Realism, Confirmation, and Explanation: Studies in Models and Model-Based Science

By

Gareth Fuller

MA., University of Kansas, 2018

MA., University of Texas El Paso, 2011

BA., Guilford College, 2009

Submitted to the graduate degree program in Philosophy and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Chair: Dr. Armin Schulz

Dr. Eileen Nutting

Dr. Sarah Robins

Dr. John Symons

Dr. Raymond Pierotti

Date Defended: 26 April 2022

The dissertation committee for Gareth Fuller certifies that this is the
approved version of the following dissertation:

Realism, Confirmation, and Explanation: Philosophical Studies in Models and Model-Based Science

Chair: Dr. Armin Schulz

Date Approved: May 3rd, 2022

Abstract

Over the last several decades, the philosophy of science has been enamored of the role that models play in scientific practice. The methods and applications of models, their role in the production of knowledge, and many other features challenged previously held assumptions in the philosophy of science. That models are poor representations of their targets, that they are limited in scope, and that it is not always seen as a problem that they conflict with other models of the same target caused problems for older accounts of the epistemology of science. For instance, given the features of models, and their prominent role in scientific practice, the classic nomological-deductive (covering-law) account of explanation–where explanation is made by logical deduction from a universal statements–did not seem particularly applicable. The role of many different types of models in various aspects of scientific work, whether explanation, reasoning, prediction, or something else, has taken much of the attention of many philosophers of science. In the chapters that make up this dissertation, I too concern myself with models.

The first chapter examines how the "modeling turn" in the philosophy of science impacts the scientific realism debate. The two classic arguments in this debate–the "No Miracles Argument" in favor of realism and the "Pessimistic Meta-Induction" against realism–have often been presented in relation to a view of science as built out of theories. Theories, in the philosophers' sense, are typically treated like logical languages. Important in this theories-based account of science was that theories were intended to be accurate representations of the world. The NMA claims that theories were generally accurate that this accuracy provided for the success we often attribute to science. The PMI, in response, would point to the graveyard of failed theories in an attempt to show that we had little reason to buy into the truth of our current scientific theories. I argue that these two arguments are not easily applied to a model-based view. Models are not intended to be accurate representations of their targets, often idealizing,

abstracting, fictionalizing, or misrepresenting in some way. Further, these misrepresentations are not seen as flaws of the model, like they would be in a theory, but are often central to the goals of the model. Further, models are not discarded because of their lack of truthful representation, but are often discarded because the role that they can play is no longer needed. Given these features of models, the central role that models play in the philosophy of science, and the focus of the NMA and PMI on truth, I argue that the realism debate needs to move beyond these arguments.

The second chapter takes up a related concern around the role models play in confirming hypotheses. That models employ idealizations has been a concern for their role in producing knowledge, explanations, and other epistemic achievements. One proposal for getting around the potential concerns introduced by idealizations has been robustness analysis. This is a method of constructing several models with a shared core assumption but different sets of idealizations. If all of these models produce the same result, this is assumed to show that the results of these models are driven by shared assumption and not the idealizations. This, then, is meant to confirm the shared assumption of the models. The confirmatory power of robustness analysis has been questioned in several ways, one of which focuses on the fact that all of the models in a robust set are idealized, and therefore false. Given this, even if all of the models agree, they cannot provide confirmation since none of them accurately reflect the target system. In response, I argue that this is a misunderstanding of the role of some idealizations in models. Idealizations can be incorporated into models for a variety of reason, and sometimes might play the role of controlling some causal influence that is not of interest to the modeler. In this way, idealizations might play a role similar to that of experimental conditions, where laboratory conditions are contrived to control causal influences that might interfere with the causal relation being studied. Robustness, under these conditions, can be analogized with experimental replication. Since it is unclear how laboratory conditions might influence the results of an experiment, replicating the experiment in a variety of ways increases our belief in the result of each experiment. Similarly, it is unclear how a particular idealization, as a means of control, might influence the result of a model, so robustness is needed.

Finally, in the third chapter, I take up an account of functional kinds derived from model-based science. Daniel Weiskopf (201a, 2011b, 2017) has argued for the explanatory role of functional kinds and has developed an account of functional kinds derived from model based science. This view, however, has faced criticism from mechanistic-based philosophers. One such criticism is that functional kinds, or at least some of them, cannot be considered true scientific kinds since they are explanatorily weak when compared to similar mechanistic kinds. These arguments, however, often allow that functional kinds might be multiply realized. I argue that, granting this multiple realizability, there is a unique range of explanatory counterfactuals that functional kinds can capture. Given this, I argue that such kinds should count as true scientific kinds.

Acknowledgments

I owe deep gratitude to Armin Schulz, without whom the completion of this dissertation would have been impossible. If nothing else, the fortitude he exhibited in reading through and constructively commenting on drafts of the papers that wound up in this dissertation is nothing short of superhuman.

I owe further gratitude to my family. My wife for providing the support and consolation needed to make it through the struggles and to making happy times greater. To my children for providing important and comforting distractions. To my mom for worrying about my completing the dissertation, and to my brother for being a good friend.

Finally, in loving memory of Dad and Amos. Both gone far too soon and greatly missed.

Contents

Introduction

The role of models in science has been a vogue topic in the philosophy of science for several decades now. Often seen as part of a "practice turn" in the philosophy of science, where the focus is on analyzing the practices of scientists, the focus on models has brought with it many wrinkles in the philosophy of science. Most significant of these is the heavy use of idealizations in models.[1]

Through the early and mid-twentieth century, with the logical positivists and those related or inspired by them, studying the philosophy of science focused greatly on logical languages and reconstructions. Science was often seen as in the business of developing theories about the world, and theories were treated as logical languages. This focus on logical languages then influenced all aspects of the analysis of scientific practice. Explanation and confirmation, for instance, were treated in terms of this logical language. A classic account of explanation, provided by Hempel (1965a), holds that an explanation requires a generalized statement, a statement capturing some specific conditions, and that the combination of these statements is supposed to deductively imply the explanandum. Further questions of intertheoretic reduction were similarly handled in terms of logical languages and relations.

There are many reasons that models have taken the prominent role that they have.[2] One is that it seems that many sciences are not unified in a way that lends themselves to an overarching

---

1 It has been argued by some that theories are themselves idealized (Cartwright 1983). However, models clearly wear their idealizations and seem to take advantage of these idealizations in a way that was not so obvious for theories.

2 It is important to note that the use of model here is distinct from the use of model forms some philosophers of science who hold that theories are classes of models (van Fraassen 1980). These classes of models are a logician's model and these accounts of scientific practice are closer the theory-based view. See Godfrey-Smith (2006) for a discussion of this distinction.

theory. Related to this, it does not seem that the goal or methods of these sciences lead themselves to developing such a theory. Further, it is argued that, even when there is sufficiently robust theoretical backing, models are constructed in a way that makes their epistemic reach beyond what can be supplied by a theory itself. Models, on this picture, present a *unique* means of investigating the world. Finally, there is the fact that many scientists themselves see their work as prominently constructing models, and not any sort of theory that resembles the philosopher's notion. We can see this readily in works in the development of ecology and population biology. Richard Levins, for instance, not only set out a method of constructing models in population biology that made no reference to some overarching theoretical background, but also held that a theory was simply a set of models describing the same phenomenon (1966).

