outward. This will account for the early hump on the outward current which is seen at all but the lowest strength of depolarization in the centre column of Fig. 2.
>
> Hodgkin and Huxley, 1952, p. 454.

By "the early hump" Hodgkin and Huxley mean the downturn in the current at the left of the center curves, most pronounced in the depolarizations of -70, -84, -98, and -112 mV. If we take "account for" to be synonymous with "explain," then there is reason to think that Hodgkin and Huxley understand themselves to be doing something like abductively inferring that sodium ion fluxes are responsible for the observed transient inward currents. They conclude that sodium ions are responsible for the inward currents because that would explain the early hump.

1.2 Abduction and Constitutive Abduction

In recent philosophical work, abductive reasoning is widely understood to be a form of defeasible inference to some hypothesis based on what that hypothesis explains. Consider the following descriptions: abduction is "inference to an explanatory hypothesis" Magnani, 2001, p. xxi.; abduction is "reasoning from an observation to its possible explanations," Aliseda, 2005, p. xii; "By 'abductive' inference I shall mean an inference where a central component of that inference is the fact that the inferred (purported) facts provide a putative explanation of the evidence or some part thereof" Bird, 2005, p. 5. So understood, abductive inference differs from deductive inference in being defeasible. Further, abductive inference differs from simple enumerative induction insofar as the former, but not the latter, implicates explanation. Douven, 2017, has a more detailed exposition of this.

Some philosophers distinguish among different types of abductive inference by appeal to the explanation the abductive inferences invoke. Psillos, 2007, and Schurz, 2008 articulate such a view. So, for example, etiological abductions invoke etiological explanations. Just so, we propose that constitutive abductions invoke constitutive explanations. What we propose is distinctive of constitutive abductions, as opposed to, say, etiological abductions, is the reliance on constitutive explanations as opposed to etiological explanations. By a constitutive explanation, we here mean a representation of an asymmetric, many-one, synchronic ontological dependence relation between the activities of many entities (a set of x_i's φ_i-ing, one might write) and the activity of a whole those entities constitute (an S Ψ-ing, one might write). This description is meant to capture Hodgkin and Huxley's explanation of the initial inward current of the action potential (an S Ψ-ing) in terms of the inward flux of sodium ions (a set of x_i's φ_i-ing). The idea that a representation is involved captures the idea that Hodgkin and Huxley came up with this explanation in the sense that they developed a representation of this ontological dependence relation that pre-existed scientific research. The asymmetry of the relation captures the fact that Hodgkin and Huxley proposed to explain the initial inward current in terms of the inward movement of sodium ions, but not the inward movement of the sodium ions in terms of the axonal current. Every feature of this proposal—the role of representation, the asymmetry of the ontological dependence relation, the synchronicity, the existence of an ontological dependence relation distinct from causation, and more—has been debated within the New Mechanist literature. For a small sample of the debate on these issues, see Bechtel & Abrahamsen, 2005, Craver, 2007, 2014, Schindler, 2013, Leuridan, 2012. Although we have defended this account of constitutive explanation in other works, we shall not review that defense here. See, for example, AUTHOR 2. The current work is meant to build on what we have set out before.

This paper is dedicated to a "primary project" of describing one way in which scientists try to confirm mechanistic constitutive hypotheses, such as that sodium carries the initial inward current of the action potential. We shall not, however, engage in any number of "secondary projects" of presenting other philosophical accounts, how they differ from the present account, and what reason(s) there might be to prefer our account. (We feel compelled to make one exception to this rule. We briefly discuss Baumgartner & Casini's "abductive theory of constitution" from Baumgartner & Casini, 2017. Thus, we will have nothing to say about the views of, for example, C. S. Peirce, Gilbert Harman, Peter Lipton, Alexander Bird, Lorenzo Magnani, or John Norton. Limitations of space preclude sufficiently serious attention to any one of these secondary projects. Moreover, it makes sense to develop the primary project, before turning to the secondary one.

Our primary project is to provide an account of how scientists, such as Hodgkin and Huxley, sometimes use experimental results to support constitutive hypotheses, such as that the initial inward current of the action potential is carried by sodium ions. Our account, in brief, is that they used what we have described as constitutive abduction. To put the matter another way, the proposal is that Hodgkin and Huxley, among others, used abduction with the constitutive mechanistic model of explanation.

For the record, we believe that, as framed above, this is a novel proposal. Of course, maybe some philosopher has published something on this idea, or something like it, and we have overlooked it. Nevertheless, that work, if it exists, has yet to make a significant impact.[2] Schurz, 2008, attempts to provide a classification of patterns of abduction that is as complete as possible, but he does not refer to abductive inferences that involve constitutive mechanist hypotheses. That paper has no references to the New Mechanist literature. The recent McCain-Poston anthology on inference to the best explanation has one sentence mentioning mechanistic explanation. (See McCain & Poston, 2017.) Frank Cabrera's recent overview of inference to the best explanation mentions various models of explanation—the deductive-nomological model, the causal model, and the unificationist model—but not the constitutive mechanistic model. (See Cabrera, 2022.) We do not, however, propose to provide further support for this belief, since this would require us to make the case that our account of constitutive abduction differs from Peirce's account of abduction, Lipton's account of inference to the best explanation, and so forth. In other words, making the case for the novelty of our proposal would involve us in numerous secondary projects. Those secondary projects, however, cannot be covered in a single paper.

