An Investigation of Scientific Phenomena

by

David Colaco

Bachelor's in Philosophy and Sociology, Rutgers University,

2012

Submitted to the Graduate Faculty of the

Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2019

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

David Colaco

It was defended on

June 3, 2019

and approved by

Mazviita Chirimuuta, Associate Professor, History and Philosophy of Science

Kenneth Schaffner, Professor Emeritus, History and Philosophy of Science

James Woodward, Distinguished Professor, History and Philosophy of Science

William Bechtel, Professor, Philosophy (UCSD)

Dissertation Director: Edouard Machery, Distinguished Professor, History and Philosophy of Science

An Investigation of Scientific Phenomena

David Colaco, PhD

University of Pittsburgh, 2019

To determine how things work, researchers first must determine what things occur. Such an idea seems simple, but it highlights a fundamental aspect of science: endeavors to theorize, explain, model, or control often result from first determining and adequately characterizing the targets of these practices. This dissertation is an investigation of how researchers determine one important kind of target: *scientific phenomena*. In doing so, I analyze how characterizations of these phenomena are formulated, defended, revised, and rejected in light of empirical research. I focus on three questions. First, what do characterizations of scientific phenomena represent? To answer this, I investigate what it means to characterize a phenomenon, as opposed to describing the results of individual studies. Second, how do researchers develop these characterizations? This question relates to the logic of discovery: I examine how researchers use existing theories and methods to explore systems, search for phenomena, and develop representations of them. Third, how do researchers evaluate these characterizations? This question relates to the logic of justification: I investigate how empirical findings serve as defeasible evidence for the characterizations of phenomena and in light of what evidence we should accept, suspend judgment about, or reject them.

Table of Contents

List of Figures

Preface

I thank Edouard Machery, my dissertation advisor, for his persistent guidance and unwavering support in developing this work. I also thank my dissertation committee members Mazviita Chirimuuta, Kenneth Schaffner, James Woodward, and William Bechtel for their consistently insightful feedback and recommendations on the essays included in this work. Much of this work was developed with insights from my graduate cohorts, especially Nora Mills Boyd and Aaron Novick. Further, Chapter 2 is largely indebted to my collaboration with Kevin Jarbo at Carnegie Mellon University. I am eternally grateful to him for agreeing to interact with a philosopher. Finally, I thank my partner Jen Liu for her support throughout this process.

1.0 Introduction

To determine how things work, researchers must first determine what things occur. Such an idea seems simple, but it highlights a fundamental aspect of science: endeavors to theorize, explain, model, or control often result from first determining and adequately characterizing the targets of these practices. This dissertation is an investigation of how researchers determine one important kind of target: *scientific phenomena*. In doing so, I analyze how characterizations of these phenomena are formulated, defended, revised, and rejected in light of empirical research. Three questions organize my account of scientific phenomena:

What do representations of scientific phenomena represent?

How do researchers develop these representations?

How do researchers evaluate these representations?

Talk of phenomena has a long history in philosophy, but significant analysis of scientific phenomena by philosophers of science has a more recent origin. Philosophical investigations of the notion and its fit into the domain of science was largely set in place in the 1980s, with the publication of two important works. These works serve as a starting point for my own investigation, so it is worth introducing them here.

The first work is Ian Hacking's *Representing and Intervening*. This work serves as a key starting point for philosophically investigating the nature of experimentation and how the results of experiments help us to learn about the world. In describing the targets of scientific investigation, Hacking notes:

> The word 'phenomenon' has an ancient philosophical lineage. In Greek it denotes a thing, event, or process that can be seen, and derives from the verb that means, 'to appear'. From the very beginning it has been used to express philosophical

> thoughts about appearance and reality. The word is, then, a philosopher's minefield. Yet it has a fairly definite sense in the common writings of scientists. A phenomenon is *noteworthy*. A phenomenon is *discernable*. A phenomenon is commonly an event or process of a certain type that occurs regularly under definite circumstances (1983, 221).

Hacking makes salient the fact that philosophical analysis of scientific phenomena has been borne out of the use of the notion by scientists themselves. My dissertation continues this tradition. My analysis of scientific phenomena is informed by the practices by which scientific researchers learn about the phenomena that they aim to investigate.

The second work is James Bogen and James Woodward's "Saving the Phenomena," in which the two lay out the distinction between data and phenomena. Set out in opposition to the idea that scientific theories explain and predict observables, Bogen and Woodward note:

> Our argument turns on an important distinction which we think has been ignored in most traditional analyses of science: the distinction between data and phenomena. Data, which play the role of evidence for the existence of phenomena, for the most part can be straightforwardly observed. However, data typically cannot be predicted or systematically explained by theory. By contrast, well-developed scientific theories do predict and explain facts about phenomena (1988, 305-306).

Bogen and Woodward clearly identify that phenomena, rather than the data produced via experiments and observations, are the targets of both theorization and explanation amongst scientists. In other words, and in contrast to earlier works in the philosophy of science, theories do not predict or explain observables but instead explain phenomena. These observables are the data, which serve as evidence for a phenomenon.

Though the two accounts differ in both commitments and (to some degree) aims, the insights from Hacking as well as Bogen and Woodward have been widely influential in the philosophy of science. Both of these works provide insight into the characteristics of phenomena and how researchers learn about them from their empirical research. In part illustrating the influence of these works, scientific phenomena are now counted as targets of explanation (e.g.,

Machamer, Darden, & Craver 2000; Woodward 2003; Bechtel and Richardson 2010), modeling (e.g., Cartwright 1983; Batterman 2005), representation (e.g., Suárez 2004; van Fraassen 2008), and intervention (e.g., Franklin 1986; Woodward 1989). While all of these accounts are also widely influential in the field, they provide limited insight into what constitutes a phenomenon. In fact, it is unclear whether or not these accounts agree on what counts as a phenomenon, despite the fact that the notion is used by both philosophers and scientists.

Thus, several popular accounts of different scientific activities all take scientific phenomena to be the targets of these respective activities. Despite this fact, most of these accounts provide no additional insight into what these phenomena are like or how researchers determine their characteristics. My research fills this lacuna: my aim is to account for how representations of phenomena are formulated and evaluated in light of the evidence researchers collect in empirical studies.

The scientific cases that I investigate in this dissertation are drawn from the fields of neuroscience and psychology. This choice is due to the fact that significant advances in these fields are often centered around the discovery and characterization of previously unknown phenomena via the results of observation and experimentation. In these cases, the characterization of a phenomenon of interest often precedes any explanation or theory of it, sometimes for many years. Take the placebo effect: though identified in the western scientific canon over one hundred years ago (de Craen et al. 1999), psychologists and neuroscientists still struggle to explain why it occurs (see Benedetti et al. 2005). Despite this, characterizing the placebo effect has had a major positive impact on both the theory and practice of science; through characterizing it, researchers can now determine when it might occur and what effects it might have in research and everyday life.

1.1 Outline

This dissertation consists of five chapters, which cover the formulation and evaluation of representations of scientific phenomena, along with a chapter dedicated to the first question above: namely, answering what representations of scientific phenomena represent. Chapters 1 through 4 highlight different experimental practices in psychology or neuroscience, which are used to characterize interesting phenomena. Chapter 5 shifts gears, as it provides a more general analysis of what representations of phenomena represent, an issue that is not specifically tied to the practices of psychology or neuroscience.

In Chapter 1, I address how researchers develop representations of phenomena following their discovery. Specifically, I examine how researchers detect phenomena with their experiments, and I discuss the role of theory in this process of discovery. I investigate the discovery of phenomena with *exploratory experiments*. Because these kinds of experiments are not designed to test hypotheses, philosophers like Ken Waters and Laura Franklin-Hall have argued that they involve no direction from existing theory about the system under investigation. I argue that this view is mistaken, which results in an inaccurate conception of exploration and the search for phenomena by scientists. Rather than lacking input from theory, I argue that existing theories of the system and theories that underlie researchers' methods direct exploratory experiments. These theories direct exploration by serving as the basis for *auxiliary hypotheses*, akin to those first identified by Pierre Duhem. To defend this thesis, I examine cases in which researchers use transcranial magnetic stimulation to explore the relation between brain activity and behavior. While exploratory, these cases rely on auxiliary hypotheses derived from existing theories about brains and behavior that guide the search for causal relations. This, in turn, can provide insight into stable and repeatable occurrences of brain activity. By measuring these occurrences, the

characterization of an interesting phenomenon can be formulated. Together, my analysis in this chapter captures how researchers use theory and their understanding of their experimental techniques to discover phenomena.

In Chapter 2, I continue to discuss how characterizations of phenomena are formulated and how experimental practices are used to evaluate these characterizations. These practices allow researchers to determine what consistently occurs whenever the phenomenon occurs and in what alternative contexts this phenomenon can occur. This chapter also reports the philosophical upshots of empirical research that I have done in collaboration with cognitive psychologist and neuroscientist Kevin Jarbo of Carnegie Mellon University. I analyze a case in which Jarbo and colleagues have characterized a scientific phenomenon, following the reports of numerous interesting findings about human spatial reinforcement learning in individual experiments. Based on these interesting findings that are thought to provide evidence for this phenomenon, I discuss a set of experimental strategies that were used by Jarbo and colleagues to formulate and evaluate a characterization of spatial reinforcement learning. These strategies allowed researchers to develop, modify, and improve this characterization. Throughout the identification process, the strategies and their products *vindicated* one another, which ultimately allowed Jarbo and colleagues to develop a characterization of a phenomenon from the observations of a collection of individual occurrences.

In Chapter 3, I examine how representations of phenomena are evaluated by addressing the question of when new evidence provides reason to recharacterize phenomena. As mentioned earlier, popular accounts of scientific explanation, most notably mechanistic explanation, take phenomena to be explananda. On these accounts, to explain, one must first characterize a phenomenon. Researchers aim to accurately characterize a phenomenon's features and the

conditions of its occurrence so that their explanations are accurate and their interventions are effective. Philosophers who defend mechanistic explanation also claim that the discovery of mechanisms leads to the recharacterization of the phenomena that they underwrite. However, the exact epistemic role of these discoveries has not been properly analyzed, resulting in a lacuna in the literature regarding whether or not explaining a phenomenon can provide sufficient reason to revise its characterization. Informed by the work of epistemologists Jane Friedman and Scott Sturgeon, I argue that mechanistic explanations that are discrepant with an existing characterization of a phenomenon provide reason to *suspend judgment* about the phenomenon but may not provide reason to recharacterize it. As I show with the neurobiological phenomenon long-term potentiation, suspending judgment about a phenomenon's characterization results in the further investigation of this characterization, while rejecting it results in researchers moving on to new ones. Discovering that distinct mechanisms underlie instances characterized as one phenomenon suggests that there are differences that may be relevant to its characterization, leading to the suspension of judgment. Yet, this discovery need not provide information about the phenomenon's features or the conditions of its occurrence, which is all that is relevant to determining the characterization's accuracy.

In Chapter 4, I continue my analysis of the evaluation of a phenomenon's characterization by discussing the reasons why it might be rejected in light of empirical evidence. In doing so, I examine how evidence supports a phenomenon's characterization and how this evidence is *defeasible*. I discuss the alleged phenomenon of "memory transfer" under investigation in the 1960s and 70s. Scientists allegedly transferred memories from one organism to another via the (cannibalistic) transfer of tissue. Their contemporaries rejected memory transfer, but historians and philosophers of science have questioned why they did so. I argue that memory transfer was

rejected because replication failures and the identification of confounds both undercut the evidential relationship between the alleged phenomenon's characterization and the findings thought to support it. This form of scientific inquiry, which I call a *defeater strategy,* involved phenomena being rejected following the defeat of the evidence for their characterizations. Drawing on John Pollock's account of defeasible reasoning, I show how this strategy provides reason to reject characterizations of phenomena, even if there is no study that conclusively provides evidence against it. Thus, this chapter introduces a distinction between providing evidence against a scientific claim, following which positive and negative evidence are weighed against one another, versus demonstrating that a study is flawed and thus cannot provide evidence in the first place. Without this distinction, one cannot explain why a phenomenon like memory transfer was rejected, despite there being evidence in its favor.

In Chapter 5, I address what representations of scientific phenomena represent. Though many philosophers accept that scientific phenomena are an important target of investigation, there remain questions about what phenomena are and, correspondingly, what representations of phenomena represent. In this chapter, I introduce a *nominalist* account of scientific phenomena. According to my account, one need not commit to the idea that phenomena are something over and above their individual occurrences. Nonetheless, representations of phenomena are distinguishable from representations of their manifestations, and researchers adopt different epistemic stances towards these respective representations. With this account, I provide reason to think that previous accounts of phenomena – namely, those that claim that phenomena are abstract objects, ideal types, or various senses of 'pattern' – are inadequate, due to their inability to explain the causal role phenomena are alleged to play or their inability to explain how representing phenomena is not equivalent to representing their manifestations.

Together, these essays constitute a new account of scientific phenomena and their role in the theory and the practice of science. The essays cover what phenomena are, why researchers want to represent them, how these representations are formulated, and how they are evaluated. To achieve this, I have taken the approach of integrating research from the history of science as well as different areas of philosophy. I have focused in great detail on both historical and contemporary cases of scientific practice. In addition, I have drawn several connections between my work in the philosophy of science and others' work in epistemology.

2.0 Discovering Scientific Phenomena

To explain their role in discovery and contrast them with theory-driven research, philosophers of science have characterized *exploratory experiments* in terms of what they lack: namely, that they lack direction from what have been called "local theories" of the target system or object under investigation. I argue that this is incorrect: it's not whether or not there is direction from a local theory that matters, but instead how such a theory is *used* to direct an experiment that matters. Appealing to contemporary exploratory experiments that involve the use of experimental techniques – specifically, examples where scientists explore the interaction of neural activity and human behavior by magnetically stimulating brains – I argue that local theories of a target system can inform *auxiliary hypotheses* in exploratory experiments, which direct these experiments. These examples illustrate how local theories can direct the exploration of target systems where researchers do not aim to evaluate these theories.

2.1 Introduction

Recent years have seen a growth of interest in experiments that involve the manipulation of systems and exploration of their characteristics, which have been dubbed *exploratory experiments*. This interest is not a quirk of philosophers, as scientists are keen to perform experiments that they explicitly classify as exploratory. For instance, neuroscientists perform exploratory experiments in which brain areas are stimulated to discover and characterize relations

between brain activity and behavior. This kind of experimentation is not merely common: exploratory experiments are important for discovering new and interesting scientific phenomena, when theoretical frameworks are still under construction (O'Malley 2007; Feest 2012). Though initially characterized from a historical perspective (Burian 1997; Steinle 1997), philosophers recognize that exploratory experiments are not a vestige of science past; these experiments are just as important to contemporary science as they have been throughout history (Burian 2007; O'Malley et al. 2009).

While historians and philosophers initially argued that theory has no role to play (e.g., Steinle 1997), many now agree that, particularly in contemporary examples, exploratory experiments can be guided by theories. To make sense of how there can be theoretical guidance despite a lack of the aim to evaluate theories, many philosophers who study exploratory experimentation have embraced a distinction between *local theory* and the *theoretical background*, which was popularized by Franklin-Hall (Franklin 2005). This has resulted in what I call the *no-local-theories* conception of exploratory experimentation: while they may involve direction from the theoretical background, exploratory experiments lack direction from local theory. This conception has been widely adopted in subsequent discussions of exploratory experimentation (Franklin 2005; Elliott 2007; O'Malley 2007; Feest and Steinle 2016).

I argue that the no-local-theories conception mischaracterizes the role of theory in exploratory experiments and thus fails to adequately capture what researchers take these experiments to be. It equates the idea that exploratory experiments are not designed to test hypotheses with the idea that these experiments lack direction from local theories. However, it is not the locality of theories that matters. Rather, it is the *use* of theories that matters. We cannot make sense of the aims of many exploratory experiments unless we acknowledge that local

theories can direct exploratory experiments, which they do in the form of *auxiliary hypotheses*. This is for two reasons. First, experimental techniques are often used in exploratory experiments, and, when they are used, it is necessary to appeal to theories of the system in order to determine that the system is an appropriate candidate for the application of a technique. Second, contemporary exploratory experiments are often performed on systems where researchers already have a local theory and need to use this theory in tandem with their techniques to take the next step to explore more details of the system.

With a better characterization of the role of theory in exploratory experimentation, I provide a better understanding of when researchers perform these experiments and for what reasons they do so. While philosophers have suggested that exploratory experiments are performed when researchers lack theory, or to efficiently investigate many aspects of a target system, neither of these reasons capture the aims of much of the exploratory research on biological systems such as the neural systems found in the brain. In these fields, researchers often already have a theory about the system they investigate, but this theory represents the system's components and their causal relations in insufficient detail to generate predictions of interest to the researchers. Thus, one of the main aims of exploratory experimentation in research on these systems is to use the theories and methods available to researchers to direct exploration of these systems in greater depth and detail. This use of exploratory experimentation is incompatible with the no-local-theories conception. This use also explains why this kind of experimentation is pervasive in fields like neuroscience.

To start, I introduce the no-local-theories conception and review what is meant by *local theory* and *background theory* in Section 2.2. In Section 2.3, I characterize how theories direct experiments and discuss the role of auxiliary hypotheses. In Section 2.4, I discuss a class of

techniques called transcranial magnetic stimulation (TMS), in which magnetic pulses are used to influence brain activity. In Section 2.5, I draw upon examples of TMS experiments to show how exploratory experiments involve the direction of local theory and why this local theoretical direction is critical to the aims of these experiments. I also address two objections to my characterization of the role of theory in exploratory experimentation: (1) that I am mistaken to call theories that inform auxiliary hypotheses 'local theories,' and (2) that there is no sharp distinction between exploratory and theory-driven experiments.

2.2 Accounts of the role of theory in exploratory experimentation

To understand what exploratory experimentation is, I contrast it with a conception of the scientific method that was once dominant. Traditional accounts of the scientific method describe what is called hypothesis- or theory-driven research, in which experiments are designed to test a hypothesis (e.g., Hempel 1966). Confirming or disconfirming this hypothesis is a means for researchers to evaluate the theory from which this hypothesis is derived (Giere, Bickle, & Mauldin 2006, 25). Thus, theory-driven research involves confirmatory experiments, or experiments that are designed to test hypotheses.[1] By contrast, exploratory experiments are used to explore systems and search for phenomena, without the aims of testing hypotheses or evaluating existing theories.[2]

[1] Philosophers agree that experiments are empirical studies performed to investigate causal relations, typically involving manipulations and comparisons to controls (Feest and Steinle 2016). When I talk about exploratory experiments, I have in mind individual experiments, rather than research programs (see Waters 2007 for more on this distinction).

[2] Exploratory experiments were initially characterized solely in terms of what they lack (Elliott 2007, 322). More recently, Elliott (2007), O'Malley (2007), and Feest (2012) have provided positive characterizations, by identifying these experiments' roles in discovery, concept formation, and the development of theory.

There are differences between the confirmatory experiments associated with theory-driven research and exploratory experiments. Franklin-Hall argues that exploratory experiments are "not guided by hypothesis" (Franklin 2005, 888). However, theories may *direct* them: an experiment "that is directed by theory need not test theory, but only be planned, designed and performed from the perspective of theories of the object" (Franklin 2005, 891).[3] To unpack this claim, Franklin-Hall makes a distinction between the theoretical background and local theories. The *theoretical background* corresponds to the systematic knowledge of a scientific field and is a collection of *background theories*.[4] By contrast, *local theory* represents "the behavior of the particular objects being measured" (Franklin 2005, 891). Franklin-Hall states that exploratory experiments "lack a local theory, rather than the lack of a theoretical framework altogether" (Franklin 2005, 893). I call this the *no-local-theories* conception of exploratory experimentation.

According to the no-local-theories conception, the theoretical background can "direct inquirers to the kinds of properties that could possibly have a causal role in their local investigations" (Franklin 2005, 893), but it "need not direct the explorer to one group of… objects over another" (Franklin 2005, 894). The theoretical background guides "the explorer to look for certain classes of objects whose activities are known, as a class, to relate to one another" (Franklin 2005, 894). Franklin-Hall is not the only philosopher who endorses such a view: Waters has argued for an analogous conception of the role of theory in exploratory experimentation. He argues that researchers perform exploratory experiments to "generate significant findings about phenomena without appealing to a theory about these phenomena for the purpose of focusing experimental

[3] 'Theory' is used thinly in this literature to mean the representations and claims about a topic.
[4] This is my reading of "theoretical background" and "background theory." While Franklin-Hall uses both terms, she does not make explicit what she takes to be the relation between them.

attention on a limited range of possible findings," even though these experiments are "embedded within scientific inquiry that relies on a lot of theory" (Waters 2007, 279).[5]

According to Franklin-Hall, the difference between local and background theories lies in what these theories represent. Let me make this distinction more precise. When experimenting on system **A**, theory $\mathbf{T_1}$ counts as a background theory if and only if $\mathbf{T_1}$ represents **A** in virtue of the fact that $\mathbf{T_1}$ is applicable to a class of systems of which **A** is a member. For example, a theory about neuronal signaling is a background theory when experimenting on a particular brain region (say, the basal ganglia), in virtue of the fact that the basal ganglia is a system that is made up of neurons. The background theory does not represent any characteristics of the basal ganglia that are unique to this brain region. Conversely, when experimenting on system **A**, theory $\mathbf{T_2}$ is a local theory if and only if $\mathbf{T_2}$ represents **A** in virtue of the fact that $\mathbf{T_2}$ represents the components of **A** and their causal relations: that is, if $\mathbf{T_2}$ is a theory about the particular object of investigation, **A**. For example, a mechanism schema of the basal ganglia counts as a local theory when investigating the basal ganglia. The no-local-theories conception leads us to expect that, in an exploratory experiment, researchers might choose to manipulate the basal ganglia directed by theories of neuronal signaling, but a model of the system itself will not direct them. Once one has a model of the system, this conception suggests, one does not need to further explore this system, as this model can direct the experiment's aims and interpretation in a theory-driven manner.

---

[5] Other philosophers accommodate the no-local-theories conception into their accounts of exploratory experimentation, albeit with qualifications. Elliott accepts that "background theory plays an important role" in exploratory experiments, while there is no "test or analysis of a specific, local theory" (2007, 324), though local theories may serve "as a starting point or foil" (2007, 324) to determine "what has gone wrong with the old paradigm" (2007, 327). O'Malley notes, "background knowledge and general concepts frame… data-gathering" (2007, 350), though, between theory-driven and exploratory experiments, those that appeal to theory are on the theory-driven side (2007, 352). Burian departs from this conception to some degree: he suggests that exploratory experiments can be "integrated with highly specific knowledge [or] 'local' theories," though it is unclear that he has in mind cases where local theories direct individual experiments (2007, 308).

Without a local theory, there is little to direct researchers to focus on specifics of the target system. This explains why those who endorse the no-local-theories conception claim that exploratory experiments involve methods that do not require specific targets of intervention or measurement. The methodology of exploratory experiments has been described as the variation of "a large number of different experimental parameters" (Steinle 1997, S70; 2002, 429; Elliott 2007, 323) or "wide instrumentation," which allows "scientists to assess many features of an experimental system" (Franklin 2005, 896). The choice of these methods is a consequence of the no-local-theories conception. Without a local theory, there is no reason to vary one parameter as opposed to another. Thus, according to these accounts, exploratory experiments involve manipulating or measuring many parameters. We might call these *brute force* strategies. Varying or measuring virtually every parameter is a means to acquire the results researchers would have obtained, had they directed their experiment with any local theory. With these strategies, there is no need to specify a restricted set of features to manipulate or measure.

There are a number of virtues to the no-local-theories conception. First, the conception captures the difference between the role of theory in exploratory experimentation versus theory-driven research: the former does not involve the evaluation of theory or aim to test hypotheses endemic to confirmatory experiments. This conception provides an explanation of how it can be the case that exploratory experiments are not designed to evaluate theories, but nonetheless involve theoretical direction. Second, the conception introduces the distinction between local and background theories, which seems to me to be a genuine and important distinction: not all theories relevant to the direction of experimentation are about the target system uniquely. Third, this conception helps us to understand why exploratory experiments are not performed only at the early stages of research on a target system. It may be the case that researchers have theories that are

relevant to the system that they investigate but need exploratory experiments to formulate local theories about this system. Because of this, exploratory experiments may be needed well into the investigation of a system. Yet, despite its virtues, I argue that the no-local-theories conception of exploratory experimentation does not hold up to scrutiny, as one cannot capture the role of theory in these experiments while also committing to the idea that they lack direction from local theory. To motivate my challenge, I address the roles of theory in experimentation.

2.3 Theories, experiments, and techniques

That exploratory experiments are not designed to test hypotheses seems to be one thing everyone in the literature agrees upon (Elliott 2007, 322). But, theories can be relevant to experimentation in other ways. Franklin-Hall mentions that theories "direct" experiments: the commitments of a theory lead researchers to investigate certain kinds of things. Likewise, theoretical direction consists in how researchers' conceptions of the target system determine how they choose to manipulate and measure components of this system. Thus, even though confirmatory experiments are driven by the evaluation of the theory from which a hypothesis is derived, background theories also can direct these experiments. Exploratory experiments involve the direction of background theory as well, according to the no-local-theories conception.

The best way to make sense of a theory directing an experiment whose aim is not to evaluate this theory, I argue, is to accept the idea that *auxiliary hypotheses* are derived from theories that merely direct experiments. Auxiliary hypotheses are not hypotheses that the experimenters aim to test, but they are falsifiable claims that represent facets of the experiment and the target system. Stemming from the work of Duhem (1906), any theory-driven, confirmatory

experiment involves the incidental test of a bundle of auxiliary hypotheses, in tandem with the test of a *main hypothesis*, which is the hypothesis researchers aim to test.[6]

Traditionally, an auxiliary hypothesis is characterized as being auxiliary to a main hypothesis (whence its name). One might even argue that the invocation of 'auxiliary hypothesis' is too heavily anchored to theory-driven, confirmatory experimentation to be usefully applicable to exploratory experimentation. I disagree. I argue that the theoretical claims that I call 'auxiliary hypotheses' play an equivalent role in confirmatory and exploratory experiments, which is why it is helpful to call them by the same term and relate their uses in these different kinds of experiments. This is because, regardless of whether we are looking at exploratory or confirmatory experiments, auxiliary hypotheses are falsifiable theoretical claims, where the aim of these experiments is not to test them.

With an understanding of auxiliary hypotheses on the table, we might ask what it is that these hypotheses are auxiliary *to* in an exploratory experiment. To answer this, I introduce a positive characterization of auxiliary hypotheses, which captures what these hypotheses represent. Auxiliary hypotheses are derived from the theories of how an experiment is designed, how experimental tools interact with the system, what the class of system is like, and what else about the system and its components has been theorized. These auxiliary hypotheses are needed to make inferences about characteristics of the target system based on an experiment's results. If a background theory directs an experimenter to investigate certain kinds of objects of the target system, then a claim about the features of those kinds of objects is an auxiliary hypothesis in this

[6] This fact is what makes it challenging (if not impossible) to perform a crucial experiment, which decisively disconfirms any individual hypothesis: because a researcher incidentally tests auxiliary hypotheses bundled with a main hypothesis they aim to test, they only determine that something about the bundle of hypotheses from ostensibly disconfirming results.

experiment. While they may be incidentally confirmed or disconfirmed in the process, researchers do not aim to test what count as auxiliary hypotheses in any kind of experiment. This characterization explains that exploratory experiments can involve what are best conceived of as auxiliary hypotheses, without abandoning the idea that the aim of these experiments is not to test hypotheses or evaluate theories. My invocation of the use of auxiliary hypotheses in exploratory experiments is not merely a different way of stating that these experiments are "theory-laden," even if these experiments are, in general, theory-laden (Karaca 2013). Rather, these auxiliary hypotheses actively direct these experiments, by providing reason to investigate target systems in certain ways, in order to explore certain characteristics of these systems.

With my characterization of auxiliary hypotheses, I reformulate the no-local-theories conception of exploratory experimentation. Auxiliary hypotheses are derived from background theories. This is because a background theory applies to a class of systems, whose individual features may differ from one another and therefore are not inferable from background theories that cover the whole class. Background theories direct these experiments when auxiliary hypotheses are derived from them, which are needed to determine what about the system to investigate and what to infer about the target system from the experimental results.

Does the no-local-theories conception capture the role of theory in exploratory experimentation? I argue that it does not. This is because, once we acknowledge the fact that exploratory experiments can involve theoretical direction in the form of auxiliary hypotheses, it becomes clear that these experiments can involve many auxiliary hypotheses derived from various theories, some of which are background, some of which are local, and some of which are not strictly about the target system. This issue becomes salient when we look at exploratory experiments that involve the use of experimental techniques.