This focus on models brought with it a focus on a new approach to understanding science, as well as many new wrinkles for the philosophy of science. There has been the rise of what is called "model-based" science, which is distinct from previous accounts of scientific practice, which focused on theory building. Model-based science is a method of investigating the world by constructing idealized models and "experimenting" on these to gain insight into how the world functions. The most important difference for the work that follows is the heavy incorporation of idealizations. Idealizations, particularly as they will be understood below, are the known (and often intentional) misrepresentation of relevant aspects of the real-world system being modeled. What counts as a relevant aspect of the real-world system is open to interpretation, but a classic example of an idealized representation is that of a frictionless plane. Further examples include representing populations as infinite, or fluids as continuous. That models are highly idealized throws a wrench in many common analyses of scientific realism and epistemic achievements since there is often a requirement for truth in the representation before we can, say, have an explanation. Models seem to buck this since they so readily flaunt their misrepresentations. How it is that such obviously flawed representations as models could be used to achieve knowledge, understanding, or some other positive epistemic goal has been an open question.

Not only are idealizations in models common, but they are generally very stubborn. It is not often clear how to go about removing them, or if it is even possible. Scientists do not generally try to refine their models to remove the idealizations, filling them in as we learn more or as computing power develops.[3] Once again, Richard Levins points out (1966), and others have echoed (Plutynski 2008, Morrison 2015), it is often that we outgrow the questions which a certain model can answer and move on to another model designed to answer another concern.

I take up some philosophical concerns related to this focus on models in the philosophy of science and the compounding concerns that models seem to introduce. The first chapter in this dissertation looks at the scientific realism debate and model-based science. The turn towards models and their many idealizations has led to some questions about how to understand scientific realism. Some (Cartwright 1983, Odenbaugh 2011) have argued that idealizations undercut arguments for scientific realism.[4] I take up a similar concern and examine how the main arguments in the literature fare in relation to model-based science. The two arguments I consider are the "No miracles argument", in favor of scientific realism, and the "Pessimistic Meta-Induction" (PMI) against scientific realism. The NMA claims that it is the approximate truth of

3 To be sure, the advancement of computers has allowed for models to include far more information in them, but this does not mean that they have removed idealizations and falsehoods.

4 It should be noted that Odenbaugh has moved on from the view he expressed in 2011.

scientific theories that best explain their success, while the PMI argues that the history of science is filled with scientific theories once considered to be useful that we now accept to be false.

Both of these arguments have played a central role in debates about scientific realism, where various formulations of scientific realism and anti-realism need to make sense of both of these intuitions. In turn, I argue that the rise of model-based science throws a wrench in this debate, as model-based science is not amenable to evaluation in terms of these arguments. Models are not designed to be true, thus, arguments for or against their truth is beside the point. From this, I argue that the NMA and PMI only work on theory-based accounts of science, where what counts as a theory is the philosopher's notion. This leads to a dilemma for the scientific realism debate, where we must save the main motivating arguments in the debate by giving up on model-based science, or we need to give up on the NMA and PMI. Neither of these options is particularly enticing.

The next chapter deals with confirmation of highly idealized models. In particular I focus on the practice of robustness analysis. This is a method of confirming the results of an idealized model by developing further, differently idealized, models of the same phenomenon in an attempt to show that the idealizations of our models do not drive the results. In general, the idea is to develop several models with a shared core set of assumptions but different idealizations. If all of these models produce the same result, a "robust theorem", then this is meant to show that the idealizations are not influencing the results, but it is the shared core set of assumptions. Insofar as this shared core is representing something that is supposed to be in the model, this should confirm our models.

The confirmatory power of robustness analysis has come under scrutiny. I answer a particular dilemma, which holds that, since all of the models in the robust set are idealized, and therefore false, the robust set cannot provide confirmation. A set of false models is not evidence, even if they all agree. The only way a robust set could be confirmatory is if at least one model is de-idealized, or made true. But then there is no need for a robust set since there are no idealizations to "discharge" like a robust set is meant to do. So, when robustness analysis has a role to play, it cannot provide confirmation, and the conditions where a model can provide confirmation, robustness analysis is not needed. Robustness analysis, therefore, provides no confirmatory power.

I argue that this dilemma turns on two misunderstandings. First, I believe that a model, or experiment, or some other scientific investigation can be confirmatory and still require robustness analysis. Second, I believe that this argument has a mistaken view of the role of idealizations. I argue that at least some idealizations play a role of controlling certain causal factors that allow scientists to investigate some other causal relation in the system of interest. Idealizations, then, can play a role very similar to tightly controlled experimental conditions, where many of the causal influences of the world are removed, so a better understanding of the relationship between other causal factors can be derived. I then draw an analogy between robustness analysis and experimental replication, in that they are means of ensuring that certain experimental conditions are not influencing the results in an unexpected way. This then allows some idealized models to be confirmatory while still retaining a distinctive role for robustness analysis.

The final chapter takes up questions about the explanatory power of functional kinds and explanations. Daniel Weiskopf (2011a, 2011b0 has developed an account of functional kinds where these are functionally defined categories that find use in explanatory models across a range of target systems. Considering functional kinds as proper scientific kinds have long been criticized by new mechanistic philosophers, and Weiskopf's account is no different. In particular, they are criticized as being explanatorily inferior to mechanistic kinds, and are, at best, sketches of mechanisms to be replaced once more details of the mechanism have been discovered.

I argue that such criticisms miss an important explanatory power held by functional kinds. It is admitted that such functional kinds can, and often are, multiply realizable. Given this, functional kinds are able to answer a range of counterfactuals that mechanisms cannot by virtue of being multiply realizable. In particular, if some phenomenon is multiply realized, then any explanation in terms of a mechanism is indexed to that particular mechanism. This is problematic for a range of counterfactuals about the multiply realized phenomenon, as they will be independent of any particular mechanism. This generality is not something that can be captured by a mechanistic account of kinds. Thus, there is a distinctive role for functional explanations to play.

This defense of functional kinds is not new, and I argue two further points. One is that this defense is often overlooked because there is a general assumption of the mechanistic framework in the arguments against functional kinds. If this framework is not accepted, however, these arguments are far less persuasive. Second, I argue that this defense of functional kinds undercuts some further aspects of the new mechanistic framework. In particular, it is often pressed that explanation proceeds by elucidating a mechanism, but functional kinds often cannot play this role. This defense of functional kinds, then, is a problem for the general mechanistic framework.

Each chapter in this dissertation addresses just a few questions that have stemmed from the role of models in science. It is important to recognize that there are many follow up questions to consider, such as how to go about formulating the debate over scientific realism, what other ways idealizations and falsehoods might be manipulated to generate knowledge, and whether or not we can ever rid ourselves of functional kinds and explanations and the sciences that use them, such as psychology.

