CHAPTER 10

Manipulationist New Mechanism

The Peircean and IBE traditions have had little interest in theories of explanation. The New Mechanists, for their part, focus on a theory of explanation, but have had only a limited interest in abduction or IBE. Rather than considering an abductive/IBE approach to confirmation, some New Mechanists – the manipulationist New Mechanists (MNM) – have proposed a "manipulationist" approach. Here, I will contrast my account of abductive confirmation with the latest iteration of this approach, namely, Carl Craver, Stuart Glennan, and Mark Povich's "matched interlevel experiments" (MIE) account.

In rough outline, this chapter has the following organization. MIE is meant to describe actual scientific practice, or it is not. If it is, then there are features of scientific practice that it often fails to capture. This opens the door to an abductive approach that captures those features. If MIE is not meant to describe actual scientific practice, then again this opens the door to an abductive approach that captures those features. Thus, for my purposes, it is immaterial whether MIE is meant to describe actual scientific practice or not.

Section 10.1 will provide a barebones introduction to the MIE account of interlevel experiments along with an imagined example. Section 10.2 will describe two alternatives to the MIE interpretation of the imagined interlevel experiments: HD confirmation and abductive confirmation. This section also further develops the idea of singular interlevel abduction as a contrast with singular compositional abduction. Section 10.3 will move from the imagined example to two actual instances of the scientific use of interlevel experiments. One is borrowed from Kaplan (2012), and the other from Prychitko (2021). One might think that these instances illustrate MIE in action in science, but in point of fact the scientists instead relied upon a combination of interlevel and intralevel experiments. The MIE approach does not allow this, whereas the abductive approach does. Section 10.4 will offer a diagnosis of why the manipulationist approach is

unsuited to addressing the combination of interlevel and intralevel experiments found in actual practice. It will argue that there is no easy fix for MIE. Sections 10.3 and 10.4 examine MIE as a description of how scientists sometimes confirm compositional hypotheses. In contrast, Section 10.5 examines various proposals that align in claiming that MIE is not meant to describe actual scientific practice at all. For what it is worth, these strike me as dubious interpretations of GC&P's goals for MIE, but my official view here is noncommittal. What matters to me is that these approaches leave open the door to an abductive approach to the scientific confirmation of compositional hypotheses.

I might note, at the outset, that my presentation will be unlike the standard fare in the MNM literature. First of all, I will largely stick with my own terminology, rather than that offered by the New Mechanists. I do not think the differences make a difference, so there is no need for a reformulation. Second, although I will refer to Craver's earlier works at multiple points, I will not provide a serious account of the historical development of MIE. I will not review Craver's initial description of interlevel experiments in Craver (2002), nor the "mutual manipulability" version of the approach in Craver (2007), nor the subsequent critique of mutual manipulability centering on Baumgartner and Gebharter (2016). There is already an abundance of writing on that.[1] Finally, I will not be concerned with the details of ideal interventions. There is a lot of literature on this, most of which is irrelevant to my comparison of the MIE approach with my abductive approach.

10.1 Matched Interlevel Experiments

My version of a science-first philosophy of science began with the idea that the sodium hypothesis is a compositional hypothesis that explains the initial inward current in a voltage-clamped axon. I then examined how Hodgkin and Huxley supported this hypothesis. A central observation of this case is that they manipulated an axon by voltage clamping it, then measured the current in that axon, and then postulated an unmeasured flux of sodium ions. They performed an intralevel experiment. Moreover, they postulated the flux of sodium because that would compositionally explain the initial inward current. Reflecting on this case, one can see the affinities between it and the interpretation of any number of other experiments in the cognitive science literature. One can present a native speaker

[1] See, for example, Harinen (2018), Krickel (2018b), and Prychitko (2021).

https://doi.org/10.1017/9781009435710.011 Published online by Cambridge University Press

of a language with a string of words and measure some response, and then venture some hypothesis regarding a putative mental grammar that does, or does not, generate that string. One can have an infant observe some sequence of events and measure some response, and then venture some hypothesis regarding the infant's cognitive endowment.

With my science-first approach – with my philosophical preconception – I have found scientists using intralevel experiments to confirm compositional hypotheses. However, CG&P have a different preconception. CG&P propose that "philosophers should be guided in their thinking about constitutive relevance by attending to the *interlevel experiments* scientists use to test whether something is a component in a mechanism" (Craver et al., 2021, p. 8809, italics in original). Maybe this is a reasonable proposal, but it clearly has a limitation. By taking interlevel experiments as a guide, the CG&P approach overlooks the intralevel cases I have been writing about. In this chapter, I take up the CG&P approach and examine how scientists interpret interlevel experiments. Rather than adopt the MIE account of these experiments, I outline a singular interlevel abductive account.