An experimental technique consists in a procedure through which tools are utilized according to a set of instructions to achieve a certain kind of outcome in an experiment.[7] Designing an experiment involves developing its protocol, or the instructions that direct the investigator to perform certain actions to produce certain effects (Sullivan 2009, 513). An experimental protocol may specify experimental techniques, the use of which constrain an experiment: the decision to use a technique informs what other components of an experimental protocol can be chosen, as they must complement or at least be compatible with what this technique can do. What a technique can do also limits what the experiment's findings can address. Techniques predictably affect certain components of a system and not others. Any experimental aim must be formulated in light of what can be learned about a target system from the use of a chosen technique.

The constraining role of experimental techniques is a general feature of laboratory biology. The technique researchers choose to use determines what features of the target system researchers can measure and how they must set up the system in order to use this technique. If, for instance, one chooses to use a staining technique on a cell to investigate it with microscopy, one will not be able to investigate any components of the cell that are destroyed by the staining process. If the stain modifies a component of the cell one wants to investigate, one must have some understanding of how the stain changes the component's properties, which allows the researcher to differentiate what about the result of the investigation is indicative of the cell and what about the result is merely an artifact of the staining technique.

An experiment may be directed by different theories, which can be distinguished in terms of what they represent. A technique's use in an experiment is informed by a theory of this

[7] Thus, an MRI scanner is a tool, while fMRI is a technique. This definition emphasizes the materials (the tools) and the methods (the instructions) of techniques.

technique, which represents a technique's capabilities, if the technique is applied to a candidate target system with a specified set of prototypical features.[8] One can contrast this with a theory of the system, which represents the target system and its components. These two theories may overlap, and the theory of a system can be used to develop a theory of a technique.[9] Nevertheless, while they are not equivalent to one another, there is an interaction between the direction of a theory of the system and the use of an experimental technique informed by its theoretical underpinnings. One must appeal to a theory of a system in order to know if the use of a particular technique is applicable in an experiment on this system and to determine if the features of the system under investigation correspond to the prototypical features of the kind of target system to which the technique can be applied, which is specified in the theory of this technique.[10]

Techniques can be used in exploratory experiments. In these experiments, the technique is applied to the system, and researchers measure the effects of this technique on the system. Thus, the protocol of exploratory experiments is constrained by the use of a technique, in addition to auxiliary hypotheses derived from whatever theories of the target system predict that the system is a candidate to be investigated with this technique. Even though not all exploratory experiments involve the use of an experimental technique, many of these experiments do. These experiments are particularly prevalent in fields like neuroscience and neurobiology, where researchers use them to investigate the components and causal relations in neural systems in greater depth and detail.

[8] The technique's use likely is informed by researchers' tacit or procedural knowledge as well, but this fact is orthogonal to my theses.
[9] For example, CRISPR-Cas9 is a technique adapted from a theory of a system (Horvath and Barrangou 2010).
[10] The idea that theory of the experiment plays a role in research is not in itself novel. Hacking (1992, 45) and Rheinberger (1997, 28) each address this idea. Neither clarifies the roles these theories play in experiment and how using them relies on appeal to theories of the system.

Because of this fact, a conception of the role of theory in exploratory experimentation ought to be able to account for exploratory experiments that use techniques and are directed by local theory.

Just as having direction from background theories does not make an experiment less exploratory according to the no-local-theories conception, I argue that the mere direction of local theories is not sufficient to change the aims of exploratory experiments into confirmatory ones. In other words, the direction of local theories does not entail that an experiment is theory-driven or confirmatory in any relevant sense. There may be more input from theory, for sure, but there is no aim of evaluating this theory. Rather, the local theory is needed to explore other aspects of the target system in greater detail or sophistication than would be possible without the local theoretical direction. Let me defend this point concretely.

2.4 Exploring with the direction of local theories and techniques

To illustrate how local theory directs exploratory experiments that involve the use of experimental techniques, I introduce a popular class of techniques used in brain research: transcranial magnetic stimulation (TMS). TMS is a class of experimental techniques in which a magnetic field is used to induce electrical activity in human tissue. TMS allows researchers to manipulate brain activity so that they can investigate the causal role of brain regions. When using TMS, a brain area is chosen, a form of TMS is chosen, TMS zaps the brain, and measurements are made of its effects. The efficiency, noninvasiveness, and relative ease of use of TMS make it ideal for exploratory research on relations between brain and behavior, buoyed by TMS's validation and theoretical underpinnings.

While researchers attempted to use magnetism to induce effects on behavior as early as the 19th century (Walsh and Pascual-Leone 2003, 28), the development of TMS is due to Anthony Barker, who demonstrated that nerves in the peripheral nervous system could be stimulated through the delivery of magnetic pulses (Polson, Barker, & Freeston 1982). This result led to the refinement of TMS, which was successfully used to activate the motor cortex of a human subject, causing muscle movements following magnetic pulses (Barker, Jalinous, & Freeston 1985). While all uses of TMS are based on the same fundamental theories, there are different forms, which involve differently shaped stimulating coils, and they can involve single pulses, paired pulses, or repetitive pulses. The differences between forms change the shape, intensity, and penetrability of the magnetic field. They also can affect whether the stimulation inhibits or excites neurons in the targeted portion of the brain.[11]

The development and use of TMS is based on theories from the fields of electromagnetism and neurobiology. The idea to use magnetism to induce biological changes was informed by research on the nature of the relation between electricity and magnetism. In TMS machinery, a capacitor is discharged into a stimulating coil. The current running through the coil induces a magnetic pulse. This pulse results in a change in the magnetic field, which generates an electrical field (Walsh and Pascual-Leone 2003, 39-40). This exploits the relation between electric current and magnetic fields specified in the Maxwell-Faraday equation: "a change in magnetic field induces a flow of electric current in nearby conductors that… include human tissue" (Sauvé and Crowther 2014). This relation turns us to theories about the electrical properties of neurons.

[11] Forms of TMS have been validated, involving pairing TMS pulses with brain activity measurements via fMRI and EEG, allowing researchers to determine the regularity of changes in activity caused by magnetic stimulation (Walsh and Pascual-Leone 2003, 62). This adds the theory behind TMS and in what circumstances its use is applicable.

TMS exploits the fact that neurons exhibit electrochemical properties that allow them to maintain electrical charges through their membrane channels, which are mediated by the movement of ions across their membranes. The movement of ions causes a membrane potential to build up in neurons, the strength of which can be modified by the movement of current across the cell. The change in current induced by the magnetic field causes part of the neuron's membrane to have a change in electrical potential, which can precipitate the firing of the neuron. How quickly the current is generated, its duration, and the orientation of the neuron in relation to the magnetic field all affect how the current induces changes in neuronal activity; all of these methodological parameters can be taken into account to produce neural effects with different forms of TMS (Barker 1999). Thus, the use of TMS is based on existing theories about neural systems and the electrical changes that can be induced by the TMS intervention, which together inform its underlying theory.

Here is an example where TMS was used to test a hypothesis about the relation between brain activity and human behavior. Menkes and colleagues applied slow frequency repetitive TMS on the right frontal lobe to test whether or not stimulation of this brain region effects the symptoms of depressive patients (1999; I will refer to this as *Example 1*). In this experiment, the researchers hypothesized that their intervention would inhibit neural activity and thus would produce antidepressant symptoms: slow frequency TMS would affect the imbalance of "frontal lobe function wherein depression occurs with a hypofunctioning left frontal lobe" (Menkes et al. 1999, 113). Thus, they hypothesized that inhibiting the right frontal lobe would balance out the activity of the lobes, thereby reducing symptoms of depression. To test their hypothesis, Menkes and colleagues performed TMS on patients who met the DSM criteria of depression and non-depressive individuals as controls. Before and after the intervention, each participant was given a

depression inventory scale. Following TMS, there was a significant improvement of depressive patients on the depression inventory scales, while no change was reported for the controls. In light of these findings, the researcher claimed to have confirmed their hypothesis.

TMS is also used to explore relations between brain activity and behavior. For example, Levkovitz and colleagues developed an exploratory experiment to explore the effect of deep TMS on the prefrontal cortex of depressive patients (2009; I will refer to this as *Example 2*). This experiment was motivated by a desire to explore the relation between brain stimulation and depression in more detail so that researchers could determine if they could make TMS a more effective clinical treatment for depression. The researchers did not formulate a hypothesis about the relationship between excitation induced by deep TMS and change in depressive symptoms. The researchers applied deep TMS using four distinct protocols, using distinct TMS coil designs and different stimulation intensities (Levkovitz et al. 2009, 188). Depressive patients were assigned to the different protocols, and their depressive symptoms were measured prior to and following stimulation via the Hamilton Depression Rating Scale. The aim of the researchers was to "examine and perhaps identify potential factors associated with successful outcome of brain stimulation that used the DTMS [deep transcranial magnetic stimulation]" so that these factors could be tested in future experiments (Levkovitz et al. 2009, 195). Thus, the researchers explored the system in greater depth, literally and figuratively, to discover more about the prefrontal cortex and its relation to depressive symptoms than previously had been identified.

Not all exploratory TMS experiments involve the use of many forms of the technique, however. For example, Salih and colleagues applied paired-pulse TMS to partially epileptic patients to affect neural activity in the hemisphere of their frontal lobe that was seizure prone to explore the relation between non-REM sleep and epileptic seizures (2007; I will refer to this as

Example 3). The experiment was designed to produce findings that would contribute to the formulation of hypotheses about the relation between seizure induction and intracortical brain activity. Researchers applied a figure eight TMS coil to each patient and a control, and paired-pulse TMS was induced during both sleep states and wakeful states. Following stimulation to promote inhibitory or excitatory cortical activity, the intracortical inhibition and excitation of the participants were measured via EEG. Researchers determined that intracortical inhibition was greatly decreased by TMS stimulation in the experimental patients, when compared to the normal control participants. Likewise, they determined that intracortical inhibition was decreased by TMS stimulation in patients during non-REM sleep, when compared to patients in a wakeful state. The discoveries made in this study informed the formulation of a hypothesis about how intracortical inhibition operates and how its disruption can lead to the excitation of cortical areas, leading to seizures.

A trend appears amongst these examples, which casts doubt on the validity of the no-local-theories conception of exploratory experimentation. The researchers explicitly characterized Examples 2 and 3 as exploratory experiments, and these experiments did not involve the aim of testing hypotheses or evaluating theories. That being said, the consistent reference to components of the respective target systems and their causal relations certainly makes it look like local theories directed all three examples, exploratory experiments included. This local theoretical direction cannot be squared with the no-local-theories conception.

2.5 The role of theory in exploratory experiments

I have reviewed two exploratory TMS experiments, and a confirmatory experiment as a comparison. The trend amongst the examples suggests that the roles of theories in Examples 2 and 3 do not align with an account that suggests that only background theories direct exploratory experiments. Local theories directed each example, regardless of the fact that two are exploratory experiments. This, I argue, is for two reasons. First, the use of a technique involves appeal to a theory of the system to which the technique can be appropriately applied. Second, contemporary exploratory experiments are often designed to investigate systems for which there already are local theories, and these existing theories direct further exploration. These reasons are related: researchers use techniques to explore systems in greater depth and detail, directed by the theories of these systems that they already have. Also, these reasons are not unique to TMS experiments: modern experimental research, which often makes use of powerful, theoretically complex techniques, can be used to explore systems in more depth and detail than is represented in existing theories of these systems.

Starting with the most fundamental difference, Example 1 involved the explicit aim to test a hypothesis, while Examples 2 and 3 did not. However, in Examples 2 and 3, there was theoretical motivation for the experiments: previous research on brain and behavior provided reasons to claim that there are causal relations between the activity of certain brain regions and particular behaviors. These are empirically supported theoretical claims about the respective target systems, even though these self-described exploratory experiments were not, by all accounts, designed to evaluate these existing theories.

In Examples 2 and 3, the researchers expressed interest in developing hypotheses about the components of the target system in question and their causal relations to one another. In each

example, the use of TMS was possible because TMS would affect the target system in a predictable way. This is due to the fact that the target systems under investigation have the right electrophysiological characteristics for a technique like TMS to be applied to it. In Example 2, there was direction from theories of depressive symptoms and the activity of deep neural tissue, and there was the use of TMS and depression measurement techniques that relied on theories of brain activity and depression as well. In Example 3, there were theoretical claims about brain activities in sleep stages and the relation between activity in certain brain areas and epileptic seizures. In each example, reference to specific components and causal relations in the target system, which provided reason to think these systems were candidates for intervention via TMS, amount to nothing less than local theories.

If local theories directed these exploratory experiments, how do they differ from Example 1? What was missing in Examples 2 and 3 is a specification and aim to test what effects the researcher's manipulations and measurements would have on the target system. We instead find claims of this sort: *intervening on **X** may have some effect on feature **Y***. Beyond the claim that **Y** is sensitive to **X**, there is no characterization of the relation between **X** and **Y**. Is this a hypothesis? Well, it is a falsifiable claim about the components of the target system that the researchers aimed to investigate. This fulfills the criteria for a hypothesis. Does that mean that the aim of either Example 2 or 3 was to test such a hypothesis? The answer is no: the experiments were performed to discover more about the causal relations of components of the systems, in order to formulate more substantive theories of the respective systems under investigation. In other words, this is one instance of the auxiliary hypotheses that are present in Examples 2 and 3. In these examples, researchers did not set out to confirm or disconfirm a hypothesis about the causal relation between **X** and **Y**. Rather, the theoretical direction of auxiliary hypotheses facilitated the exploration of the

possible ways in which this relationship manifests, in order to further develop a local theory that represents the relation between **X** and **Y**.

These auxiliary hypotheses were integral to the direction of Examples 2 and 3 in two ways. First, these auxiliary hypotheses capture what the researchers performing these exploratory experiments already knew about the systems they investigated. Simply confirming that there is some kind of causal relationship would not have helped the researchers to learn much new about the system, as there was already evidence of such a relationship from previous research. However, with TMS, researchers recognized that they could manipulate the systems in more sophisticated ways to reveal the dynamics of their causal nature, which was not known prior to these experiments. Hence, they aimed to use their experimental techniques to discover more about the respective systems under investigation, which their local theories did not represent.

Second, auxiliary hypotheses were integral because the exploration of causal relations between brain activity and behavior in these cases relied upon the use of TMS, which required an understanding of its effect on neural tissue whose activity was manipulated. Because they chose to use TMS, its theoretical underpinnings directed these examples. Using this technique is only appropriate if the system under investigation corresponds to what the theory of TMS specifies are the features of candidate systems for the applicability of the technique. Researchers determined that the systems have the appropriate characteristics for TMS to be an appropriate technique to be used to explore these systems. Theoretical claims about each system in question amount to local theory, from which auxiliary hypotheses were derived.

My characterization that auxiliary hypotheses are derived from local theories in exploratory experiments explains why exploratory experiments typically have different methods when compared to theory-driven, confirmatory experiments. Recall earlier when I introduced the idea of

brute force strategies, which avoid the need to formulate a hypothesis about the target system in order to determine what interventions to perform or what things to measure. Even these strategies require theoretical support, in the form of auxiliary hypotheses, though they do not require the formulation of main hypotheses to guide the experiment. The key to the methods of exploratory experiments is that one cannot use a method that requires having a hypothesis about the system to test. However, one can be directed to certain aspects of the target system, based on what has already been theorized about this system.

My characterization of the role of theory also provides a better understanding of why researchers perform exploratory experiments like Examples 2 and 3, when compared to previous accounts. At a basic level, researchers perform these experiments in order to discover things about a target system, where not enough about the target system is known to generate precise, interesting predictions of the results of researchers' interventions (Burian 2007, 286). This is consistent with the prototypical aims of exploratory experimentation, as suggested by previous accounts. But, this is not sufficient to explain these cases. My characterization also explains that, even if researchers already have a theory of the target system, this theory may not represent the right characteristics to predict how certain interventions will affect the system's components. Once researchers reach the point of investigating these components, they must explore the system in greater depth and detail, using what theories of the system they already have to direct them. In addition, to identify their results as aspects of a system about which they have not yet theorized, researchers must have a theory about what their techniques do to the system. They must be able to determine what about the interaction of their techniques and the system is responsible for the effects that they produce.

With theory and method aligned, researchers can determine whether or not they can use a technique to perform manipulations on the system and discover more about the system's

components. This is paramount to research on biological systems like those found in the brain. In this research, enough has been theorized about the systems to identify what to explore, due to the direction of local theory. Likewise, techniques available to researchers and their theoretical underpinnings help researchers to determine how to search for phenomena in the system. However, researchers cannot predict what they will find in their search, which is why the exploratory experimentation is necessary. The results of these experiments help researchers to formulate new theories that represent the system in greater detail.

Thus, my characterization explains why neuroscientists perform exploratory experiments that are directed by local theory, as evidenced by Examples 2 and 3. These experiments are not performed for mere convenience or efficiency; they are needed to explore the nature of the causal relations between components of neural systems and their relation to behavior. Even with a local theory of these systems in hand, researchers cannot infer how these relations will manifest. However, with the local theory directing them, researchers can determine how to search the system and what techniques to use in order to discover the nature of the components of a neural system and their interactions. This is why neuroscientists use local theories to direct their exploratory experiments.

The different roles of theories in these examples are not well captured by the distinction between local and background theories. Recall that Franklin-Hall characterizes local theory as a theory that represents the particular objects being manipulated or measured. In Examples 2 and 3, the choices of what brain regions to stimulate and what effects to look for were directed by theoretical claims about the activity of those brain regions and their relation to particular behaviors. This fulfills the criteria for local theoretical direction. Yet, researchers performing the experiments in Examples 2 and 3 did not aim to test hypotheses about the character of the effect the stimulation

of the chosen brain area would have on the behavior of the participant. Thus, there was a local theory but no main hypothesis.

Should my characterization count as a local theory directing an exploratory experiment? A defender of the no-local-theories conception might object that 'local theory' is coextensive with 'theory from which a hypothesis to be tested is derived,' and any theory that directs an experiment but researchers do not aim to evaluate is a background theory. However, this objection is unsatisfactory. 'Local theory' is redundant if it is merely another way to refer to main hypotheses. Calling it 'local theory' is also misleading: a hypothesis one aims to test is not the only theoretical claim one can have about the specifics of the target system or object of investigation. Finally, this defense does not align with the differences between local and background theories that Franklin-Hall introduces. The difference between local theories and the theoretical background, Franklin-Hall suggests, is a difference between what these theories represent. The relevant difference is not whether or not hypotheses are derived from them. Even if a main hypothesis is derived from a local theory in a confirmatory experiment, this does not entail that what counts as a local theory is determined by whether or not a main hypothesis is derived from it. Thus, the distinction between local theory and background theory is not the same as the distinction between main hypothesis and auxiliary hypothesis.

A defender of the no-local-theories conception also might object that there is no sharp distinction between theory-driven and exploratory experimentation, and the fact that there are "exploratory-esque" experiments that are directed by local theory merely confirms the fact that there is a middle ground between the two. Waters notes, "the distinction between exploratory and theory-driven experimentation does not mark a sharp division," and the differences between the two are better understood as a continuum than a dichotomy (2007, 279). O'Malley concurs: she

argues that theory-driven and exploratory experimentation lie at the ends of a "continuum of practices," to make sense of experiments that seem to be exploratory but involve greater theoretical direction (2007, 351). According to this objection, exemplary exploratory experiments do not involve the direction of local theory, fulfilling the no-local-theories conception, but real-world cases may depart from these exemplars.

I do not doubt that there are experiments that do not fit into existing characterizations of either the confirmatory or exploratory varieties. However, this objection does not explain away the presence of local theoretical direction in exploratory experiments. This objection boils down to the idea that the only reason we might find local theoretical direction in something we might otherwise believe to be an exploratory experiment is because it is more a theory-driven, confirmatory experiment than it would be if it did not have local theoretical direction. These kinds of examples are closer to the theory-driven side of the continuum of practices, this objection would suggest.

However, there is no reason to accept this claim. This is because local theories play the equivalent role that the background theories play in exploratory experiments, and there is no aim to test either. It is possible that researchers might even disconfirm their auxiliary hypotheses derived from local theories, but this still does not make the experiment in question any less exploratory in its aims. Imagine a case in which researchers had a local theory of the target system that implicated some kind of causal relation between target system components **X** and **Y**, where researchers used this theory to derive an auxiliary hypothesis and explore how the causal relation manifests. In this case, researchers could find that they cannot identify any causal relation between **X** and **Y** with their manipulations. In this case, there is something wrong with some or all of the researchers' auxiliary hypotheses: their local theory is empirically inadequate, their theory of the technique is incorrect, their measurement instrument in inaccurate, etc. However, this is not more

confirmatory than an exploratory experiment that has no local theoretical direction. Background theories can also be empirically inadequate, and thus we equally might worry that they can be disconfirmed with exploratory experiments that have only background theoretical direction. Hence, even in a case where auxiliary hypothesis bundles are disconfirmed, there is still no difference between the role local and background theories can play in exploratory experiments. Thus, this objection fails; local theoretical direction need not make experiments any less exploratory than they would be without this direction.

2.6 Conclusion

With the issues of the no-local-theories conception addressed, I present my characterization of the role of theory in exploratory experimentation. An exploratory experiment is not designed to test a hypothesis or evaluate an existing theory. Nevertheless, a local theory can direct an exploratory experiment. If researchers choose to use an experimental technique, a theory of this technique can play a role in an exploratory experiment as well; this theory provides justification for inferences about the target system based on use of this technique, if and only if the system has the characteristic features specified in the theory of this technique. Auxiliary hypotheses are derived from theories of the target system, which direct researchers' exploration of the target system. With direction from these theories, researchers are able to explore things unknown of target systems.

Much remains to be discovered in biological systems like those found in the brain, which is why contemporary researchers continue to perform exploratory experiments. Though they are not designed to evaluate theories, exploratory experiments can be directed by local theories,

enabling the appropriate use of powerful experimental techniques, which come with sizable theoretical baggage. The no-local-theories conception of exploratory experimentation is ultimately flawed because it is focused on whether or not certain theories direct experiments, rather than how these theories are used to direct experiments.

I conclude with a general theme that becomes apparent following my analysis of the role of theory in exploratory experimentation. In various ways, previous accounts have internalized the idea that exploratory experiments are used in times when researchers lack a theory about a target system. At first glance, this makes sense: why explore if one already has a theory to evaluate? But, this doesn't get the picture right. It need not be the case that researchers lack a theory from which they could derive hypotheses to test, but only that they cannot derive any interesting hypotheses from the theory that they do have. In these cases, researchers use what theories and techniques they already have developed to explore further so that they can discover new things about the system.

3.0 Characterizing a Scientific Phenomenon

In this chapter, I analyze a case in which researchers characterized a scientific phenomenon, following the reports of numerous interesting findings about human spatial reinforcement learning in individual experiments. Based on these initial findings thought to evidence this phenomenon, I discuss a set of experimental strategies that were used by researchers to formulate and evaluate a characterization of this phenomenon. These strategies allowed researchers to develop, modify, and improve this characterization. Throughout the identification process, the strategies and their products *vindicated* one another, which ultimately allowed researchers to develop a characterization of the phenomenon from the observations of instances of its occurrence.

3.1 Introduction

In the previous chapter, I investigated how the initial discovery of interesting phenomena can be achieved through exploratory experiments. The results of experiments like these serve as a precursor to characterizing these interesting phenomena, which can then lead to theorizing about and controlling these phenomena in future research. However, simply recording the occurrence of a phenomenon in one of these experiments is not equivalent to characterizing this phenomenon. From these recordings, researchers must determine what occurs in these experiments and the conditions under which this occurrence is induced, inhibited, or modulated. With these characteristics specified, researchers can then apply this characterization to other contexts. Thus,

the central question addressed in this chapter is as follows: how do researchers get from interesting results in individual studies to the characterization of a phenomenon?

One of the aims of scientific research is to characterize interesting scientific phenomena to be explained, theorized about, and controlled. This is because "discovering phenomena, describing them appropriately…, and exploring [their] robustness and generality" is integral to the acquisition of scientific knowledge (Goldin-Meadow 2016). Despite the importance scientists place on characterizing phenomena, little research in the philosophy of science presents, let alone analyzes, the strategies that make this possible. With few exceptions, philosophical accounts have treated characterizing a phenomenon as a given, often as a stepping-stone to the discussion of explanation (e.g., Machamer, Craver, & Darden 2000) or modeling (e.g., Batterman 2005). This has led some philosophers to point out that "different aspects of the experimental practice have not yet received much attention from the side of philosophers" (Feest & Steinle 2016, 290).

In this chapter, I remedy this situation. With the example of the learning phenomenon spatial reinforcement learning (SRL), I document how researchers use experiments to determine the constitutive features of a phenomenon's occurrence along with the conditions under which this phenomenon occurs. I proceed as follows. In Section 3.2, I discuss what is means to characterize a scientific phenomenon, using the placebo effect as a prototypical example. In Section 3.3, I discuss the aims of the characterization process. I also address the challenges of formulating a characterization of a phenomenon from individual studies by highlighting concerns presented by Steinle (1997) and Sullivan (2009). In Section 3.4, I introduce a number of experiments designed to investigate SRL in the fields of cognitive psychology and neuroscience. I discuss the strategies that researchers used to characterize SRL in these experiments in Section 3.5. It is by these means that researchers started with the results of individual studies and developed a characterization that

can be applied to other instances, such as alternative experimental contexts or the natural world. While these strategies cannot guarantee that researchers have exhaustively characterized the phenomenon of interest, the interaction between these strategies and their products results in the process of characterizing phenomena possessing an iterative, self-vindicatory character, which I discuss in Section 3.6.

3.2 What is being characterized?

In order to analyze how phenomena are characterized, let me first discuss what is being characterized. Phenomena consist in occurrences that have certain features, manifest under a certain of precipitating conditions, and can be inhibited or modulated by a distinct set of conditions (Craver & Darden 2013, 56). Phenomena are repeatable: the same phenomenon can, in principle, occur again. Phenomena are also stable: the same phenomenon can occur in different contexts.

An example is helpful to illustrate these characteristics. The placebo effect is a phenomenon that is well known to researchers in psychology and medicine. It occurs when individuals receive no active treatment for a condition but manifest a treatment response.[12] In the Western medical canon, systematic observations of the placebo effect date back to the turn of the nineteenth century, when Haygarth reported the reduction of patients' undesirable symptoms in concert with the use of an inert treatment during a medical trial (de Craen et al. 1999). The identification of this effect allowed researchers to refine placebo-controlled research and understand the effect of active treatment when compared to the patient whose placebo 'treatment'

[12] The response that scientists most frequently investigate is analgesia, the reduction of undesirable conditions, though the effect is characterized more broadly (Price, Finniss, & Benedetti 2008).

either did or did not result in improvement. This involved determining when and how often the effect occurred, who is susceptible to it, and under what conditions it occurs. Further, its characterization led psychologists and later neuroscientists to theorize about why this phenomenon occurs in humans.[13]

This example highlights the following characteristics of phenomena. First, the placebo effect occurs when individuals manifest a response without treatment, regardless of whether or not the researchers are aware of this. History illustrates that discovering and characterizing the placebo effect has been integral to the development of placebo-controlled trials, as it has allowed researchers to control for the confounding effect it can have in the test of a medical treatment. Thus, a phenomenon can affect the results of research before researchers have characterized it, which illustrates the independence of phenomena from scientists' knowledge of it. Because of their causal efficacy, identifying them is critical from both a theoretical and practical standpoint.

Second, the placebo effect is repeatable.[14] If the proper set of conditions, either experimental or natural, is in place, then it will occur. Thus, when characterizing a phenomenon, it is key to identify both its constitutive features and the set of conditions or factors under which it occurs (Steinle 1997, S70; Craver & Darden 2013, 56). These conditions can be highly contrived or unusual, as they may be in a laboratory context, or common, as they may be if the phenomenon occurs in nature without the aid of human intervention. The conditions also may be difficult to reproduce at a given time, due to, for example, the practical limitations of experimental design. It is because of a phenomenon's repeatability that researchers expect that, in principle, the

[13] Despite success in characterizing the placebo effect, there has been comparatively little success explaining why it occurs (see Benedetti et al. 2005).
[14] It is consistent with what Bogen and Woodward's claim that phenomena "have stable, repeatable characteristics" (1988, 317).

phenomenon can occur (and, by consequence, can be detected) again. Conversely, if whatever conditions that induce it are not in place in a particular context, then there is an expectation the phenomenon will not occur.

Third, the placebo effect is stable.[15] It occurs in a variety of different contexts, from medical experiments to everyday use of ineffective drugs. That being said, the placebo effect does not occur everywhere: there are boundaries to the contexts in which it is expected to occur. Phenomena are stable in the sense that they occur in different contexts, so long as, in these contexts, a certain set of conditions is in place. In other words, they exhibit insensitively to some factors of a context (Radder 1996, 28). This is analogous to the sense of 'stability' Woodward describes in causal relationships: a phenomenon is stable in the sense that it occurs in multiple circumstances that involve different background conditions (2010b). It is because of this characteristic that researchers can make claims about a phenomenon that are applicable to multiple contexts and determine what conditions or factors are shared by these contexts.