Scientific Realism at the Cross-Roads: Model-Based Science and the Traditional Arguments About Scientific Realism

Introduction

The No Miracles Argument (NMA) and Pessimistic Meta-Induction (PMI) have traditionally been, and are still taken to be central arguments in the debate over scientific realism. The NMA presents an intuitive insight in favor of scientific realism, whereby the truth (or approximate truth) of theories best explains their success. The PMI presents a counter-intuition that the history of failures in science should give us plenty of reason to doubt the truth of our current theories. These arguments have been central enough to the debate that significant work is done to either find a way to satisfy both intuitions (Worrall 1989, Wray 2018), or to explain away one of them.

Since the formulations of these arguments, a distinctive approach to understanding some of the working of science has found prominence in the philosophy of science through model-based approaches (Godfrey-Smith 2006). This is a particular method often taken in opposition to the methods of theory-based science, where the task of science is developing and investigating idealized models to learn about the world. While many questions regarding, say, what counts as a model, the representational relationship between models and their targets, the relationship between models and theories, and many others are still open issues, there is a generally accepted point about models, which is that they tend to be *poor* representations of their target.

In this chapter, I will argue that some aspects of model-based science raise concerns for both the NMA and PMI. I will argue that we are left with a dilemma: either give up on understanding science as (sometimes) model-based, or to give up on both the NMA and PMI.

Given the significant role model-based science has played in making sense of the practices of science, I argue that it is best to move on from the NMA and PMI as expressions of scientific realism.

In section 1, I begin by laying out the basics of scientific realism, insofar as this is relevant here. In section 2, I describe model-based and theory-based science. In section 3, I present the NMA and PMI. In section 4, I argue that both the NMA and PMI require a theory-based view of science to carry any weight. Finally, in section 5 I argue that this leaves us with an unsavory dilemma.

Before continuing, it is important to note that the target of my discussion is philosophy and philosophers of science. The arguments that philosophers of science have produced for and against scientific realism seem to now conflict with how philosophers treat some of the practice of science. Philosophers now distinguishing model-based science as a distinctive aspect of scientific practice is what produces this conflict. Nothing here bears on how scientists treat their work, excpet in-so-far as focus on model-based science is meant to capture a “practice-turn” in the philosophy of science.

1. The NMA and PMI

The No Miracles Argument (NMA)—an argument in favor of realism—and Pessimistic Meta-Induction (PMI)—an argument in favor of anti-realism—are still two of the most significant arguments in the debate about scientific realism. For present purposes, a very general understanding of scientific realism is all we need.[5] Scientific realism can be seen as incorporating two claims, i.e., a metaphysical and an epistemic claim, with a third semantic

5 From here on out, I will use realism to mean scientific realism unless noted otherwise.

claim closely connected (Chakravartty 2007 Ch. 1 Psillos 1999, Ch. 1). Scientific realism is the belief that there is a mind-independent world (metaphysical), some of which is unobservable, and that we are justified in believing in some of the posited unobservable aspects of science (epistemic). The methods of science are able to provide strong enough evidence for the unobservable features that we are justified in believing in their existence or the truth of some claims about them. We are justified in believing that, say, electrons, protons, and neutrons exist, are mind-independent, and generally have the properties we ascribe to them, despite not being able to directly observe them.[6]

The NMA has been presented in various versions, but generally carries the same basic intuition, which is the starting assumption of the NMA is that science has been successful. Success in science can be identified in various ways, but I will work with a notion of success as providing accurate and novel predictions (Worrall 2007).[7] The NMA then focuses on providing an explanation for why science has been so successful and argues that a realist position has the upper-hand.[8]

The realist can explain the success of science by pointing out that many of our theories are true, or at least approximately true. Given that our theories have gotten things close-to-right,

---

6 There is much more to say about realism than this. For instance, some hold that realism is not a question of whether we are justified in believing the posits of our science, but a question of whether or not science even aims at truth (van Fraassen 1980). Further, some hold that there are specific conditions or aspects we should believe in, such as entity realists or structural realists. I will take realism to be the view presented above, leave out many of the details, but, as it turns out, my discussion will not turn on which account is accepted.

7 We might want to include explanations as part of this success, but explanation often requires that something be true and so to cite explanation as a success would then be question begging because you have to assume the truth of science.

8 There are some, notably van Fraassen, who do not find this question particularly puzzling. Van Fraassen finds that our theories are successful because those are simply the ones that we keep, making a comparison to natural selection. I will not consider the details of his position here, though will comment on various aspects of it throughout.

it seems to follow that their predictions will also be correct. Further, an explanation of why science progresses, and provides further novel predictions, is that science is getting closer to the truth as we progress. True theories provide for accurate, novel, predictions.

The anti-realist, however, cannot lean on the truth of theories to explain the success of science. While the anti-realist does not need to accept that our theories are *actually false*, they must at least claim we are not in a position to say that we are justified in believing in them. Some other explanations for the success of science are required—however it is not clear how to provide these. Many anti-realists try to drive a wedge between truth and success, but with little to replace the role truth plays in the realists' explanation (see e.g., van Fraassen, 1980).[9] Putnam (1975) goes so far as to claim that the anti-realists' only option is that the success of science is a miracle (where the NMA gets its name). Realism, then, is the option that does not make the success of science a miracle, and we are to prefer the non-miraculous to the miraculous, at least in this case.

The PMI is an argument in favor of anti-realism. It is typically presented as an inductive argument based on the past failings of science. It begins with the assumption that our previous scientific theories, those we have abandoned, were abandoned because they were discovered to be false. Although they might have been successful, it was discovered that they were empirically inadequate, and either posited or predicted phenomena or entities that do not exist, or failed to account for some phenomena or entities. Examples of past theories that have been discarded for such failures are plenty, and this is the crux of the PMI.

---

9 Wray (2018) is a notable exception in trying to show that anti-realist provides a *better* explanation for the success of science. However, even if he is correct, this does not change anything for our present purposes.

As noted above, the strategy of many anti-realists is to show that truth and success do not go hand-in-hand. These many discarded previous theories were often successful but all failed to be true. A common example is that of Newtonian Mechanics, which was a successful theory, so much so that it still plays a significant role in many scientific explanations, but has been judged to be limited when we examine quantum physics. Even if one does not want to specifically target the connection between truth and success, a basic point of the PMI is that contemporary scientific theories are not likely to be distinct from the past failures, or at least some account of *how* they are distinct is needed in order for realist intuitions to work carry weight.

Both of the arguments discussed-the NMA and the PMI- bring with them an intuitive weight. The NMA makes clear that science has been successful, and this needs to be taken seriously. The PMI makes clear the fact that some of our theories have failed up until now, and we know our current theories are not perfect. Realist positions need to be sensitive to the fact that they cannot make the claim that *whatever* our current theories say is true.

These two intuitions pull strongly on both realists and anti-realists. There have been various realist positions that split the gap between these two arguments, claiming that we carry over some aspects of our theories and discard others aspects as we progress (Psillos 1999, Worrall 1989). Anti-realists, as well, have attempted to make sense of progress outside of truth (Wray 2018). So, these two arguments carry significance for the general debate.