Interlevel experiments might be divided into two types: top–down and bottom–up. In a top–down experiment, one manipulates an activity instance of some individual, an S Ψ-ing, and then measures an activity instance of one of its associated parts, an x_i φ_i-ing. In a bottom–up experiment, one manipulates an activity instance of some individual, an x_i φ_i-ing, and then measures an activity instance of a whole of which it is a part, an S Ψ-ing. Bottom–up experiments may also be subdivided into two types: those in which activating x_i φ_i-ing activates S Ψ-ing and those in which inhibiting x_i φ_i-ing inhibits S Ψ-ing.

Imagine a set of experiments based on an example from Craver et al. (2021). I propose these as imagined experiments, since CG&P do not cite actual experiments from the scientific literature. Let S Ψ-ing be an instance of a worm's turning when touched and let x_i φ_i-ing be an instance of the firing of an ALML neuron. Let Ψ_{in} be a touching of the worm's head and let Ψ_{out} be the worm's turning. Here are the three experiments:

1) There is a bottom–up experiment in which the worm's head is touched while the firing of an ALML neuron is inhibited, and then the worm's turning is inhibited. (This is a bottom–up inhibition experiment.)
2) There is a bottom–up experiment in which an ALML neuron is stimulated to fire and then the worm's turning is activated. (This is a bottom–up activation experiment.)

https://doi.org/10.1017/9781009435710.011 Published online by Cambridge University Press

3) There is a top–down experiment in which the worm's head is touched and then the ALML neuron is found to fire and the worm's turning is activated. (This is a top–down activation experiment.)

Here, there is an imagined set of experiments, an instance of an imagined experimental practice.

How should philosophers of science interpret this imagined practice? What theory should be offered of it? CG&P propose MIE:

(**MIE**) To establish that an entity X and its activity φ are constitutively relevant to a mechanism that Ψs, the following experimental results and matching condition are jointly sufficient:

(CR1i) If an experiment initiates conditions Ψ_{in} while a bottom–up intervention, **I**, prevents or inhibits X's φ-ing, alterations to or prevention of Ψ's terminal conditions, Ψ_{out}, are detected.

(CR1e) If a bottom–up intervention, **I**, stimulates X's φ-ing, Ψ's terminal conditions, Ψ_{out}, are detected.

(CR2*) If a top–down experiment initiates conditions Ψ_{in} and detects Ψ's terminal conditions Ψ_{out}, X's φ-ing is also detected.

(Matching) The activities φ_i activated or inhibited in bottom–up experiments (CR1i and CR1e) must be of the same kind as, and occur within quantitatively overlapping ranges with, the activities φ_i detected in top–down experiments (CR2*). (Craver et al., 2021, p. 8822)

It is reasonable to read MIE as mere joint sufficiency conditions for establishing that an x_i φ_i-ing implements an S Ψ-ing, while being entirely noncommittal about whether scientists ever use MIE. That is how the conditions are formulated. Further, it would also explain why CG&P never refer to scientific journal articles that describe scientists performing the experiments described in MIE. This would fit the description of the experiments as imagined experiments.

Then again, looking at the broader context of the CG&P paper, there are reasons to think that the conditions of MIE must be descriptive of actual scientific practice. In the Preface to Craver (2007), Craver wrote,

> My project is both descriptive and normative. My descriptive goal is to characterize the mechanistic explanations in contemporary neuroscience and the standards by which neuroscientists evaluate them. This cannot be accomplished without attention to the details of actual neuroscience. I illustrate my descriptive claims with case studies from the recent history of neuroscience. . . . For philosophers, I limit myself to the details required to demonstrate that the view corresponds to real neuroscience. (Craver, 2007, p. vii).

https://doi.org/10.1017/9781009435710.011 Published online by Cambridge University Press

I am uncertain how one reconciles these claims with the idea that MIE offers mere joint sufficiency conditions for establishing compositional hypotheses. Maybe Craver intended the theory of explanation to be descriptive and normative, but the conditions of MIE to be neither. They are mere joint sufficiency conditions. Or maybe the theory of explanation is supposed to be descriptive and normative, but the conditions of MIE are meant to be merely prescriptive. Or maybe Craver thinks that MIE provides a set of jointly sufficient conditions *that scientists actually use*.

Fortunately for me, my project does not depend on how one interprets Craver's comments. My goal is to understand how scientists try to confirm compositional hypotheses in the primary experimental literature. For that goal, it does not matter what goals CG&P have for MIE. I face MIE using a constructive dilemma. Either MIE is meant to describe actual practice, or it is not. If it is meant to describe actual practice, there are features of that practice that it does not capture, namely, the role of intralevel experiments. This invites an approach that does assign a role to intralevel experiments. An abductive approach can handle these cases. If MIE is not meant to describe actual scientific practice, then, again, that leaves open the door for an abductive approach that does.

10.2 Alternatives to MIE

A first step in thinking clearly about MIE is to realize that there are alternative philosophical interpretations of the imagined experimental practice. A logical empiricist might propose that the practice be interpreted as instances of HD-confirmation. Each of the experiments could be interpretated as the testing of the hypothesis that the firing of ALML neurons implements that worm's turning behavior. One might give more complicated formulations, but for simplicity I will limit myself to the Hempelian version.