Thus, the case of the placebo effect provides a starting point to exemplify what goes into characterizing a phenomenon. We expect the occurrence of a phenomenon to be repeatable and for these repeated occurrences to all be represented by our characterization of this phenomenon. We also expect phenomena to occur in contexts that are different from those in which they were originally observed, so long as certain conditions are in place in these contexts where it occurs.

[15] This is consistent Cartwright's claim that "I use the word 'phenomenon' to indicate that we are interested in studying events and processes to which we attribute a high degree of stability: the same phenomenon recurs across a variety of contexts" (1991, 146).

3.3 The aims and challenges of characterizing phenomena

With an understanding of what scientific phenomena are, I turn to the aims of characterizing them. The aims are to answer the following kinds of questions:

- *What occurs?*
- *Where does it occur?*
- *On what timescale does it occur?*
- *In what contexts does it occur?*
- *Does it occur again if the conditions are reproduced?*

Thus, the aim is to accurately characterize a set of co-occurring features and the conditions relevant to their co-occurrence. In answering these questions, researchers determine if what they describe has the stability and repeatability that is expected from something that can be called a 'phenomenon.' While these questions apply to characterizing phenomena in general, I focus on *experimental practices* in this chapter. As such, I focus on studies that involve the induction, manipulation, and measurement of a phenomenon of interest under experimental conditions.

These aims set up the challenge of characterizing phenomena. On one hand, phenomena are expected to occur in different contexts. However, phenomena are expected to only occur when certain conditions are in place. There can be a lot of conditions that must be in place for a phenomenon to occur, and these conditions themselves can be contrived, unusual, or practically impossible to reproduce. While balancing these two aspects of phenomena results in challenge in the process of characterizing phenomena, these aspects themselves are not in tension. A phenomenon occurs only when certain conditions are the case, but it is stable in the sense that these conditions themselves can occur in a variety of otherwise different contexts. Thus, the challenge of characterizing phenomena, which stems from their repeatability and stability, is that researchers characterize a phenomenon based on recordings of manifestations of this phenomenon

from individual contexts. This is a challenge for two related reasons, which have been identified in the philosophical literature.

First, there is the issue that not all factors present in an experiment will be present every time a particular phenomenon's occurrence is induced. As a result, researchers do not include the sum total of experimental conditions as the conditions that induce the phenomenon of interest. Rather, they must determine what subset of these factors induce the phenomenon's occurrence, which will then be counted as its precipitating conditions in its characterization. Thus, they must have a means for "determining which of the different experimental conditions are indispensable" for the occurrence of the phenomenon in question (Steinle 1997, S70). This can be illustrated with the placebo effect example. When a patient exhibits the placebo effect after taking an inert pill with water, researchers do not think that drinking the water was a precipitating condition for the phenomenon's occurrence, even though, in this example, the drinking of the water preceded the occurrence of the phenomenon. This is because there is an expectation that the phenomenon can occur without the patient drinking water.

The first issue leads into the second. Experimental contexts can (and often do) have a variety of differences between them. For example, researchers might use different materials and methods to induce and measure what they expect to be occurrences of the same phenomenon in different experimental contexts (Sullivan 2009). As a result, even if a phenomenon is induced in one of these contexts, one's expectations that this phenomenon will occur in a second context depends on this second context having both the factors that induce the phenomenon as well as no factors that will inhibit its occurrence. In order words, one can characterize the results of an experiment, but, if one wants to characterize a phenomenon that is repeatable and stable across different contexts, one must determine what factors about this context are expected to induce the

phenomenon in other contexts. Further, if one wants to determine that the factors of a context serve as inhibiting conditions for a phenomenon's occurrence, one must investigate these inhibiting conditions by developing an experimental context in which the precipitating conditions are in place but the phenomenon nevertheless does not occur. This means that researchers must have a way to investigate differences between contexts and determine in which contexts their characterization of a phenomenon is applicable.

This second issue can be illustrated with the placebo effect. Even if researchers have reports of cases in which a treatment response that occurs without receiving any treatment, this alone does not tell researchers in what other contexts they should expect the placebo effect to occur. To determine the conditions for the placebo effect's occurrence, researchers must determine what factors of the context induce its occurrence and what other factors might inhibit or modulate its occurrence.

While these issues constitute a challenge when it comes to characterizing phenomena, they are not insurmountable. In fact, these issues help to illustrate why researchers adopt specific strategies when formulating and refining characterizations of phenomena. These strategies all come down to attempting to induce or not induce the same phenomenon under the same or different conditions.

3.4 Characterizing spatial reinforcement learning

With the challenge of characterizing phenomena presented, I introduce an example of the study of the phenomenon spatial reinforcement learning at the intersection of cognitive neuroscience and psychology. Spatial reinforcement learning (SRL) is a learning phenomenon that

occurs when participants improve their performance on in a reward-based spatial decision-making task. Informed by the study of reinforcement learning, or "learning what to do – how to map situations to actions – so as to maximize a numerical reward signal" (Sutton & Barto 1998, 3), several researchers have identified cases in which human participants improve their performance on spatial tasks over a number of trials, based on the rewards these participants receive for their decisions (Colby & Goldberg 1999; Behrmann, Geng, & Shomstein 2004; Gottlieb 2007). Based on cases like these, researchers have taken interest in this learning phenomenon as it manifests in humans. However, these researchers have also faced issues with how the occurrence of SRL in humans relates to existing models of human decision-making. Jarbo and colleagues investigate this phenomenon; I present two of their works on this topic. The first investigates a possible neural model that underwrites SRL. The latter analyzes how to best characterize the conditions under which SRL does and does not occur. In other words, the second work is focused on how to characterize the phenomenon of interest and determine its repeatability and stability.

Based off of the reports from previous cases – taking the phenomenon to involve the improvement of performance in a task under the condition that the participant receives numerical rewards – Jarbo and Verstynen used two brain imaging methods to develop a neurologically plausible network that may be responsible for SRL (2015). These methods consisted in both structural and functional connectivity analyses, which together suggest that there is a convergence of projections between striatum and cortical areas: specifically, the orbitofrontal cortex, the dorsolateral prefrontal cortex, and posterior areas of the parietal cortex. These brain areas are thought to be associated with reward processing, attention, and spatial perception, all of which are relevant to the learning phenomenon of interest.

The researchers then used two methods to determine if the pathways between these cortical areas and the striatum form a network. The first method, fiber tractography on diffusion spectrum imaging (DSI), reveals structural relations between brain regions. DSI is a technique in which the diffusion of water along fiber bundles is measured via magnetic resonance imaging. Because water diffuses more quickly along fiber bundles that run parallel to one another, the diffusion reflects fiber tracts. Diffusion imaging thus provides information about structural relations between brain regions. In this study, the tractography allowed the researchers to visualize the fiber tracts. This revealed the structural connections between the cortical areas and the rostral end of the striatum. The second method, resting state functional magnetic resonance imaging (fMRI), reveals functional relations between brain regions. This technique measures the BOLD effect, in which oxygen-rich blood flows in response to local oxygen depletion. Resting state fMRI measures the effect when the participant is not asked to perform a task; rather, the participant is at wakeful rest. This putatively captures the task-independent relations between brain regions. In this study, fMRI data suggests a functional relation between the activity in the rostral area of the striatum and the activity of the cortical regions. The pattern of activity is strongly reminiscent of the structural relations modeled with DSI data.

Using these techniques, the researchers suggest that there are two regions of the striatum, the caudate nucleus and the putamen, that are structural and functional convergence zones for the cortical areas. The researchers concluded their article with a hypothesis that these zones serve as sites for the integration of the processes to which the regions correlate. Thus, this work suggests a possible neural network that may underlie SRL.

While this neural network seems plausible given what has been specified about SRL, researchers encountered another issue. SRL does not always occur. That is, humans do not always

improve their spatial decision-making behavior over trials when they receive numerical rewards. This called into question models of decision-making behavior like maximum expected gain theory, which suggests that the individual making the decision implicitly calculates sensory input information and produces motor outputs (Trommershäuser, Maloney, & Landy 2003). This model also suggests that individuals calculate the variation of both the input and the output. For example, the models include parameters that estimate the exogenous sensory variability of the environment and the endogenous variability of the individual's motor ability. Furthermore, these models suggest that the individual's decision-making behavior is optimal. That is, they suggest that the individual's implicit psychological processes correctly calculate the exogenous and endogenous features, which allow the individual to perform an optimal movement.

Given that empirical evidence indicates that human beings do not have optimal gain-maximizing performance in all spatial decision-making tasks (e.g., Meyer et al. 1988), there was a concern that these models cannot adequately model SRL. But, more fundamentally, these kinds of findings suggest that there are more conditions under which SRL does and does not occur, which had not been identified. As developing a better model of decision-making behavior that could represent SRL relies on better understanding these conditions, researchers turned to characterizing the phenomenon of interest in more depth.

Following the development of a neural network model that could potentially underwrite the occurrence SRL, the second work from Jarbo and colleagues moved to the more fundamental project of characterizing SRL's occurrence in different contexts (Jarbo, Flemming, & Verstynen 2018). This set of studies involved the use of behavioral tasks in a psychophysical paradigm so that the researchers could identify in what contexts participants improved their performance when reward conditions were held consistent. Participants in three sets of experiments were asked to

pick out pixels closest to the center of a collection of white dots (the Target), while avoiding the center of a collection of red dots (the "Danger Zone"). The closer to the center of the Target, and the further from the center of the Danger Zone, the more reward points the participant received. With this paradigm, the researchers were able to modify the spatial relationship between the Target and Danger Zone and observe how participants' spatial decision-making actions would change given the variance, penalty, and reward feedback information the participant could learn in the process.

In the first experiment, the researchers manipulated the variance of the Target pixels as well as those of the penalty. This allowed the researchers to test to what extent the severity of the penalty the participants received affected their Target choice and how this related to the difficulty of selecting the Target as its variance increased. In the second experiment, the ratio of the variance of Target and Danger Zone remained static, and the penalty severity and Danger Zone proximity to the Target were manipulated. This allowed for a test of participants' Target choice in closer proximity and to what extent variation in Target choice is due to the presence and severity of penalty. In the third experiment, picking a pixel near the Danger Zone incurred no penalty, while the proximity between the Target and Danger Zone was varied. This allowed researchers to determine what effect the Danger Zone's presence had on participants when it did not incur a penalty in the process.

A related set of results can be inferred from the three experiments. The participants improve their performance in comparatively low-variance cases, which elucidates the circumstances under which SRL occurs. From the effect of the presence and proximity of the Danger Zone present in all experiments, they show that learning and visual biases are at odds with one another during decision-making tasks. Thus, while reward optimization will increase as the participant acquires

experience with the experimental paradigm, this increase is affected by the presence of a penalty. This effect is greater (1) the more severe the penalty is and (2) the more challenging it is to choose the target and avoid the penalty.

Each of the studies provides insight into the stability of SRL, while also showing that this phenomenon is repeatable. The researchers were able to achieve this by attempting to pull apart the phenomenon's features and conditions via selective manipulation, in order to determine the context in which the specified features of SRL occur and in which contexts they do not. Given the comparatively underdeveloped characterization of SRL presented in previous works, this project can be seen to greatly add to how SRL should be characterized. Inferring from their data, and their relationship to variations in the experimental protocol used in the three experiments presented in the work, the Jarbo and colleagues provide more information about the contexts in which SRL occurs. This work elucidates how to characterize the learning phenomenon, which allowed the researchers to return to investigating how SRL relates to both models of decision-making and the neural network that might underwrite this phenomenon's occurrence.

3.5 Strategies for characterizing phenomena

With the SRL case introduced, the nature of the strategies used to characterize a phenomenon can be assessed. We start with a number of interesting features measured to co-occur in a number of experiments. While these features are interesting, researchers must establish that they co-occur again (establishing repeatability) and determine in what different contexts they co-occur (establishing stability). When both are achieved, researchers can formulate a characterization

of the phenomenon of interest that is consistent with the empirical evidence that they have acquired. As mentioned in Section 3.3, the central challenge of characterizing phenomena revolves around both determining which factors that are present when its induced are indispensable and determining whether or not these features co-occur in different contexts so that the conditions that induce, inhibit, or modulate the phenomenon's occurrence can be specified as well.

How did the researchers studying SRL's characterization meet this challenge? In this case, the researchers developed new experiments to induce the phenomenon in new contexts, in order to determine what features consistently co-occur and the different contexts in which these features co-occur. From these experiments, they identified what factors are present in each context that serve as precipitating conditions for the phenomenon's occurrence and what other factors might induce or modulate its occurrence. In other words, researchers determined if the thing that they had identified is stable and repeatable, by determining whether or not the specified features could be reproduced and whether and to what extent these features could be reproduced under different test conditions.

Based on previous studies whose results seem to evidence the occurrence of SRL, these researchers developed a new experimental context in which they could measure participants' performances on spatial decision-making tasks in light of the numerical rewards these participants received over a series of trials. In these new experiments, the phenomenon again was induced and measured: participants improved their performed following receipt of rewards. I call this an *induction test*. Upon measuring this change in performance, researchers provided additional evidence of the repeatability of spatial reinforcement learning: in contexts that involved decision-making tasks and reward information, the specified improvement of performance by individuals was measured.

That being said, this experimental context was new: the researchers' experimental protocol, while informed by previous research, was not a direct replication of any previous study. As a result, the fact that the phenomenon's features could be induced in this case provides some insight into what aspects of experimental contexts are dispensable to the occurrence of SRL. In this sense, the researchers successfully induced the phenomenon that had previously been identified under different experimental conditions. This can rule out many conditions of the previous experiments as dispensable conditions, which should not be included when characterizing the phenomenon.

While the researchers were able to reproduce the phenomenon and provide evidence for which conditions are indispensable to inducing its occurrence, this alone does not address the second issue brought up in Section 3.3. After all, there were reports of some cases in which SRL did not occur, despite these contexts seeming to have in place the indispensable conditions for inducing this phenomenon. This suggested to the researchers that there might be factors in these experiments that inhibit SRL's occurrence, despite the fact that the factors for inducing it are also in place. For this reason, Jarbo and colleagues picked two possible factors that might inhibit the phenomenon's occurrence based on previous research (Trommershäuser, Maloney, & Landy 2003) – factors related to penalty and variance of spatial stimuli – and designed their experiments so that these factors could be manipulated and their effect on the phenomenon's occurrence could be measured.

In the three experiments, the researchers varied the severity of penalty and the variance of spatial stimuli, in order to determine how each individually and the two jointly might affect the participants' performance on the task. Other aspects of the experiment were kept the same: participants stilled received reward information about the spatial decision-making task they performed. These experiments provide evidence that SRL does not occur when penalty or variance

conditions reach a certain level of severity. Jarbo and colleagues were able to "overwhelm" participants with penalties and variance so that the learning phenomenon did not occur. From this, they determined that, while SRL occurs in a variety of contexts, there are limits to its stability.

This case illustrates that researchers can deliberately aim to inhibit a phenomenon's occurrence with certain conditions, in order to better understand the contexts in which the phenomenon does not occur. I call this an *inhibition test*. Researchers can aim to inhibit the features thought to be constitutive of the phenomenon's occurrence. This is different from the control a researcher performs to determine if they have a positive result. An inhibition test involves the variation of test conditions in an experiment, with the goal of determining the change in conditions that results in a non-occurrence of the features they have previously identified. This strategy provides a means to fill out the intuition that "if an effect disappears when you predict it will, then it is valid" (Franklin 1989, 453). If the features thought to constitute the phenomenon's occurrence do not occur, researchers can determine some of the boundary conditions of the phenomenon; this provides evidence for the limits of its stability.

It is also through an inhibition test that researchers can determine what set of features consistently do or do not occur concurrently. If some but not all features occur, researchers can investigate these outlier features. It is by these means that researchers can eliminate systematic artifacts. If a feature persistently occurs despite the fact that no other features thought to constitute the phenomenon's occurrence occur with it, researchers can investigate whether or not this persistent feature is endemic to the system they investigate but otherwise not relevant to the phenomenon of interest. In this way, negative results can provide insight into whether or not the co-occurrence of features is due to the fact that they are constitutive of a phenomenon. Researchers can pull these features apart and determine which co-occur across a number of experimental

contexts. Thus, this strategy provides evidence for which features to include in the phenomenon's characterization and which conditions to specify.

Despite the fact that researchers were able to inhibit the occurrence of SRL with a certain level of penalty severity or spatial variance of stimuli, it is important to recognize that this was not an all-or-nothing affair. Rather, when penalty and variance severity were increased, researchers observed that the learning phenomenon still occurred, but the rate at which improvement of performance occurred was correspondingly slower. This is not an outright inhibition of the phenomenon's occurrence. Rather, it is a modulation of its occurrence. By varying the experimental conditions, the researchers determined the relation between manipulating these conditions and the rate of SRL. These conditions included demonstrating that SRL does not occur at the same rate in contexts in which penalty biases the individual. The changes between manifestations of SRL provided a test of how the changes in the constitutive features of SRL and the conditions of its occurrence relate to one another.

This case illustrates that researchers can deliberately aim to modulate a phenomenon's occurrence, in order to better understand how changes in conditions for the phenomenon's occurrence correspond to changes in the way in which the phenomenon manifests. I call this a *modulation test*. A modulation test involves the use of an experiment in which conditions are varied, in order to determine how the constitutive features of the phenomenon are affected by these variations in conditions. This provides a means to investigate the stability of the phenomenon, as it provides evidence for the different contexts in which the phenomenon occurs. However, and more importantly, it provides researchers with evidence about how a phenomenon's occurrence can vary depending on the presence of certain modulating conditions in an experiment.

Unlike an inhibition test, which can provide evidence of which features thought to constitute a phenomenon's occurrence consistently co-occur, a modulation test can provide evidence about how manifestations of a phenomenon of interest can differ from one another in predictable ways, which, in turn, can be captured via the phenomenon's characterization. With this evidence, a phenomenon's characterization becomes, in a sense, less static. Rather than simply specifying which features co-occur under which conditions, specifying modulation conditions allows researchers to formulate expectations about how, systematically, manifestations of a phenomenon will differ from one another in different contexts. This can be achieved without having to formulate separate characterizations of separate phenomena for each manifestation that is unlike the next in a predictable way.

Together, the SRL case illustrates how performing induction tests, inhibition tests, and modulation tests each helped researchers to formulate and evaluate a phenomenon's characterization, starting from the insights from a number of interesting results of previous experiments. Consistent with the issues I presented in Section 3.3, researchers have a means of both evaluating which conditions of an experimental context are indispensable to the induction of a phenomenon's occurrence and determining in what other contexts they can expect the phenomenon to occur. Overall, these tests give the researchers a better understanding of what occurs and the circumstances in which it occurs. It is not enough to discover a single occurrence of a phenomenon; one must also evaluate the features and conditions of its occurrence to characterize a phenomenon.

3.6 Iterative self-vindication

After reviewing these strategies for characterizing the phenomenon SRL, the reader might be struck with the following kind of worry. While it is no doubt an improvement, the characterization that can be developed from these induction, inhibition, and modulation tests will not be exhaustive of the contexts in which the phenomenon may occur. Further, one might argue, no number of these tests could ever conclusively prove that we have characterized all and only the indispensable conditions for a phenomenon's occurrence, let alone all of the currently unknown factors that might inhibit or modulate its occurrence. This argument leads to the following conclusion: despite using the strategies that I described in Section 3.4, the SRL researchers have not resolved the issues I presented in Section 3.3.

While it is correct to say that the strategies that I have analyzed can never fully overcome the issues I have presented, this should not be taken to entail that these strategies are not beneficial for assessing what features constitute the phenomenon's occurrence and under what conditions these phenomena occur. Furthermore, and more importantly, the characterization the phenomenon that can be developed in light of these strategies can lead to more experiments to evaluate this characterization. By improving our characterization of a phenomenon, researchers can develop more thorough tests, which, in turn, allow for a better characterization of it. Herein lies the *self-vindicatory* character of the process. My use of the language of 'self-vindication' mirrors a point made by Hacking regarding the relation between theory and method in the laboratory sciences. He notes that theories and technology develop "that are mutually adjusted to each other," and, throughout the affair, they become "self-vindicating in the sense that any test of theory is against apparatus that has evolved in conjunction with it" (Hacking 1992). The adjustment allows the work

completed in the laboratory to improve the theory, as well as improve the methods by which researchers can detect the occurrence of phenomena in the laboratory.

Characterizing phenomena involves an analogous self-vindicatory character. While an initial characterization of a phenomenon likely will be incomplete and indicative of relics of particular experimental conditions, it is improved through additional experimental tests. These tests, in turn, are improved once the researchers have a better understanding of what tests they need to perform in order to provide evidence for a phenomenon's stability and repeatability. Thus, improvement of a characterization of a phenomenon can be garnered iteratively. This self-vindicatory character thus reflects the fact that the process to improve a phenomenon's characterization can lead to an improvement to the tests of it. This reflects one of the reasons why researchers characterize phenomena: not only do they create a better characterization of a phenomenon of interest, but they also increase their skills to characterize them further.

In the case of SRL, a better characterization of the phenomenon has allowed Jarbo and colleagues to develop additional experiments that are designed to test how SRL can be induced, inhibited, and modulated by factors that had not been investigated previously. For example, Jarbo and colleagues now aim to determine how factors related to the presentation of a spatial decision-making task – specifically, the contextual factors used to frame the task as representing a morally-loaded scenario like a military drone bombing – might affect the way in which participants learn from the rewards that they are provided when performing this task (Jarbo, Colaço, & Verstynen under review). This new set of experiments has been developed in light of SRL's characterization: these experiments rely on the expectation that penalty and variance factors can modulate or even inhibit SRL from occurring. Thus, by developing the characterization of SRL, these researchers are now able to test the occurrence of SRL in new experiments. This, in turn, will likely lead to

further developments of the characterization of SRL. Hence, through iteration, the experiments and the characterization of the phenomenon inform one another, leading to new and exciting research on the phenomenon.

3.7 Conclusion

In this chapter, I have detailed a case in which researchers characterized the phenomenon spatial reinforcement learning, drawing from reports of this phenomenon's manifestation in previous studies. I have introduced the main aims of characterizing phenomena, as well as two issues that pose a challenge for this process. In seeing how researchers in the SRL case met this challenge, I have analyzed three distinct strategies that were employed: induction tests, inhibition tests, and modulating tests. While these strategies cannot completely resolve the issues I have described, new research on this phenomenon's characterization and explanation is now possible because of these strategies, highlighting in no small way that the process of characterizing phenomena is both iterative and self-vindicatory.

4.0 Recharacterizing Scientific Phenomena

In this chapter, I investigate how researchers evaluate their characterizations of scientific phenomena. Characterizing phenomena is an important – albeit often overlooked – aspect of scientific research, as phenomena are targets of explanation and theorization. As a result, there is a lacuna in the literature regarding how researchers determine whether or not their characterization of a target phenomenon is appropriate for their aims. This issue has become apparent for accounts of scientific explanation that take phenomena to be explananda. In particular, philosophers who endorse mechanistic explanation suggest that the discovery of the mechanisms that explain a phenomenon can lead to its recharacterization. However, they fail to make clear how these explanations provide warrant for recharacterizing the phenomena that they explain. Drawing from cases of neurobiological research on potentiation phenomena, I argue that attempting to explain a phenomenon may provide reason to suspend judgment about its characterization, but this typically cannot provide warrant to recharacterize it. This is because an explanation cannot provide warrant to recharacterize its explanandum phenomenon if researchers cannot infer a phenomenon's characteristics from this explanation. To explicate this, I go beyond explanation – mechanistic or otherwise – to capture why researchers recharacterize scientific phenomena.

4.1 Introduction

A substantial body of literature in the philosophy of science identifies *scientific phenomena* as a target of scientific reasoning. A phenomenon is taken to consist in causal interactions; it is

stable and repeatable, allowing scientists to explain why they acquire the empirical results they do in individual studies and to theorize about what is common to each of its occurrences (Bogen & Woodward 1988; Woodward 1989). Several popular accounts of scientific explanation now take phenomena (so understood) to be what scientists explain. This includes mechanistic explanation (Machamer, Craver, & Darden 2000; Glennan 2002; Craver 2007; Bechtel & Richardson 2010), whose adherents accept as a "platitude" that "each mechanism has a phenomenon" (Garson 2017, 104). These accounts propose, roughly, that a phenomenon is explained by schematizing its underlying causal structure. For these accounts, "to characterize a phenomenon correctly and completely is a crucial step" for formulating an "acceptable mechanistic explanation" (Craver 2007, 128).

Characterizing phenomena is important to achieve practical aims as well. Phenomena are causally efficacious; their occurrences have measurable effects (Hacking 1983). In fields like biology, where researchers aim to develop therapies for disease, researchers want to exploit the occurrences of phenomena to control the systems in which they occur (Craver 2007, 1). Determining the characteristics of a phenomenon allows researchers to determine how inducing it affects other components of the system. Researchers aim to determine what *causal role* a phenomenon plays in the circumstances in which it occurs, by determining what can cause this phenomenon to occur and what effects its occurrence can have.

Despite the importance of phenomena, until recently, little has been said about how their characterizations are *evaluated*. One exception to this comes from philosophers who endorse mechanistic explanation. While these "mechanists" accept that characterizing a phenomenon is necessary for explaining it, many also argue that a phenomenon can be recharacterized based on its explanation. At first glance, scientific practice seems consistent with this argument. As the

proliferation of memory phenomena suggests (Squire 2009), discovering that there are distinct mechanistic explanations for what was initially characterized as a single phenomenon can lead to its recharacterization. Thus, there is a question regarding why it is often the case that phenomena – the targets of explanation – are recharacterized following researchers' attempts to explain them. The mechanists suggest that this is due to the "feedback" between explaining and characterizing: mechanistic explanations provide insight into the adequacy of the target phenomenon's characterization (Bechtel & Richardson 2010, 238).

I disagree. In order to determine what role explaining phenomena can play in recharacterizing them, I address a more basic question: *why do researchers recharacterize phenomena?* In answering this question, I explicate why researchers often do not recharacterize phenomena based on how they explain them. This because researchers typically cannot infer the characteristics of a phenomenon from its explanation. That being said, explaining a phenomenon is not entirely irrelevant to evaluating its characterization. This highlights a portion of scientific reasoning that has been underexplored in this literature. When researchers devise explanations of a phenomenon that raise concerns about the adequacy of its characterization, they *suspend judgment*, rather than recharacterize it. These epistemic attitudes are distinct: researchers proceed differently when they suspend judgment versus when they recharacterize.

To defend my position, I present a case study of the neural phenomenon long-term potentiation (LTP), a popular case amongst the mechanists. I introduce a mechanist account of characterizing phenomena in Section 4.2. In Section 4.3, I address the evaluation of a phenomenon's characterization, and I identify the difference between recharacterizing a phenomenon and suspending judgment about it. In Section 4.4, I discuss how LTP was initially characterized. I discuss how its characterization was evaluated following researchers' attempts to

explain LTP in Section 4.5. In Section 4.6, I explain why some evidence provides warrant to recharacterize a phenomenon, while other evidence provides reason to suspend judgment about it. I show that evaluating a phenomenon's characterization requires researchers to directly determine the adequacy of the characteristics that they specify. I conclude by examining the limited sense in which explaining a phenomenon could provide warrant for recharacterizing it.

4.2 Characterizing a phenomenon

To characterize a phenomenon, researchers identify a set of co-occurring features and the conditions under which these features occur. They formulate a representation of these features along with these conditions, which I refer to as the *characterization of a phenomenon.* An account of this process comes from Craver and Darden.[16]

According to Craver and Darden, researchers characterize a phenomenon to develop a "description of the behavior or product of the mechanism as a whole" (8), where mechanisms are "entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions" (Machamer, Darden, & Craver 2000). Importance is placed on characterizing phenomena in the project of mechanistic explanation, as "to discover a mechanism, one must first identify a phenomenon to explain" (Craver & Kaplan 2014, 275). A phenomenon's characterization "prunes the space of possible mechanisms": it constrains the search for *mechanistic details*, which causally underwrite this phenomenon's occurrence (52). The mechanists distinguish between characterizing a phenomenon and explaining it. If one

[16] Unless otherwise specified, all references are to Craver & Darden 2013.

characterizes a phenomenon, one need not be able to explain how it occurs or why it occurs under certain conditions. For mechanists, explaining a phenomenon consists in schematizing what causally underwrites its occurrence. Because they are explanatory and not descriptive, mechanistic details are not included in a phenomenon's characterization.

Craver and Darden take characterizing a phenomenon to have two parts. First, researchers determine the features that constitute the phenomenon of interest (the *features* of a phenomenon). Second, researchers establish the conditions under which these features occur (the *conditions* of a phenomenon). This includes precipitating conditions, which are "all of the many sets of conditions sufficient to make the phenomenon come about" (56), and inhibiting conditions, which are "the conditions under which the phenomenon fails or is blocked from occurring" (57). Characterizing phenomena in this way is not unique to the mechanists. Other accounts of explanation (Woodward 2003) and modeling (Batterman 2005) appeal to phenomena (so understood). All of these philosophers take phenomena to have "stable recurrent features which can be produced regularly by some manageably small set of factors" (Woodward 1989, 395). These qualities of phenomena make them "potential objects of explanation and prediction by general theory" (Woodward 1989, 393).