These two arguments are also often taken to be the two *most significant* arguments. Howson (2013 pg. 205) says, “Putnam’s ‘positive argument’ [i.e. the NMA]...has become the argument of choice of many, probably most, philosophers anxious to promote a realist philosophy.” Henderson (2017 pg. 1295) points out that the NMA has been referred to as, “the ‘ultimate argument’ for scientific realism.” Dawid and Hartmann (201 pg. 4063) state that, “The

No Miracles Argument (NMA) is arguably the most influential argument in favor of scientific realism." When discussing the debate between realists and anti-realists, in a paper defending the NMA, Jenger (2015 pg. 174) says that, "a major player in this debate [about realism] is the No Miracles Argument (NMA)." Finally, Wray (2018 pg. 143) states that, "Realists claim that they have one important advantage in the debate with anti-realists...they have an explanation for the success of science." Worrall (1989 pg. 101) similarly says that, "The main argument…likely to incline someone towards realism I shall call the 'no miracles' argument.[10] In short, realists and anti-realists alike currently cite this argument as a primary argument for the realist position.

The PMI is of similar importance to the debate. Rowbottom (2019) points out that, for the PMI (and NMA), "discussion on each argument remains remarkably vigorous." Wray (2018 pg. 68) has claimed that, "this line of argument [the PMI] has figured prominently in the contemporary realist/anti-realism debate...Some have suggested that the Pessimistic Induction [the PMI] is the anti-realist's strongest argument." Devitt (2011 pg. 285) points out that he had, "labeled the meta-induction "the most powerful argument against scientific realism"," and, when discussing a particular version formulated by Kyle Stanford, that, "his version of the meta-induction is indeed the most powerful challenge."[11] Doppelt (2007 pg. 106), argues that an important test for his version (and any other version) of, "scientific realism is its ability to rebut the pessimistic meta-induction. This ability will immeasurably enhance its intuitive plausibility."

10 Worrall also points out that there are only two main arguments in the realism debate, and they seem to be quite old. These arguments, according to Worrall, are the NMA and the PMI.

11 Stanford's argument, known as the unconceived possibilities argument, is often treated differently from the standard PMI (as it should be). It does, however, take a very similar form. At any rate, the differences are not so relevant for present purposes.

Just as with the NMA, the PMI is still of current significance in the realism debate, such that both realists and anti-realists take its claim seriously.

In sum, the debate over scientific realism plays host to two opposing intuitions, expressed in the form of the NMA and PMI, which have played a significant, central, role in the dispute over scientific realism. Accounts of both realism and anti-realism are formulated with these two arguments in mind, and significant work is either done to appease both of these intuitions or explain one away.

2. Two Views of Science: Some Preliminaries

In recent philosophy of science, there has been a very general split recognized in understanding the representational and epistemic approaches to the workings of science, a theory-based view and a model-based view. Very generally, I will hold that a theory-based view sees science as primarily in the business of constructing and evaluating theories, understood in a particular philosophical sense. A model-based view holds that science is in the business of constructing models that are distinct from any theories that might be employed. In a theory-based view, then, it is theories that carry the brunt of the work, while in a model-based view, this load is offloaded to models.

To make this clearer, consider the classic example of a theory-based view of science: a stereotypical mid-twentieth century positivist-type view of science.[12] At the bedrock of this view was formal logic, where a theory was a logical calculus. This calculus was the set of axioms that defined the syntax of the language. From there, a particular language was interpreted with the

12 There are many subtleties and interesting aspects of each individual positivists view of science. Still, the remarks in the text sketch the key outlines of the major positivistic views of science.

domain of a particular science, so as to be applied. The variables, predicates, and terms of the language would pick out the objects and properties of the domain. In this way, a theory directly describes the world.

We can get a sense of the logical flavor requirement for literal description assumed in the philosophical literature in several ways. We might look to van Fraassen's (1980) criteria of acceptance for a theory. To accept a theory is to accept the literal empirical consequences of the theory. Tied up in this evaluation of the literal empirical consequences of a theory is that the theory must be consistent. If a theory is inconsistent, then it would imply too much and fail this measure of adequacy.

Further, it is in providing literal descriptions of various aspects of the world that the epistemic achievements of science are captured. Explanation, on some famous accounts (Hempel 1965a), was a case of logical inference. A phenomenon was explained when it could be inferred, within the logical calculi, from a generalized statement and a statement providing specific conditions. However, something was only an explanation if the generalized statement and specific statement were true. This truth was determined, as with many linguistic entities, in terms of reference, such that the terms had to pick out real parts of the world. A similar point can be drawn for prediction.

There are a few key points here. First is that the theories directly represent the world. It is by picking out the objects and properties of a particular scientific domain (e.g., the particles of physics) that a theory can then be used. Science is also viewed as constructing and testing theories, understood as something like the positivist-style account above, through their ability to predict and explain. Further, theories are presented in terms of logical calculi and are generally taken to be linguistic in nature.

We might compare this to model-based science, and to several prominent accounts of constructing models. Godfrey-Smith (2006) outlines model-based science as a distinctive method of doing science where a model is constructed and investigated. It is through constructing and manipulating a model that we try and gather information about the world. Belot (2007 pg. 279) puts the model-based view somewhat differently: "The central thesis of the models-based approach is that a theory's laws do not determine the models that scientists use to represent the phenomena." In both of these cases, the scientists go about representing the world, and what then carries significant epistemic weight, are models.

Models have been understood in several slightly different ways. Giere (1988) presents models as an idealized structure used to represent the world. An idealization is an intentional misrepresentation incorporated into the model. These distortions can be introduced for various reasons, whether for pragmatic and practical concerns, as a result of a particular focus, or out of necessity. Distortions can include altering size, in the case of physical scale models, intentionally removing certain aspects of the target system from the model, adding aspects to a model that do not appear in the target system, and a plethora of other alterations.[13]

A slightly different account of how modeling functions is presented by Weisberg (2013), and his account of target-directed models. Target-directed models represent a target system which is an abstraction of a phenomenon in the world. In this way, a target-directed model does not represent the world directly, but some abstracted target system. In the development of the model, the relevant aspects of a real world system are determined while other aspects of the target system are abstracted away to develop the model's target.

13 See Downes (1992) for a discussion of the many types of models, and how some of their idealizations are introduced. See Jones (2005) and Weisberg (2013) for a discussion of idealization and some of the varieties of idealization.

An important step in the development of a model on Weisberg's account is the use of two kinds of fidelity criteria. Dynamic fidelity is how closely the predictions of a model must fit the real-world phenomenon. Representational fidelity is how closely the structure of the model must represent the structure of the real-world phenomenon (Weisberg 2013). These fidelity criteria are determined in various ways, but are somewhat malleable, and do not require direct representation. Similar to Weisberg's account, Giere (2006) evaluates models in terms of "similarity" where a model must satisfy some appropriate criteria of similarity to its target.

There is an important reason for moving to a representational relationship such as similarity. This is because there is a heterogeneity attributed to the class of scientific models that is not attributed to traditional theories. Weisberg, for instance, outlines mathematical models, physical models, and computational models. Each of these models could be ontologically distinguished from each other. Further, the way that these types of models represent, explain, and predict are somewhat different. A physical model is ontologically quite distinct from a mathematical model. Further, a computational model, according to Weisberg, goes about explaining, predicting, or retrodicting in a way quite different from a mathematical model.[14] In an attempt to capture this heterogeneity in models, and to try and group the studying of various types of models and their role in science together, a somewhat looser representational goal is needed. So, the turn towards similarity.