Experiment 1):

If the firing of ALML neurons implements the worm's turning behavior and the worm's head is touched and the ALML firing is inhibited, then the worm's turning is inhibited.

The worm's turning is inhibited.

= = =

Therefore, the firing of ALML neurons implements the worm's turning behavior and the worm's head is touched and the ALML firing is inhibited.

https://doi.org/10.1017/9781009435710.011 Published online by Cambridge University Press

Therefore, the firing of ALML neurons implements the worm's turning behavior.

Experiment 2):

If the firing of ALML neurons implements the worm's turning behavior, then the worm turns when the ALML neuron fires.
The worm turns when the ALML neuron fires.
= = =
The firing of ALML neurons implements the worm's turning behavior.

Experiment 3):

If the firing of ALML neurons implements the worm's turning behavior, then when the worm's head is touched and the ALML neuron fires, the worm's turning is activated.
When the worm's head is touched and the ALML neuron fires, the worm's turning is activated.
= = =
The firing of ALML neurons implements the worm's turning behavior.

The HD confirmation interpretation understands each experiment as essentially backing an instance of affirming the consequent. Thus, on this account, each experiment provides a measure of confirmation of the compositional hypothesis.

The point in reviewing the HD account is not to recommend it, but to make a conceptual point. There are theoretical rivals to MIE. The HD account also serves as a springboard to my abductive account. In Chapter 4, I proposed an abductive approach as a successor to HD confirmation for intralevel experiments. In Chapter 7, I showed how the abductive approach contrasts with Peirce's version of HD confirmation. Here, I will suggest how to develop a successor to HD confirmation for interlevel experiments.

To abductively confirm the compositional hypothesis that an activity instance of a firing of an ALML neuron implements an activity instance of a worm's turning, one needs this compositional hypothesis to be included in an explanans. Further, to tie the abductive inference to interlevel experimental results, one needs the experiment to generate an explanandum for the hypothesis to explain. With these assumptions in place, let us return to the three types of imagined experiments.

The first type of experiment is a bottom–up inhibition experiment wherein the worm's head is touched while the firing of an ALML neuron is inhibited, and then the worm's turning is inhibited. This experiment

raises the explanation-seeking why-question, "Why did the inhibition of the ALML neuron firing after the head touch inhibit the worm's turning?" Answer: The firing of the ALML neuron implements the worm's turning, so inhibiting the firing inhibits the turning. Further, insofar as this explanation is used to support the compositional hypothesis that the firing implements the turning, there is an abductive inference.

What sets explanations like this apart from the compositional explanations of Chapters 2 and 3 is that the explanandum is interlevel, rather than intralevel. What is to be explained is a co-occurrence of an instance of a lower level inhibition of an ALML neuron and an instance of an upper level behavior of the worm, whereas in the compositional explanations considered in Chapters 2 and 3 what is to be explained is, say, an instance of a rat navigating a maze or an axon manifesting a current. Just to have a name, call these "singular interlevel explanations" and the corresponding abductive inferences "singular interlevel abductions."

It is, of course, possible that something else explains why the inhibition of the ALML neuron firing inhibits the worm's turning, but that is just a reminder that abductive inferences are defeasible and that they are often defeated by rival explanatory hypotheses. Peirce noted this feature of abduction and Harman made it a key element in his proposal that warranted abductive inference is IBE. The New Mechanists have often drawn attention to this by saying that bottom–up experiments alone are not sufficient to establish compositional relations.[2]

The foregoing sketch of a bottom–up interlevel abductive account sets out the pieces for the interlevel abductive approach. This sketch can straightforwardly extend to the other interlevel experiments. The second type of experiment is a bottom–up excitatory experiment wherein an ALML neuron fires and the worm turns. This experiment raises the explanation-seeking why-question, "Why did the firing of the ALML neuron co-occur with the worm's turning?" Answer: The firing of the ALML implements the worm's turning, so when the neuron fires, the worm turns. Insofar as this explanation is used to support the compositional hypothesis that the inhibition implements the turning, it is used in an abductive inference.

As before, there are other possible explanations, but that does not undermine the abductive approach. Indeed, as just noted above, a central issue in abductive confirmation is the exclusion of rival explanations. Here again, the explanation is not one of the familiar types of compositional

[2] See, for example, Craver (2002, p. S92f).

explanation, since the explanandum is a putative interlevel fact. So, again, it is an abductive account that is not a singular compositional abductive account. It is a singular interlevel abduction.

Finally, there is a top–down activation experiment, wherein the worm's head is touched and then the ALML neuron fires and the worm turns. This experiment raises the explanation-seeking why-question, "Why was touching the worm's head followed by the ALML neuron firing and the worm turning?" Answer: The firing of the ALML implements the worm's turning, so that when the head is touched the neuron fires and the worm turns. The rest of the abductive story applies as before. This gives us an opening into an abductive approach to interlevel experiments.