On Craver and Darden's account, "a purported phenomenon might be recharacterized or discarded entirely as one learns more about the underlying mechanisms" (62). This is because the discovery of mechanistic details can reveal one of two errors: either there has been a "lumping together [of] two separable phenomena produced by different mechanisms" or a "splitting [of] one phenomenon into many," where a phenomenon is mischaracterized as multiple phenomena (60). In either case, these details are *discrepant*: there is not a single characterization of a phenomenon and a single mechanistic explanation for its occurrence. On this account, a discrepancy between

the characterized phenomenon and its mechanistic explanation should result in the phenomenon's recharacterization. *Recharacterization* occurs when a phenomenon's characterization is rejected, revised or changed to specify different characteristics, and subsequently accepted in this new form. This includes *splitting recharacterizations*, in which a characterization is rejected and subsequently fractured in its revision, resulting in characterizations of distinct phenomena.

The idea that an explanation can result in recharacterizing the explanandum phenomenon is found elsewhere in the mechanist literature. Craver suggests that "dissociating realizers mandates splitting kinds," suggesting that a phenomenon's characterization should be split if its manifestations are explained by distinct mechanisms (Craver 2004, 962). Craver and Darden's account echoes Craver's norms of explanation, which suggest that "lumping errors" are resolved when phenomena initially thought to be unitary are shown to be "composed of a variety of distinct and dissociable processes" (2007, 124). Other mechanists like Bechtel and Richardson claim that phenomena should be "reconstituted" following their explanation, as "our conception of what needs explaining… is shaped by the explanations and the models we develop" (2010, 194). They argue that mechanistic explanations require that the explananda phenomena "must themselves be understood" in different terms (Bechtel & Richardson 2010, 239). In addition, they argue that "there is feedback between phenomena and explanatory models" (Bechtel & Richardson 2010, 238). Thus, several mechanists have suggested that an explanation provides warrant for recharacterizing its explanandum, with some suggesting that discrepant mechanistic details provide warrant to split this explanandum.

An alleged example of this is the split of procedural and declarative memory following research on patient H.M. After a surgery that removed his hippocampus, H.M. could not form new memories of events. However, H.M. could learn new tasks – he could form new procedural

memories – though he would have no explicit memory of them (Scoville & Milner 1957). Mechanists suggest that H.M. provided evidence that "the mechanisms for procedural memory and declarative memory are distinct," and treating memory as a unitary phenomenon amounts to "lumping together what we now know to be two distinct kinds of memory" (Craver 2004, 961). Along with other research, the study of H.M. "seemingly removed any doubt" that there are multiple memory phenomena, which ought to be characterized differently (Craver 2004, 962).

As a summation of the mechanistic perspective on scientific phenomena, Craver and Darden's account has several virtues. They correctly assert that characterizing a phenomenon includes specifying the conditions under which it can be induced or inhibited, as well as specifying what the phenomenon itself is like. Determining a phenomenon's features and how it is induced or inhibited are both needed to determine its causal role. Craver and Darden are also right that characterizing a phenomenon is distinct from explaining it, as explanatory questions about the phenomenon – why questions and how questions – cannot be answered with its characterization. And, most importantly for this article, they rightly identify that phenomena are often recharacterized following their mechanistic explanations.

Indeed, these mechanists might be onto something. If what we characterize as manifestations of the same phenomenon turn out to be underwritten by distinct causal structures, it seems reasonable to think that these structures may cue us into differences that should lead us to question whether or not we ought to characterize them as distinct phenomena, which we should theorize about and control differently. But, does this mean that mechanistic details provide warrant for recharacterizing explananda phenomena? To answer this question, we must go beyond phenomena as they relate to mechanisms and instead analyze their characterizations.

4.3 Epistemic attitudes towards a characterization of a phenomenon

To investigate when a phenomenon should be recharacterized, I first discuss the reasons why these characterizations are accepted. As mentioned in the introduction, philosophers have rightly identified that researchers characterize phenomena for theoretical and practical aims. To fulfill these aims, researchers seek to accept accurate characterizations of phenomena. A phenomenon's characterization is accurate to the extent that (1) the features specified in this characterization are consistent with the features that co-occur, and (2) the conditions specified to precipitate or inhibit the characterized phenomenon are consistent with the conditions under which the occurrence is observed. Accuracy in this sense is obtainable even though a phenomenon's characterization is abstracted away from many of the details of its occurrences. This is because this phenomenon can potentially occur in different contexts, so long as these contexts include the conditions of its occurrence that are specified in its characterization. In other words, because a phenomenon is characterized to represent what is common to each of its occurrences, the idiosyncratic characteristics of each of its manifestations are not represented. All that is needed is a characterization that describes what features will occur under a specified set of conditions. With an accurate characterization, researchers can explain and predict what happens when the phenomenon manifests, reason about the characteristics every manifestation has, and control its occurrences. An inaccurate characterization of a phenomenon will not explain or predict what consistently occurs, and it will also lead to ineffective interventions, limiting researchers' control over the target system.

Researchers assess the accuracy of a phenomenon's characterization in light of evidence related to the phenomenon's constitutive features, along with evidence related to the conditions under which it occurs. If the evidence for a phenomenon's characterization is consistent with what

is specified in this characterization, researchers should accept it. If they accept it, they should theorize and experiment in a way that is consistent with the characterized phenomenon's occurrence.[17] If more evidence is produced, then researchers should reassess their acceptance of this characterization.

It is not necessary that researchers either accept or reject a phenomenon's characterization upon reassessment, as there are more than two epistemic attitudes one could take towards it. There is a third kind of attitude, if 'attitude' is the right word for it: upon assessment of evidence, researchers may *suspend judgment* about a phenomenon's characterization.[18] Suspending judgment involves more than merely neither accepting nor rejecting, as it seems wrong to say that researchers suspend judgment about claims that have not been considered (Friedman 2013, 168). Rather, judgment is suspended when researchers consider a characterization of a phenomenon, and they cannot determine what other attitude to adopt towards it. Instead, they adopt a "neutral state" towards this characterization (Friedman 2017, 307).

One might question the relevance of the difference between rejecting a characterization of a phenomenon and suspending judgment about it, given that, whatever other differences there are between them, they both amount to not accepting it. The answer lies in what researchers *do* in light of holding these respective attitudes. When deliberating on a phenomenon's characterization, a researcher "reflects on his evidence (or lack thereof) for and against" the characterization (Friedman 2013, 179). But, if the researcher "doesn't take his evidence to settle the matter," then

[17] This is an idealization; researchers do not have strict criteria when they have sufficient evidence to accept a phenomenon's characterization nor is there a sharp distinction between accurate and inaccurate characterizations. Nevertheless, researchers converge upon a judgment of what evidence is sufficient and what degree of accuracy is desired.

[18] These epistemic attitudes can be conceived of as continuous rather than as a trichotomy. Something like tweaking could be thought of as a practice based on an attitude that lies between suspending judgment and rejection.

they suspend judgment (Friedman 2013, 179). This is not where the inquiry ends, however. Suspending judgment has a "push towards its own demise: in suspending we ask a question and (at least in some minimal sense) seek an answer" (Friedman 2017, 316). This is because researchers need to characterize phenomena to achieve their aims. Thus, researchers want to resolve their suspended judgment. To do this, they perform more research.

Herein lies the difference between rejecting and suspending judgment. When researchers suspend judgment, they continue to inquire about a phenomenon's existing characterization. If this occurs, researchers investigate the putative phenomenon further so that they can evaluate its characterization's accuracy in light of more definitive evidence. When they reject this characterization, they move on to new ones, having reason to think that it is not accurate and therefore cannot be "saved" in its existing formulation. It is the latter that is the first step of recharacterization: rejecting a phenomenon's existing characterization as deficient. Thus, while, in practice, researchers may not explicitly adopt epistemic attitudes about phenomena, we can learn about researchers' reasoning from their response to acquiring new evidence.

With an understanding of the differences between rejecting and suspending judgment, the nature of the disagreement between the mechanists and myself becomes salient. If researchers should reject a phenomenon's characterization based on their explanation of it, then mechanistic details provide warrant for recharacterization. If they instead should suspend judgment, then, whatever epistemic role mechanistic details play, they do not provide this warrant. Mechanists suggest the former: researchers should reject a phenomenon's characterization if they discover that distinct mechanisms underwrite its manifestations. This constitutes a "lumping error." Further, they should revise its characterization so that they have characterizations of distinct phenomena, each of which is consistent with a distinct mechanistic explanation. As these together constitute

recharacterizing a phenomenon, this suggests that mechanistic details not only provide warrant to reject a phenomenon's existing characterization but also to revise it. If this position is correct, then researchers should recharacterize a phenomenon in light of what they learn from explaining it, following which they should theorize and experiment in a way that is consistent with the recharacterized phenomenon's occurrence.

My position, by contrast, is that the distinction between rejecting and suspending judgment provides a way of understanding how explanation can be relevant to evaluating the characterization of the phenomenon to be explained, without committing to the idea that explanation provides warrant to recharacterize it. When researchers have the theoretical and practical aims that I describe, I expect them to perform additional studies to evaluate the accuracy of a phenomenon's characterization in light of discovering discrepant mechanistic details. This is because these discoveries typically do not provide warrant to reject these characterizations, let alone to recharacterize them. I defend my position by showing that researchers inquire about a phenomenon's characterization following the discovery of discrepant mechanistic details, which is best explained by the idea that researchers suspend judgment about this characterization in light of these discrepancies. This not only undermines the mechanists' claims, but it also provides the basis for determining, more systematically, what provides warrant to recharacterize a phenomenon.

4.4 Formulating and evaluating a phenomenon's characterization

To show how my position captures how researchers evaluate characterizations of scientific phenomena, I describe the initial characterization of long-term potentiation (LTP). LTP is a

neurobiological phenomenon that involves the potentiation of a neural synapse that lasts on a timescale of minutes or more. When a synapse is potentiated, there is an increase in the strength and number of excitatory responses of postsynaptic neurons following the activation of the presynaptic neurons. Its discovery is credited to Terje Lømo (Craver 2003).

Lømo demonstrated that the stimulation of presynaptic neurons in the hippocampi of anesthetized rabbits resulted in minutes-long potentiation of postsynaptic neuron populations (1966).[19] He applied high-voltage inputs to the presynaptic cells in quick succession. Changes in the responses of the population of postsynaptic cells were recorded following stimulation. The intensity and number of excitatory postsynaptic responses increased when the presynaptic cells were stimulated, while the latency of the responses shortened. Lømo described the excitation of the presynaptic cell population, stimulated by his electrode, as a precipitating condition of this potentiation phenomenon. He also described how varying stimulation affects the intensity of the potentiation, as well as the time needed for the system to return to its initial state. Together, his specification of the experimental conditions that induced the potentiation, along with the description of the potentiation itself, consisted in the initial characterization of the phenomenon.

Following its formulation, this characterization of LTP was tested to determine its accuracy. In collaboration with Tim Bliss, Lømo determined the strength of the stimulation required to induce the effect (1973). Lømo and Bliss modified the experimental protocol they used to induce potentiation, in order to understand the effect of the placement of the stimulating electrode. The researchers found no evidence that LTP resulted from these aspects of the experiment. This gave researchers a clearer sense of when LTP occurs. In distinct experiments

[19] Research on LTP was motivated by research – such as studies on H.M. – that suggested that the hippocampus is responsible for memory formation. However, Lømo himself did not test the relation between LTP and memory (Lømo 2003).

with Tony Gardner-Medwin, Bliss performed experiments that were identical to Lømo's, except for the fact that the rabbits were not anesthetized. They observed the features of LTP "without anesthesia and under the more stable conditions," thereby precipitating the phenomenon in a way unspecified by Lømo initially (Bliss & Gardner-Medwin 1973, 358). Thus, Bliss and Gardner-Medwin provided evidence for the accuracy of Lømo's characterization of LTP – Lømo correctly identified some of conditions of its occurrence – but also identified other conditions that precipitate the phenomenon. With these findings, the researchers revised the characterization of LTP to convey that the phenomenon occurs under "approximately normal physiological conditions," specifying that LTP can be precipitated in physiologically-active organisms (Bliss & Gardner-Medwin 1973, 358).

The work from Lømo, Bliss, and Gardner-Medwin informed what I call the *standard characterization* of LTP, which includes its precipitating and inhibiting conditions. Thus, initial research on characterizing LTP took the following trajectory. Lømo formulated a characterization of a phenomenon. Bliss and Gardner-Medwin induced the phenomenon under different but consistent conditions. The researchers tested what conditions were needed to precipitate the features specified in LTP's characterization: their findings provided reason to accept that Lømo's initial characterization was accurate in the sense that none of features and conditions that Lømo specified failed to occur when the phenomenon was induced, though it could be made more detailed. Overall, in this initial phase, evaluating LTP's characterization amounted to the direct test of the specified characteristics of this phenomenon.

4.5 Recharacterization and explanation

The previous section shows how researchers use evidence about a phenomenon's features and the conditions under which it occurs to formulate its characterization. But, what about explanation? Moving forward in time, I address the evaluation of LTP's characterization following its mechanistic explanation. In attempts to mechanistically explain LTP, researchers discovered two kinds of discrepancies. These kinds of discrepancies are similar in the sense that they each involved the discovery that distinct sets of mechanistic details underwrite manifestations of what were thought to be the same phenomenon. I contrast the evidential role of empirical findings related to the features of LTP and the conditions under which it occurs with these two cases, in each of which there was a discovery of discrepant mechanistic details underwriting LTP. While they both influenced the evaluation of LTP's characterization, neither mechanistic explanation alone provided reason to recharacterize it.

In the first case of a discrepancy, the molecular details of LTP in the region investigated by Lømo and colleagues were discovered, and the role of the NMDA glutamate receptor was identified (Collingridge, Kehl, & McLennan 1983). In the hippocampus, LTP has been explained by being causally underwritten by the increase of glutamate receptors in the postsynaptic neurons. As a result of stimulation, more glutamate is released into the synaptic cleft. The release of glutamate results in AMPA and NMDA receptors reacting. It is the activation of NMDA receptors that sets off a chain of reactions that bring additional AMPA receptors to the dendrites of the postsynaptic cells. This results in the potentiation of the neuron, which is called early-phase LTP. If researchers continue to stimulate, the postsynaptic cell will undergo changes in transcription enzymes, which result in the production of the AMPA receptors, which is what is called late-phase

LTP. Thus, researchers discovered that distinct sets of mechanistic details underwrite what were thought to be all manifestations of LTP, and the sets do not always occur in tandem.

A second kind of discrepancy was discovered concurrent with research on the phases of LTP. In brain areas other than the hippocampus, and thus outside of the purview of Lømo and colleagues' investigations, researchers discovered potentiation effects that were initially thought to be consistent with the standard characterization of LTP. However, researchers found that these potentiation effects are underwritten by an inconsistent set of mechanistic details, which involve the interaction of distinct neurotransmitters and receptors (see Sullivan 2017 for a historical review). Calling the standard characterization "NMDAR-dependent LTP" after the receptor in part underwriting its occurrence, researchers aimed to determine whether or not potentiation effects in other parts of the brain were manifestations of the same phenomenon as described by the standard characterization. Again, different mechanistic details were found to underlie what were initially thought to be manifestations of the same phenomenon in different areas of the brain.

Given that these discoveries suggest that different manifestations of LTP are best explained according to distinct mechanistic schemata, we can examine what role these discoveries played in evaluating LTP's characterization. One set of discrepant mechanistic details relates to the phases of LTP; the other set relates to the manifestations called 'LTP' in other areas of the brain. If the discovery of discrepant mechanistic details provides warrant for recharacterizing a phenomenon, we should expect that these discrepancies provided warrant to split the standard characterization of LTP. In other words, researchers should have characterized the phases as distinct phenomena and characterized the manifestations in other parts of the brain as distinct phenomena.

However, this is not what occurred. Despite agreement that discrepant mechanistic details were discovered, and thus that distinct occurrences of LTP are explained differently, there was

disagreement about whether (and if so how) LTP should be recharacterized. The discovery of both kinds of discrepant mechanistic details led to a new round of experiments so that researchers could determine whether or not there are differences between the features of the instances of what were initially thought to be manifestations of the same phenomenon or the conditions under which they occur. While different decisions were made in light of the different discrepancies, at the heart of each of these debates, we find the same reasoning at play.

Related to the first discrepancy, the differences between the mechanistic details associated with the respective phases were investigated so that researchers could determine if there are differences between the features of LTP or the conditions of its manifestations when LTP is underwritten by one set of mechanistic details as opposed to the other. Some neuroscientists thought that the standard characterization of LTP was consistent with the fact that different mechanisms underwrite LTP's phases. Both sets of mechanistic details were experimentally determined to play a role in the occurrence of LTP. While explained differently, LTP increases the number of receptors in the cell membrane in all instances, which results in increased signaling between neurons. Thus, in the aftermath of the discovery of the different phases of LTP, many neuroscientists still accepted its standard characterization.

Even the researchers who have argued that LTP's phases should be characterized as distinct phenomena do not take the discovery of discrepant mechanistic details alone to provide warrant for their position. Researchers who challenge the orthodoxy of the standard characterization, and argue that it is problematic to "ignore [the] possible distinction" between the phases, do not base their challenge on the by all accounts agreed-upon different mechanistic explanations schematized for the phases (Huang 1998). Rather, they argue that "the induction requirements of the two phases… differ," and potentiation persists to different lengths of time (Huang 1998). Thus, debate

over the standard characterization of LTP qua phases of LTP is a debate over whether or not there are differences between the features of instances counted as LTP. In other words, the debate is over the characteristics of different instances called 'LTP,' rather than whether or not they are explained differently.

Related to the second discrepancy, research on potentiation that occurred in different neural regions led researchers to determine that the standard characterization of LTP did not accurately characterize the features or conditions that can be measured in all of the contexts in which potentiation was induced. Researchers have since demonstrated that potentiation phenomena in other brain regions are different from one another: there are forms of LTP that "may share some, but certainly not all, of the properties and mechanisms of NMDAR-dependent LTP" (Malenka & Bear 2004, 5). Because of these differences, some argued that researchers should "define at which specific synapses these phenomena are being studied… and how they are being triggered," indicating that there are differences between the characteristics of what were initially counted as manifestations of the same phenomenon (Malenka & Bear 2004, 5; see also Blundon & Zakharenko 2008).

Following the discovery of the second kind of discrepancy, LTP was recharacterized. Specifically, it was split, leading to characterizations of distinct phenomena that occur throughout the brain. These distinct phenomena are explained in terms of distinct mechanistic schemata: some forms of LTP do not involve the NMDA receptor or involve activities in the presynaptic cell rather than just in the postsynaptic cell, as is the case in NMDAR-dependent LTP (Malenka & Bear 2004). However, like for the first kind of discrepancy, the discovery of discrepant mechanistic details alone did not provide reason to recharacterize LTP. In light of the discovery of mechanistic details, researchers performed additional experiments to determine how the forms of LTP are

different from one another, beyond simply being explained differently. Based on the results of these experiments, researchers determined that different instances once called 'LTP' have different precipitating and inhibiting conditions. For example, the stimulation that triggers NMDAR-dependent LTP can inhibit other potentiation phenomena (Kullmann & Lamsa 2008). The potentiation strength varies between these phenomena as well (Abraham 2003).

Thus, regarding potentiation phenomena throughout the brain, researchers determined that there are differences between the characteristics of instances initially thought to be all manifestations of LTP. With their experiments, researchers demonstrated the differences between what were once thought to be manifestations of the same phenomenon. These splitting recharacterizations were no mere tweaks: the features and conditions were both mischaracterized in different cases. Many of these revisions were radical, reformulating precipitating conditions into inhibiting ones and vice versa.

In overview, the discovery of both sets of discrepant mechanistic details led researchers to evaluate how they had characterized LTP and investigate the manifestations of what were initially characterized as the same phenomenon. For the first discrepancy, there remains debate about whether or not the phases of LTP ought to be characterized as distinct phenomena. For the second discrepancy, most researchers agree that there are distinct potentiation phenomena that should be characterized as such. In both cases, however, the reasoning was the same. Researchers discovered discrepant mechanistic details, which led them to evaluate its once-accepted characterization. LTP was recharacterized only when researchers found differences between the features or conditions of phenomena explained by distinct mechanisms.

Contemporary discussions of the standard characterization of LTP reflect this reasoning as well: debates regarding whether the phases of LTP are the same phenomenon, are parts of a single

phenomenon, or are distinct phenomena are in part *motivated* by the differences between the mechanistic details that underwrite each phase, but these debates are not *resolved* by the discovery of these details. Instead, researchers aim to resolve these debates by investigating (1) the conditions that precipitate or inhibit the potentiation, (2) the persistence and strength of the potentiation, and (3) the effect of the potentiation in the neural systems in which it occurs. These correspond to the phenomenon's conditions, features, and causal role respectively.

4.6 Recharacterizing a phenomenon

The case of LTP – starting from the formulation of its characterization and continuing to researchers' attempts to mechanistically explain it – illustrates how a scientific phenomenon's characterization is evaluated. To draw normative implications from this example, I reflect on how the evaluation of a phenomenon's characterization relates to the theoretical and practical aims of researchers who characterize it. I examine why researchers typically reexamine, rather than reject, their characterizations in light of the discovery of discrepant mechanistic details. By aligning a phenomenon's characterization with the researchers' aims, I show that my conclusions are not unique to the case of LTP.

4.6.1 Reasons to Recharacterize a Phenomenon

The discovery of mechanistic details preceded the splitting recharacterization of LTP, but this fact fails to capture what provided warrant for this recharacterization. Following each discovery, researchers performed additional experiments; with the findings from these

experiments, they evaluated their characterization of LTP. With this case in mind, a clearer picture of when phenomena should be recharacterized can be discerned.

Based on the actions of the researchers, what provided warrant to recharacterize LTP? In its evaluation following attempts at its explanation, LTP was recharacterized based on evidence produced in experiments that had one of the following forms: researchers either observed the features of LTP under conditions that are inconsistent with those specified in the characterization of LTP or produced the conditions of LTP and observed features that are different from those specified in the characterization of LTP. The results of these kinds of experiments relate directly to the components of a phenomenon's existing characterization. They serve as a test of the specified characteristics of the phenomenon. In the case of LTP, the evidence provided warrant to reject the existing characterization and to recharacterize the phenomenon so as to accommodate the features or conditions that they observed.

More generally, when researchers have theoretical and practical aims, a phenomenon should be recharacterized when the descriptions researchers make about its occurrences based on what is specified in its characterization are not consistent with the evidence that they collect. Either inconsistent features occur under the specified conditions of the phenomenon's occurrence, or the features occur under conditions that are inconsistent from those that are specified. It is under these circumstances that a phenomenon's characterization should be rejected. If this evidence provides reason to think that these specified characteristics should be changed, so that the new characterization is consistent with the available evidence, then a phenomenon's characterization should be revised and accepted in its new form. Combined, this is when a phenomenon should be recharacterized.

What role did the discovery of mechanistic details play in recharacterizing LTP? On the negative side, the explanation of instances called ‘LTP’ did not provide evidence inconsistent with the specified characteristics of LTP. As a result, the differences between mechanistic details that were discovered did not result in researchers giving up on its characterization as it was formulated at the time. On the positive side, the discoveries that resulted from attempts to explain LTP led researchers to evaluate the standard characterization’s accuracy. What accounts for this change from accepting to reexamining a phenomenon’s characterization? The answer lies in the fact that differences between the entities or activities of distinct mechanistic details suggests that their respective initiations might be sensitive to different external causes or that different external effects might be sensitive to their respective terminations. If either of these possibilities were the case, then manifestations of the phenomenon in question underwritten by distinct mechanistic details would have different features or conditions. The fact that researchers discover mechanistic differences that are suggestive of potential inconsistencies between instances of this phenomenon led them to evaluate this characterization.

Why does the discovery of discrepant mechanistic details provide reason to reexamine the characterization of a phenomenon they explain but not provide warrant for its recharacterization? The answer is based on the limitations of what can be inferred about phenomena from the mechanistic details that underwrite them. The fact that researchers determine that distinct mechanistic details underwrite instances characterized as the same phenomenon means that there may be contexts in which the characterization inaccurately describes the occurrence under investigation. The differences between the mechanistic details suggest that the underlying mechanisms may initiate under different conditions or may produce different features through the distinct activities of their entities. When discrepant mechanistic details are discovered, what the

mechanistic details underwrite may be different from one another in ways that are relevant to a phenomenon's characterization.

However, the discovery of discrepant mechanistic details is not in and of itself inconsistent with a single characterization of the phenomenon. This is because distinct mechanisms may be initiated by the same conditions or may produce the same outcome. Given that, for theoretical and practical aims, a characterization of a phenomenon *only* describes the features of a phenomenon and the conditions under which it occurs, the characteristics of the mechanistic details that underwrite it may make no difference to its characterization. Thus, if there is no difference, the phenomenon as characterized may occur even if it is best explained by distinct mechanistic schemata.

This is why it is useful to distinguish between suspending judgment and rejecting a phenomenon's characterization. Some findings provide reason to think that a phenomenon's characterization is deficient from the perspective of their aims – in the LTP case, deficient in terms of accuracy – and therefore should be rejected or even recharacterized. Other findings are merely suggestive: they provide reason to suspend judgment and evaluate a phenomenon's characterization. thus providing warrant to either accept or reject it. Thus, the mechanists correctly identify that explaining a phenomenon can lead to recharacterizing it. However, the limitation of their accounts results from a failure to identify the nature of this relation. Even if, as a matter of historical fact, findings that are suggestive of issues with a phenomenon's characterization often precede research that reveals that it is deficient, this need not be the case. So long as it is possible to acquire more evidence, researchers can suspend judgment about scientific claims and inquire into their adequacy. Following this inquiry, researchers can acquire evidence that provides warrant to recharacterize a phenomenon.

4.6.2 The Aims of Characterizing a Phenomenon

My position that researchers suspend judgment about a phenomenon's characterization following the discovery of discrepant mechanistic details captures what researchers investigating LTP actually did. This history is best explained by the fact that suspending judgment about a phenomenon's characterization in this kind of situation is more in line with researchers' aims when compared to simply rejecting it. In other words, if researchers have the theoretical and practical aims that I have described, then they care about the accuracy of their phenomenon's characterization, and, as a result, the discovery of mechanistic details alone typically will not provide warrant to recharacterize it.

Suspending judgment but not rejecting a phenomenon's characterization in light of discrepant mechanistic details aligns well with researchers' theoretical aims. To evaluate a general theory about the phenomena that manifest in a target system when investigating "objects of research" (Feest 2017), researchers identify the phenomena present in the system under investigation, and they determine whether or not this theory predicts that the phenomenon as characterized will occur. If discrepant mechanistic details are discovered, researchers must investigate whether or not the differences between the mechanistic details are relevant to the causal role of what they each underwrite. However, from a theoretical point of view, making a typological distinction between phenomena with the same conditions and features – and thus play the same causal role – is counterproductive, even if instances are explained differently. This is because the aim is to theorize about a type-level characterization of what constitutes the phenomenon and what it can do in the target system. The ability to theorize about this type-level characterization is undermined if researchers split this characterization along mechanistic lines, even though the phenomenon in question plays the same causal role.

The fact that researchers suspend judgment rather than reject a phenomenon's characterization in light of the discovery of discrepant mechanistic details also aligns with their practical aims. Phenomena are characterized so that they can be exploited to induce changes in the systems in which they occur. To induce the desired changes, researchers must determine what conditions precipitate or inhibit the phenomenon of interest and what effects the phenomenon can have. It is these characteristics that researchers are interested in when trying to control phenomena. Characterizing as different what can play the same causal role is counterproductive to inducing changes to fulfill these practical aims. Thus, from a practical perspective, a phenomenon should only be recharacterized to reflect differences in features of the phenomenon and conditions under which it occurs, as these differences are what make a difference when exploiting the phenomenon's occurrence to cause other changes in the system.

The point of all of this is not that these are the only aims that researchers could have, nor is it the case that there could be no other criteria for characterizing phenomena. Rather, what is important is that these aims are common and intuitive for researchers to hold, and they are fulfilled via accuracy.[20] Given that explanation typically does not provide evidence about an explanandum phenomenon's characteristics, it is not because of evidence of mechanistic details that researchers recharacterize phenomena.

[20] LTP is not the only example of recharacterization in the way that I describe. The case of H.M. is not one in which a phenomenon's characterization was split based on mechanistic explanations. This is because, short of determining that procedural memory did not involve the activity of the hippocampus, the distinction of the two was determined before either were explained. It seems more plausible that procedural and declarative memory were split because the phenomena associated with them were distinguished from one another.