A few key distinctions can be drawn from this. First is that, on the theory-based view, representation is carried out in terms of a language, and is ultimately linguistic in nature. For the model-based view, models are not necessarily linguistic, and might be physical objects that

14 See Weisberg 2013, ch. 2 for more details.

instantiate an appropriate structure.[15] Following from this linguistic nature, the means of evaluating theories requires an, ultimately, semantic criteria. The terms of the theory must refer to objects in the world. Models, on the other hand, not necessarily have a representational relationship amenable to semantic analysis. Further, even when models might be considered linguistic in nature, since there are logical and mathematical models of some phenomena, they are not evaluated in semantic terms. Similarity, for instance, is not ultimately a semantic notion.

Further, both accounts of modeling do not require that models be literal descriptions of the world. On Giere's account, idealizations are expected and it is only required that a model be similar in some appropriate way, determined by the purpose of the model. In Weisberg's account, models are representing a particular abstracted version of some real-world system.

Now, as implied by the Belot quote, we can make room for both models and theories in our sciences, however, according to Belot, a model-based view requires that models be somewhat independent of theories. A similar view is presented in Morgan and Morrison (1999), where models are somewhat "autonomous" from theories. So, the distinction we end up with in terms of theory-based and model-based views will ultimately be that models, constructed and outlined as above, play a distinctive role that cannot be brought in line with any theories.[16]

Before closing out this section, it will be useful to discuss some examples of models and theories from the literature. While the necessary and sufficient conditions for theories and models is debated, there are some common examples that are suggestive of the difference.[17]

---

15 Weisberg (2013) allows that models may be physical or computational as well as mathematical.

16 This might be simply because there are no theories.

17 See Frisch (2004) and Belot (2007) for a brief exchange about how to distinguish theories.

An example of a theory treated in the literature is that of classic electrodynamics (Frisch 2004 and Belot 2007), but we might also approach Classic Mechanics as a theory. The laws of motion in classical mechanics, as expressed by Newton's three laws of motion and pertinent updates, specify the relationship between the motion of objects and the forces acting on the objects. These laws, taken together, imply many propositions about, say, the relationship between mass, force, and acceleration. A theory, can be, very generally, tested by whether or not the world satisfies the laws of the theory. So, a theory may fail if it requires an ontology that the world does not satisfy or if it implies some properties not found in the world.

This brief presentation of mechanics opens up the discussion for the distinctive role of models. In many applications of classical mechanics, objects are treated as point particles. This is something that is not clearly specified in the laws of mechanics, nor is it something that clearly reflects how the world is. For some (Frisch 2004), it is these sorts of misrepresentations introduced that distinguish models as distinct from a theory. Interpreting and understanding these sorts of misrepresentations in application are not clearly outlined in the laws of the theory, and therefore draw on resources that are not part of laws or axioms of the theory. So, we might have the theory of classical mechanics that is to be a literal description of the world, but the application of such a theory may require some misrepresentations that require resources not made available within the theory.

The literature is also full of potential examples of models that buck some of the linguistic assumptions of theories. Weisberg (2013) for instance provides for both physical models and computational models. An example of a physical model is that of the San Francisco Bay and Estuary Model constructed by the Army Corps of Engineers. This is a physical model, built of concrete, water, made to represent tides and currents in the San Francisco Bay, all-be-it considerably smaller than the Bay itself. Being physical, it, plausibly, is not itself a linguistic construction.

3. No Miracles, Pessimistic Meta-Induction, and Truth

In this section I return to the NMA and the PMI and discuss them in relation to theory-based and model-based accounts of science. In particular, I argue that the NMA and PMI cannot be applied to a model-based view. I consider several possible formulations and show that they still do not directly apply.

Historically, it is clear that the primary target of these two arguments has been theories. Given the general timeframe of the development of these arguments and the general trends in the philosophy of science, it is theory-based views that were at the forefront. The common presentations found in many of the works cited at the end of section II specifically discuss these arguments in regard to theories. Further, these have often been presented with a semantic tone, focusing on the referential successes and failures of theories. Reference, typically understood as a linguistic property, easily applies to theories understood in the philosopher's sense. Because theories are understood as a literal description, this provides a better grounding for determining truth (van Fraassen 1980).

These lines of reasoning, however, do not readily transition over to a model-based view. Let us look at the NMA first. Models are idealized representations, and recognized as being false representations of any particular real-world system. Therefore, it cannot be the truth of the models that directly explains their success. A brief argument has been developed to make this clear (Odenbaugh 2011, Wheeler 2020). The NMA takes truth be the best explanation of the success of science. But all models are known to be false, and so it cannot be the truth of models that explains their success. After all, no amount of success is going to convince us of the truth of something we know to be false.[18]

A further point lurks, which is that it is hard to apply truth directly to the accounts of model evaluation. In the case of Giere's account of evaluation, a model must be similar to its target. In Weisberg's account, a model must have structural and predictive similarities that are within certain fidelity criteria. While we might conclude that it is true that a certain model is similar to its target in an appropriate way, it does not follow that the model necessarily carries a semantic notion of truth often assumed in presentations of both the NMA and PMI. The similarity of a model to its target is not necessarily evaluated in terms of referring to objects and their properties.[19] Thus, on a semantic understanding of truth, as the NMA is often applied, it fails to appropriately capture the representational aspect of models.

There are a couple of reasons why we might see the NMA as taking a semantic reading. First, many accounts of realism adopt a semantic tenet (Psillos 1999, Chakravartty 2007). Given this, truth is then spelled out in terms of reference. Further, many defenses of the NMA, and scientific realism, often focus on defending accounts where reference to the objects of a theory are carried over between theories (Putnam 1975).[20]

The PMI has similar issues. It is often applied with a focus on semantic failure which does not directly carry over to accounts of models. However, there are further concerns. The PMI works not just because our scientific history is filled with theories that have been discarded, but have been discarded for being false. However, models are constructed with the explicit

18 In Bayesian terms, our priors for the truth of the models is 0, so our belief in their truth will not increase.

19 This is not to say that no models can function this way, but simply that it is not a requirement.

20 A classic example would be Putnam's causal theory of reference (1973, 1975).

understanding of being false and are thus not discarded simply for being false. Further, these falsehoods that are incorporated into a model are often leveraged to provide further insight into the target system, and so it is by the idealizations and falsehoods that models can be used to understand the world. This causes two problems for the PMI. First, the PMI, formulated under this understanding, cannot make sense of why some false models are kept and why some false models are discarded. A further distinction is needed about being false in the right or wrong way. Second, is that looking to the history of models does not provide the same intuitive motivation for anti-realism because models are not discarded because they are false. They are all false, but importantly they tend to be discarded when the role they play in investigation is no longer needed. As Levins (1966) puts it, we move on from models when we are no longer interested in the questions they can answer.