1.3 Four Features of Scientific Constitutive Abduction

In what follows, we focus on scientific abductive inferences. We, therefore, set aside abductive inference as it may be used in other domains, such as computer science, legal

[2] As of this draft, we know from personal communication that Maria Serban is working on a paper on abduction and the Hodgkin-Huxley model.

reasoning, and philosophical arguments for scientific realism. Constitutive abductive inference in science, we propose, has four features that will figure into our subsequent discussion. (Although we take constitutive abduction to be a novel type of abduction, the novelty does not arise from these four features. These four features are not unique to constitutive abduction.) 1) Scientific constitutive abduction is sometimes meant to offer some degree of confirmation for the hypothesis in the explanans. 2) Scientific constitutive abductive inference may be used to postulate entities that are qualitatively distinct from the entities cited in the supporting evidence. 3) Scientific constitutive abductive inference may be used to postulate entities that are not directly empirically detected. 4) Scientific constitutive abductive inference may rely on background beliefs. We review these points in order, showing how they are illustrated by the Hodgkin-Huxley example. In section 2., we will show how they are illustrated in the Tolman and Baumgartner examples.

In the abduction literature, there is some debate about the extent to which abduction provides confirmation. Some philosophers, such as C. S. Peirce, have proposed that abduction does not confirm a hypothesis but merely introduces some hypothesis that is worthy of further examination. (See, for example, Peirce, CP 8.209, Hanson, 1961, Hoffmann, 1999, Niiniluoto, 1999, Magnani, 2001, Schurz, 2008, Plutynski, 2011, Misak, 2017.) According to Peirce, other methods are required to confirm an abductively suggested hypothesis.

There seem to be three reasons to doubt a concept of abduction along these lines characterizes Hodgkin and Huxley's reasoning. First, they provided three constitutive abductive arguments in support of the hypothesis that the initial inward current was carried by sodium. Assuming that the first argument sufficed to introduce the hypothesis, why would they provide two more arguments to do the same work? A confirmational interpretation of abduction, however, can at least offer an approach to an answer. Each argument further supports the hypothesis. How and why this is plausible merits further attention, but at least we have a start. Second, Hodgkin and Huxley did not stop giving abductive arguments, then switch to some other form of argument to confirm the sodium hypothesis. The arguments were consistently abductive. Third, after reviewing the principal experimental results of their paper, Hodgkin and Huxley comment

> These results support the view that depolarization leads to a rapid increase in permeability which allows sodium ions to move in either direction through the membrane. These movements carry the initial phase of ionic current, which may be inward or outward, according to the difference between the sodium concentration and the electrical potential of the inside and outside of the fibre. (Hodgkin & Huxley, 1952, p. 471).

Suppose we understand "support" to mean confirmation. If this is right, the only results they could take to confirm the sodium hypothesis are those used in the constitutive abductive arguments.[3] So, to recap, there is reason to think that Hodgkin and Huxley use constitutive abduction for confirmation.

Move on now to feature 2). In a simple enumerative inductive inference, a scientist might cite as evidence that copper sample 1 has a melting point of 1085°C, copper sample 2 has a melting point of 1085°C, and copper sample 3 has a melting point of 1085°C. This

[3] Although we have not drawn attention to the point, our proposal is that, as a matter of descriptive fact, scientists sometimes try to confirm hypotheses abductively. We leave open the question of whether, in giving abductive arguments, scientists really do confirm hypotheses. For all we argue in these papers, scientists' efforts may be in vain.

might form the basis of the inference that all copper has a melting point of 1085°C or that the next sample of copper will have a melting point of 1085°C. In this case, the copper and the melting point mentioned in the conclusion are not qualitatively distinct from the copper and the melting point mentioned in the evidence. In the case of Hodgkin and Huxley's experiments, the currents measured in axons are qualitatively distinct from the movements of sodium and potassium across the cell membrane. There is another way to make the point. New Mechanists often propose that an explanandum is an entity, such as an axon, engaged in an activity, such as firing an action potential, while the explanans is a set of entities, such as sodium ions, engaged in their own distinctive activities, such as moving across the neuronal membrane. (See, for example, Craver, 2007.) The point is that, in this case, what appears in the explanandum is distinct from what appears in the explanans. This illustrates the point that scientific constitutive abductive inference may be used to postulate entities that are qualitatively distinct from the entities cited in the supporting evidence.

Consider 3). In scientific constitutive abduction, scientists sometimes postulate entities, not based on direct measurement or detection of them, but because they are explanatory. Sometimes this point is made by historical examples drawn from archeology, geology, or evolutionary biology, but the Hodgkin and Huxley experiment bears it out just as well. Given the technology of their day, Hodgkin and Huxley could measure the overall current across the cell membrane, but they could not track the movements of specific ionic species or individual ions. The best they could do was infer that, during action potentials, certain types of ions were moving according to their concentration gradients and reversal potentials, based on what such movements would explain.