4.6.3 What Would it Take for Explanation to Provide Warrant for Recharacterization?

I have argued that mechanistic details typically do not provide warrant to recharacterize a phenomenon, at least for the commonly-shared aims that I have described. What would need to be the case for mechanistic details to provide this warrant? The answer is that mechanistic details would provide warrant to recharacterize an explanandum phenomenon if the *characteristics of the schematized mechanisms* could be independently demonstrated to be inconsistent with its specified features or the conditions of this phenomenon's occurrence.

This claim rests on characterizing a mechanism's components as well as its set-up and termination conditions. In establishing these set-up and termination conditions, researchers may determine that their characterization of the conditions that precipitate the occurrence of the phenomenon are inconsistent with the conditions that initiate the mechanism that causally underwrites the phenomenon, and they may determine how to better characterize these conditions in light of what is known about the mechanism's initiation conditions. For instance, researchers might discover that the conditions that initiate one of the mechanisms cannot concurrently occur with the precipitating conditions specified in the characterization of the phenomenon. This would provide evidence inconsistent with a phenomenon's characterization, which would provide reason to reject it, given the differences between what initiates each mechanism and the precipitating conditions that are specified in its existing characterization.

However, pointing out this way in which mechanistic details could provide warrant for recharacterizing phenomena reveals why it is not typically the case that they do so. The discovery of discrepant mechanistic details need not tell researchers about the relationship between the mechanism's input and the conditions specified in the phenomenon's characterization. Likewise, these details need not tell researchers about the relationship between the mechanism's output and

the phenomenon's features. Furthermore, rejecting an existing characterization of a phenomenon resolves only part of the inquiry. To *recharacterize* a phenomenon, researchers must also revise this existing characterization. The fact that distinct mechanistic details causally underwrite a single characterized phenomenon does not provide evidence for how the characterization should be revised, unless it also provides details about how researchers should modify their specification of its features and conditions.

One might object to this conclusion by arguing that differences between mechanistic details entail that they must explain distinct phenomena. If one were committed to the view that every phenomenon is underwritten by a unique mechanism, then schematizing distinct mechanisms would be sufficient for recharacterizing the phenomenon they underwrite or at least would be sufficient for rejecting its existing characterization.

This objection can be read in two ways, and neither reading is defensible. First, this objection could suggest that phenomena should be differentiated whenever distinct mechanisms causally underwrite them. However, it is unclear why such a commitment is warranted. Researchers can determine the causal role phenomena play, and how they can exploit their causal properties, fulfilling their theoretical and practical aims. These aims are not fulfilled if researchers treat instances with the same causal role as distinct phenomena. Second, the objection could suggest that, if there are differences between mechanisms, then the instances that are characterized as the same phenomenon must be different in ways relevant to their characterization. However, that there is a difference does not entail that this difference must be relevant to counting something as a manifestation of a phenomenon. This is because a phenomenon's characterization describes what set of conditions and features are shared amongst its manifestations. The shared features of the phenomenon may not vary if distinct mechanisms underlie them. This is why researchers

identify the shared characteristics of a phenomenon's manifestations: what underwrites the specified features that co-occur under the specified conditions is not relevant unless they make a difference to their co-occurrence or causal role.

What if one objects in the other direction? Craver has expressed conventionalist leanings on the topic, suggesting that "human perspective enters into decisions about which mechanisms matter for splitting or lumping a putative kind" (2009, 585). He also argues that "there is no objectively appropriate degree of abstraction for typing mechanisms" (Craver 2009, 589). If one were committed to this view, then convention is what matters both when determining whether or not mechanistic details provide warrant for recharacterizing explananda phenomena and when determining whether or not mechanistic details are discrepant with this characterization.

My emphasis on theoretical and practical aims should make it clear that I do not deny that convention matters when characterizing phenomena. These aims are "conventions" – popular conventions, which favor accuracy – but conventions nonetheless. That being said, it is important to recognize that the conventions regarding how phenomena are characterized need not be determined by the conventions regarding how they are explained. This is because characterizing a phenomenon is not always, let alone solely, done in the effort of explaining it. This is why there are cases, like LTP, where there is a characterization of a unitary phenomenon according to certain conventions but it is explained (according to different conventions) by distinct mechanistic schemata that constitutively underlie its respective manifestations. In these cases, there is nothing about the conventions of the researchers that should lead us to expect that these explanations will provide warrant to recharacterize the explanandum.

This fact supports the fundamental thesis of this chapter. Characterizing phenomena is an important scientific process, which has its own conventions. Its aims cannot be properly

understood, let alone satisfied, solely from the perspective of explanation. While characterizing and explaining phenomena are linked in the ways that I have discussed in this chapter, a satisfactory account of why phenomena are recharacterized cannot be obtained without investigating this process unto itself. This is what is missing in the philosophical accounts according to which phenomena are the targets of theory and control, which is why the literature on the topic can benefit from a reorientation from solely discussing how phenomena are explained and modeled to discussing how they are characterized in the first place.

4.7 Conclusion

I have provided an account of how characterizations of phenomena are evaluated in light of evidence related to the empirical commitments specified in these characterizations. In doing so, I have identified the sense in which phenomena can be recharacterized following the discovery of mechanistic details that causally underwrite the phenomenon's manifestations. What is in question is what role these details play. I have argued that the discovery of discrepant mechanistic details typically does not provide warrant for recharacterizing its explanandum phenomenon. Occurrences with different schematized mechanistic explanations can be characterized as manifestations of the same phenomenon if the differences between these schematized mechanisms do not result in differences between the features of the phenomenon or the conditions under which the features occur. Whether or not there are differences and what those differences are only can be determined with additional studies.

Nevertheless, the discovery of mechanistic details can provide reason to suspend judgment and reexamine a phenomenon's characterization. The differences between the mechanisms may

relate to differences between the features of a phenomenon or the conditions under which it occurs. This conclusion provides more general insight into when researchers should suspend judgment about empirical claims. When researchers have evidence that suggests that their claims are not adequate, they should suspend judgment and inquire into them. This inquiry is achieved by doing more research.

I do not deny that schematizing the mechanistic details that causally underwrite a phenomenon's occurrence is a viable way to explain it. Rather, I point out that determining a discrepancy in its mechanistic details typically does not tell researchers the right things about a phenomenon to warrant recharacterization. Evidence about the features of a phenomenon and the conditions under which it occurs is what matters to evaluating its characterization. This makes clear how researchers evaluate a phenomenon's characterization, which many philosophers, mechanistic and otherwise, view as integral to theorization, explanation, and control.

5.0 Rip It Up and Start Again: The Rejection of a Characterization of a Phenomenon

In this chapter, I investigate the nature of empirical findings that provide evidence for the characterization of a scientific phenomenon and the defeasible nature of this evidence. To do so, I explore an exemplary instance of the rejection of a characterization of a scientific phenomenon: *memory transfer*. I examine the reason why the characterization of memory transfer was rejected and analyze how this rejection tied to researchers' failures to resolve experimental issues relating to replication and confounds. I criticize the presentation of the case by Harry Collins and Trevor Pinch, who claim that no sufficient reason was provided to abandon research on memory transfer. I argue that skeptics about memory transfer adopted what I call a *defeater strategy*, in which researchers exploit the defeasibility of the evidence for a characterization of a phenomenon.

5.1 Introduction

The identification of phenomena is a critical scientific research activity, as it is responsible for the discovery and characterization of the types of events to be explained by theory. To fulfill their theoretical and practical aims, researchers set out to accept characterizations of phenomena when empirical findings are put forward in their favor. When a characterization of a phenomenon is accepted, researchers theorize and experiment in a way that is consistent with the existence of the phenomenon. However, many episodes in the history of science involve the abandonment of characterizations of phenomena that were once empirically promising. This raises a question:

under what circumstances do researchers reject a characterization of a scientific phenomenon, despite evidence that appears to support it?

In this chapter, I analyze the rejection of a phenomenon through the lens of two philosophical topics. The first topic relates to how empirical findings can serve as evidence for a characterized phenomenon and the defeasible nature of this evidence. The second topic relates to the strategies through which researchers test an existing characterization of a phenomenon. In this chapter, I investigate what I call the *defeater strategy*. A collection of experiments can be used to undercut the empirical findings thought to support the characterization of a phenomenon. By defeating all evidence, any empirically motivated reason for accepting a phenomenon as characterized is eliminated. With this strategy, researchers do not simply provide evidence to challenge a characterization of the phenomenon; they also demonstrate the faultiness of the experiments whose findings are thought to support the characterization.

I explore a case in which the characterization of a phenomenon was rejected: the research on memory transfer. This alleged phenomenon was described as the transfer of learned behavior by the insertion of tissue from a trained donor organism to an untrained receiver. It received a great deal of attention from scientists and the public alike, due to its implications and to researchers' use of sensational experiments involving cannibalism. Formulated and defended in light of empirical findings, the characterization of memory transfer was considered by some to be accurate; this led to a cottage industry about its characterization, its theoretical significance, and its underlying mechanisms. The research program was abandoned, and contemporary scientists generally consider the "phenomenon" to not exist.

The case of memory transfer has generated controversy in the history and philosophy of science. Sociologists and historians of science have questioned the motives of the scientific

community that abandoned research on memory transfer. For instance, Harry Collins and Trevor Pinch argue that there was no "decisive technical evidence" that disproved the existence of memory transfer and that research was abandoned due to disinterest in the purported phenomenon (1998, 25). Collins and Pinch present a powerful challenge to the alleged justification researchers had in rejecting memory transfer. They argue that there were deficiencies in the evidence put forward in opposition to memory transfer, and there is a discrepancy between the perceived decisiveness of evidence in opposition to memory transfer and the actual decisiveness of this evidence. They base their challenge on the fact that no evidence against memory transfer applies to all experiments whose findings were thought to provide support for the characterization of the alleged phenomenon.

Collins and Pinch's challenge is, at its core, one about the evidence required to reject a characterization of a phenomenon. They argue that there was no decisive evidence for the abandonment of research on memory transfer. I argue that the evidence provided against memory transfer was decisive for the rejection of the characterization of the alleged phenomenon. The only way to understand why there was decisive evidence, I argue, is to recognize the fact that scientific evidence is defeasible, and the defeat of evidence for memory transfer eliminated all reason to accept the characterization of memory transfer. By exploiting this fact, the defeater strategy provides a way to undermine evidence in favor of scientific claims, including characterizations of phenomena.

I will proceed as follows. In Section 5.2, I introduce the process of identifying phenomena, and I discuss how empirical findings can serve as defeasible evidence for the characterization of a phenomenon. In Section 5.3, I present three projects in which researchers attempted to provide evidence for memory transfer. In Section 5.3.1, I examine the work of James McConnell on

planarians. I continue in Section 5.3.2 with the development of memory research in mammals. In Section 5.3.3, I discuss the work of Georges Ungar. For each project, I review the dissenting opinions in the scientific community at the time. In Section 5.4, I analyze why researchers were justified in abandoning memory transfer. I discuss the defeater strategy, which was used to test the accuracy of the characterization of this alleged phenomenon. This strategy applies to the case of memory transfer, but it has the potential to be applied to other instances of scientific practice as well. With an account of this strategy, I rebut the claims about memory transfer presented by Collins and Pinch.

5.2 Identifying a scientific phenomenon

When researchers identify a scientific phenomenon, what is it that they identify? For this chapter, a *scientific phenomenon* is a type of event whose characteristics exhibit repeatability and stability (Bogen & Woodward 1988). This distinguishes phenomena from data, which are the empirical findings collected in experiments that are used to infer the characteristics of a phenomenon. Phenomena are discovered in the world or created in the laboratory (Hacking 1983, 221). Researchers aim to measure the features of phenomena that manifest in observational or experimental contexts, and they use their empirical findings to accurately describe them by characterizing the features of their manifestations and the conditions under which they occur. Accepting a characterization of a phenomenon amounts to accepting that the phenomenon as characterized occurs. Conversely, rejecting this characterization amounts to rejecting that a phenomenon occurs in the way that it is characterized.

Characterizing phenomena is important for scientific practice. Theories are tested against the phenomena that are discovered or created; researchers seek to determine if the predictions they derive from their theories correspond to the phenomena they characterize (Bogen & Woodward 1988). In addition, characterizing a phenomenon "guides the construction" of hypotheses for the investigation of mechanisms and aids in the construction of theoretical models (Craver & Darden 2013, 52; see also Woodward 2003).

The identification of scientific phenomena is a process through which researchers discover and accurately characterize a phenomenon of interest. At the start, researchers record the features constitutive of a phenomenon from instances of its occurrence. To evaluate this initial characterization of a phenomenon, researchers produce the set of characterized features in the same and different contexts to determine the conditions under which these features co-occur. These strategies allow the researchers to determine the features that constitute the phenomenon of interest, as well as the conditions that precipitate, inhibit, and modulate its occurrence (Craver & Darden 2013, Chapter 4). When accurate, a characterization of a phenomenon corresponds to a general type of event, which occurs in the instances in which the experimental data were collected. If findings that support the characterization are produced, there is reason to accept that this phenomenon occurs and its characterization is accurate. This is because researchers take the findings to be causally related to the occurrence of an instance of the phenomenon.

Empirical findings provide evidential support for a characterization of a phenomenon, but this evidence is *defeasible*. Findings that serve as evidence for a characterization of a phenomenon do not entail its accuracy. The evidential status of findings may alter if additional findings are collected. Certain findings, known as *defeaters*, provide reason to think that the initial findings do not provide evidence for the characterization of the phenomenon. If evidence is defeated, it no

longer provides reason to accept the characterization of a phenomenon. This depiction of defeasibility hints at a strategy that researchers can adopt. If researchers sever the relationship between findings and the characterization of the phenomenon the findings are thought to provide evidence for, they provide reason to no longer accept it. Undefeated evidence is required to accept a characterization of a phenomenon, and researchers can actively seek to defeat evidence in order to provide reason to reject it. This strategy, which I call the defeater strategy, played a crucial role in the abandonment of research on memory transfer.

An alternative strategy to provide reason to reject a characterization of a phenomenon is to develop new experiments whose findings provide evidence of the inaccuracy of the characterization. However, this strategy presents an additional challenge: evidence against a characterization of a phenomenon does not cause any evidence for it to lose its epistemic status. To dismiss findings as evidence for a characterization of a phenomenon, researchers must formulate an alternative hypothesis, inconsistent with the characterization of a phenomenon in question, which is better supported by the sum of the reported findings. In research on memory transfer, this kind of strategy was not used. This is because, as I will illustrate, there is no single reason why researchers produced findings they thought supported memory transfer as characterized.

5.3 Memory transfer

Research on memory transfer developed out of research on learning in planarian worms in the 1950s, with a report of a transfer effect in 1962. By the end of the 1970s, most researchers agreed that the phenomenon did not exist. In this section, I track memory transfer research through

three research projects (see Figure 1 for articles that support or challenge memory transfer). Rather than present a strict temporal ordering of the projects, I present them based on their research questions and character of the controversy surrounding them. I start with the work by McConnell on planarians. I then discuss research on mammals by the Fjerdingstad group, the Jacobson group, and the Rosenblatt group. I finish by discussing Ungar's research on memory transfer and its underlying mechanism. I do not intend to provide an exhaustive history of the case. Rather, I focus on the researchers who contemporaries and historians agree played the most significant role in the discovery, characterization, theorization, explanation, and defense of memory transfer, and I demonstrate how, in each set of research projects, there were unresolved experimental issues (Irwin 1978; Travis 1981; Setlow 1997).

Citation	Empirical Findings	Relation to Memory Transfer
McConnell (1962)	Report of successful transfer of learned memory in planarian worms, using an associative learning paradigm and cannibalism	Supports
Halas, James, & Stone (1961)	Report of potential issues with memory association paradigm from McConnell 1962	Ambiguous
Halas, James, & Knutson (1962)	Report of failure to engage planarian worms in associative learning paradigm from McConnell 1962	Challenges
Bennett & Calvin (1964)	Report of failure to engage planarian worms in associative learning paradigm from McConnell 1962	Challenges
Babich et al. (1965)	Report of memory transfer in rats, using a Skinnerian learning paradigm and injection of neural tissue	Supports
Nissen, Røigaard-Petersen, & Fjerdingstad (1965)	Report of memory transfer in rats, using a reinforcement learning paradigm and injection of neural tissue; identification of reversal effect	Ambiguous
Walker (1966)	Report of confound of sensitization in protocol presented in McConnell 1962	Challenges
Walker & Milton (1966)	Report of confound of sensitization in protocol presented in McConnell 1962	Challenges
Rosenblatt, Farrow, & Herblin (1966)	Report of memory transfer in rats, using a Skinnerian learning paradigm and injection of neural tissue	Supports
Rosenblatt, Farrow, & Rhine (1966a)	Report of memory transfer in rats, using a Skinnerian learning paradigm and injection of neural tissue; replication of Babich et al. 1965	Supports
Rosenblatt, Farrow, & Rhine (1966b)	Report of memory transfer in rats, using a Skinnerian learning paradigm and injection of neural tissue	Supports
Byrne et al. (1966)	Report of failure to replicate memory transfer protocol presented in Babich et al. 1965	Challenges
Røigaard-Petersen, Nissen, & Fjerdingstad (1968)	Report of memory transfer in rats, using a reinforcement learning paradigm and injection of neural tissue; modification to protocol in response to Byrne et al. 1966; replication of reversal effect	Ambiguous
Ungar (1970)	Summary of reports of memory transfer, using a reinforcement and Skinnerian learning paradigm and injection of neural tissue	Supports
Goldstein, Sheehan, & Goldstein (1971)	Report of failure to replicate memory transfer protocol presented in Ungar 1970	Challenges
Krech & Bennett (1971)	Report of failure to replicate memory transfer protocol presented in Ungar 1970	Challenges
Ungar, Desiderio, & Parr (1972)	Report of memory transfer in rats, using a reinforcement learning paradigm and injection of neural tissue	Supports

Figure 1 Articles that Support or Challenge the Characterization of Memory Transfer

The research on memory transfer was driven by experiment. The major proponents of memory transfer were experimentalists, who did not develop a characterization of the phenomenon in light of a well-established theory of memory. Instead, research was based on speculations from earlier experimental findings. While the research was informed by background theory on memory, experiment, rather than theory, drove subsequent experimentation. Thus, rejecting memory transfer was not considered to entail the refutation of an otherwise successful theory that predicted its occurrence.

Each memory transfer experiment shares a protocol template. This template is as follows:

Donor Training: An organism is trained to display a behavior consistent with the association of stimuli, which indicates that the organism has learned.

Transfer from Donor to Receiver: Tissue is excised from the donor organism and inserted into the receiving organism.

Receiver Training: The receiving organism is trained following a learning paradigm. Researchers determine if the receiver demonstrates the behavior consistent with having the memory or learns more quickly than an organism that has not received tissue from the donor.

Thus, in a memory transfer experiment, there must be an associative memory, and there must be a transfer of that memory, which allows the receiver to do something that could not be achieved without having that memory.

Debate over memory transfer revolved around two kinds of issues with the experiments whose findings were thought to serve as evidence for the characterization of the phenomenon. The first was the failures to replicate the results. These failures occur when researchers are unable to reproduce the findings reported by the supporters of the alleged phenomenon, despite reproducing the reported conditions of the original experiment. The second was the identification of confounds. In the present context, a *confound* is any feature not included in the characterization of the

phenomenon that plays a role in producing the empirical findings in the phenomenon's purported experimental demonstration.

5.3.1 McConnell and the Worms

Research on memory transfer developed from research on the planarian as a model organism for neural experimentation. With Robert Thompson, McConnell putatively demonstrated that planarians could be classically conditioned (1955). To demonstrate conditioning, an organism must display behavior consistent with the association of an unconditioned stimulus and a conditioned stimulus, following tandem presentation of the stimuli to the organism. This amounts to the organism displaying a behavioral response to the unconditioned stimulus when the conditioned stimulus is presented. Thompson and McConnell had the unconditioned and conditioned stimuli played by electrical shock and light exposure respectively: each planarian was trained to react to light exposure, with its unconditioned reaction to electrical stimulation. These empirical findings indicated that planarians were capable of acquiring memories, which would make them amongst the simplest organisms known to be able to do this.

The conditioning paradigm filled the Donor Training step of the protocol template. In the study, the planarians were divided into four groups: an experimental group (which received conditioned and unconditioned stimuli), a light control group (which received only conditioned stimulus), a shock control group (which received only unconditioned stimulus), and a control group (which received neither stimulus). Following their exposure to the stimulus (or lack thereof), the researchers observed the behaviors of the planarians. Only members of the experimental group demonstrated a significant increase in the number of behavioral responses. Organisms in this group

received a three-second light stimulus, in which their behavior was recorded during the first two seconds, and they were shocked during the last second. That they engaged in the behaviors associated with the shock when exposed to the light only suggests that they had learned to associate the stimuli.

With a learning paradigm developed, McConnell began to investigate if a planarian would retain its memory following segmentation and regeneration of its body (McConnell 1965). Planarians regenerate if cut transversely: each half will grow back a head or tail to result in two distinct bodies. The researchers determined that if the planarians were trained to demonstrate the conditioned response, then cut in half and allowed to regenerate, each new planarian demonstrated the conditioned response (McConnell, Jacobson, & Kimble 1959). This was considered remarkable. Neural structures are largely absent in the planarian tail, as they regrew after segmentation; it was "as if their new brains were created with the old 'learning' already 'wired in'" (McConnell 1965, 6). These findings sparked considerable debate and led McConnell to develop a journal dedicated to planarians, called the *Worm Runner's Digest*. Though the journal included poems, comics, and facetious articles as part of McConnell's desire to introduce zaniness into an austere scientific community (Travis 1981), the journal also included serious articles, including the article that presented the first characterization of memory transfer.

If planarians could regenerate memories, McConnell speculated that "the 'engram' must be stored throughout the planarian's body," and these 'engrams' might be transferrable between organisms (McConnell 1965, 6). To determine if transference was possible, he developed an infamous protocol that exploited the cannibalistic nature of planarians:

> We trained some 'victim' worms to criterion (using the now-standard light/shock conditioning technique). We then fed the trained victims to starved, untrained cannibals. At the same time, a set of untrained 'victim' worms was fed to a different group of cannibals. Both sets of cannibals were then given light/shock training. The

> results were clear-cut: the cannibals that had ingested 'trained' victims were, on the very first day of training, significantly superior to the cannibals that had eaten untrained animals (McConnell 1965, 7).

First published in the *Worm Runner's Digest* in 1961, then in the *Journal of Neuropsychiatry* in 1962, 'Memory transfer through cannibalism in planarians' includes the first characterization of memory transfer: "learning seems to be transferrable from one animal to another via cannibalistic ingestion" (McConnell 1962, S48). The characterization includes specific features of the experiment, but it also indicates the generalizability of the type of memory event. Through the cannibalistic transfer of tissue, the memories of the donor organism are transferred to the receiving organism.

The characterization of the phenomenon and the learning paradigm both drew controversy. Failures to replicate the learning in planarians were reported (Bennett & Calvin 1964). Confounds were also reported. Researchers suggested that McConnell may not have demonstrated associative learning and that the protocol instead may provoke the response of sensitization (Halas, James, & Stone 1961). Sensitization is non-associative, so it is not the kind of learning McConnell had depicted in his characterization of memory transfer. If light provoked a response – a circumstance McConnell attempted to control for with the light control group – then the experiment would not have shown the association of the stimuli. Skeptical researchers found "a significant difference between [the light control group] and [the normal response control group]" (Halas, James, & Knutson 1964, 791). They state that there is "no adequate explanation for this discrepancy other than to point out that there was a trend toward a significant difference between these groups in the Thompson and McConnell study": they point out that McConnell's 1962 article references a non-significant increase in behavioral responses in the light control group, suggesting that planarians

may be responding to light as though it were an unconditioned stimulus (Halas, James, & Knutson 1964, 791).

Other researchers criticized McConnell's protocol in his 1962 article. Researchers demonstrated a significant relationship between the shocking of planarians and their light response but argued that this relationship was due to the fact that the cannibal planarians had ingested shocked tissue (Walker & Milton 1966). Researchers were able to demonstrate behaviors consistent with McConnell's results without training the donor worms. This criticism was extended by research in which conditioning was extinguished – that is, the association was formed through training and then dissociated through further training (Walker 1966). There was no significant difference in the behavior between the worms that had consumed tissue from conditioned planarians when compared to those that had been conditioned and then extinguished.

Thus, McConnell's research was undermined for two reasons. First, his findings had failed to replicate, raising concerns for the characterization of McConnell's experimental conditions. As critic Edward Bennett states, no one could "point to a 100% procedural replication," which made it "impossible in the absence of further experiments to determine if the relevant factors necessary for reproducible and reliable training of planarians have been described" (1970, 150). Second, several research groups provided reasons to suspect the protocol did not have the sophistication to rule out confounds like sensitization. This meant that McConnell's experimental findings were equivocal between demonstrating memory transfer and demonstrating an effect of a very different kind. Consequently, empirical challenges defeated the evidence McConnell provided for his characterization of memory transfer.

5.3.2 The Inclusion of Mammals and the Theorization of Memory Transfer

Following popularization by McConnell, other researchers became interested in memory transfer. Respecting the issues with McConnell's experiments, rodents were used "to avoid any discussion of whether... [the] experimental subjects could learn" and to determine if "the phenomenon would prove so general as to be found in mammals" (Fjerdingstad 1971, xvi-xvii). Following the transition to mammals, several research groups putatively demonstrated memory transfer. However, given that a principal question posed to researchers was "Does the phenomenon really exist…?", it is clear that the studies were contentious (Fjerdingstad 1971, xiv). I discuss three groups that investigated memory transfer in mammals: Fjerdingstad's group, Jacobson's group, and Rosenblatt's group.

Following up on a pilot study that appeared to demonstrate memory transfer, researchers in the laboratory of Enjar Fjerdingstad developed a protocol to investigate transfer in mammals (Nissen, Røigaard-Petersen, & Fjerdingstad 1965). Rats were trained via a reinforcement paradigm, meaning that the organisms behave consistently with the association of a behavior and a reward stimulus. Following training, their brains were extracted using a phenol solution, chemically treated in an isopropanol solution, and injected into receiving organisms. The chemical treatment increased the concentration of the RNA from the tissue, a possible chemical substrate for memory transfer. Their experiment followed the protocol template: donor rats were conditioned either to light or to darkness. Receiving rats were injected with tissue from light-conditioned or dark-conditioned rats, and they were then were trained to reinforce dark preference or light preference. The publication reports a significant difference between the learning quickness of rats injected with light-condition donor tissue and those who were injected with dark-conditioned donor tissue. However, the direction of difference is the opposite of what was predicted: light-

conditioned rats gave better performance under reinforcement of dark preference and vice versa. This 'reversal effect' is inconsistent with the association of behavior and reward. Individual differences of rats and "specific inhibitory effects" are cited as potential explanations for the reversal effect (Nissen, Røigaard-Petersen, & Fjerdingstad 1965, 271).

Concurrent with this research, members of the laboratory of Allan Jacobson began to investigate memory transfer. The researchers used a Skinnerian training paradigm to teach rats to associate clicking sounds with a behavior: rats would approach a food bowl when a click occurred, even if there was no food in the bowl (Babich et al. 1965). Once learning was established, the rats' brains were extracted and then injected into rats in an experimental group. The researchers also included a control, with untrained donor chemicals being injected into a group of rats. When comparing the behavior of the experimental and control groups, there was a significant difference in the tendency the rats had to approach the bowl when the clicking occurred. This study was reproduced, and the findings were replicated. The receiving organisms did not need to undergo training to display the behavior; they simply 'had' the memory following the injection.

Researchers in the laboratory of Frank Rosenblatt also began to study memory transfer. Reproducing Jacobson's protocol, Rosenblatt and colleagues reproduced the experiment and replicated the findings consistent with memory transfer (Rosenblatt, Farrow, & Herblin 1966a). In addition to a direct replication, the researchers also modified the Donor Training step of the protocol template and showed that associations formed by different training paradigms could be transferred. However, the researchers also uncovered individual differences within their subject population. They state that "it appeared from these data that the injection had 'taken' on some of the rats, but not on others" (Rosenblatt, Farrow, & Herblin 1966b, 48). In another replication of Jacobson, the laboratory researchers confirmed issues with the injection 'taking': "we find

repeated examples of this phenomenon," and, though "the reasons… are still subject to speculation," the authors include variation of dosage, subjects' sensitivity to the extract, and "the transfer of an adaptation effect which tends to reduce activity that would otherwise result from normal curiosity" as potential explanations for the individual differences in the subjects' behavior (Rosenblatt, Farrow, & Rhine 1965, 553-554).