There are further understandings of the PMI and NMA, however. For instance, within the theory-based view there has been a split between those who treat theories as logical languages and those that treat theories as sets of models that satisfy the axioms of a theory. While the move is not too significant, the evaluation of a theory is in terms of isomorphism rather than a direct referential relation.[21] The general structure of the NMA and PMI would then be that which best explains the successes of our science is that the models of our theories are isomorphic to the world, or perhaps, the history of science is filled with predictively successful isomorphic failings.[22] However, even this account cannot make sense of the practice of modeling, since

21 What kind of difference this amounts to in understanding scientific realism is debated.

22 van Fraassen 1980 provides a strong account of this understanding of scientific realism. However, see van Fraassen 2008 for an update to this view that moves away from isomorphism as an account of scientific representation.

models are not isomorphic with real-world phenomenon.[23] For instance, similarity does not require, and often explicitly is not, expressed in terms of isomorphism (Giere 2006). Similarly, Weisberg's target-directed account focuses on abstractions of real-world phenomenon; thus a reduction to isomorphism between a model and the world would similarly not hold.

There are other views that could be considered. For instance, it is granted that we do not have to choose between explicitly treating only models or theories as the sole means of scientific progress. Theory-based accounts allow for the use of models, and model-based accounts allow for theories. It could be argued that the NMA and PMI apply to the theories of our science and have little to say about the models.

While model-based views make some room for theories, there is more to model-based science than what theories alone can provide. Models are taken to be a distinct way that we study the world, and, on the model-based view, they are autonomous of theories in their construction. Models regularly deviate from the principles of theories in ways that cannot be brought back in line with the theory (Morrison 2007). What would be required for the NMA and PMI to adequately address the model-based view is that theories have to be taken as accurate descriptions of the world, and that the use of models be derived from *this accuracy*. The model-based view does not need to make these concessions, and in fact often denies this (Morgan and Morrison 1999, Godfrey-Smith 2006). Thus, the NMA and PMI leave a significant aspect of scientific progress untouched, namely that of the work of models.

In short: both the NMA and the PMI fail to apply in a useful fashion when we move from a theory-based view of science to a model-based view. To be fair to those quoted at the end of

23 Some (da Costa and French 2003) have tried to spell out modeling in terms of partial isomorphism. However, they emphasize the pragmatic features of their account that would lead to similar concerns as I have discussed.

section II, none were pushing for a model-based view. Also, those who present a model-based view of science do not, as far as I have seen, go on to apply either the PMI or the NMA. However, what we see here is a separation from what is taken as a standard approach to the scientific realism debate and some significant work on the practice of modeling in science.

4. The Dilemma for Philosophy of Science

If the NMA and PMI do not work with a model-based view, this creates a dilemma about how to proceed with the scientific realism debate. The NMA and PMI are meant to capture the main intuitions behind the two respective positions, and much of the debate over scientific realism has focused on these two arguments. However, we now have a significant account of the model-based practice of science that is not amenable to such standard evaluations. While there are proponents of both the model-based and theory-based view, giving up the model-based view is a difficult prospect given the role it has played in making sense of some parts of science.

One option to resolve this problem may be to limit the realism debate to those parts of science analyzable primarily in terms of a theory-based view. In this case, we can maintain the focus on the NMA and PMI. There is even some plausibility in this proposal, given the contentious nature of interpreting what models actually say about the world. This solution, however, seems rather dubious. In particular, although it may be contentious how exactly to interpret models in terms of a realism, it is not denied that models can provide knowledge, or at least justification for beliefs about the world. Further, since models are often used to investigate unobservable parts of the world, e.g., in modelling chemical processes, this move may be just *ad hoc*. Some further reason would need to be provided as to why we should think that realism should be limited to just theory-based accounts.

Another resolution may be to move the focus of the realism debate from scientific representations (i.e., theories and models) to a more "cognitive" account of realism, where the focus is on the development of true beliefs, understanding, or some other epistemic achievement. Rice (2019), for instance, argues that idealizations and falsehood in models can be used to develop modal information about a target, and we then use this modal information to develop understanding. Understanding, as Rice sees it, is a grasping of how sets of facts relate to each other.[24]

However, such cognitive reformulations of realism may be of little use to resolving the dilemma. One can question whether or not our scientific representations provide the appropriate justification for our beliefs or understanding. This returns to the same issues as the ones sketched above. Models are deliberately oversimplified, and thus not true. Therefore, it is not their truth that best explains their success. The fact that models are false does not, by itself, undercut their use in justifying and exploring beliefs about the world. In turn, this invalidates appeals to the NMA and PMI.

Given the considerations I have presented, it seems that the NMA and PMI may not be the best options to express the realist and anti-realist positions. Given the distinctive account of model-based science, and its use in capturing the practices of science, this undercuts the general application of the NMA and PMI. While they might maintain use in some applications, their role as the main arguments for and against realism seems limited.

24 This "grasping" metaphor is common in the literature on understanding. How this metaphor is fleshed out depends on the particular account, but it usually implies that there is some explicit recognition, sense, or experience that comes with understanding

5. Conclusion

In this paper I have argued that the debate around scientific realism encounters a dilemma, because the main arguments both for and against realism do not function well, concerning model-based science. This is a problem since model-based science has taken a prominent role in the philosophy of science. I considered several reformulations for scientific realism, or at least the debate about realism, that might allow us to maintain both the NMA and PMI and model-based science. However, I found that each of these options do not lend themselves to maintaining a central role for the NMA and PMI.

That the NMA and PMI may not be the main arguments for and against realism does not mean that there can be no debate about realism. For instance, rather than debating about whether or not models are true or false, debates about the methods of interpretation, means of construction, or some other aspect of the use of models may become central. We might question the results of an experiment, or a class of experiments, due to some concern about its methodology, and could similarly question whether, say, differential equations are appropriate in some applications. This would, generally, undercut the universal aspect of realism debates, as models would need to be dealt with on a more limited scale, but we can still have debates about realism.

Robustness and Replication: Models, Experiments, and Confirmation

Introduction

In the past several decades, robustness analysis has received its fair share of philosophical attention. Garnering much initial interest from the work of biologist Richard Levins, the focus of robustness analysis was as a method of confirming models (Levins 1966 and 1993). Robustness analysis of models is a significant practice in many sciences, such as biology, ecology, climate science, and economics, where other, more directly empirical means of confirmation may be unfeasible or impossible.

However, the confirmatory power of robustness analysis has been questioned (Orzack and Sober 1993, Sugden 2001, Odenbaugh, and Odenbaugh and Alexandrova 2011). In this paper I attempt to resolve a dilemma for the confirmatory power of robustness analysis initially presented in Orzack and Sober (1993) and expanded upon by Odenbaugh and Alexandrova (2011). The dilemma holds that robustness analysis cannot be confirmatory since it deals with false, idealized, models and false models cannot provide confirmation. However, if one of the models involved in robustness analysis is de-idealized, then robustness analysis is no longer needed. In either case, robustness analysis has nothing to add in terms of confirmation. I agree with these critics that many claims of the confirmatory power of robustness analysis are overstated. However, I also contend that these criticisms overstate their result, arguing that there are conditions where robustness does provide confirmation of a modeled result, given we properly understand the role of the idealizations involved. I defend this claim by drawing an analogy between robustness in modeling and replication in experiments.