Finally, there is 4). Scientific constitutive abductions may rely on background beliefs. This feature is discussed in Day & Kincaid, 1994, Niiniluoto, 1999, and Schurz, 2008. In a familiar non-scientific example, one might see marks in the snow, from which one concludes that a deer walked through the area. The conclusion might be supported since a deer's walking through the area would explain the marks in the snow. Among the background beliefs that support this conclusion might be the belief that no elk live in the area. Returning to a scientific case, Hodgkin and Huxley's abductive arguments about the movements of sodium and potassium ions relied on background beliefs about the relative concentrations of sodium and potassium ions inside and outside of the cell and their relation to their reversal potentials according to the Nernst equation.

Why are these four features noteworthy? They indicate why constitutive abduction is an apt alternative to manipulability accounts. Both manipulationism and abduction may be taken to offer some degree of confirmation for some compositional hypothesis. Both may be used to relate qualitatively distinct entities. Constitutive abduction, unlike manipulationism, may be used to postulate entities that are not directly empirically detected. Constitutive abduction may rely on background beliefs.

2.0 Two Case Studies of the Scientific Use of Constitutive Abduction

Suppose that a philosopher of science is interested, not just in how one might confirm constitutive hypotheses, but in how scientists actually do this. A natural approach is to look to historical cases. More than this, one does not merely want a single, perhaps outlying, example of a philosophical proposal. Instead, one wants a philosophical account—in this case an account of constitutive abduction—that at least plausibly applies to a range of cases. Thus, we present two more cases of scientific constitutive abduction that share the four features illustrated by the Hodgkin-Huxley example.

2.1 Tolman on Latent Learning

In a rough outline, Tolman, 1948, offers a theoretical interpretation of experiments on rats navigating mazes of various designs. From these experiments, Tolman concluded that the rats navigating the mazes were not learning the maze through appetitively reinforced stimulus-response pairings. Instead, through experience with mazes, the rats built up cognitive maps they could use as might be advantageous to them. As Tolman presents it,

> the central office itself is far more like a map control room than it is like an old-fashioned telephone exchange. The stimuli, which are allowed in, are not connected by just simple one-to-one switches to the outgoing responses. Rather, the incoming impulses are usually worked over and elaborated in the central control room into a tentative, cognitive-like map of the environment. And it is this tentative map, indicating routes and paths and environmental relationships, which finally determines what responses, if any, the animal will finally release. (Tolman, 1948, 192).

An example of a “latent learning” experiment—one of the three types of experiment discussed in Tolman, 1948—will illustrate what figured into Tolman’s thinking. In one such experiment, Group I was the control group in which rats were released in the start box and had to navigate their way to the food reward. (See Figure 3.) Over the course of a week, these control animals learned to move from the start box to the food with few wrong turns. This is illustrated by the solid line steadily decreasing in Figure 4. In Group II, the animals did not receive food at the end of the maze for the first six days. They were, instead, given food two hours after their trial. On the seventh day (highlighted by the “x” above the dashed line in Figure 4), and each day thereafter, they found the food at the end of the maze. In Group III, the animals did not receive food for the first three days (highlighted by the “x” above the dotted line in Figure 4) but then received it thereafter. Commenting on this Figure, Tolman writes,

> It will be observed that the experimental groups as long as they were not finding food did not appear to learn much. (Their error curves did not drop.) But on the days immediately succeeding their first finding of the food their error curves did drop astoundingly. It appeared, in short, that during the non-rewarded trials these animals had been learning much more than they had exhibited. This learning, which did not manifest itself until after the food had been introduced, Blodgett called "latent learning." Interpreting these results anthropomorphically, we would say that as long as the animals were not getting any food at the end of the maze they continued to take their time in going through it—they continued to enter many blinds. Once, however, they knew they were to get food, they demonstrated that during these preceding non-rewarded trials they had learned where many of the blinds were. They had been building up a 'map,' and could utilize the latter as soon as they were motivated to do so. (Tolman, 1948, p. 194).

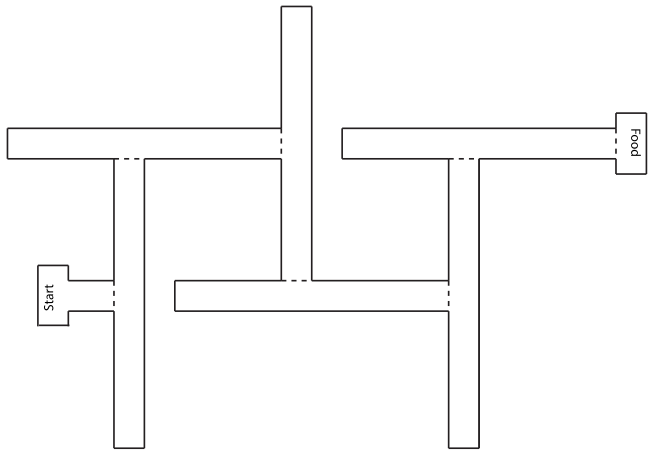

Figure 3. Six-Unit Alley T-Maze, Based on Tolman, 1948, Figure 4.