Despite complications, Rosenblatt states that his experiments convinced him "that the phenomenon of 'memory transfer' is a real one, which must be taken into account in any theoretical approach to biological memory" (1967, 34). However, given the lack of evidence about the chemistry that underlies transfer, he suggests that "it must be recognized that we are, in fact, entering the realm of science fiction; the present experiments, although suggestive, leave us completely in doubt as to the mechanism at work" (Rosenblatt 1967, 34). Thus, Rosenblatt accepted the characterization, even with no knowledge of what mechanistic details might underwrite memory transfer's occurrence. With no evidence for the mechanism, Rosenblatt admits that any effort to model what causally underwrites memory transfer amounts to speculation. This frustrated more eager modelers, with one saying that "we are inclined to feel that it is no more science fiction than was, for example, Einstein's derivation of his famous $E=mc^2$ relation at a time when there was not a single shred of experimental evidence for it" (Rashevsky 1968, 342). Modelers who were convinced of memory transfer modeled it, and they assumed that subsequent research would vindicate their models.

The empirical results of these groups were questioned during this time. In an 18-experiment replication performed by researchers in a number of different laboratories, there was failure to replicate the memory transfer phenomenon in a direct reproduction of Jacobson's protocol (Byrne et al. 1966). The failures suggest complications with Jacobson's protocol – and, by implication,

Rosenblatt's protocol – though the authors state that "it would be unfortunate if these negative findings were to be taken as a signal for abandoning pursuit of" memory transfer, as "failure to reproduce results is not, after all, unusual in the early phase of research when all relevant variables are as yet unspecified" (Byrne et al. 1966, 658). Thus, the failures to replicate provided reason to think that Jacobson's findings could not serve as evidence for memory transfer. However, those who worked on memory transfer continued to iterate their practices to develop an experiment that did not include confounds and could be replicated.

Responding to the confounds identified in their previous work and the failures to replicate, Fjerdingstad and colleagues modified their protocol "in order to make the 'transfer effect' more reproducible," which was "considered to be of primary importance before further experiments concerning the degree of specificity of this effect and [its] exact chemical nature" (Røigaard-Petersen, Nissen, & Fjerdingstad 1968, 1). To control for the individual differences of the organisms, the researchers picked only rats that were explorative. Fjerdingstad and colleagues developed a novel light-dark reinforcement experiment. Prior to injection, the researchers kept the chemicals at consistent temperatures to control for any damage that may occur to the chemical under improper holding conditions. While the researchers report some effect on learning from the injection of control organisms – those injected with tissue from donors who were not trained – when compared to noninjected ones, they report a significant difference between the learning of the experimental and control injected groups.

However, the researchers replicated the reversal effect – the rats with the dark reinforcement injections demonstrated improved light behavior and vice versa – which is likely why the researchers mention that their findings "must therefore be interpreted with some reservation until these relations have been further investigated" (Røigaard-Petersen, Nissen, &

Fjerdingstad 1968, 12). The researchers again cite individual differences of the experimental organisms and inhibitory effects as potential explanations for the reversal effect. They did not rule out confounds related to inhibitory effects, even though their results suggest that something that is not specified in the characterization of memory transfer likely played a role in their experiments.

In response to the failures to replicate, the researchers state the following:

> During the last year the proportion of the number of 'positive' to the number of 'negative' reports seems to have changed. No doubt this is due to the realization of the potential importance of even seemingly trivial variables, both of behavioral methods, extraction and injection. Too little attention to these factors characterize [sic] many of the early negative reports (Røigaard-Petersen, Nissen, & Fjerdingstad 1968, 14).

While the researchers do not make clear which factors the negative reports missed, their intent is clear: negative reports, but not positive ones, fail to take into account key experimental features. It is due to this issue with the negative reports, they suggest, that replication does not occur.

Thus, the three laboratory groups were unable to resolve issues relating to replication failures or confounds. Many were unable to replicate the findings presented by the supporters. This was summed up in the comments from one critic that "it is essential that several laboratories can replicate and extend the primary observations" (Bennett 1970, 150). Critics did not accept the speculation that confounds were responsible for only failures to replicate. Furthermore, the supporters did not address which features of their experiment led to the inconsistent findings presented by Fjerdingstad and colleagues. For these reasons, the findings were not taken by the community to evidentially support the characterization of memory transfer.

5.3.3 Ungar, Mammals, and the Mechanism of Memory Transfer

One other researcher was a key investigator of memory transfer in mammals. Georges Ungar started working on memory transfer with the investigation of the transfer of habituation from mice to rats. In this experiment, the Donor Training involved exposing the donor organisms to a stimulus that would elicit a startle response in the organism (Ungar & Oceguera-Navarro 1965). Once the donor organisms were acclimated to the stimulus, they were killed and their tissue was transferred. Along with this, a control group was injected with the brains from naïve organisms. Following injection, the researchers report a significant difference in the habituation rates of the two groups, with the experimental group habituating much more quickly. This led Ungar to investigate associative learning.

In his research, Ungar typically exposed the experimental injection to a battery of chemical reactions, and, from this, he determined that a key component of the injection was a protein or peptide chain. As such, he took failures to replicate memory transfer (such as Byrne et al. 1966) to be irrelevant to his work, as Byrne and colleagues did not concentrate peptides. However, he explained the previous successes by Jacobson's group to be due to poor design and the failure to remove peptides from their injection.

Ungar continued to perform experiments to determine the specificity of the transfer of associated memory. With a publication that summed up his research, Ungar concluded that he had produced sufficient evidence for the phenomenon's existence (Ungar 1970). The features relating to chemical composition, dose, interval between injection and testing of receivers, and duration of donor training were taken by supporters to be features that had not been controlled for in failures to replicate. However, Ungar also states:

> The reliability of the method is limited because of the multiplicity of the factors determining success and failure. The experience accumulated in the last five years in many laboratories explains most of the past failures (1970, 162).

This hedging of claims about memory transfer came at the same time as another series of failures to replicate: one in a new experimental context (Krech & Bennett 1971) and the other a direct replication of Ungar's protocol (Goldstein, Sheehan, & Goldstein 1971). With critics demanding "an experimental procedure which will yield consistently positive results in the hands of all qualified experimenters and in different laboratories," skeptics would not settle for "weak positive answers, only fitfully obtained, and only among a chosen few experimenters" (Krech & Bennett 1971, 161). For skeptics to endorse Ungar's results, he had to produce a more sophisticated experiment or else modify his characterization of memory transfer.

To convince the skeptics, Ungar developed a large sample study to demonstrate the phenomenon (Ungar, Desiderio, & Parr 1972). Training rats and mice on dark avoidance, the project involved 4000 experimental subjects trained on the paradigm, whose brains were then prepared to isolate a peptide. Ungar believed this peptide to underwrite associative memory, and he further believed its isolation and transmission from one organism to another via injection to be the mechanism underwriting memory transfer. The researchers report a significant difference in the dark-avoidance behaviors of the experimental group from both the group that received injections from untrained donors and those who received no injections at all. Individual differences between the subjects are not reported nor are controls on the experimental protocol except when relating to the development and transfer of the peptide chemical. There was no investigation of the issues with previous memory transfer experiments.

Ungar's 1972 article was published with commentary from the reviewer Walter Stewart, as Ungar was not able to satisfy Stewart's reservations with the experiment. Stewart agreed that

Ungar demonstrated the synthesis of the peptide that could potentially causally underwrite the alleged phenomenon but that isolation of the brain material from the rodents was "so grave… that the authors' conclusions are more likely false than true" (Stewart 1972, 209). Furthermore, Stewart raises concerns about whether or not "the bioassay [can] be successfully repeated outside of [Ungar's] laboratory" and what conditions are "necessary to minimize its variability" (Stewart 1972, 209). In his commentary, Stewart challenges Ungar to reproduce the phenomenon and have other laboratories successfully use his protocol. He also mentions that Ungar's protocol is quite different from previous research, resulting in a worry that Ungar has not demonstrated if his work relates to previous successes and failures to demonstrate memory transfer. Ungar was never able to meet Stewart's challenge. He was never able to develop a protocol in which he or anyone else could consistently demonstrate memory transfer. This failure, in result, led researchers to not accept that he had provided evidence for the alleged phenomenon.

5.3.4 Aftermath

No supporter of memory transfer was able to provide a clear characterization of the phenomenon and the conditions under which it occurs. While modelers theorized about the phenomenon's implications, and experimentalists investigated its underlying biochemical mechanisms, these projects were hampered by challenges to the characterization of memory transfer. Supporters shared their frustrations. Louis Irwin states, "it was not confusion over the biochemical identity of the transfer factors or the debate over the behavioral specificity of the effect which most damaged the credibility of the transfer paradigm, but simply the unreliability of the phenomenon" (1978, 486). Thus, it was not an issue with determining the mechanistic underpinnings of memory transfer that doomed the research program; it was the persistent defeat

of any evidence for the alleged phenomenon. Irwin goes on to say, "the phenomenon was by no means universally replicable, and many labs… failed to obtain satisfactory evidence for any form of transfer," and "even for the proponents of the transfer phenomenon, the magnitude of the effect was always marginal" (1978, 486).

McConnell turned out to be right that planarians can learn (Rilling 1996), though his protocol was not sophisticated enough to demonstrate it. Those working on the biochemistry of memory started again to investigate memory without the baggage of memory transfer. Progress in this field was made, but later researchers did not return to the memory transfer research of the 1960s and the 1970s. Today, most researchers think that the study of memory transfer was flawed (Setlow 1997) and any transfer phenomenon that might exist would likely have a very different characterization.

5.4 The rejection of a characterization of a phenomenon

Issues with the experiments whose findings were thought to provide support for the characterization of the purported phenomenon precipitated the abandonment of the investigation of memory transfer. In this section, I analyze what issues like these indicate about the accuracy of a characterization of a phenomenon. I account for the evidential role of empirical findings in characterizing a phenomenon and explain why many experiments were used to defeat evidence for the characterization of memory transfer.

5.4.1 Why Did Researchers Reject the Characterization of the Phenomenon?

Proponents of memory transfer faced the challenge of reported failures to replicate the finding when the reported conditions of the experiment were reproduced. In response, the supporters developed new experiments but were unsuccessful: additional failures to replicate were reported, which themselves replicated. The supporters did not determine if the phenomenon occurs under a more constraining set of conditions than originally specified: there was no test of whether or not memory transfer only occurred in certain organisms, with certain learning paradigms, with only non-associative forms of memory, or with only certain associated stimuli. It may be the case that these variables must be held fixed for memory transfer to occur and thus should have to be included in the characterization to delimit the conditions under which the phenomenon occurs.

Failures to replicate – meaning that researchers reproduce the reported conditions of the experiment but not the result – suggest that there is issue with the characterization of the conditions under which a phenomenon occurs. The findings produced in the original experiment may be due to conditions that were not reported and not reproduced in the replication attempt. As a result, researchers infer that the characterization fails to specify the conditions that precipitate the phenomenon, or it fails to specify some conditions that inhibit the phenomenon. As unexplained failures become more common, researchers' skepticism of the accuracy of characterization of the phenomenon grows, which is one reason why researchers abandoned memory transfer. The characterization of memory transfer may have had too great a scope, and it was not recharacterized to indicate the narrow set of conditions under which the phenomenon may occur.

In addition, there was the issue of the identification of confounds in memory transfer research. These confounds were not thought to be constitutive of the phenomenon, but they nevertheless appeared to be responsible for the findings thought to support the phenomenon's

characterization. McConnell was unable to develop an experiment in which he could eliminate sensitization, Fjerdingstad was unable to eliminate the features responsible for the reversal effect, and Ungar was unable to eliminate individual differences between his experimental subjects. Skeptics developed experiments that suggested that these features might be responsible for the findings obtained in memory transfer experiments. McConnell, Fjerdingstad, and Ungar may have discovered real phenomena, but these phenomena were not accurately described by the characterization of memory transfer.

The identification of confounds reveals a problem with the characterization of the features thought to constitute the phenomenon. In normal circumstances, researchers take data, the empirical findings from individual experiments, to provide a basis for inferring a phenomenon (Bogen & Woodward 1988). The occurrence of the phenomenon of interest is taken to be causally responsible for the data, and thus the data reflect the phenomenon's features. However, if there is a known confound that might be responsible for the findings, there is an alternative explanation for why researcher acquire the data that they do, undermining the inference from data to the characteristics of the phenomenon of interest. Researchers must determine if the findings are indicative of the occurrence of the phenomenon of interest or of the feature that acts as a confound. Once it is identified, if researchers cannot produce consistent findings while simultaneously eliminating the presence of the confound, the initial findings are equivocal.

In the memory transfer case, the skeptics produced findings consistent with those from McConnell's experiments without the involvement of memory. This trend continued in research on mammals: confounds were identified, sometimes by the supporters themselves. To restore evidential support to the characterization of the phenomenon, supporters had to either eliminate the confounds that had been identified or recharacterize the phenomenon so as to not rule out that

something aside from memory transfer may be responsible for their findings. The supporters were unable to do the former and unwilling to do the latter. In general, as a result of the identification of actual confounds relevant to experiments in which positive results are reported, researchers' skepticism of the accuracy of the characterization of the phenomenon grew. This is another reason why memory transfer was abandoned. The characterization of memory transfer may have been too specific; it did not indicate that the phenomenon might involve the interaction of different sets of features, each of which was consistent with the experimental results.

Thus, findings from multiple experiments provided different insight regarding the accuracy of the characterization of memory transfer. The proponents of the phenomenon neither resolved the issues that were put forward nor modified the characterization of memory transfer. With no experimental support and no independent theoretical support, there was no reason for members of the research community to accept the phenomenon. Thus, the characterization of memory transfer was rejected.

5.4.2 Were Researchers Justified?

The issues I have presented led researchers to reject memory transfer. However, one might question whether or not researchers were epistemically justified in their rejection. Collins and Pinch deny that there was sufficient reason to abandon memory transfer. They challenge the decisiveness of three articles they argue are cited as providing decisive evidence against memory transfer, each of which I mention in Section 5.3: the failure to reproduce planarian learning by Bennett and Calvin (1964), the failures to replicate by Byrne and colleagues (1966), and the reservations of Ungar's protocol by Stewart (1972). Collins and Pinch note that these publications

"seemed decisive at the time" but "in retrospect they seem much less decisive" (1998, 24). Based on the evidential deficiencies of these articles, they conclude:

> A determined upholder of the idea [of memory transfer] would find no published disproof that rests on decisive technical evidence. For such a person it would not be unreasonable or unscientific to start experimenting once more. Each negative result can be explained away while many of the positive ones have not been. We no longer believe in memory transfer but this is because we tired of it, because more interesting problems came along, and because the principal experimenters lost their credibility. Memory transfer was never quite disproved; it just ceased to occupy the scientific imagination (Collins & Pinch 1998, 25).

While Collins and Pinch do not characterize what counts as a "disproof that rests on decisive technical evidence," their position can be inferred from their discussion of the limitations of the three articles. They note that each article targets only a subset of the empirical research on memory transfer. They claim that Bennett and Calvin's criticisms of planarian learning are moot, because it was later determined that planarians can learn, and their criticisms do not apply to experiments that involve model organisms that uncontroversially express associative learning. Likewise, they claim that the failures to replicate presented by Byrne and colleagues are irrelevant to research by Ungar because his protocols were different than those developed by the laboratory of Jacobson, which Byrne and colleagues investigated. They address no specific limitation of Stewart's article, but it can be inferred that their argument is based on the fact that Stewart's reservations only apply to issues with Ungar's experiments.

Collins and Pinch's conclusion can be read in two different ways. It may be the case that their argument rests on the fact that some evidence for memory transfer was not "explained away." However, this is inconsistent with the history of memory transfer research I have presented: every major experiment thought to provide support for memory transfer was questioned in light of the failures to replicate or the identification of confounds, and other reported experiments were based on the designs of the major ones. Alternatively, it may be the case that their argument rests on the

fact that the criticisms expressed in each article apply only to certain experiments whose findings were thought to provide evidence for the phenomenon as characterized. No criticism applies to the total sum of empirical research on memory transfer. Those working with mammalian model organisms "explained away" the issues with planarian experiments. Ungar "explained away" the issues with experiments that did not involve the concentration of the peptides that he believed to underlie the memories that were transferred. It can be inferred that, to count as "disproof that rests on decisive technical evidence," researchers must demonstrate an issue that applies to all experiments whose findings are thought to provide evidential support for memory transfer. Alternatively, the skeptics must put forward an alternative hypothesis that is equally well supported by the proponents' empirical findings and can be tested against claims about memory transfer in new experiments. This corresponds to the second strategy to provide reason to reject a characterization of memory transfer I introduced in Section 5.2.

I agree with this reading of Collins and Pinch's argument: the challenges to the characterization of memory transfer targeted specific features of individual experiments, none of which can be applied to research in the field as a whole. Likewise, even though Collins and Pinch fail to discuss many articles critical of memory transfer, I agree that no critical publication describes an issue that applies to all memory transfer experiments and no alternative hypothesis was presented that is equally well supported by the empirical findings. However, I disagree with the conclusion that Collins and Pinch draw from these facts.

Researchers were justified in their rejection of the characterization of memory transfer. Skeptics of memory transfer targeted the individual experiments that were used to produce the results that ostensibly demonstrated the phenomenon. The goal of the skeptics was to raise issues with each experiment whose findings were thought to provide evidence for memory transfer. Their

strategy thus followed: reproduce the experiments, demonstrate issues with the experiments, and sever the evidential relationship between the findings from these experiments and the characterization of the phenomenon the findings were taken to support.

The skeptics reproduced the experiments in order to determine the relationship between the findings that were initially reported and the phenomenon's characterization. In the iterated experiments, they determined two kinds of things related to the issues in Section 5.4.1. First, they determined that the experiments did not produce findings consistent with the features described in the characterization, despite the fact that the experiment was consistent with the conditions specified in the characterization. Second, they determined that the experiments involved features that could be responsible for the findings produced in the experiment, despite the fact that those features were not included in the characterization. Both issues suggest that the inference thought to be warranted by the findings from the original experiments – that they had demonstrated the phenomenon – is not warranted.

Aspects of the skeptics' strategy can be compared to the role of internalist epistemic defeaters in accounts of human reasoning and the analysis of knowledge. Epistemologists have claimed that knowledge claims are defeasible, meaning that the justification for a knowledge claim can be defeated by the acquisition of new evidence about the relationship between the claim and what is thought to justify it (Chisholm 1966). Likewise, cognitive scientists characterize the reasons individuals have for belief and the defeasibility of these reasons. A characterization of defeaters in reasoning comes from John Pollock:

> R is a defeater for P as a prima facie reason for Q if and only if P is a reason for S to believe Q and R is logically consistent with P but (P & R) is not a reason for S to believe Q… R is an *undercutting defeater* for P as a prima facie reason for S to believe Q if and only if R is a defeater and R is a reason for denying that P wouldn't be true unless Q were true (1987, 484-485, my emphasis).

If an individual believes a conclusion for a certain reason, an undercutting defeater is one that is consistent with the reason but attacks "the connection between the reason and the conclusion" (Pollock 1987, 485). For example, the belief that an object is red because one perceives it to be red is defeated by the discovery that the object is illuminated by a red light. The undercutting defeater does not entail that the object is not red, but it is a reason to deny that it would not look red unless it were actually red. In this way, it defeats the evidential relationship between the conclusion and the reason provided to support the conclusion.

The idea of undercutting defeaters can be applied to the strategy of characterizing phenomena. Researchers accept characterizations of phenomena upon the assessment of empirical findings that are taken to support these characterizations. This is because the occurrence of the phenomenon is thought to be responsible for the findings that researchers acquire. Undercutting defeaters undermine the evidential relationship between findings and a characterization. They provide reason to think that the occurrence of the phenomenon may not be responsible for the findings initially taken to support its characterization. Researchers who are skeptical of a phenomenon's characterization can actively seek to determine whether or not there are undercutting defeaters for the findings that are thought to support the characterization. To the extent that they can discover defeaters, they can undermine inferring the phenomenon from the empirical findings thought to support it. This is the defeater strategy.

Skeptical researchers employed the defeater strategy to systematically undermine all empirically motivated reason to accept the characterization of memory transfer. The undercutting defeaters in this case were the findings related to the failures to replicate and the findings related to the identification of confounds. Each set of negative findings is an undercutting defeater for evidence for the characterization of memory transfer, as each provides reason to think that the

features or conditions specified in the characterization of memory transfer may not be responsible for the experimental findings thought to support its characterization. Every major memory transfer experiment was challenged, and every reported positive finding was undercut by identified confounds and failures to replicate. Thus, for every finding thought to provide evidence for memory transfer, an undercutting defeater was presented. Each defeater relates to a particular experimental attempt to demonstrate memory transfer. However, as a collection, the defeaters provide sufficient reason to think that there was no undefeated empirical evidence to accept the characterization of memory transfer. It was not merely the possibility of defeaters that undercuts evidence for the characterization of the memory transfer, as all empirical evidence is defeasible. It was the actual collection of findings that are undercutting defeaters of evidence for a characterization of memory transfer that plays this role.

The defeater strategy was appropriate in a case like memory transfer. This is due to the fact that many experiments were performed in the various memory transfer projects, whose findings were thought to provide evidence for memory transfer. The experiments were different, involving both differences in protocol and model organism. Skeptics suggested alternative explanations for the positive findings, but these explanations were sensitive to particular aspects of the protocol or organism involved. Thus, the issues related to McConnell's protocols were very different than the issues related to the work of Ungar, Jacobson, Rosenblatt, or Fjerdingstad. The fact that there was not a single factor that was responsible for all previous findings means that skeptical researchers were not able to develop an experiment that proved that any single factor, rather than the occurrence of memory transfer, was responsible for all findings. Instead, they challenged each experiment thought to demonstrate the phenomenon individually and undercut every empirical finding thought to support its characterization.

Collins and Pinch are right that it would not be unscientific for researchers to begin again to search for memory transfer. Analogous to the fact that an undercutting defeater does not entail the falsity of the conclusion whose reason is undercut, defeaters do not rule out the possibility that there exists a phenomenon with features in the vague vicinity of what had been characterized. However, there is no reason to accept that the phenomenon occurs as characterized. More importantly, it would be unscientific to use the very same experiments described by supporters of memory transfer to rekindle investigation, due to issues with the reported conditions and identified confounds. It would take new techniques and protocols to renew the search for a phenomenon that is something like memory transfer.

The defeater strategy employed in research on memory transfer reflects a more general characterization of when researchers have reason to reject the characterization of a scientific phenomenon despite evidence that initially appears to support it. A characterization of a phenomenon ought to be accepted if there are empirical findings that provide reason to support it. If undercutting defeaters challenge the evidential role of the findings for the characterization of the phenomenon in question, and experiments that were thought to demonstrate the phenomenon are equivocal, then the empirical support for the characterization is reduced. If the empirical findings provide reason to accept the phenomenon as characterized, then it ought to be rejected if this reason is undermined. The effectiveness of the defeater strategy rests on challenging a characterization of a phenomenon by providing reason to think that the experiments whose findings are thought to provide support for the characterization are faulty and lack the requisite sophistication to infer the occurrence of a phenomenon from its findings. More than providing evidence against the accuracy of the characterization of a phenomenon, this strategy provides a means to challenge evidence put forward in the characterization's favor.

5.5 Conclusion

In this chapter, I have analyzed an episode from the history of science in which there was a rejection of the characterization of a scientific phenomenon despite initially promising empirical findings. Proponents of memory transfer produced findings that were thought to support the characterization of the phenomenon. The issues with the experiments in which the findings were produced ultimately precipitated the rejection of the phenomenon as characterized. The experimental strategy employed by skeptics of the reality of the alleged phenomenon exploited the defeasible evidential relationship between the characterization of memory transfer that proponents accepted and the empirical findings that served as reasons to accept it. My analysis of the memory transfer case provides a novel way to think about the assessment of scientific evidence. New experimental findings can defeat the evidence provided for a characterization of a phenomenon. This provides reason to reject the characterization, even if, as Collins and Pinch note, no one of the new experiments is individually decisive.

6.0 What do Representations of Scientific Phenomena Represent?

Though many philosophers accept that scientific phenomena are an important target of investigation, there remain questions about what phenomena are and, correspondingly, what representations about phenomena represent. In this chapter, I introduce a *nominalist* account of scientific phenomena. According to my account, one need not commit to the idea that phenomena are something over and above their token manifestations. Nonetheless, representations of phenomena are distinguishable from representations of their manifestations, and researchers adopt different epistemic stances towards these respective representations. With this account, I provide reason to think that previous accounts of phenomena – namely, those that appeal to abstract objects, ideal types, or patterns – are inadequate, due to their inability to explain the causal role phenomena are alleged to play or their inability to explain how representing phenomena is not equivalent to representing their manifestations.

6.1 Introduction

It is common for scientists to call the things that they investigate *scientific phenomena*. Take the placebo effect. Characterized as the effect on an individual "that cannot be attributed to the properties of the placebo itself entirely (since it is inactive)," (Zis & Mitsikostas 2018, 444), this "well-known phenomenon" (Everitt & Skrondal 2010, 327) is a target explanandum for neuroscientists who aim to answer why the placebo effect occurs. Further, it is a target of control for medical researchers who aim to inhibit or mitigate its occurrence when testing drug treatments.

This scientific interest in phenomena is reflected in philosophical accounts of explanation and theorization:

> What is explained is a generic pattern… what… I call a phenomenon (Woodward 2003, 17).
>
> Mechanisms are sought to explain how a phenomenon comes about (Machamer, Darden, & Craver 2000, 2).[21]
>
> Physical applied mathematics is in the business of constructing and investigating models of physical phenomena (Batterman 2009, 1).[22]

Despite widespread appeal to phenomena by scientists and scientifically-minded philosophers, it remains unclear what a phenomenon is supposed to be. In light of this lack of clarity, I raise the following question: when we represent scientific phenomena, *what* do we represent?

Answering this question requires recognizing that representations of phenomena are often type-level representations, rather than representations of individual occurrences. The placebo effect illustrates this: its representation is not of an individual occurrence, such as a case of the occurrence of placebo effect in a drug trial. In fact, many characteristics of this individual are not equivalent to what is represented when representing this phenomenon. This non-equivalence of phenomena with their manifestations is one reason why philosophers have defended accounts according to which phenomena are abstract objects (Brown 1994), ideal types (Teller 2010), or are, in some sense, patterns (Woodward 1989; McAllister 1997; Glymour 2010; Apel 2011; Feest 2011). On its face, the motivation for these accounts is clear: if representations of phenomena are type-level and distinct from those of individual occurrences, then they must represent some distinct type of thing.

---

[21] See also Bechtel & Richardson 1993.
[22] See also Cartwright 1983.

Despite this motivation, I argue that these accounts cannot explain how phenomena factor into causal claims and how their representations relate to those of token manifestations. As a replacement, I introduce a *nominalist* account of phenomena:

> *To represent a scientific phenomenon is to formulate a representation of characteristics of an occurrence, which one holds, in principle, to be applicable to other occurrences, in virtue of which these occurrences are manifestations of this phenomenon.*

My account gives up on the idea that a phenomenon is something distinct from its manifestations. Nevertheless, my account explains both how a phenomenon's representation is distinguishable from a representation of its manifestations and how the epistemic stance researchers adopt towards a phenomenon's representation is distinct from the stance they may adopt towards a representation of its manifestations. In essence, rather than assuming that there are kinds or types things that are phenomena that are represented, I instead argue that developing representations of phenomena is a means to represent expectations about characteristics of different occurrences uncovered through scientific investigation.

I proceed as follows. In Section 6.2, I introduce two desiderata of an account of phenomena gestured at in previous philosophical analyses, which relate to the causal character of phenomena and the non-equivalence of phenomena and their manifestations. With that, I introduce accounts according to which phenomena are abstract objects, ideal types, or patterns. In Section 6.3, I introduce my nominalist account of phenomena. I defend the adequacy of my account in Section 6.4. In Section 6.5, I show how existing accounts of phenomena either fail to satisfy these desiderata or are best explained by my account.

6.2 Attempts to account for phenomena

Contemporary philosophical discussion has identified key causal and evidential characteristics of scientific phenomena, resulting in several popular philosophical accounts positioning phenomena as the targets of theorization, explanation, and control. I review two influential accounts and introduce two desiderata for an account of phenomena that can be derived from them. The first is from Hacking; the second is from Bogen and Woodward.

6.2.1 What Scientific Phenomena are Like

According to Hacking, phenomena are discernable events or processes that occur with regularity under certain circumstances, which researchers aim to represent and manipulate (1983, 225). An example of what Hacking has in mind when he talks of phenomena is the Hall effect.[23] The Hall effect occurs when a voltage difference results from an electric current moving through a conductor in a magnetic field. This voltage difference appears because a magnetic force perpendicular to the electric current pushes charge carriers to one side of the conductor. According to Hacking, the Hall effect can be consistently induced, though it may be practically impossible to induce it unless under a set of specified experimental conditions. The Hall effect requires the induction of a magnetic field and the lack of external electrical and magnetic influence on the conductor.

[23] Hacking sometimes contrasts phenomena that occur naturally with "effects" that occur with human aid (1983, 221). However, he does not consistently maintain this distinction. My account makes no distinction between representing a phenomenon and representing an effect.