The paper will proceed as follows. In section 1 I present the basics of robustness analysis and the proposed dilemma. In section 2 I consider the dilemma in relation to replication in experiments and argue that replication undercuts one of the horns of the dilemma. In section 3 I then draw an analogy between replicating experiments and robustness in models. In section 4 I consider some concerns over this analogy. In section 5 I present the concessions and upshot of my position.

1. Robustness and The Confirmatory Dilemma

The general philosophical impetus for robustness analysis is drawn from biologist Richard Levins (1966, 1993) and the response by Orzack and Sober (1993), as well as, somewhat separately, William Wimsatt (1981, 2007). Levins developed his account of robustness analysis within a more general discussion of the method of model building in population biology. His discussion was generally brief, and since then many different types of robustness have been distinguished (Woodward 2006). I focus on robustness analysis that most closely resembles what Levins had in mind.

The starting point for robustness analysis is that models used in sciences incorporate many falsehoods. Models are generally considered to rely upon sets of assumptions, with some such assumptions meant to be accurate, whereas other assumptions are obviously simplified to allow mathematical tractability. Some of these less accurate assumptions are generally viewed as rather harmless.[25] However, some assumptions remove or falsify causally relevant factors in the

25 What counts as inaccurate but innocent assumptions is not agreed upon, but it is generally agreed that there are some such assumptions. Even those who argue against idealized models as being confirmatory do not hold that *all* false assumptions are problematic (Alexandrova 2008, Odenbaugh and Alexandrova 2011).

target system and are viewed with greater suspicion. These will be referred to as idealizations and will be the focus of the discussion below.

These idealizations are often necessary for a model to function. We cannot derive a result from our model without the idealizations. This, on its own, is not problematic, but these idealizations are often incorporated into somewhat complex models. Given the general complexity of many models, it is not clear whether some result derived from the model is a result of the assumptions of the model that are meant to be accurate or if they are driven by some idealization (Levins 1966).[26] This is problematic since, if we want our models to tell us about the world, we want the results to be derived from the parts of our models that are like the world, or at least not be driven by the clearly false parts. The strategy of robustness analysis is to develop several models with a shared core set of assumptions, but different idealizations. If all of the models produce the same result, then this is supposed to lead to the conclusion that it is the shared core of the models that is responsible and not the idealizations. Such a result shared by all the models in the robust set is known as a "robust theorem".

Let us consider some model specifying an unconfirmed causal mechanism that implies some empirically confirmed result. The, hopefully accurate, representation of the causal mechanism does not imply the result on its own, and perhaps implies nothing. Without the idealized parts the model does not work at all. Given the assumption about the complexity of models discussed above, this makes it unclear if the result was produced by the, hopefully, accurate assumptions stipulating the core mechanism, or if it was driven in some key way by the

26 Levins 1966 compares the idealizations in the models of population biology to idealizations or simplification found in a street map. It is clear, say, that color is included in maps to assist in ease of reading and that the color selected does not interfere with the purpose of the map.

idealizations. Since the idealizations are known to be false, this means that producing the empirically confirmed result cannot provide confirmation that the model accurately represents the causal behavior of the target system. Since the idealizations are necessary, the next best option is to replace them with different idealizations while keeping the causal core the same. If the empirically confirmed result turns out to be a robust theorem, this is meant to show that the result is driven by the shared core and therefore is supposed to provide evidence for the mechanism specified therein.

There have been several criticisms of robustness analysis. For example, one criticism questions the possibility of the kind of non-empirical confirmation robustness analysis seems to provide (Orzack and Sober 1993).[27] Another criticism is that models in robust sets cannot be epistemically independent from each other enough to provide the epistemic boost required (Cartwright 1991, Orzack and Sober 1993, Odenbaugh and Alexandrova 2011, Justus 2015). Here, however, I focus on what is arguably a more central criticism of robustness—namely, the existence of a dilemma for its confirmatory power.

To begin, it is important to distinguish robustness analysis from the accumulation of evidence. The two are similar in that both are held to increase support for some proposition. When it comes to accumulating evidence, everything else being equal, if we have two bits of evidence for some proposition, we have more reason to believe it than if we have one (Achistein 2001 Ch. 2). A key reason for this is that evidence is fallible. So, if we have two bits of evidence supporting some proposition, and it turns out that one of the bits of evidence is flawed, then we

27 There have been several responses to this criticism as well. See Weisberg 2006 or Kuorikoski *et al*. for accounts of how models might incorporate appropriate empirical content.

still have one bit of evidence remaining.[28] If we only have one bit of evidence and it is flawed, then we have no evidence for our proposition in question.

Robustness is not meant to work this way, though. First, each model in the robust set is false, and so no single model in the robust set can thus be taken as evidence on its own. If the purpose of any of the models is to confirm the causal core, no individual model can do this.[29] It is the complete set of models, each showing that the idealizations of the others are irrelevant, that is intended to confirm the causal core. So, we start with several bits of non-evidence and by combining them together in confirmational alchemy we get evidence.

This sets up the key dilemma about the truth of the models in the robust set (Orzack and Sober 1993, Odenbaugh and Alexandrova 2011). Since all of the models in a robust set are false, combining them can provide no confirmation. Since each model is idealized, none of them specify a causal mechanism, and certainly not a causal mechanism as it appears in the world (Odenbaugh and Alexandrova pg. 764). Even if the robust set shows that the shared core drives the results of the model, the causal mechanism has still not been modeled since causally relevant factors have been idealized. In order for the robust set to provide confirmation, at least one of the models needs to be de-idealized. However, once a model is de-idealized there is no need to show that idealizations are driving the results. Therefore, this model can be confirmatory without a

28 This may work in the case of two experiments performed separately, both confirming some hypothesis, but it is found out that one of the experiments used flawed methodology. This flaw in one of the experiments does not influence the results of the other, and so we still have that single experiment as evidence. Compare this to a single experiment that turns out to be flawed. Now we lack any evidence whatsoever.

29 Kurikoski *et al*. discuss some reasons why this is the case in economics, particularly when compared to other sciences like physics. Given the nature of the parameters, the fact that they are constantly varied for instance, there are no ways to perform an analysis of how much a certain value of the parameter throws off the results, outside of the proposed robustness analysis.

robust set. In either case, the robust set is unable to provide any confirmatory power, since it either involves all false models or cannot play the role of "discharging" idealizations.

As a result, we end up with a dilemma for robustness analysis. On one horn of the dilemma a robust set provides no confirmation because each model in the set is idealized. To avoid this, a model is de-idealized and therefore can provide confirmation. However, once a model is de-idealized, the robust set adds nothing, creating the other horn. Therefore, the argument goes, robustness has nothing to add in terms of confirmation.[30]

2. Robustness and Replication

I take the confirmatory dilemma to be based on a false dilemma, however. In particular, I hold that idealizations can play an important epistemic role in confirming some claims about a target, but still introduce an uncertainty that is resolved by discharging idealizations. I look to another application of robustness reasoning, that of multiple experiments and replication to show that some scientific endeavor might be confirmatory on its own but still require robustness.