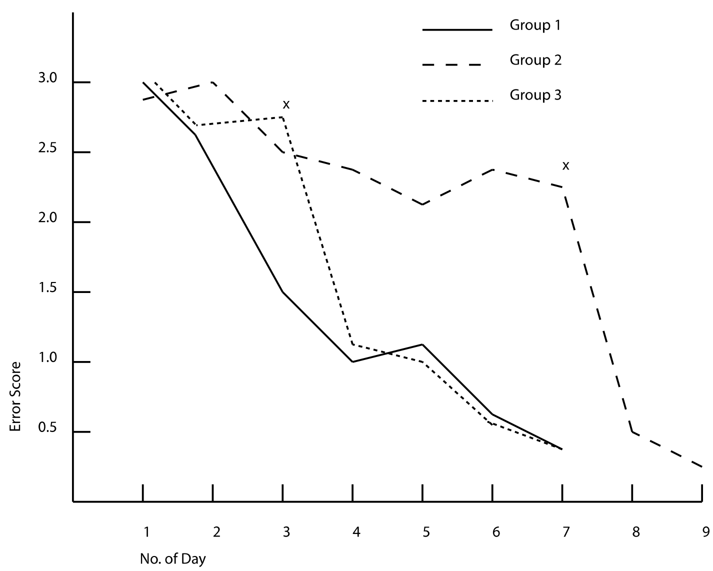

Figure 4. Redrawn from Tolman, 1948, Figure 4, (From H. C. Blodgett, The effect of the introduction of reward upon the maze performance of rats. Univ. Calif. Publ. Psychol., 1929, 4, No. 8, p. 120.)

Tolman and a colleague later performed another version of this experiment with a larger 14-unit maze (as opposed to the six unit maze in Figure 3) and with a longer training regimen. One group received food throughout, another never received food, and one received food for the first time on the eleventh day and every day thereafter. As expected, after the eleventh day, the rats' average errors declined precipitously.

Notice that these experiments involve placing a rat in a maze, allowing it to navigate the maze, and then measuring activities of the rat. In this case, the measure was the number of wrong turns the rat took. The rat navigating the maze looks to be an S ψ-ing and Tolman postulates some structure in the rat's brain that functions as a cognitive map. This looks to be a matter of some unknown entity engaging in an activity of φ-ing. What went on in Tolman's mind connecting the results of this experiment to the idea that the rat has a cognitive map? Tolman does not say in the text discussing the example, but a plausible conjecture is that he was thinking that the activity of a rat's cognitive map (some unknown entity φ-ing) would explain the rat's maze navigation (an S ψ-ing). In other words, a plausible conjecture is that he was using constitutive abduction. Indeed, it has the four features we have previously ascribed to abductive arguments.

To begin with, Tolman takes these experiments to support the cognitive map hypothesis. In introducing the experiments, Tolman comments that "The [experiments], out of many, which I have selected to report are simply ones which seem especially important in reinforcing the theoretical position I have been presenting" (Tolman, 1948, p. 193.) If we understand "reinforcing" as similar in meaning to confirming, then Tolman interprets his experiments to offer confirmation of the cognitive map hypothesis. Note as well that after the Blodgett six-unit experiment, Tolman performed a 14-unit variant of it. He apparently did not take the Blodgett experiment to have introduced a plausible new hypothesis that had to be confirmed by some other method. He replicated the experiment, thereby attempting to provide further confirmation of the cognitive map hypothesis.

The example illustrates other features of our account as well. The explananda and the explanantia are qualitatively distinct. While the rat was engaged in an activity of walking through a maze, turning this way or that at T-junctions, Tolman postulated some structure in the rat brain engaged in an activity of “working over” and “elaborating” inputs into a cognitive-like map. Tolman could not directly detect the relevant brain structure or its activity. All he manipulated or measured in the experiments was the rat or its behavior. Finally, the design and interpretation of one experiment was shaped by background beliefs. Most obviously, the design and interpretation of the second experiment was guided by the results of the first experiment.