Bogen and Woodward also introduce an account of scientific phenomena, and they contrast them both with data and with theory. They contend that data are evidence for phenomena, rather than theories. Conversely, theories explain and predict phenomena, rather than data. On their account, data are the recordings of the results of empirical procedures, as they are the products of particular experimental or observational contexts. Because of this, data can differ from one another as a result of differences between these contexts, can evidence factors of these contexts independent of the phenomenon's occurrence, and may not evidence all characteristics of a phenomenon. By contrast, phenomena are insensitive to some aspects of the contexts from which they are inferred. Bogen and Woodward illustrate the distinction between data and phenomena with an example: the melting point of lead. According to their account, one does not determine how to represent the phase transition of lead from solid to liquid from a single manifestation of lead melting. Rather, one takes a number of measurements. Characteristics of the phase transition of lead can be determined from these measures, and a representation of the phenomenon can be formulated.

From this example, we can infer the characteristics of phenomena according to Bogen and Woodward's account, the core of which is consistent with Hacking's account. Bogen and Woodward "think of phenomena as … belonging to the natural order of the world itself and not just to the way we talk or conceptualize that order," suggesting that phenomena are distinct from their representations (1988). They involve causal interactions, suggesting that phenomena involve or consist in causal activities. Further, phenomena have stable and repeatable characteristics. In this context, *stability* means that the same phenomenon can occur in different contexts, as the manifestation of a phenomenon is invariant to some aspects of the context in which it occurs. *Repeatability* means that the same phenomenon can manifest again and therefore is not a singular

occurrence. Woodward suggests that a phenomenon "has stable recurrent features which can be produced regularly by some manageably small set of factors" (1989, 395). These factors, which others have called "conditions," induce, inhibit, or modulate the manifestation of the phenomenon (Craver & Darden 2013, 56).

The work from Hacking as well as Bogen and Woodward has been widely influential on how philosophers think about scientific phenomena. Two desiderata of an account of phenomena are apparent from these accounts. First, phenomena have causal characteristics that make them candidates for discovery, creation, and manipulation. Phenomena are causally efficacious, as they can cause other things to occur. Further, they themselves can be causally induced. Thus, researchers aim to make causal claims about phenomena, even if what causes a phenomenon is due to human intervention. I call this the *causal desideratum*.

Because of these causal characteristics, researchers can induce the occurrence of a phenomenon with curating a laboratory context in which includes those factors that induce its occurrence and eliminates those that might inhibit its occurrence. However, researchers do not need to induce a phenomenon for it to occur. When a phenomenon manifests without the deliberate induction by researchers, factors of a context are responsible for a its occurrence. In either case, there is a consistency to a phenomenon's occurrence: given its repeatability, we expect that, in principle, a phenomenon will occur again if the same conditions are in place. This is why researchers represent the factors or conditions that induce, inhibit, or modulate the occurrence of the phenomenon. The Hall effect illustrates this point: even if its manifestation is the product of human aid, it nevertheless is a phenomenon of interest due to its repeatable and stable causal characteristics.

Second, representations of manifestations of a phenomenon may be adequate for representing these individuals but not be adequate for representing the phenomenon in question. This is because representations of its manifestations may include the occurrence of characteristics that are idiosyncratic to that manifestation or else do not occur every time that the phenomenon manifests. Thus, representations of manifestations of phenomena are *not equivalent* to representing the phenomenon. I call this the *non-equivalence desideratum*.

There are three senses of 'non-equivalence' to take into account, each of which can be illustrated by a different interpretation of Bogen and Woodward's lead example. First, data that evidence the same phenomenon may be non-equivalent due to measurement error and differences in instrumentation. Thus, according to this sense, measures of a phenomenon (i.e., the data) can be non-equivalent to one another. For example, data evidencing the phase transition of lead may deviate from one another due to biases in thermometric instrumentation, even if these are different measures of the same occurrence of lead melting. Thermometers do not always read the same temperature when measuring the melting of lead; rather, a set of temperatures values typically are recorded. This non-equivalence results from the measurement, not the melting of lead itself. In principle, a phenomenon could manifest exactly the same way again and again, but the data could nonetheless be non-equivalent for this reason.

This sense of 'non-equivalence' is exemplified in the study of spatial reinforcement learning – the phenomenon that humans improve their performance on spatial decision-making tasks over a series of trials when provided with a numerical reward (Jarbo & Verstynen 2015) – which has been measured in tasks where participants use a mouse to move a cursor to the center of a target that appears on a computer screen (Jarbo, Flemming, & Verstynen 2018). In this case, the data are values of the distance between the cursor and the target's center. Due to the muscle

movement of the participant, as well as characteristics of the mouse, monitor, and code that underwrite the task, the data produced in this task will not be identical between participants' or their respective trials. Thus, the variability of these data is not due to the phenomenon's occurrence; the variation is caused by other factors present in this experimental context.

Beyond the data, it is also the case that manifestations of the same phenomenon are not always identical to one another, and the sense of 'non-equivalence' relevant to the desideratum relates to the relation between a phenomenon's representation and those of its manifestations. The second sense of 'non-equivalence' is that representations of a phenomenon's manifestations may be non-equivalent in ways that are irrelevant to representing the phenomenon in question. This is because some characteristics of a phenomenon's manifestation might be idiosyncratic to this manifestation or the context in which is occurs. For example, the color of a lead sample melting might be unique due to particular trace elements in the sample. Though some level of lead purity may be specified in the phenomenon's representation, the quirks of the lead sample melting due to trace elements may not be not included in the phenomenon's representation. Likewise, the occurrence of spatial reinforcement learning sometimes occurs concurrently with a change in bodily movement to make decisions, but these changes are not always present, and their presence or absence has no effect on the occurrence of the phenomenon of interest.

Third, representations of manifestations of a phenomenon may be non-equivalent in ways that are relevant to representing the phenomenon in question. For example, the melting point of lead can be predictably modulated by the pressure under which a sample of lead melts. Thus, the differences between manifestations of lead samples melting due to pressure differences is a characteristic to capture when representing this phenomenon. Likewise, the occurrence of spatial reinforcement learning can be modulated by increasing the size of the target, which systematically

results in the phenomenon occurring at a slower rate. This this a characteristic that researchers aim to capture when representing spatial reinforcement learning.

6.2.2 Attempts to Account for Phenomena

Several accounts of phenomena have been presented following the insights from Hacking, Bogen, and Woodward. While not explicit, each account I will discuss can be interpreted as an attempt to address one of these desiderata. Some philosophers have developed accounts according to which phenomena go beyond the concrete or physical. For instance, Brown defends a view according to which "phenomena are abstract entities which are (or at least correspond to) visualizable natural kinds" (1994, 125). Brown suggests that phenomena are akin to abstract mathematical objects like geometric figures, as he conceives of them: while the figures themselves are not concrete, they "correspond to" concrete particulars. Thus, according to Brown, when researchers conceive of phenomena, they conceive of abstract entities, rather than any particular manifestation. Correspondingly, the representation of a phenomenon represents this abstract object. Teller puts forward a similar account of phenomena, according to which phenomena are best thought of as "ideal types to which real world instances approximate" (2010, 817). According to Teller, the phenomena are types, which are represented by idealizations to which no token corresponds, but only approximates.

By contrast, some philosophers conceive of phenomena as patterns. There are two types of accounts that fit this conception, though debate in the philosophical literature on phenomena has resulted from a failure to disentangle the two. First, a number of philosophers have argued that phenomena to consist in data patterns (e.g., McAllister 1997; Glymour 2000). These accounts fit what has been called a "pattern view" of phenomena, according to which "phenomena are or

correspond to particular patterns in data sets" (Apel 2011, 27-28). Feest presents this kind of account, where "surface phenomena" are "equated to empirical data patterns" and "hidden phenomena" are "removed from particular regularities" (2011, 63). On these accounts, a representation of a phenomenon represents a data pattern. While they appeal to information theory (e.g., McAllister 1997) or causal modeling (e.g., Glymour 2000) as the basis of their use of the term 'pattern,' none of these accounts characterizes what a data pattern is.

Second, there are accounts of phenomena as patterns that are evidenced by data. Woodward is one of these philosophers. He argues that detecting phenomena consists in identifying a "stable and invariant pattern" (1989, 376). This account of phenomena is also represented in his account of causal explanation: to explain the motion of blocks down an inclined plane, he suggests, one could either provide "an explanation of any specific episode of a block sliding down a plane" or explain "generic pattern in the motion of blocks sliding down planes" (Woodward 2003, 17). For Woodward, the latter is an explanation of a phenomenon, this "generic pattern." Though Woodward does not elucidate what he means by 'pattern,' one thing is clear: phenomena are not data patterns on this account (2010a).

Thus, on one side, we find philosophers talking about phenomena as abstract objects or ideal types, and, on the other, we find them talking about phenomena as patterns. These accounts are similar in the sense that, according to them, phenomena can be construed as something that is distinguishable from an individual manifestation. Whether explicit or implicit, each account distinguishes between a phenomenon and its manifestations and, correspondingly, distinguishes between representing a phenomenon and representing its manifestations.[24]

[24] Bogen and Woodward state that phenomena "fall into many different traditional ontological categories… [including] objects, objects with features, events, processes, and states" (1988, 321). Kaiser and Krickel suggest that

I argue that none of these accounts is satisfactory. This is because each account fails to fulfill either the causal or the non-equivalence desideratum. In addition, these accounts implicate a metaphysics of phenomena that is ultimately unnecessary to fulfill these desiderata and make sense of how representations of phenomena factor into the theory and practice of science. I provide reason to think that appeal to abstract objects or ideal types cannot satisfy these desiderata, while appeal to patterns is best explained by my nominalist account.

6.3 Going nominal with regards to phenomena

In accounting for what representations of phenomena represent, I draw inspiration from conceptions of *nominalism*. In its most basic form, nominalism is a stance that consists in the lack of commitment to the reality of "nonindividuals" (Goodman 1966, 37), with the aim of using talk of nonindividuals in a way that is "significant in context but naming nothing" (Goodman & Quine 1947, 105).[25] This lack of commitment is satisfied if "one consistently refuses to interpret the language of [nonindividuals] and provides a formulated syntax for manipulating that language like an abacus," and this talk of nonindividuals "recommends itself only where… translation is so difficult as to seem hopeless" (Goodman 1966, 35). Essentially, this stance saves appeal to nonindividuals, as it makes sense of this talk's significance within inquiry. At the same time, it does not commit to the idea that this nonindividual talk represents anything, except when reformulated into the language of individuals.

---

token phenomena are "object-involving occurrents," which, "at least according to one reading of events," "are simply events" (2017, 769).

[25] This does not involve an overt rejection of anything that is not an individual but merely no commitment "as to whether anything else exists" (Goodman 1966, 37).

6.3.1 Applying Nominalist Ideas to Scientific Phenomena

To apply nominalism to phenomena, I start by identifying as individuals the manifestations of a phenomenon. These are individuals can be induced, inhibited, and discovered as per the causal desideratum. In addition to this, we have representations of phenomena that are distinct from representations of their manifestations as per the non-equivalence desideratum, which are the representations that are involved in type-level claims. They are representations of nonindividuals. Thus, the central aim of using nominalism make sense of phenomena is to have a position where one does not commit to the idea that phenomena (the nonindividuals) are something distinct from their manifestations, yet make sense of representations of phenomena in terms of individuals. While the nominalism of Goodman was designed to focus on lexical or propositional representations, the stance applies to representations more generally. This is because what matters is the ability to formulate representations of nonindividuals into representations of individuals, regardless of the format of these representations.

To unpack my nominalist account, I count any representation of a manifestation of a phenomenon as a token-level representation and any representation of a phenomenon *simpliciter* as a type-level representation. Correspondingly, manifestations of phenomena are phenomenon-tokens, which are distinct from whatever might count as a phenomenon-type – such as abstract objects, ideal types, or patterns. My account provides a way of recognizing that "it is the types we can do without," as "actual discourse… is made up of tokens that differ from and resemble each other in various important ways" (Goodman 1966, 360). This eliminates the need to explain what makes individuals tokens of a particular type or address what these types are. This sets the stage for determining how we should we think of the relation between phenomena and their manifestations.

6.3.2 Reformulating Representations of Phenomena

On my account, representations of phenomena are reformulated into representations of an individual occurrence. Let me begin with propositional formulations, from which we can extrapolate to representations more generally. Claiming that:

> *There is a phenomenon P, which is characterized as having features X, that occurs under conditions Y*

Can be formulated as:

> *There occurs O, which, under conditions Y, is characterized as having features X*

With this formulation, we move from phenomena to individual occurrences. While this claim captures the characteristics of a single manifestation, it does not represent all of its characteristics. Rather, the formulation only includes some of its characteristics, which researchers accept will be shared by other manifestations of a phenomenon. What is included are the co-occurring features and what factors or conditions are necessary to bring about the set of co-occurring features. The representation can be more or less detailed, either in terms of the features thought to constitute the phenomenon of interest (*X*) or the factors or conditions that must be in place to induce, inhibit, or modulate it (*Y*).

Why is this formulation applicable to more than just the individual it describes? This is answered by the fact that the characteristics included are formulated to represent only the characteristics that are expected of any manifestation. Researchers *project* representations of these characteristics on to other instances. In this way, the representation does not specify what researchers expect to be idiosyncratic characteristics endemic to one of a phenomenon's manifestations. Conversely, representing a manifestation of a phenomenon consists in representing some of the characteristics of that individual instance, regardless of whether or not those

characteristics are expected of other manifestations. In this way, the representation of the phenomenon and the representation of a manifestation of this phenomenon can be distinguished. Nevertheless, there is nothing that amounts to a type.

To explain how this formulation can be applied to multiple manifestations, we must also consider the *epistemic stance* one adopts when one accepts a representation of a phenomenon. When one accepts this representation, one accepts that:

> *If conditions Y are in place, then, despite any differences between the contexts, I expect that features X will occur*

Unless there is some inhibiting condition that is also in place, as per the components of the representation in question. This expectation is a prediction about what a manifestation will be like. Thus, accepting a representation of a phenomenon is to expect that manifestations of this phenomenon will occur if the specified in this representation are present. Any characteristics not included in this representation need not be the same, as accepting the representation of a phenomenon does not commit one to accepting anything about unspecified characteristics.[26]

Note that this formulation relates to expectations regarding the occurrence of a phenomenon, rather than any expectations about the data produced from measuring this occurrence. This relates back to the first sense of 'non-equivalence.' Even if one has expectations about the occurrence of a phenomenon, the nature of data collecting will invariably result in a distribution of values, due to other aspects of the context in which the phenomenon occurs. This is in line with Bogen and Woodward's claims: data reflect the context in which they were produced. As a result, a phenomenon's representation must be paired with some understanding of

[26] This explains why there is a pragmatic equivalence between representing data and representing phenomena (Nguyen 2016). The difference is that one holds the epistemic commitment to a representation of a phenomenon but need not hold this commitment to a representation of data.

the measurement and instrumentation used in a study in order to determine what observations should be expected when the phenomenon occurs. This explains why researchers may never observe (say) lead melting at 600.61 degrees Kelvin, even though the phase transition is represented as having this characteristic. Error and variability are the result of the measurement process, which leave their mark on the data that result from this process.

Let me illustrate my commitments with an example. Accepting that 'the melting point of lead is 600.61 degrees Kelvin under normal atmospheric pressure' amounts to accepting that 'at normal atmospheric pressure, if exposed to a temperature of 600.61 degrees Kelvin, I expect that a sample of lead will melt.' In including pressure as a modulating condition of the phenomenon's occurrence, one accepts that 'if the pressure is higher or lower than one Atmosphere, then I expect that a sample of lead will melt at a lower or higher temperature than 600.61 degrees Kelvin, respectively.' With this reformulated representation, researchers can determine how to induce, inhibit, or modulate the melting of lead samples. They can also theorize about the physical properties of lead that explain why lead samples melt at this temperature and why pressure levels affect the melting point of lead.

Are there worries to adopting my account of phenomena? One might worry that representing phenomena involves appeal to other types of things. Types that have a complex relation to their tokens are certainly implicated in some scientific domains – one obvious example is talk of types in evolutionary biology (see Novick 2019). That being said, my account is nominalist about phenomena. It is not a general account of nominalism, and it does not address other types of things that scientists may represent. However, as a heuristic, not committing to types that are distinct from their tokens seems prudent in any case in which commitment to these types is not necessary to account for the theory and practice of science. In other words, if one wishes to

commit to types, one better have a good reason for doing so. For many philosophers, the reason why there is appeal to phenomena as types seems to stem from what I have identified as the non-equivalence desideratum. However, I have shown that accounting for type-level representations of phenomena does not require any positive metaphysical commitment to phenomenon-types. Thus, such a commitment is unnecessary.

My account thus amounts to the following. To represent a phenomenon, one represents what a manifestation of this phenomenon is like, but one only represents the characteristics that they expect to co-occur in other manifestations. When this representation is accepted, one expects that, in any context the circumstances of which are consistent with the specified conditions of the phenomenon's induction, the specified features will co-occur. In formulating a representation of some characteristics of a manifestation, along with an expectation about the characteristics specified in this representation, one can adequately represent everything that is important about phenomena without committing to phenomena as distinct from their manifestations.[27]

6.4 Adequacy of my account

With my nominalist account introduced, let me turn to the benefits of endorsing it. My account satisfies the two desiderata of an account of phenomena I examined in Section 6.2.1. It also makes sense of scientific inquiry about phenomena and the data-phenomena distinction.

[27] While nominalists who portray their accounts through language inform my account, my account is more general. On my account, any representation of a phenomenon in any format can be formulated in terms of the characteristics of an instance, along with the expectation that the characteristics that are specified will co-occur in other instances.

6.4.1 Satisfaction of Desiderata

My account satisfies the causal desideratum. The representation is formulated to include some characteristics of its manifestations. These manifestations can be induced, inhibited, or modulated, if researchers develop or discover a context that has the requisite conditions or factors in place. In such a context, researchers can predict that a manifestation of the phenomenon that has the specified characteristics will occur, even if a well-regulated environment of the laboratory is the only place in which its characteristics co-occur. A phenomenon is causally efficacious in the sense that the occurrence of its manifestations can have the same effects. In light of Hacking's account of phenomena, the features that are shared by the manifestations of a phenomenon are the features that are causally responsible for certain effects in the systems in which they occur. Thus, because the manifestations share these features, and these features have stereotypical causes and effects, the phenomenon can be said to have a causal role that corresponds to the causal character of its manifestations. For these reasons, researchers can formulate type-level causal claims in light of their representations of a phenomenon of interest, without needing to commit to the reality of a type.

The example of the Hall effect illustrates this point. Recall that the Hall effect occurs when a voltage difference is produced in a conductor as the result of a magnetic force applied perpendicularly to the current flow in the conductor. For this effect to occur, the conductor must be isolated from all magnetic forces aside from that which induce it. According to Hacking, this is extremely unlikely to occur in nature without human aid. Why this is the case is simple: magnetic fields stemming from the earth, other natural systems, and human instrumentation permeate us all of the time. Thus, without appropriate apparatus and technique, it is (probably) not possible to have a context in which the conditions specified in the representation of the phenomenon can be

in place. However, it is possible to achieve this magnetic isolation in laboratory contexts. Thus, researchers can induce the Hall effect, maintain their epistemic stance towards its representation, and determine what effects the Hall effect can have in the systems in which it manifests. As all we care about are individual manifestations of the Hall effect, there is no issue with making causal claims about these manifestations, and, because the manifestations share the right characteristics, causal claims about these manifestations can be made.

My account also satisfies the non-equivalence desideratum. A phenomenon's representation does not need not be equivalent to the representation of a single manifestation of a phenomenon. As Bogen and Woodward note, manifestations of phenomena – as well as the data that result from the measurement of these manifestations – may be idiosyncratic and thus have characteristics that are not expected to occur every time the phenomenon in question manifests. Because of this, a representation of a manifestation may include characteristics that are not included in the representation of the phenomenon. In addition, researchers do not take the same epistemic stance towards representations of manifestation of phenomena that include characteristics they do not expect to occur in other contexts. In other words, they do not predict that the manifestation of a phenomenon, inclusive of its non-shared characteristics, will occur again if the specified conditions in the representation are in place in a new context. As a result, representations of individual manifestations of a phenomenon are not equivalent to the representation of a phenomenon.

We have already seen how the first sense of 'non-equivalence' is fulfilled. The data researchers produce in measuring a phenomenon's occurrence are variable due to factors relating to the instrumentation and measurement procedure. As illustrated by melting lead, samples of lead melting measured with different thermometers might yield different data, due to differences in

measurement error. Likewise, data that evidence spatial reinforcement learning are variable because of factors of the experimental context that are independent of this phenomenon's occurrence. However, in both of these cases, researchers have expectations about what will occur, which are captured in their representations of these respective phenomena. To determine what they should expect to observe when these phenomena occur, researchers must pair these representations with some understanding of how the phenomenon is measured and how their instrumentation works. From this, they can determine a distribution of values of the data that they should expect when these phenomena occur.

By contrast, according to the second sense of 'non-equivalence,' individual manifestations of lead melting may differ from one another, due to differences about the cases that are not specified in the representation of the phenomenon in question. As mentioned earlier, the composition of a lead sample may include trace elements that change what happens when this sample melts, which may be captured in a representation of this manifestation. Likewise, spatial reinforcement learning is sometimes paired with changes in bodily movement, which may be captured when representing solely an individual occurrence of this phenomenon. However, these changes need not be something that researchers make predictions about when representing the phenomenon. Thus, attempts to represent these manifestations will not be equivalent to the representation of the phenomenon in question. However, with the representation of the phase transition or spatial reinforcement learning *as a* phenomenon, researchers commit to a prediction of what will be common to each manifestation.

What about the third sense of 'non-equivalence'? This is where the specification of modulation conditions, or conditions which change how the phenomenon is manifest, are needed in the representation (Craver & Darden 2013, 52). Within some set of boundary conditions, the

phenomenon's manifestations may vary from one to another, but these variations are important to capture via the representation of the phenomenon in question. For example, modulating conditions relating to the pressure and level of purity of the lead are included when representing the phenomenon of the phase transition of lead. This is why the melting point of lead is typically specified at standard atmospheric conditions and to a certain degree of lead purity. As a result of these modulating conditions, predictions can be made about variance in the temperature of the melting of a sample of lead. Researchers expect that the phase transition will occur at 600.61 degrees Kelvin at standard atmospheric pressure – again, this is not equivalent to saying that they will *observe* lead samples melting at this exact temperature – but, if the pressure is increased, the melting point correspondingly increases. Likewise, researchers expect that, systematically, spatial reinforcement learning will be slower if the target's size is increased. Thus, representing the phenomenon in this case includes specifying predictions about both what is consistent and what is inconsistent between manifestations as a result of differences in modulating conditions. This is why, unlike the examples earlier in this chapter, representations of phenomena often explicitly include a litany of inducing, inhibiting, and modulating conditions, whose inclusion in a representation allow researchers to predict under what conditions a phenomenon of interest will manifest and what features each manifestation will have.

6.4.2 Elucidation of Discovery, Evaluation, or Rejection of Phenomena

With the desiderata satisfied, we can examine the additional benefits of my account. First, my account explains what it means for researchers to discover a phenomenon. In fact, two distinct senses of 'discovery' can be explained with my account. If one adopts a fairly thin sense of 'discovery,' one can be said to discover a phenomenon if one discovers one of its manifestations.

That is all that is needed to formulate a representation of the phenomenon, according to my account. By contrast, one could adopt a stronger sense of 'discovery,' according to which one can be said to discover a phenomenon if one formulates a representation of a manifestation of a phenomenon and has evidence to determine to what contexts to apply this representation. A stronger sense of discovery here reflects that it typically requires the identification of multiple manifestations of a phenomenon to formulate an adequate representation of the phenomenon (Kordig 1978). Identification of multiple manifestations will provide evidence regarding what should be included in a representation to project to other manifestations. While one might get lucky and include only the common features of manifestations of phenomena from the discovery and formulation of a representation of one of its manifestations, this is typically not the case, and researchers should not assume that it is the case. With multiple manifestations, researchers can determine what expectations to have about a phenomenon's occurrence. In doing so, they can formulate the representation so that only projectable characteristics of a manifestation of the phenomenon are included, and they can also determine if their measurements are biased.

Turning to the second benefit, my account explains what researchers do when they evaluate representations of a phenomenon. All evaluations are tests of researchers' expectations about a phenomenon's occurrence. In other words, evaluating a representation of a phenomenon consists in testing the hypothesis consistent with this representation. If they disconfirm this hypothesis, this provides evidence against the accuracy of the representation. This can provide evidence to revise the representations as researchers learn more about what characteristics are common to manifestations of the phenomenon of interest and in what contexts these manifestations occur. This iterative revision of a representation of a phenomenon may require either subtracting features

constitutive of the phenomenon's occurrence from the representation or adding conditions or factors that, when in place, induce the phenomenon's occurrence.

Finally, my account explains what it means for researchers to reject a representation of a phenomenon, or, more colloquially, for them to claim that a phenomenon isn't real. Rejecting a representation of a phenomenon amounts to claiming that no manifestation of the phenomenon occurs: nothing occurs that has the characteristics specified in the phenomenon's representation. No manifestations can be induced, though conditions specified in its representation are in place. In such a case, researchers either revise the characterization in light of new evidence or reject it outright. To sum up, I argue that each of these components of scientific practice – discovery, evaluation, and rejection – can be understood while maintaining no commitment to phenomena as types.

6.4.3 Elucidation of Data-Phenomena Distinction

My account also explains why it is challenging to formulate a representation of a phenomenon from data, as initially described by Bogen and Woodward. Data, the recordings made in empirical contexts, evidence manifestations of the phenomenon of interest and thus may evidence the characteristics that researchers do not expect will be common to other manifestations. This is not surprising, given the non-equivalence desideratum and the example of spatial reinforcement learning. However, this problem is compounded due to the nature of collecting data. Data do not provide evidence of every single characteristic of these manifestations. This is because, in an empirical context, there may be an unbounded number of factors that can be recorded. As a result, whatever data are collected will not exhaustively evidence everything about that context and the phenomena that occur therein.

The partiality of data also explains why data plays a different epistemic role than representations of phenomena. This is a concern raised by Leonelli, who rejects the idea that data are "local" when compared to claims about phenomena (2009, 737). Leonelli challenges the idea that "data are intrinsically local" – in the sense that researchers must be acquainted with the means of their production in order to use the data as evidence for scientific claims – by illustrating the methods that can be used to package data in a format in which researchers can appeal to it across contexts (2009, 746). To explain the differences between their epistemic roles, Leonelli suggests that the feature of claims about phenomena that is critical is their format; phenomena are represented in the form of propositions, while data are not (2016, 87).

The distinction between the epistemic roles of data and representations of phenomena is neither due to their locality nor due to their format; instead, it is due to the difference between what they evidence. Phenomena need not be represented in a propositional format and often times are not. Phenomena can be represented mathematically, in the form of phenomenological models (Cartwright 1983, 148), or pictorially, in the form of diagrams (Sheredos et al. 2013). Correspondingly, researchers can (and often do) propositionally represent data. Nevertheless, these differences in representational format need not change what they evidence. Rather than their format, what matters to the difference between data and representations of phenomena is that the former evidence characteristics of one of several manifestations of a phenomenon. The latter represent the shared characteristics of all manifestations of a phenomenon, which serve as evidence for systematic theory (Bogen & Woodward 1988). This difference is made transparent by my account. Data partially evidence characteristics of a manifestation of a phenomenon, some of which may not be expected to occur every time the phenomenon manifests. Representations of phenomena do not.

6.5 Doing away with alternative accounts of phenomena

With the adequacy of my account presented, I now turn to the deficits of alternative accounts of representing phenomena. I examine the problems with appeals to phenomena as abstract objects or ideal types, which stem from their inability to capture the relation between type and token as well as the causal characteristics of phenomena. I also provide reason to accept that best way to explain what it means to call a phenomenon a "pattern" is to adopt my nominalist account, which explains the commitments of these accounts and makes clear the issues that come with suggesting that patterns are metaphysically real in a way that is distinctive of what instantiates these patterns.

6.5.1 Accounts of Phenomena in Terms of the Abstract or Ideal

On the positive side, accounts of phenomena as abstract objects satisfy the non-equivalence desideratum. The abstract object, which "corresponds" to the tokens, is distinct from any concrete manifestation, and their representations are distinguishable because of this fact. Likewise, accounting for phenomena as ideal types also satisfies the non-equivalence desideratum. Representations of phenomena are idealized due to the fact that they represent ideal types, while representations of token manifestations of this phenomenon do not.