10.3 Understanding Actual Interlevel Experiments

An ambitious critic of the MIE account might try to show that it has some fatal flaw that prevents it from accounting for any actual scientific practices. I am not such an ambitious critic. I am not even a critic in the sense that I will not argue that the account is mistaken. As far as I am concerned, there may exist cases of actual scientific reasoning in which the MIE account applies. Or MIE might have some other virtue, aside from characterizing actual scientific practice. My more modest goal is to draw attention to a feature of MIE that apparently limits the range of cases to which MIE is applicable: MIE has no role for intralevel experiments. There is no clause in MIE that refers to intralevel experiments. Indeed, insofar as the conditions for MIE are supposed to be jointly sufficient for establishing a compositional relation, this implies that the use of intralevel experiments is otiose. As against that, scientists sometimes combine interlevel experiments with intralevel experiments to confirm compositional hypotheses. Let me review two illustrations.

10.3.1 The Experience of "Flutter"

Consider a set of experiments originally noted by Kaplan (2012) to illustrate the scientific use of the Craver (2007) mutual manipulability account.[3] In these experiments, let S Ψ-ing be an experience of tactile

[3] Kaplan seems to think that the manipulability approach is used by scientists: "To motivate the mutual manipulability account, consider how neuroscientists typically go about determining components in neural mechanisms underlying cognitive capacities" (Kaplan, 2012, p. 557). Of course, CG&P may, or may not, agree with Kaplan's assessment.

https://doi.org/10.1017/9781009435710.011 Published online by Cambridge University Press

"flutter." This is an experience – an activity instance – that is experimentally produced by the application of rapid sinusoidal indentations to the skin of the hand. Importantly, we should distinguish between the experience of flutter and the process of perceiving the flutter. We first touched on this experience/perception distinction in Chapter 2. The perception might begin with, say, compression of receptor cells in the hand, then continue to some point in the cortex; the experience, however, is a cortical event. In this case, the scientific focus is on the experience of flutter, rather than the perception of flutter. Further, we should also draw a contrast between perceptual behavior and perceptual experience.[4] The behavior will be triggered by the movements of a probe on the surface of the skin and may end with a bodily response, such as a verbal report or key press. By contrast, the experience begins a short time after the impression on the hand – after nerve impulses travel from the hand to the brain – and may, or may not, end before the muscular movements of the verbal report or key press. The experience is a cortical event that is typically triggered by the movement of the probe and that, in turn, may trigger a verbal report or key press. We might relate experience, perception, and perceptual behavior by saying that a behavior is the most temporally extended activity instance, the perception the next most temporally extended activity instance, and the experience the least temporally extended activity instance.

In one experiment from Talbot et al. (1968), scientists stimulated the skin of monkey participants. Let this stimulation be a Ψ_{in}. Let the participant's experience of flutter be the S Ψ-ing. The scientists then recorded activity instances in a subset of S1 neurons – so-called quickly adapting (QA) cells – that responded with action potential trains that matched the frequency of the vibrating stimulus. Let the firings of these QA neurons be the x_i's φ_i-ing. Thus, this was a top–down activation experiment. Scientists concluded that instances of the firing of QA neurons implement instances of the experience of flutter.

In a second experiment cited by Kaplan, Romo et al. (1998) used microstimulation in area 3b of S1 to induce an experience of flutter in monkeys. Reading the details of the experiment closely, the stimulus was not applied to single QA-cells, but to parts of columns:

> As S1 is organized into modules of neurons with similar functional properties (columns), we aimed to drive a cluster of 3b neurons—mostly of the

[4] This distinction is often ignored or denied in the embodied cognition literature.

https://doi.org/10.1017/9781009435710.011 Published online by Cambridge University Press

> QA type, hopefully—in a way that matched the periodic responses seen with mechanical stimuli. (Romo et al., 1998, p. 388)

Here, the x_i's φ_i-ing was an activity instance of a column. The response in this case was for the monkey to press one of two buttons indicating which of two stimuli had a higher frequency.

Although Kaplan does not discuss the introductory section of Talbot et al. (1968), the section is important. Talbot et al. begin by announcing that their paper seeks to combine two experimental designs: single-unit analysis and psychophysics. Single-unit analysis involves peripheral stimulation followed by measurements of single units or cells. As Talbot et al. put it: "Electrophysiological studies, particularly with the method of single-unit analysis, can now provide precise measures of the neural encoding in first-order nerve fibers of the parameters of peripheral stimuli, and of the successive relay and transformation of that neural replication from periphery to cerebral cortex" (Talbot et al., 1968, p. 301). One of the limitations of these methods, however, is that they "have so far provided little understanding of those cerebral mechanisms which, operating upon the transformed replication of the peripheral event in the primary receiving areas of the cerebral cortex, are thought to lead to subjective sensory experience and its overt behavioral counterparts" (Talbot et al., 1968, p. 301). Note that Talbot et al. here embrace the experience/behavior distinction: "Psychophysical studies ... seek to establish lawful relations between those experiences and certain physical aspects of the stimuli which evoke them" (Talbot et al., 1968, p. 301). As repeatedly noted, psychophysical experiments are intralevel experiments.[5] There were many illustrations of this in the case study of the biological basis of the Hermann grid illusion. This last comment is especially important. A major implication of Talbot et al.'s introduction is that they take a combination of intralevel and interlevel experiments to be more revealing than top–down experiments alone. Those who are guided by the idea that scientists use interlevel experiments to confirm compositional hypotheses risk overlooking this contention.