2.2 Baumgartner’s Retinal Ganglion Cell Theory of the Hermann Grid Illusion

In 1870, Ludimar Hermann published a brief description of an optical illusion. (Hermann, 1870.) In one version, there are intersecting vertical and horizontal white bars on a black field wherein scintillating gray smudges appear at the intersections somewhat peripheral to the viewer’s fixation point. (See Figure 5.) During the 1950s, neurophysiologists recording the response properties of cells in the optic nerve discovered retinal ganglion cells (RGCs) that were activated by lights shown in their central regions but suppressed by lights shown in a surrounding annulus. (See Kuffler, 1953.) These came to be described as ON-center/OFF-surround cells. In 1960, Baumgartner proposed that these cells can be part of an explanation of the illusion. When an ON-center/OFF surround cell falls at an intersection, there is more OFF-surround stimulation, thereby making for a weaker cell response and a “smudge.” (See the right half of Figure 6.) When an ON-center/OFF-surround cell falls along a line away from an intersection, however, there is less OFF-surround stimulation, thereby making for a stronger cell response and no “smudge.” (See the left half of Figure 6.) Here is a translation from Baumgartner’s German, “The darkening of the light bars at the crossings is then explained by the here more pronounced edge zone inhibition, which reduces the activation of the on-neurons in comparison to the other bar sections.” (Baumgartner, 1960, p. 22)[4]. Baumgartner apparently believed that RGCs form the biological basis of the Hermann grid illusion since the activities of these cells would explain the perception of the illusion.

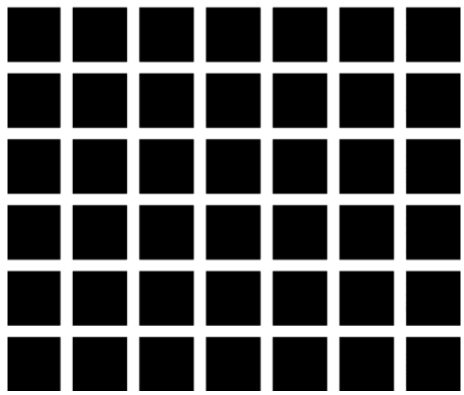

Figure 5. The Hermann Grid Illusion.

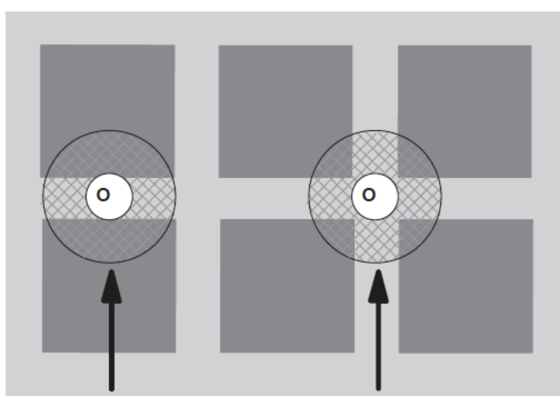

Figure 6. The RGC Explanation of the Hermann Grid illusion.

[4] [FRIEND] kindly checked this translation. For those who are interested, here is the original German: “Die Verdunkelung der Hellbalken an den Kreuzungen erklärt sich danach durch die hier ausgeprägtere Randzonenhemmung, welche die Aktivierung der on-Neurone im Vergleich zu den übrigen Balkenabschnitten vermindert.” (Baumgartner, 1960, p. 22)

An earlier version of this example claimed that the explanandum is the conscious experience of the illusion. A reviewer, however, raised the following objection. "The explanans is the activity of cells with specific receptive fields. The conscious experience arises only after the activity of these cells and only if there is activity in prefrontal regions .... Thus, there does not seem to be a synchronic ontological dependence relation that is required for constitution. Therefore, this may be an example of etiological explanation rather than constitutive explanation." The gist of the point is that the activity of the RGCs takes place in the retina, but it is activities of downstream cortical cells that constitutively explain the perceptual experience. This point invites an extended digression.

This is an apt point, which requires a reformulation of the example. We should understand the explanandum in this example as the perception of the illusion, where the perception of the illusion is the entire process beginning with phototransduction and continuing up until the conscious experience of the illusion. Part of this process is constitutively explained by phototransduction, another part is explained by color opponency processing, and another part, proposed by Baumgartner, is explained by RGC processing.

On its face, this reformulation may seem entirely *ad hoc*, but it is, in truth, necessary to render the example just like the familiar constitutive explanation of a rat's maze navigation. (See, for example, Craver, 2002, Figure 4.) In the rat maze navigation story, the rat might turn its head, which then triggers hippocampal activity. That hippocampal activity can, in turn, trigger muscle movements that take place after the hippocampal activity. In other words, there is neural processing upstream and downstream of the hippocampal processing. So, a standard (but, typically, unarticulated) assumption in the

New Mechanist literature is that an explanandum S Ψ-ing can take place during the interval t_0-t_3, but this can be constitutively explained by x_1 ϕ_1-ing, x_2 ϕ_2-ing, … x_p ϕ_p-ing taking place during the interval t_1-t_2, where t_0 precedes t_1 and t_2 precedes t_3. So, what the reviewer is pointing to in the Hermann grid case is just an instance of this. We need to distinguish between a) an overarching activity of perceiving the illusion which takes place during the interval t_0-t_3 and b) a narrower activity of perceiving the illusion taking place during the interval t_1-t_2. It is the latter that is explained, according to Baumgartner, by RGC activities.

In this point, we can see a further synergy among the examples. In their first experiment, Hodgkin and Huxley voltage clamped the axon to -65 mV and then measured the "total current." Part of their achievement was to show that the total current at any given instant is the sum of a capacity current, a sodium current, a potassium current, and a leak current. In offering "the initial inward current" as an explanandum, Hodgkin and Huxley were gesturing toward part of the overall total current that is constitutively explained by sodium influx. By contrast, in the Hermann grid example, there is as yet no accepted decomposition of the perception of the illusion.

