

David Colaço

Accepted for publication in *Biology and Philosophy*. Please Cite Final Version

Rethinking the Role of Theory in Exploratory Experimentation

Abstract: 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.

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, discussion of the role of theory in exploratory experimentation has been informed by a distinction between *local theory* and the *theoretical background*, 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, but rather the *use* of theories. 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 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. In Section 3, I characterize how theories direct experiments and discuss the role of auxiliary hypotheses. In Section 4, I discuss a class of techniques called transcranial magnetic stimulation (TMS), in which magnetic pulses are used to influence brain activity. In Section 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. Accounts of the Role of Theory in Exploratory Experimentation

To understand what exploratory experimentation is, it is helpful to 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, and 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

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.

the purpose of focusing experimental 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 **T₁** counts as a background theory if and only if **T₁** represents **A** in virtue of the fact that **T₁** 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 **T₂** is a local theory if and only if **T₂** represents **A** in virtue of the fact that **T₂** represents the components

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).

of **A** and their causal relations: that is, if **T₂** 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.

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, the 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 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, let me address the roles of theory in experimentation.

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

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.

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 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

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.

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 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

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).

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, 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. 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 to make them 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

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.

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.

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, to 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, and 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, and 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 can involve single pulses, paired pulses, or repetitive pulses. The differences between forms change the shape,

intensity, and penetrability of the magnetic field, and 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.

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

¹¹ Forms of TMS have been validated, involving pairing TMS pulses with brain activity measurements via fMRI and EEG, 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.

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 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, to determine if researchers 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.

5. Rethinking 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 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 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 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 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 is 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.

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.

References

- Barker AT (1998) The history and basic principles of magnetic nerve stimulation. *Electroencephalography and clinical neurophysiology*. Supplement 51:3-21
- Barker AT, Jalinous R, Freeston I L (1985) Non-invasive magnetic stimulation of human motor cortex. *The Lancet* 325(8437):1106-1107
- Burian RM (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
- - - (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
- 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
- Feest U (2012) Exploratory experiments, concept formation, and theory construction in psychology. *Scientific concepts and investigative practice* 3:167-189
- Feest U, Steinle F (2016) Experiment. In: Humphreys P (ed) *Oxford Handbook of Philosophy of Science*. Oxford University Press, New York, pp. 274-295
- Franklin LR (2005) Exploratory Experiments. *Philosophy of Science* 72(5):888-99
- Giere R, Bickle J, Mauldin R (2006) *Understanding Scientific Reasoning*, 5th edn. Thomson Wadsworth, Toronto
- Hacking I (1992) The Self-Vindication of the Laboratory Sciences. In: Pickering A (ed) *Science as Practice and Culture*. Chicago: University of Chicago Press. pp. 29-64
- 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
- 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

Levkovitz Y, Harel EV, 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

Menkes DL, Bodnar P, Ballesteros RA, Swenson MR (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

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

O'Malley MA, Elliott KC, Haufe C, Burian RM (2009) Philosophies of funding. *Cell* 138(4):611-615

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

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 WM, Crowther LJ (2014) The science of transcranial magnetic stimulation. *Psychiatric Annals* 44(6):279-283

Steinle F (1997) Entering New Fields: Exploratory uses of experimentation. *Philosophy of Science* 64:S65-S74

- - - (2002) Experiments in history and philosophy of science. *Perspectives on Science* 10(4):408-32

Sullivan J (2009) The Multiplicity of Experimental Protocols: A challenge to reductionist and non-reductionist models of the unity of neuroscience. *Synthese* 167(3):511-39

Polson MJ, Barker AT, Freeston IL (1982) Stimulation of nerve trunks with time-varying magnetic fields. *Medical and Biological Engineering and Computing* 20(2):243-244

Waters CK (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