MIE clearly has no clause referring to a role for intralevel experiments. Further, by formulating MIE as a set of jointly sufficient conditions, the

[5] In an intralevel experiment from Talbot et al. (1968) that Kaplan does not discuss, the scientists stimulated the skin of human subjects and measured the subjects' reports of what they perceived. Let these reports be Ψ^*_{out}'s, to distinguish them from the monkey key presses Ψ_{out}. In ignoring this example, Kaplan again risks cherry picking his examples, selecting ones that are favorable to his case, but excluding those that are not.

implication is that the use of intralevel experiments is unmotivated. The very concept of "joint sufficiency" implies that the conditions are sufficient without the addition of anything more. This interpretation might be reinforced by comments from Prychitko (2021), which presents a precursor to MIE. Prychitko argues that top–down experiments are not sufficient to establish "constitutive relevance" (more is needed?) and bottom–up experiments are not sufficient to establish "constitutive relevance" (more is needed?), but the combination of top–down and bottom–up experiments is sufficient to establish constitutive relevance (nothing more is needed?).[6] But, if the combination of top–down and bottom–up experiments is jointly sufficient and nothing more is needed, then why did Talbot et al. use the intralevel experiment? The MIE account assigns no role to interlevel experiments, and the "joint sufficiency" reasoning suggests that an intralevel experiment is unnecessary. The implication here is that scientists must be using something other than MIE in this case.

In an earlier iteration of the foregoing argument, one reply was that the scientists did not use the intralevel experiment. Instead, the scientists only used a conclusion or assumption based on the experiment. This is an inadequate reply. In the first place, it is entirely *ad hoc*. My proposal concerns the history and philosophy of science; it is concerned with actual historical events. What historical basis is there for saying that Talbot et al. (1968) used a conclusion or assumption based on the experiment, but not the intralevel experiment? Further, what historical basis is there for saying that the Talbot et al. (1968) used the interlevel experiment, but not the intralevel experiment? Second, the reply is pointless. Even if Talbot, et al., only used a conclusion or assumption based on the experiment, rather than the experiment, MIE still has no place for a conclusion or assumption based on an intralevel experiment. Even the *ad hoc* interpretation of Talbot et al. (1968) does not enable the MIE account to correctly characterize the historical episode.

To step back a bit, the New Mechanist framing of the issue in terms of what experiments are, or are not, sufficient can be misleading. When arguing that top–down and bottom–up experiments alone are not sufficient, Craver and Prychitko point out that while it might be the case that x_i φ_i-ing implements S Ψ-ing, there are rival hypotheses about what might be going on.[7] What they look to be saying is that the interlevel experiments are defeasible and that they can be defeated by rival hypotheses. Once the

[6] Prychitko's discussion is similar to Craver's: Craver (2007, p. 154f).
[7] See Craver (2007, chapter 4) and Prychitko (2021, p. 1832).

situation is framed in terms of defeasibility by rival hypotheses, however, it is easy to see an opening for the abductive approach. A top–down interlevel experiment may provide a defeasible reason to think that x_i φ_i-ing implements S Ψ-ing; a bottom–up interlevel experiment may provide a defeasible reason to think that x_i φ_i-ing implements S Ψ-ing. Most importantly for the compositional abductive approach, an intralevel experiment may also provide a defeasible reason to think that x_i φ_i-ing implements S. What might defeat these pieces of evidence? A rival hypothesis. This framing reveals the relevance of an abductive approach. Combinations of interlevel and intralevel experiments involve combinations of distinct sets of abductive inferences. The abductive approach, thus, correctly describes the work by Talbot et al. (1968), whereas MIE does not.

10.3.2 Motion Perception and MT

Prychitko (2021) offers three experiments regarding motion perception in monkeys that *prima facie* fit the MIE account. Prychitko proposes that Newsome and Paré (1988) offers a bottom–up inhibitory experiment in which lesions to portions of MT impair motion detection. In these experiments, Prychitko's idea appears to be that MT cell activity, x_i φ_i-ing, implements a monkey's perceiving some motion, S Ψ-ing. Salzman et al. (1990) provides a bottom–up excitatory experiment in which stimulation of directionally tuned MT cells facilitates motion detection. Finally, Britten et al. (1992) offers a top–down experiment that measures activity in single MT cells when monkeys perceive motion. One attractive feature of these experiments is that they involve the same method for measuring motion perception thresholds. In these experiments, the input stimulus, Ψ_{in}, consisted of a stream of randomly positioned dots in which the percentage of dots moving together could be varied. The more dots that moved together, the stronger the movement signal. Further, an animal's response, Ψ_{out}, was to saccade to one of two targets representing the direction of motion, either up or down. Consider each of these papers in turn.