However, both kinds of accounts fail to satisfy the causal desideratum. Abstract objects and ideal types are ill-equipped to convey the causal role phenomena play in systems under investigation or the fact that phenomena are candidates for creation and manipulation. Abstract objects or ideal types are causally inefficacious and thus cannot have causal properties to be

investigated with experiments and observations. Thus, causal claims about phenomena are not coherent if phenomena are abstract objects or ideal types. Supporters of these kinds of accounts do not escape this by arguing that it is merely the manifestations of phenomena about which we can make causal claims. If this position is taken, then the phenomenon qua abstract object or ideal type becomes an "idle wheel": neither a thing that we interact with in scientific practice, yet still somehow distinct from our representations. In addition, and as a result, these accounts give no understanding of how one learns about phenomena from concrete or non-ideal manifestations in the world. This is because the correspondence or approximation relation between phenomena and their manifestations that are implicated by these accounts are underdescribed. Thus, these accounts do not provide an adequate explanation of how researchers formulate representations of phenomena from scientific practice.

Are representations of phenomena idealized? My account makes sense of the fact that their formulation invariably involves some level of what might be considered abstraction: after all, only some characteristics of a manifestation of a phenomenon are included in a representation of the phenomenon in question. Likewise, the addition of modulating conditions provides some understanding of how manifestations of phenomena might be different from one another in ways that are expected according to the phenomenon's representation. This captures that not all manifestations will be identical. Ultimately, what looks like idealization – such as saying that the melting point of lead is (exactly) 600.61 degrees Kelvin when in reality samples might melt at slightly different temperatures – shows that the conditions specified in the representation provide insight into how the features of a phenomenon might modulate in predictable ways based on variations in the context in which they occur. Because this is taken into account when formulating a phenomenon's representation, this representation can nonetheless be used to formulate type-level

claims. Thus, I accept that may be reasons to formulate representations of a phenomenon that will depart to some degree with some of its manifestations. However, this does not entail that a phenomenon is itself is best conceived of as an ideal or abstract object.

Overall, accounts of phenomena according to which they are abstract objects or ideal types are not adequate for capturing the desiderata of phenomena. In addition, any benefits of these accounts are captured by my nominalist account.

6.5.2 Phenomena and Patterns

While accounting for phenomena in terms of patterns is common in the philosophy of science, these accounts are also limited. This is because, I argue, replacing 'phenomenon' with 'pattern' passes the epistemic and ontological buck from one notion to another: the metaphysics of patterns are equally unclear when compared to the metaphysics of phenomena. As mentioned before, there are two ways to conceiving of phenomena as patterns. First, there is the idea that phenomena are data patterns. Second, there is the idea that phenomena are patterns because individuals share characteristics, and researchers can formulate representations of these characteristics. I argue that talk of patterns in either of these ways is fine, in the same way talk of phenomena is fine for the nominalist. Talking of patterns is a valuable way of organizing one's claims about different manifestations. However, there is no reason to accept that patterns are something distinct from their instances.

The idea that phenomena are data patterns could be construed as Dennett's "real patterns." Recall Dennett's image of bar codes that are comprised of black and white squares (Dennett 1991, 31). One can image that a different level of noise has corrupted each of the bar codes. Each has a

different percentage of the pixels within the image that have been randomly set to black or white. In the bar code with 1% noise, only a few pixels stray outside of their squares. However, in the pattern with 50% noise, the squares are no longer determinable. In this case, patterns can be deciphered from the bar codes. Moving beyond the example, considering phenomena to be patterns is, on this account, accepting that phenomena are the regularities that are detectable in data.

This kind of account also satisfies the non-equivalence desideratum. Representations of phenomena are distinct from representations of their manifestations because the former represent data patterns. In this distinction, we can also explain how they relate: representations of phenomena represent these data patterns. These patterns are "real" in the sense that they are distinct from a phenomenon's manifestations. Thus, if one aims to defend the idea that scientific phenomena are data patterns, then one is committed to there being a type that is a phenomenon, which we do not do without.

Such a view is not best applied to thinking about phenomena. One reason for this is that this kind of view fails to provide any account of the relation between data and phenomena. Data are recordings, and data patterns – for example, a linear regression performed on a dataset – are transformations of these data. But, as mentioned before, data are thought to be the recordings *of* and evidence *for* the occurrence of a phenomenon. Because they are recordings of it, data evidence this phenomenon. These views confuse the result of scientific investigation of a phenomenon of interest with the target of this investigation: namely, the phenomenon of interest. Thus, adopting such a view about phenomena seems to require either abandoning the idea that data evidence phenomena or committing to the idea that data evidence data patterns. At the same time, this view also results in committing to the idea that all distinct data patterns are distinct phenomena. This commitment removes the possibility that the same phenomenon can be evidenced by different data

patterns, which, in turn, brings into question whether or not these phenomena can be detected in different contexts and via distinct empirical studies. Perhaps this issue could be resolved with a more substantive understanding of what a data pattern is, but these views provide no such understanding.

Another reason to not equate phenomena with data patterns is that such a view does not satisfy the causal desideratum. Data patterns are not causally efficacious in the way that phenomena are purported to be. For example, a reason why the placebo effect is interesting is that its occurrence can cause researchers to think that a drug is an effective treatment when patients exhibit treatment symptoms. If we take the placebo effect to be a data pattern, such as a transformation of data collected from contexts in which this effect has been measured, then this reason becomes incoherent. This data pattern does not cause researchers to think that an inert drug is effective; rather, this pattern results from measuring these cases. Like the previous one, this reason is an issue for views that equate phenomena with data patterns in part due to the fact that no supporter of one of these views provides a clear description of what a data pattern is. This ultimately makes it difficult to understand what it means to say a phenomenon is a data pattern. Likewise, it leaves no indication of what the metaphysical status of these patterns are and to what degree they are dependent on the production of data by researchers.

Conversely, one could argue that the data patterns simply represent the characteristics of a phenomenon, where measures of manifestations of this phenomenon are consistent with this pattern. This may be what Apel means when he states that phenomena "correspond" to patterns in data, though this correspondence relation is also underdescribed. However, on such a reading of this view, it is no longer the case that phenomena are data patterns. If one takes this position, data patterns are merely a means of determining the characteristics of a phenomenon, which is

accommodated on my account. Talk of patterns, like talk of phenomena, is fine to nominalist ears, but it must be understood in terms of individual occurrences. Thus, while it is possible to talk of a data pattern evidencing a phenomenon, this is not equivalent to simply equating the two.

By contrast, one could adopt a view that phenomena are patterns that exist in some way distinct from instances of this pattern. Such a view suggests that "patterns of spatiotemporal property distribution exist in nature," and, "while scientists are certainly free to label different patterns of coherence as constituting different structures… it is surely not up to them what patterns there are" (Chakravartty 2011, 171). Endorsing such an account to characterize what phenomena are involves committing to the idea that patterns "exist" in some way that is dependent on the world. On this account, whether or not phenomena are patterns is not dependent on whether or not researchers identify them, unlike with data patterns. One could thus accept an account according to which phenomena are patterns of coherence along these lines: to represent a phenomenon is to represent a "pattern of coherence."

On such a view, there "exist" both a phenomenon as a pattern of coherence and manifestations of a phenomenon as instances of this pattern. The question is in what way the two relate. One way to answer this question is to suggest that patterns of coherence exist in some way independently from the manifestations that instantiate these patterns. This leads to a follow up question: what is the metaphysical status of one of these patterns? Taking patterns to be abstract objects (e.g., Resnik 1981, 530) returns us to the problems I present in Section 6.5.1. After all, we want to be able to make type-level causal claims which implicate phenomena, and abstract objects are not causally efficacious. However, if patterns of coherence are not abstract objects and nonetheless are distinct from individual manifestations that instantiate the pattern, it is unclear what these patterns are. Further, it is unclear how, whatever these patterns of coherence are, that

they can satisfy the causal desideratum, given that it appears that it is the manifestations of phenomena that play a causal role in the systems that researchers investigate.

Perhaps the better way to interpret this view is to suggest that patterns of coherence "exist" insofar as manifestations of phenomena instantiate these patterns. On this view, the "existence" of phenomena as patterns of coherence amounts to the fact that manifestations of a phenomenon are similar and that these similarities fit a pattern. In this sense, it is not up to the researchers which manifestations that have similar characteristics they will uncover through their investigations. However, if this is the case, then I argue that all discussion of a phenomenon as a pattern can be reformulated into talk about manifestations of this phenomenon. This avoids the need to characterize the metaphysics of patterns. It is a fact about observations that certain features do or do not co-occur when certain conditions are in place. Whether or not we actually determine that features co-occur is a matter of scientific inquiry. Thus, one can talk of patterns of coherence when describing what phenomena are, but there is no requirement to make sense of what it means to be a pattern, except formulated as the characteristics of a single instance along with the commitment that these characteristics will co-occur again in certain contexts.

The point of this all is simple. Either you accept that patterns exist and are distinct from instances of this pattern or you do not. If you accept, then commitment to phenomena as patterns comes with a number of metaphysical commitments that must be resolved, akin to the issues faced by accounts that regard phenomena as abstract objects or ideal types. If you do not accept, then, while typical to common parlance, appeals to patterns become *unnecessary* to adequately answer the question of what representations of phenomena represent. My account can make sense of the philosophical and scientific use of terms like 'phenomenon' or 'pattern,' but, when reformulated,

all of these terms can be removed with no issue, leaving only manifestations of a phenomenon and the expectations that scientists hold about them.

6.6 Conclusion

The purpose of this chapter has been to account for the scientific and philosophical use of the term 'scientific phenomenon,' while not committing to the idea that scientific phenomena are something beyond manifestations about which we generate type-level representations. I have achieved this by adopting a nominalist account of phenomena and identifying the importance of one's epistemic stance towards these representations when one accepts them. With my account, there is ultimately no need to account for phenomena by appealing to types or kinds of things in any meaningful way. Nonetheless, we can make sense of talk of phenomena in these terms. What matters is that researchers represent the characteristics of a manifestation and make predictions about the characteristics of other manifestations in light of what they have represented.

In the end, then, the question posed in the title of this chapter is distinct from the question "what is a scientific phenomenon?" to which I provide no answer. But, the title question is one that I answer through redirection. This is because, in the important sense, I do not commit to the idea that representations of phenomena represent *any type of thing*. However, rather eliminating all talk of phenomena from science, I instead suggest that we understand how representations of phenomena can be reformulated into representations of aspects of their manifestation. I save representations of phenomena, as they play a significant role in scientific inquiry.

Bibliography

Abraham, W. (2003). How long will long-term potentiation last? Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 358(1432), 735–44.

Apel, J. (2011). On the meaning and the epistemological relevance of the notion of a scientific phenomenon. Synthese, 182(1), 23-38.

Babich, F., Jacobson, A., Bubash, S., & Jacobson, A. (1965). Transfer of a response to naive rats by injection of ribonucleic acid extracted from trained rats. Science, 149(3684), 656-657.

Barker, A. (1998). The history and basic principles of magnetic nerve stimulation. Electroencephalography and Clinical Neurophysiology, Supplement 51, 3-21

Barker A., Jalinous R., & Freeston I. (1985). Non-invasive magnetic stimulation of human motor cortex. The Lancet 325(8437), 1106-1107

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

Bechtel, W., & Richardson, R. (2010). Discovering Complexity: Decomposition and Localization as Strategies in Scientific Research. MIT Press.

Behrmann, M., Geng, J. J., & Shomstein, S. (2004). Parietal cortex and attention. Current opinion in neurobiology, 14(2), 212-217.

Benedetti, F., Mayberg, H. S., Wager, T. D., Stohler, C. S., & Zubieta, J. K. (2005). Neurobiological mechanisms of the placebo effect. Journal of Neuroscience, 25(45), 10390-10402.

Bennett E. (1970). Comments on The planarian controversy. In Molecular approaches to learning and memory (p. 151).

Bennett, E., & Calvin, M. (1964). Failure to train planarians reliably. Neurosciences Research Program Bulletin, 2(4), 3-24.

Bliss, T., & Gardner-Medwin, A. (1973). Long-lasting potentiation of synaptic transmission in the dentate area of the unanaesthetized rabbit following stimulation of the perforant path. J. Physiol, 232, 357-374.

Bliss, T., & Lømo, T. (1973). Long-Lasting Potentiation of Synaptic Transmission in the Dentate Area of the Anaesthetized Rabbit Following Stimulation of the Perforant Path. The Journal of Physiology 232(2), 331–56.

Blundon, J., & Zakharenko, S. (2008). Dissecting the Components of Long-Term Potentiation. The Neuroscientist 14(6), 598–608.

Bogen, J., & Woodward, J. (1988). Saving the phenomena. The Philosophical Review, 97(3), 303-352.

Brown, J. R. (1994). Smoke and mirrors: How science reflects reality. New York: Routledge.

Burian R. (1997). Exploratory experimentation and the role of histochemical techniques in the work of Jean Brachet, 1938-1952. History and Philosophy of the Life Sciences, 19(1):27-45.

Burian R. (2007). On microRNA and the need for exploratory experimentation in post-genomic molecular biology. History and Philosophy of the Life Sciences, 28(3):285-311.

Byrne, W., Samuel, D., Bennett, E., Rosenzweig, M., Wasserman, E., Wagner, A., ... & Fenichel, R. (1966). Memory transfer. Science 153, 635-636.

Cartwright, N. (1983). How the laws of physics lie. Oxford: Oxford University Press.

Cartwright, N. (1991). Replicability, reproducibility, and robustness: comments on Harry Collins. History of Political Economy, 23(1), 143-155.

Chakravartty, A. (2011). Scientific realism and ontological relativity. The Monist, 94(2), 157-180.

Chisholm, R. M. (1966). Theory of knowledge. Englewood Cliffs: Prentice-Hall.

Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. Annual Review of Neuroscience, 22(1), 319-349.

Collingridge, G., Kehl S., & McLennan, H. (1983). Excitatory Amino Acids in Synaptic Transmission in the Schaffer Collateral-Commissural Pathway of the Rat Hippocampus. The Journal of Physiology, 334(1), 33–46.

Collins, H. & Pinch, T. (1998). The golem: What you should know about science. Cambridge University Press.

Craver, C. (2003). The Making of a Memory Mechanism. Journal of the History of Biology, 36: 153–95.

Craver, C. (2004). Dissociable Realization and Kind Splitting. Philosophy of Science, 71(5), 960–71.

Craver, C. (2007). Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience. New York: Oxford University Press.

Craver, C. (2009). Mechanisms and Natural Kinds. Philosophical Psychology, 22(5), 575–94.

Craver, C., & Darden, L. (2013). In search of mechanisms: Discoveries across the life sciences. University of Chicago Press.

Craver, C., & Kaplan, D. (2014). Towards a Mechanistic Philosophy of Neuroscience. In The Bloomsbury Companion to the Philosophy of Science, ed. Steven French and Jaha Saatsi, 268-292. London, Bloomsbury Academic.

De Craen, A. J., Kaptchuk, T. J., Tijssen, J. G., & Kleijnen, J. (1999). Placebos and placebo effects in medicine: historical overview. Journal of the Royal Society of Medicine, 92(10), 511.

Dennett, D. C. (1991). Real patterns. The Journal of Philosophy, 88(1), 27-51.

Duhem, P. (1906). La theorie physique. Son objet et sa structure, Chevalier et Riviere, Paris. Translated by P.P. Wiener, The Aim and Structure of Physical Theory. Princeton University Press, Princeton.

Elliott, K. (2007). Varieties of exploratory experimentation in nanotoxicology. History and Philosophy of the Life Sciences, 28(3), 313-36

Everitt, B. & Skrondal, A. (2010). The Cambridge Dictionary of Statistics. 4th Edn. Cambridge University Press.

Feest, U. (2011). What exactly is stabilized when phenomena are stabilized?. Synthese, 182(1), 57-71.

Feest, U. (2012). Exploratory experiments, concept formation, and theory construction in psychology. Scientific Concepts and Investigative Practice, 3, 167-189

Feest, U. (2017). Phenomena and Objects of Research in the Cognitive and Behavioral Sciences. Philosophy of Science, 84(5), 1165–76.

Feest, U., & Steinle, F. (2016). Experiment. In: Humphreys P (ed) Oxford Handbook of Philosophy of Science. Oxford University Press, New York, pp. 274-295

Fjerdingstad, E. (1971). Chemical transfer of learned information. Amsterdam: North-Holland Publishing Co.

Franklin, A. (1986). The neglect of experiment. Cambridge University Press.

Franklin, A. (1989). The epistemology of experiment. In D. Gooding, T. Pinch, & S. Schaffer (eds.), The Uses of Experiment. Cambridge: Cambridge University Press.

Franklin, L. (2005). Exploratory experiments. Philosophy of Science, 72(5), 888-99.

Friedman, J. (2013). Suspended Judgment. Philosophical Studies, 162(2), 165–81.

Friedman, J. (2017). Why Suspend Judging? Noûs, 51(2), 302–26.

Garson, J. (2017). Mechanisms, Phenomena, and Functions. In The Routledge handbook of mechanisms and mechanical philosophy (pp. 122-133). Routledge.

Giere, R., Bickle, J., & Mauldin, R. (2006). Understanding Scientific Reasoning, 5th Edn. Thomson Wadsworth, Toronto.

Glennan, S. (2002). Rethinking Mechanistic Explanation. Philosophy of Science, 69(S3), S342-S353.

Glymour, B. (2000). Data and phenomena: A distinction reconsidered. Erkenntnis, 52(1), 29-37.

Goldin-Meadow, S. (2016). Why preregistration makes me nervous. *APS Observer*, *29*(7).

Goldstein, A., Sheehan, P., & Goldstein, J. (1971). Unsuccessful attempts to transfer morphine tolerance and passive avoidance by brain extracts. Nature, 211: 1160-1163.

Goodman, N. (1966). The structure of appearance. Second Edition. New York: Bobbs-Merrill.

Goodman, N., & Quine, W. V. (1947). Steps toward a constructive nominalism. The Journal of Symbolic Logic, 12(4), 105-122.

Gottlieb, J. (2007). From thought to action: the parietal cortex as a bridge between perception, action, and cognition. Neuron, 53(1), 9-16.

Hacking, I. (1983). Representing and intervening: Introductory topics in the philosophy of natural science. Cambridge University Press.

Hacking, I. (1992). The self-vindication of the laboratory sciences. In A. Pickering (ed.), Science as Practice and Culture (pp. 29-64). Chicago: University of Chicago Press.

Halas, E., James, R., & Knutson, C. (1962). An attempt at classical conditioning in the planarian. Journal of Comparative and Physiological Psychology, 55(6), 969.

Halas, E., James, R., & Stone, L. (1961). Types of responses elicited in planaria by light. Journal of Comparative and Physiological Psychology, 54(3): 302.

Hempel, C. (1966). Philosophy of natural science. Prentice Hall, Englewood Cliffs

Horvath, P., & Barrangou, R. (2010). CRISPR/Cas, The immune system of bacteria and archaea. Science, 327(5962), 167-70

Huang, E. (1998). Synaptic plasticity: Going through phases with LTP. Current Biology, 8(10), R350–52.

Irwin, L. (1978). Fulfillment and frustration: the confessions of a behavioral biochemist. Perspectives in Biology and Medicine, 21(4), 476-491.

Jarbo, K., Colaço, D., & Verstynen, T. D. (under review). The contextual framing of loss on risky spatial decisions impacts harm avoidance.

Jarbo, K., Flemming, R., & Verstynen, T. D. (2018). Sensory uncertainty impacts avoidance during spatial decisions. Experimental Brain Research, 236(2), 529-537.

Jarbo, K., & Verstynen, T. D. (2015). Converging structural and functional connectivity of orbitofrontal, dorsolateral prefrontal, and posterior parietal cortex in the human striatum. The Journal of Neuroscience, 35(9), 3865-3878.

Kaiser, M., & Krickel, B. (2017). The metaphysics of constitutive mechanistic phenomena. The British Journal for the History and Philosophy of Science, 68(3), 745-779.

Karaca, K. (2013). The strong and weak senses of theory-ladenness of experimentation: Theory-driven versus exploratory experiments in the history of high-energy particle physics. Science in Context, 26(1), 93-136

Kordig, C. R. (1978). Discovery and justification. Philosophy of Science, 45(1), 110-117.

Krech, D., & Bennett, E. (1971). Interbrain information transfer: a new approach and some ambiguous data. In Chemical transfer of learned information (pp. 143-163). Amsterdam: North-Holland Publishing Co.

Kullmann, D., & Lamsa, K. (2008). Roles of distinct glutamate receptors in induction of anti-Hebbian long-term potentiation. The Journal of Physiology, 586(6), 1481–86.

Leonelli, S. (2009). On the locality of data and claims about phenomena. Philosophy of Science, 76(5), 737-749.

Leonelli, S. (2016). Data-centric biology: A philosophical study. University of Chicago Press.

Levkovitz, Y., Harel, E., Roth, Y., et al. (2009). Deep transcranial magnetic stimulation over the prefrontal cortex: evaluation of antidepressant and cognitive effects in depressive patients. Brain Stimulation, 2(4), 188-200

Lømo, T. (1966). Frequency potentiation of excitatory synaptic activity in the dentate area of the hippocampal formation. Acta Physiologica Scandinavica, 68(Suppl 277), 128.

Lømo, T. (2003). The discovery of long-term potentiation. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 358(1432), 617–20.

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

Malenka, R., & Bear, M. (2004). LTP and LTD: An embarrassment of riches. Neuron 44(1), 5–21.

McAllister, J. W. (1997). Phenomena and patterns in data sets. Erkenntnis,, 47(2), 217-228.

McConnell, J. (1962). Memory transfer via cannibalism in planaria. J. Neuropsychiat, 3, 1-42.

McConnell, J. (1965). A manual of psychological experimentation on planarians. Ann Arbor: Worm Runner's Digest.

Menkes, D., Bodnar, P., Ballesteros, R., & Swenson, M. (1999). Right frontal lobe slow frequency repetitive transcranial magnetic stimulation (SF r-TMS) is an effective treatment for depression: a case-control pilot study of safety and efficacy. Journal of Neurology, Neurosurgery and Psychiatry, 67(1), 113-115

Meyer, D. E., Abrams, R. A., Kornblum, S., Wright, C. E., & Keith Smith, J. E. (1988). Optimality in human motor performance: ideal control of rapid aimed movements. Psychological Review, 95(3), 340.

Nguyen, J. (2016). On the pragmatic equivalence between representing data and phenomena. Philosophy of Science, 83(2), 171-191.

Nissen, T., Røigaard-Petersen, H., & Fjerdingstad, E. (1965). Effect of ribonucleic acid (RNA) extracted from the brain of trained animals on learning in rats (II). Scandinavian Journal of Psychology, 6(2), 265-272.

Novick, A. (2019). Cuvierian Functionalism. Ann Arbor, MI: Michigan Publishing, University of Michigan Library.

O'Malley, M. (2007). Exploratory experimentation and scientific practice: Metagenomics and the proteorhodopsin case. History and Philosophy of the Life Sciences, 28(3), 337-360.

O'Malley, M., Elliott, K., Haufe, C., & Burian, R. (2009). Philosophies of funding. Cell, 138(4), 611-615.

Pollock, J. L. (1987). Defeasible reasoning. Cognitive science, 11(4), 481-518.

Polson, M., Barker, A., Freeston, I. (1982). Stimulation of nerve trunks with time-varying magnetic fields. Medical and Biological Engineering and Computing, 20(2), 243-244.

Price, D. D., Finniss, D. G., & Benedetti, F. (2008). A comprehensive review of the placebo effect: recent advances and current thought. Annual Review of Psychology, 59, 565-590.

Radder, H. (1996). In and About the World: Philosophical studies of science and technology. SUNY Press.

Rashevsky, N. (1968). Some possible theoretical implications of experiments on the chemical transfer of memory. Bulletin of Mathematical Biology, 30(2), 341-349.

Resnik, M. D. (1981). Mathematics as a science of patterns: Ontology and reference. Noûs, 15: 529-550.

Rheinberger, H. (1997). Toward a History of Epistemic Things: Synthesizing proteins in the test tube. Stanford University Press, Stanford

Rilling, M. (1996). The mystery of the vanished citations: James McConnell's forgotten 1960s quest for planarian learning, a biochemical engram, and celebrity. American Psychologist, 51(6), L589.

Røigaard-Petersen, H., Nissen, T., & Fjerdingstad, E. (1968). Effect of ribonucleic acid (RNA) extracted from the brain of trained animals on learning in rats (III). Scandinavian Journal of Psychology, 9, 1-16.

Rosenblatt, F. (1967). Recent work on theoretical models of biological memory. In In Computer and Information Sciences-II. New York: Academic Press.

Rosenblatt, F., Farrow, J., & Herblin, W. (1966). Transfer of conditioned responses from trained rats to untrained rats by means of a brain extract. Nature, 209(5018), 46-48.

Rosenblatt, F., Farrow, J., & Rhine, S. (1966a). The transfer of learned behavior from trained to untrained rats by mean of brain extracts. I. Proceedings of the National Academy of Sciences, 55(3), 548-555.

Rosenblatt, F., Farrow, J., & Rhine, S. (1966b). The transfer of learned behavior from trained to untrained rats by means of brain extracts. II. Proceedings of the National Academy of Sciences, 55(4), 787-792.

Salih, F., Khatami, R., Steinheimer, S., Kretz, R., Schmitz, B., & Grosse, P. (2007). A hypothesis for how non-REM sleep might promote seizures in partial epilepsies: A transcranial magnetic stimulation study. Epilepsia, 48(8), 1538-1542

Sauvé, W., & Crowther, L. (2014). The science of transcranial magnetic stimulation. Psychiatric Annals, 44(6), 279-283

Scoville, W., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. Journal of Neurology, Neurosurgery, and Psychiatry, 20(1), 11.

Setlow, B. (1997). Georges Ungar and memory transfer. Journal of the History of the Neurosciences, 6(2), 181-192.

Sheredos, B., Burnston, D., Abrahamsen, A., & Bechtel, W. (2013). Why do biologists use so many diagrams? Philosophy of Science, 80(5), 931-944.

Steinle, F. (1997). Entering new fields: Exploratory uses of experimentation. Philosophy of Science, 64, S65-S74

Steinle, F. (2002). Experiments in history and philosophy of science. Perspectives on Science, 10(4), 408-32

Stewart, W. (1972). Comments on the Chemistry of Scotophobin. Nature, 238, 202-209.

Suárez, M. (2004). An inferential conception of scientific representation. Philosophy of Science, 71(5), 767-779.

Sullivan, J. (2009). The multiplicity of experimental protocols: a challenge to reductionist and non-reductionist models of the unity of neuroscience. Synthese, 167(3), 511-539.

Sullivan, J. (2017). Long-term potentiation: One kind or many?" In Eppur Si Muove: Doing History and Philosophy of Science with Peter Machamer, A Collection of Essays in Honor of Peter Machamer, ed. Marcus Adams, Zvi Biener, Uljana Feest, and Jacqueline Sullivan, 127–40. Springer International Publishing.

Sutton, R. S., & Barto, A. G. (1998). Introduction to Reinforcement Learning. Cambridge, MA: MIT Press.

Teller, P. (2010). "Saving the phenomena" today. Philosophy of Science, 77(5), 815-826.

Thompson, R., & McConnell, J. (1955). Classical conditioning in the planarian, Dugesia dorotocephala. Journal of Comparative and Physiological Psychology, 48(1), 65.

Travis, D. (1981). On The Construction of Creativity: The 'Memory Transfer' Phenomenon and the Importance of Being Earnest. In The Social Process of Scientific Investigation (pp. 165-193). Springer Netherlands.

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2003). Statistical decision theory and the selection of rapid, goal-directed movements. JOSA A, 20(7), 1419-1433.

Ungar, G. (1970). Chemical transfer of learned behavior. Inflammation Research, 1(4), 155-163.

Ungar, G., & Oceguera-Navarro, C. (1965). Transfer of habituation by material extracted from brain. Nature, 207, 301-302.

Ungar, G., Desiderio, D., & Parr, W. (1972). Isolation, identification and synthesis of a specific-behaviour-inducing brain peptide. Nature, 238, 198-202.

Van Fraassen, B. C. (1980). The scientific image. Oxford University Press.

Walker, D. R. (1966). Memory transfer in planarians: an artifact of the experimental variables. Psychonomic Science, 5(9), 357-358.

Walker, D. R., & Milton, G. A. (1966). Memory transfer vs. sensitization in cannibal planarians. Psychonomic Science, 5(7), 293-294.

Waters, C. K. (2007). The nature and context of exploratory experimentation: An introduction to three case studies of exploratory research. History and Philosophy of the Life Sciences, 28(3), 275-284

Woodward, J. (1989). Data and phenomena. Synthese, 79(3), 393-472.

Woodward, J. (2003). Making things happen: A theory of causal explanation. New York: Oxford University Press.

Woodward, J. (2010a). Data, phenomena, signal, and noise. Philosophy of Science, 77(5), 792-803.

Woodward. (2010b). Causation in biology: stability, specificity, and the choice of levels of explanation. Biology & Philosophy, 25(3), 287-318.

Zis, P., & Mitsikostas, D. D. (2018). Nocebo responses in brain diseases: a systematic review of the current literature. International Review of Neurobiology, 139, 443-62.