This example Hermann grid example has the four features of abductive inferences we have described. To begin with, Baumgartner does not merely take the RGC theory to be worthy of further investigation. He immediately uses the RGC theory as a basis for investigating another question, namely, the size of receptive fields. He concludes his very brief paper writing, "The size of the foveal receptive fields and the size of the receptive fields of each other retinal area can be calculated from the width and distance of the object. In humans, this determination method in the foveal area yields a diameter of the receptive fields of >25μ." (Baumgartner, 1960, p. 22). Baumgartner does not articulate how he links the RGC theory to his conclusions about receptive fields sizes, but the point remains that he takes the RGC theory to be sufficiently supported that it may serve as the basis for further conclusions.

The perception of the Hermann Grid is qualitatively distinct from the RGCs. The grid is perceived to have scintillating gray smudges; the RGCs have no perception. The perception of the grid involves horizontal and vertical white bars; the RGCs have no perception. The RGCs have radially symmetric receptive fields; the perception does not. To put matters in another way, in offering a constitutive abduction in support of the RGC theory, one appeals to entities and activities that are not found in the explanandum.

One noteworthy feature of Baumgartner's conclusion is that it was not based on any experimental work that he reported in the paper. He did not present a Hermann grid to a subject. Nor did he in any way stimulate a neuron or measure the activity of a neuron. Whether or not Baumgartner had the technical ability to perform an experiment in which he measured the activity of a participant experiencing the grid illusion, he in fact did not do such an experiment.

Instead of performing a new experiment, Baumgartner simply relied on the results of prior experimental work. Although Baumgartner provides no references at all to the scientific literature, enough is known about the relevant history of science to enable us to determine the basic structure of the relevant background work. On the one hand, there was a tradition of creating visual stimuli that had some of the contrast characteristics of the Hermann grid illusion and then showing them to participants and in one way or another measuring their responses. On the other, there was a neurophysiological tradition of probing the receptive fields of neurons by projecting small spots of light onto the retina.

Kuffler, 1953, is the pre-1960 *locus classicus* for this work. It advanced the proposal that retinal ganglion cells have roughly radially symmetric ON-center, OFF-surround receptive fields. Another way of making the point might be to say that Baumgartner drew his conclusion based exclusively on his familiarity with a lot of background information.

4. Scientific Constitutive Abduction and the New Mechanist Invitation

Some New Mechanists have proposed that combinations of top-down and bottom-up experiments are sufficient for establishing constitutive hypotheses. (See, for example, Craver, 2007, Craver & Darden, 2013, Kaplan, 2012, Prychitko, 2021, Craver, Glennan, & Povich, 2021.) More recently they have emphasized that there are other possible methods for confirming constitutive hypotheses. For example, Carl Craver, Stuart Glennan, and Mark Povich claim that "there are many ways to test causal and constitutive claims" beside the "matched interlevel experiments" account they provide. (See, Craver, Glennan, & Povich, 2021, p. 8825. Cf., p. 8824. See, also, Prychitko, 2021, p. 1831.) Further, "once we recognize that [matched interlevel experiments are] but one way to establish constitutive relevance, we open up much needed avenues of research to explore other ways to discover and confirm facts about constitutive relevance" (ibid., p. 8826). These comments invite an alternative to approaches involving mutual manipulability or matched interlevel experiments.

Further, New Mechanists have noted that scientists make use of background beliefs in confirmation. Carl Craver and Lindley Darden, for example, have stated:

> The items that appear in mechanism schemas are not merely defined variables but variables that stand for entities and activities about which we have a great deal of

> background knowledge. Scientists draw on that background knowledge as they construct, evaluate, and revise mechanism schemas. (Craver & Darden, 2013, p. 11).

> [B]iologists typically begin their search for mechanisms in an intellectual cornucopia of general knowledge that restricts the space of possible mechanisms and focuses attention on a few or a handful of key live possibilities. More specifically, a discovery episode plays out against many kinds of background knowledge (ibid., p. 69,)

Craver, Glennan, and Povich add that

> Meaningful experimentation (with useful interventions and detections) can take place only against a wealth of background knowledge about the active organization of the system under study. (ibid., p. 125 See also Craver, Glennan, & Povich, 2021, p. 8808).

Putting these observations together with those of the last paragraph, the New Mechanists have, in essence, invited the articulation of a non-manipulationist account of the confirmation of constitutive hypotheses that assigns a role to background conditions. It should be clear that the account of constitutive abduction presented in section 1 and 2 fits this bill.