In their top–down experiment, Newsome and Paré determined the monkeys' motion sensitivity thresholds by varying the percentage of dots moving together. In their study, two monkeys were trained to reliably discriminate up from down motion with roughly 1.8% correlation. The monkeys were, thus, very sensitive to motion signals. After training, ibotenic acid was injected into MT in three hemispheres (m1, m2, and w1), with the fourth being spared as a control. Lesions raised the threshold

https://doi.org/10.1017/9781009435710.011 Published online by Cambridge University Press

for correlation by 400–800% for the contralateral eye, whereas the threshold for correlation was unaffected in the ipsilateral eye.

In the Salzman et al. bottom–up activation study, three rhesus monkeys were trained to discriminate the direction of motion in the random dot display. After insertion of stimulating and recording electrodes, the display was positioned in the monkey's visual field to activate MT cells primarily in a single cortical column. The display could then produce motion in the column's preferred direction or in the null direction. During the task, the column was randomly stimulated. As expected, when the direction of motion in the display was in the column's preferred direction, the stimulation enhanced the monkeys' likelihood of correctly perceiving the direction of motion. The stimulation biased the monkeys' perception.

The third study, the top–down experiment in Britten et al. (1992), showed monkeys the motion display and then measured responses in single MT neurons. Britten et al. were especially concerned to emphasize the role of single neurons. So, they write, "We compared the ability of psychophysical observers and *single cortical neurons* to discriminate weak motion signals in a stochastic visual display" (Britten et al., 1992, p. 4745, italics added) and "In fact the responses of *single neurons* typically provided a satisfactory account of both absolute psychophysical threshold and the shape of the psychometric function relating performance to the strength of the motion signal" (Britten et al., 1992, p. 4745, italics added).

Although Prychitko does not mention it, intralevel experiments play a role here. In each experiment, the monkeys had to be trained to signal the direction they perceived. This training involved presenting the dot pattern, Ψ_{in}, and measuring the monkey's response, Ψ_{out}. The interlevel experiments would not make sense without this information regarding the training. As Talbot et al. might suggest, the combination of interlevel and intralevel experiments can be more revealing than interlevel experiments alone. Philosophers of science guided only by interlevel experiments risk missing this point.

10.4 The Manipulationist "Picture" of Confirmation

On its face, one might think that the manipulationist New Mechanists can easily modify MIE so that it describes the scientific use of combinations of intralevel and interlevel experiments in the confirmation of compositional hypotheses. Just claim that MIE can use intralevel and interlevel experiments. I have given one reason to think that the manipulationist New Mechanists might resist this approach: they think about MIE in terms of

joint sufficiency conditions. Now I want to give another reason to think that the MIE approach is resistant to this move. Incorporating a clause that includes a role for intralevel experiment would constitute a deep reconsideration of the manipulationist "picture" of confirmation.

A core theoretical assumption of the manipulability approach is that to find a dependency between X and Y, one (ideally) manipulates X and measures changes in Y. In 2007, Craver proposed that "Constitutive relevance is symmetrical in a way that etiological (that is, causal) relevance typically is not" (Craver, 2007, p. 153) and that "all constitutive dependency relationships are bidirectional" (Craver, 2007, p. 153). This is an ontological presupposition noted in Chapter 2.[8] Although the point is rarely noted in the literature, Craver takes the ontological presupposition to ground his epistemological view of mutual manipulability. He claims that this bidirectionality is "the core reason why constitutive relevance should be understood in terms of mutual manipulability" (Craver, 2007, p. 153). In other words, in 2007, Craver's idea was that, in compositional cases, X ontologically depends on Y and Y ontologically depends on X, so to confirm this one manipulates X and measures a change in Y and manipulates Y and measures a change in X. The picture does not have a place for manipulating X and measuring a change in Y, and then hypothesizing that U compositionally depends on V. More concretely, the theory does not have a place for stimulating a monkey's hand and measuring a saccade, and then postulating that a perception of flutter compositionally depends on the firing of QA cells.

MIE preserves much the same picture of how to epistemically determine dependencies. On the MIE picture, Z depends on Y and Y depends on X. Or X causes Y and Y causes Z. This is an ontological presupposition. Thus, the MIE methodological approach is to have an experiment manipulating X and measuring a change in Y and two more experiments manipulating Y and measuring changes in Z. Each manipulation-measurement pair is meant to confirm a dependency. Here is another way of making the point. If one wants a set of experiments that shows that X causes Y which causes Z, one needs an experiment that links X and Y and another two (activating and inhibiting) that link Y and Z. It will not

[8] To repeat a point from Chapter 2, Schindler (2013) has an apt critique of this presupposition. I have heard it claimed that everyone has abandoned this assumption so that attending to this is a misdirection. Yet, Schindler (2013) reports that, in personal correspondence, Craver does not abandon the symmetry assumption. One reason for Craver not to abandon the symmetry assumption is that it plays an important theoretical role in the mutual manipulability approach. Moreover, Krickel (2024, p. 14) seems to endorse the symmetry assumption.

do to have one experiment that links X and Y and another two (activating and inhibiting) that link Y* and Z, where $Y \neq Y^*$. It will not do to have one experiment that links touching the worm on the head and the firing of an ALML neuron and another two experiments linking, say, the firing of AVM neurons and the worm's turning. Confirming the chain X–Y–Z has to involve X, Y, and Z.[9] If all of that is what underlies the rationale for MIE, then what space is there in this picture for intralevel experiments?