Although some New Mechanists are, in theory, open to a constitutive abductive account, New Mechanists rarely refer to abduction, even if we include "inference to the best explanation" as another term for abduction. In a bibliographic discussion, Craver and Darden devote two paragraphs to abduction wherein they present N. R. Hanson's account of retroduction. See Craver & Darden, 2013, p. 81, and Darden, 2017, p. 255. Although their brief comments are not explicit about the matter, the intent may be to treat retroduction, as Peirce treated abduction, namely, as a rational means for coming up with a new hypothesis that might subsequently be confirmed using other methods. There are also passing comments about inference to the best explanation in Craver & Kaplan, 2020, and Krickel, 2018.

The most substantive abductive proposal in the New Mechanist literature is Michael Baumgartner and Lorenzo Casini's "Abductive Theory of Constitution." Although there is an elaborate formal development of the theory and its motivation, the core idea is simple. What is to be explained, on their account, is why some phenomenon and some of its spatiotemporal parts are unbreakably coupled. Roughly speaking, the explanatory question is, "Why is it that in all these experiments a change to S ψ-ing brings about a change to x_i's φ_i-ing, or vice versa?" The explanandum here is not an entity engaged in an activity, so the explanation is not a constitutive explanation. So, this is not what we mean by a constitutive abduction. What we mean by a constitutive abduction is an abductive inference that involves a constitutive explanation. This is the principal point where the comments from section 1 come to bear. There is another, perhaps simpler and clearer, way of drawing a contrast between the Baumgartner-Casini proposal and our constitutive abduction proposal. For Baumgartner and Casini what is to be explained involves the results of many experiments, whereas in our proposal what is to be explained involves the results of just a single experiment. There is undoubtedly much more that one might say about the Baumgartner-Casini theory, but the principal point should be clear. The constitutive abductive proposal developed here differs from the Baumgartner-Casini proposal.

We noted above that we take the constitutive abductive proposal to be novel. We have not tried to support that belief over and above observing that no one has explicitly articulated such a proposal. This leaves open questions regarding how our proposal relates to a multitude of alternatives to be found in the literature. These, however, we take to be secondary projects, each of which may be quite involved in its own right. Many of them would involve detailed exposition of a rival accounts of abduction, and then explaining how that account differs from our proposal. One might then want some reason to believe that our account is descriptively superior to the rival account. Clearly, limitations of space preclude such efforts here. We have, however, made one exception to this rule. We have made the case that the constitutive abductive proposal differs from the Baumgartner-Casini proposal.

5. Conclusion

We propose that scientists sometimes make a case for a constitutive hypothesis by constitutive abduction. We have presented three sets of examples: Two of the three arguments Hodgkin and Huxley gave in support of the sodium hypothesis, one of Tolman's arguments for the cognitive map hypothesis, and Baumgartner's argument for the reginal ganglion cell theory of the Hermann grid illusion. Each of these arguments, we proposed, has four features. Contrary to Peircean claims, scientific constitutive abduction is sometimes meant to offer some degree of confirmation for the hypothesis in the explanans. Scientific constitutive abduction may be used to postulate entities that are qualitatively distinct from the entities cited in the supporting evidence and that are not directly empirically detected. Finally, scientific constitutive abduction may rely on background beliefs. These four features make scientific constitutive abduction an apt complement to manipulability approaches. Constitutive abduction has the same goal: confirmation. Like manipulability approaches, it has produces hypotheses that may relate qualitatively distinct entities. It applies where manipulability approaches do not, as when lower level interventions are technically unavailable. Finally, constitutive abduction respects the widespread belief that confirmation depends on background beliefs.

There are, no doubt, many issues that remain to be addressed in the development of an account of scientific constitutive abduction. Many, but not all, of these issues take the form of secondary projects in which one must provide a detailed account of rival theories of the confirmation of constitutive hypotheses, how they compare to a constitutive abductive account, and what reasons there are for preferring the constitutive abductive account. Nevertheless, we believe that the outlines presented here, along with the examples the account appears to accurately describe, provide a solid reason for philosophers to consider the constitutive abductive proposal in more detail.

The author(s) have no conflicts of interest.

References

AUTHOR 1

AUTHOR 2

Aliseda, A. (2006). *Abductive Reasoning* (Vol. 330): Springer.

Baumgartner, G. (1960). Indirekte Grössenbestimmung der rezeptiven Felder der Retina beim Menschen mittels der Hermannschen Gittertäuschung. *Pflüger's Archiv für die gesamte Physiologie des Menschen und der Tiere, 272*(1), 21-22.

Baumgartner, M., & Casini, L. (2017). An abductive theory of constitution. *Philosophy of Science*, *84*(2), 214-233.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 36*(2), 421-441.

Bird, A. (2005). Abductive knowledge and Holmesian inference. In T. S. Gendler & J. Hawthorne (Eds.), *Oxford Studies in Epistemology* (Vol. 1, pp. 1-31). Oxford: Oxford University Press.

Blodgett, H. C. (1929). The effect of the introduction of reward upon the maze performance of rats. *University of California Publications in Psychology, 4*(8), 113-134.

Cabrera, F. (2022). Inference to the Best Explanation: An Overview. *Handbook of Abductive Cognition*, 1-34.

Carey, S. (2009). *The Origin of Concepts*. Oxford: Oxford University Press.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT press.