The goal to which I have been driving in this section is that: One might think that it is a trivial matter for the manipulationist New Mechanists to simply add that intralevel experiments have a role to play. And as a comment about scientific practices that would seem to be the right response. There is abundant evidence that intralevel experiments are combined with interlevel experiments to confirm compositional hypotheses. The difficulty lies in reconciling some role for intralevel experiments within a manipulationist theoretical approach in which dependencies between X and Y are confirmed by a manipulation of X and a measurement of Y. In practice, scientists manipulate X and measure a change in Y, and then hypothesize that U compositionally depends on V. Scientists do things such as stimulating a monkey's hand and measuring a saccade, and then hypothesize an experience of flutter that compositionally depends on the firing of QA cells. The challenge is saying how the admission of a role for intralevel experiments does not amount to a sweeping abandonment of the manipulationist picture about how to confirm dependencies.

10.5 MIE as Something Other than an Account of Scientific Practice

Over the years, numerous reviewers have proposed that manipulationist accounts, such as MIE, should not be interpreted as a theory of actual scientific practice. MIE has other goals, so it is unfair to criticize it on that ground. In this section, I will review some of these proposals for what MIE is about. While I think that these are dubious readings of CG&P's aspirations for MIE, I do not mean to insist on that. I am officially noncommittal about what CG&P think about this. Instead, I want to emphasize, first, that if one thinks that MIE does not provide an account of how scientists confirm compositional hypotheses, then this clears the

[9] In an apparent nod to actual scientific practice, however, CG&P do not require that Y be literally the same in the X–Y experiment and in the activating Y–Z experiment. Instead, they "must be of the same kind as, and occur within quantitatively overlapping ranges."

field for an abductive account. Second, I want to point out how this interpretation of MIE is likely to be mere temporizing. Historians and philosophers of science should not ignore how scientists actually confirm compositional hypotheses.

MIE only claims that interlevel experiments are involved in the confirmation of compositional hypotheses. This seems to me such a dubious reading of the literature that I hesitate to address it, but here goes. As I read the texts, Craver (2002) and Craver (2007, pp. 144–152) describe the scientific practice of using interlevel experiments. By contrast, Craver (2007, pp. 152–160) evidently goes beyond a descriptive claim about scientific practice. Craver offers mutual manipulability as a theory of what underlies those practices. He is offering a theory of confirmation that is an alternative to, say, HD confirmation. If he is not offering such a theory, then what is he doing with the concept of mutual manipulability? Further, there is what is "mutual" that goes beyond the bare claim that interlevel experiments are involved in the confirmation of compositional hypotheses. Further, as I read (Craver et al., 2021), it is meant to be a clarification or technical revision of Craver's 2007 mutual manipulability account that preserves the core ideas of the mutual manipulability account. Further, as just described in Section 10.4, the "core reason" for mutual manipulability goes beyond merely requiring interlevel experiments. It involves a presupposition that the experiments are somehow describable as "mutual manipulability."

Suppose, however, that one insists that MIE holds only that interlevel experiments are involved in the confirmation of compositional hypotheses. This is a short-sighted defense of MIE. The defense can yield the conclusion that MIE is not wrong, but it leaves the MNM without a theory of those practices. In other words, the objection is that the manipulability New Mechanists do not commit an error of commission, but an error of omission, if you will. If MNM just describes a set of practices, then some historian and philosopher of science should provide a theory of the reasoning underlying those practices. The abductive approach, thus, begins to fill a gap in this take on MIE.

MIE only provides sufficiency conditions for establishing compositional relations. Unlike the foregoing interpretation, there seems to me some textual basis for this interpretation. After all, CG&P present the MIE conditions as jointly sufficient: "To establish that an entity X and its activity ϕ are constitutively relevant to a mechanism that Ψs, the following experimental results and matching condition are jointly sufficient." Suppose, therefore, that MIE is nothing more than mere joint sufficiency conditions.

https://doi.org/10.1017/9781009435710.011 Published online by Cambridge University Press

Here again, this is a short-sighted defense of MIE. Concede, for the sake of argument, that MIE does constitute a set of jointly sufficient conditions for establishing a compositional hypothesis. So, MIE is not wrong and not guilty of some error of commission. Yet, why should a historian and philosopher of science care about MIE, unless it has something to do with science? Why is MIE interpreted as mere sufficiency conditions anything more than a mere philosophical curiosity? Such a view of MIE leaves open how scientists actually confirm compositional hypotheses. If one is interested in how scientists confirm compositional hypotheses, then at least the abductive approach is in the game.