Craver, C. (2007). *Explaining the Brain*. Oxford: Oxford University Press, USA.

Craver, C. F., & Darden, L. (2013). *In Search of Mechanisms: Discoveries across the Life Sciences*. Chicago, IL: University of Chicago Press.

Craver, C. F., Glennan, S., & Povich, M. (2021). Constitutive relevance and mutual manipulability revisited. *Synthese*, 1-22.

Darden, L. (2017). Strategies for discovering mechanisms. In *The Routledge Handbook of Mechanisms and Mechanical Philosophy* (pp. 255-266). New York: Routledge.

Day, T., & Kincaid, H. (1994). Putting inference to the best explanation in its place. *Synthese, 98*(2), 271-295.

Douven, I. (2017). Inference to the best explanation: What is it? And why should we care. In K. McCain & T. Poston (Eds.), *Best Explanations: New Essays on Inference to the Best Explanation*. (pp. 4-22).

Gillett, C. (2013). Constitution, and multiple constitution, in the sciences: Using the neuron to construct a starting framework. *Minds and Machines*, *23*, 1-29.

Glennan, S. (2017). *The New Mechanical Philosophy*: Oxford University Press.

Glennan, S. (2021). Corporeal composition. *Synthese*, *198*(12), 11439-11462.

Hanson, N. R. (1961). "Is There a Logic of Discovery?" In H. Feigl and G. Maxwell (eds.), *Current Issues in the Philosophy of Science*. New York: Holt, Rinehart, and Winston, (pp. 20-35).

Hoffmann, M. (1999). Problems with Peirce's concept of abduction. *Foundations of Science, 4*(3), 271-305.

Hodgkin, A. L. (1951). The ionic basis of electrical activity in nerve and muscle. *Biological Reviews, 26*(4), 339-409.

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

Hodgkin, A. L., & Huxley, A. F. (1952). Currents carried by sodium and potassium ions through the membrane of the giant axon of Loligo. *The Journal of Physiology, 116*(4), 449-472.

Illari, P. M., & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science, 2*(1), 119-135.

Kaplan, D. M. (2012). How to demarcate the boundaries of cognition. *Biology & Philosophy, 27*(4), 545-570.

Keynes, R. D. (1948). The leakage of radioactive potassium from stimulated nerve. *Journal of Physiology, 107*, 35P-36P.

Keynes, R. (1949). The movements of radioactive ions in resting and stimulated nerve. *Archives des Sciences Physiologiques, 3*(2), 165-175.

Krickel, B. (2018). *The Mechanical World: The Metaphysical Commitments of the New Mechanist Approach*: Springer.

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology, 16*(1), 37-68.

Leuridan, B. (2011). Three problems for the mutual manipulability account of constitutive relevance in mechanisms. *The British Journal for the Philosophy of Science*, *63*(2), 399-427.

McCain, K., & Poston, T. (2017). *Best explanations: New essays on Inference to the Best Explanation*. Oxford University Press.

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

Magnani, L. (2011). *Abduction, Reason and Science: Processes of Discovery and Explanation*: Springer Science & Business Media.

Miller, G. (1956). Human memory and the storage of information. *IRE Transactions on Information Theory, 2*(3), 129-137.

Misak, C. (2017). Peirce and Ramsey. In K. McCain & T. Poston (Eds.), *Best Explanations: New Essays on Inference to the Best Explanation.* Oxford: Oxford University Press. (pp. 25-38).

Niiniluoto, I. (1999). Defending abduction. *Philosophy of Science, 66*, S436-S451.

Norton, J. D. (2021). *The Material Theory of Induction*. University of Calgary Press.

Peirce, C.S. (CP) *Collected Papers of Charles Sanders Peirce* (Volumes I–VI, ed. by Charles Hartshorne and Paul Weiss, 1931–1935, Volumes VII–VIII, ed. by Arthur W. Burks, 1958). Harvard University Press: Cambridge, MA.

Plutynski, A. (2011). Four problems of abduction: A brief history. *HOPOS: The Journal of the International Society for the History of Philosophy of Science*, 1(2): 227-248.

Psillos, S. (2007). The fine structure of inference to the best explanation. *Philosophy and Phenomenological Research*, *74*(2), 441-448.

Prychitko, E. (2021). The causal situationist account of constitutive relevance. *Synthese, 198*(2), 1829-1843.

Rothenberg, M. A. (1950). Studies on permeability in relation to nerve function II. Ionic movements across axonal membranes. *Biochimica et Biophysica Acta, 4*, 96-114.

Schiller, P. H., & Carvey, C. E. (2005). The Hermann grid illusion revisited. *Perception, 34*(11), 1375-1397.

Schurz, G. (2008). Patterns of abduction. *Synthese, 164*(2), 201-234.

Spillmann, L. (1971). Foveal perceptive fields in the human visual system measured with simultaneous contrast in grids and bars. *Pflügers Archiv - European Journal of Physiology, 326*(4), 281-299.

Tolman, E. C. (1948). Cognitive maps in rats and men. *Psychological Review, 55*(4), 189-208.