MIE provides sufficiency conditions that scientists actually use in establishing compositional relations. This version of the last reply seems to me more charitable than the last. After all, claiming that the conditions are jointly sufficient does not preclude saying that the conditions are jointly sufficient conditions that scientists actually use. It would reconcile the CG&P claims about sufficiency conditions with Craver's earlier statement that "My descriptive goal is to characterize the mechanistic explanations in contemporary neuroscience and the standards by which neuroscientists evaluate them. This cannot be accomplished without attention to the details of actual neuroscience. I illustrate my descriptive claims with case studies from the recent history of neuroscience" (Craver, 2007, p. vii). Moreover, It would also make sense of why Kaplan (2012) and Prychitko (2021) refer to examples from the scientific literature: the scientific experiments are supposed to illustrate the manipulationist approach in scientific action.

Unfortunately, this interpretation simply backs into the observations in Section 10.3. The scientific examples to which Kaplan and Prychitko refer do not use MIE. The examples involve a combination of interlevel and intralevel experiments, whereas MIE assigns a role only to interlevel experiments. If MIE is supposed to provide a set of joint sufficiency conditions that scientists actually use, then it is incumbent upon those who hold this view to provide an instance in which scientists actually use them. I have no knock-down argument that there are no such cases. But, even if MIE correctly describes some cases, there are many cases where scientists plausibly use abductive reasoning.

MIE is only meant to provide norms for scientific practice. CG&P do not claim that MIE only provides a norm for scientific practice, but Craver and the New Mechanists often write about scientific norms in other contexts. So, maybe this proposal is best thought of as a possible defense of the study of MIE, rather than as CG&P's defense of the study of MIE.

In either case, the proposal is short-sighted. What is this putative norm and what reason is there to think it is a *bona fide* norm for science? Some of Craver's comments are relevant: "We can address the question of whether the norms of neuroscience are justified only when we have an idea of what the norms are and of how they can be defended. The descriptive project, in other words, is the first step in a normative project: to clarify the distinction between good explanations and bad" (Craver, 2007, p. vii). So, Craver's view in 2007 was that one cannot do the normative project without the descriptive project. Further, Craver's 2007 proposal was that philosophers of science should look at clear and uncontroversial cases of good explanations and bad explanations to distill from them the norms of good explanation. For Craver in 2007, actual practice is a guide to the norms of practice. In a different context, Craver also claims that "If one hopes to understand the norms implicit in the practice of science . . . one must begin by looking at real science" (Craver, 2007, p. 235). Maybe MIE only provides norms for scientific practice rather than a description of scientific practice. Nevertheless, this is a temporizing claim since a concern for practice apparently returns as soon as one begins to take the first step in inquiring about the basis for the norms. To justify the norm, one apparently needs to appeal to actual practice.

Suppose, however, one breaks from Craver's 2007 view and allows for some *a priori* justification for MIE. The importance of actual scientific practice would still not simply disappear. If one has some *a priori* justification for MIE, one might be concerned about cases where scientists do not follow MIE. If one's *a priori* justification for MIE conflicts with actual scientific practice, one faces a substantive question. To put the matter simplistically, "Are scientists being irrational in not following the philosopher's *a priori* norms or are scientists rational and the philosopher mistaken in her *a priori* norm?" This is not a trivial question. The upshot, again, is that to say that MIE is only a norm of scientific practice, rather than a description of scientific practice, is short-sighted.

10.6 Summary

I have proposed that scientists use abductive reasoning in support of compositional hypotheses. MIE may, or may not, be taken as a rival theory. If MIE is intended to describe scientific practice, then there will be cases that are better described by the abductive approach than by the MIE approach. Maybe there are cases that are better described by the MIE

approach than by the abductive approach, but that remains to be seen. If MIE is not intended to describe scientific practice, then there is one less rival to the abductive approach. Moreover, even if MIE is meant to do something other than describe actual scientific practice, one can see that a concern for actual scientific practice is still a pressing concern.

Here is how one might think about this chapter. Indeed, here is how one might think about the first two parts of this book. CG&P proposed that MIE is one method for determining what entities stand in compositional relations. In a late section of their paper, they argue that the MIE approach cannot explicate the rationale for thinking that a bottom bracket is a component in a bicycle drive train.[10] Further, in the final section of their paper, they cite other works that present what they take to be other methods of confirming compositional relations.[11]

One might read this chapter as an expansion on CG&P's comments. They proposed that the bottom bracket example, and others like it, indicate a need for an alternative to MIE. I have proposed that combinations of intralevel and interlevel experiments used in science also indicate a need for an alternative to MIE as a candidate description of scientific practice. My case for an alternative to MIE is not their case for an alternative to MIE. Further, rather than simply posing a problem, I have also proposed a solution: the scientific practice of combining intralevel and interlevel experiments is a matter of combining compositional and interlevel abductive inferences. The first two parts of this book have been dedicated to providing an outline of how compositional abductive inferences work in science. Chapter 6 broached the idea of singular interlevel abductive inferences, whereas this chapter fleshes it out a bit more.

[10] Craver et al. (2021, p. 8824). [11] Craver et al. (2021, p. 8826).