The role of robustness reasoning using scientific methods other than modeling has received some attention (Wimsatt 1981, Cartwright 1991, Soler *et al.* 2012). In particular, discussions have included the use of multiple experiments. It is often held that multiple, independent, experiments can be run to boost confirmation of some experimental result through replication or reproduction of the experiment (Bogen and Woodward 1988. Guala 2005, Open Science Collaboration 2015, Aarts *et al.* 2015).[31] When it comes to replicating an experiment it is

30 Robustness analysis has been provided other roles to play, such as suggesting useful lines of research or some heuristic role in counterfactual reasoning, but these lack confirmatory power. See Odenbaugh and Alexandrova 2011, pp. 768-770.

31 See also Bovens and Hartmann 2004 for discussion of reproduction and replication in relation to methods of detection.

not always strictly about exact reproduction, as this is often an impossibility (Guala 2005, pp13-15, Nosek and Errington 2017). An attempt at a replication is "the repetition of what is presumed to matter for obtaining the original result" (Nosek and Errington 2017, p.17). The reason that replication is important is because uncertainty about what drives the result of a single experiment is often not clear from that experiment alone.[32]

This uncertainty comes in many forms. There may be concerns about the method of collecting and analyzing data that leads to a false positive. Or, it might be that the experimental conditions played some role in producing the results (or both) (Cesario 2014, Simons *et al.* 2014, Stroebe and Strack 2014, and Maxwell *et al.* 2015). When it comes to uncertainty of results, replication and coherence of evidence generally provide an epistemic boost under many different conditions of uncertainty.[33] For instance, exact replications, or direct replications, can confirm whether or not the initial experiment produced a false positive. Divergent replications, where the replication changes some experimental conditions, can provide insight into whether or not it is something about the initial experimental conditions that produced the result, particularly those conditions that are not part of "what is presumed to matter."

The reason there is a concern about experimental conditions influencing the result is because such conditions are often quite contrived. The conditions of an experiment are important because they allow the experimenters to try and focus on what is presumed to matter by removing the buzzing, busy world outside the laboratory and its many confounding factors. There is often a focus on singling out the causal relationship between some limited set of causes

32 It is not just about what "drives" the results in terms of what experimental conditions, but replication can also reveal if, say, the theoretical interpretation of the results are wrong, but I will put this aside. See Stroebe and Strack 2014 and Nosek and Errington 2017

33 See Bovens and Hartmann 2004 appendices C and D for proofs in a Bayesian framework for how both exact replication and indirect replication boost epistemic prospects.

in a phenomenon of interest to gain a better understanding of those causal factors; while the experimental conditions may be contrived so as to limit the other causal factors found in the target. The purpose of multiple experiments is to try and determine whether or not a causal relationship exists between some variables, where the conditions that might produce this relationship in the world are too messy. For such an experiment, “what is presumed to matter” is a limited subset of how things function; the purpose being to discover a causal relationship between a subset of the causally relevant factors. To achieve this goal, experimental conditions are constructed that control, limit, remove, or exaggerate the influence of some causally relevant aspects of the real world so that a better understanding of the actual important causative factors can be had. In this way they do not reflect the world but can provide insight into some causal relation.

We can find clear examples of this in nutritional studies. The study of a particular macronutrient (protein, carbohydrate, or fat) or some subcategory (e.g., saturated or unsaturated fat) on some health metric or outcome, such as blood lipid levels, requires that diets of experimental subjects be tightly controlled and monitored. In many studies, such as metabolic ward studies, subjects are fed specially designed diets to ensure that the subjects are in energy balance (an isocaloric diet, or eating the same number of calories as they burn) and the percentages of nutrient breakdown are tightly regulated. For instance, one diet will have 20% of total calories from saturated fat, while a comparison diet will have 10% of calories from saturated fat and replace those lost total calories with a source of unsaturated fat. Such experiments tightly control many causal influences on blood lipid levels, such as whether or not a diet is hypercaloric, hypocaloric, or isocaloric, and the activity levels of subjects that might confound evidence collected from free living populations.

There can be insecurity when introducing such experimental conditions that the results are not determined by the causal relationship of interest, but are influenced in a way that the experimenters do not anticipate by the experimental conditions. Replicating the experiment may involve changing one way of controlling a certain causal factor for another. In this way it can be determined whether or not that method of control was influencing the results in a way that undercut the scientists' study of presumed casual factors. Replication in experiments includes the ability to determine which causal processes were responsible for the results determined in the initial experiment, and whether or not the results of such an experiment can then lend support to claims about the roots of true causation.

In the case of nutrition experiments, foods do not contain only a single nutrient. There is no food that only carries saturated or unsaturated fat; often carrying with it micronutrients, anti-oxidants, or phytochemicals. Each of these may have an impact on, say, absorption of fat during digestion or production of cholesterol or blood lipids. Fiber is one such nutrient. As a result, if we are interested in knowing what would happen if saturated fat in the diet is replaced with carbohydrates, carbohydrates of different fiber content may need to be used to distinguish what impact carbohydrates themselves have on blood lipids rather than the fiber content of a carbohydrate.

There are a couple of important points to make concerning experimental setup and replication. First, is do the experimental conditions reflect the target as it would be found out in the wild. Instead, are they constructed so as to alter some of the causal factors to gain a better understanding of those causal influences of interest. More important, does such an experiment play a confirmatory role, but robustness reasoning is still needed. It is unclear whether the underlying causal process produced the results of any of the experiments on their own, because it is unclear whether the experimental conditions actually played the controlling function they were meant to. However, given replication of an experiment, or several replications, it can be made clear whether or not any particular method of controlling non-important causally relevant factors played an unwanted influence on the results of the experiment.[34]

The focus of the above discussion is that robustness reasoning can still play a distinct confirmatory role even under conditions where an experiment can provide confirmation. Further, there is no analogy to deidealization in this case, because the purpose of the experimental conditions was to provide greater clarity on a particular causal relationship. Making the experimental conditions exactly reflect the world would undercut the ability to identify single causative factors. The takeaway, at least in this case of robustness reasoning, is the fact that something might be confirmatory does not mean that there is no role for robustness reasoning.

3. Idealizations and Controls

If robustness reasoning has a role to play in the case of replication, then the concern about robustness not being applicable if something is already confirmatory is unfounded. What needs to be shown is that models can be understood in terms similar to experimental conditions.

That we might want to remove some causal confounding elements to study some causal relation of interest is not unique to experiments. This is a method common in modeling as well. Weisberg and Elliott-Graves (2014), for instance, discuss what they call “minimal models”, where much of the target system is idealized in order to get a better understanding of how the system functions. This approach is not generally questioned. What is more important is the role

34 The discussion of replication does not scratch the surface for all of the possible roles that replication can play, but is sufficient for my purposes.

that idealizations can play in achieving this goal. For this, a better understanding of the content of idealizations is necessary.

Michael Strevens (Citation) has developed a useful account (for my purposes) of idealizations that distinguishes between the literal content of an idealization and its explanatory content.[35] The literal content of an idealization is exactly as it sounds, i.e., what the idealization literally implies. The explanatory content of an idealization is more context dependent, and involves the purpose to which the idealization is incorporated into a model. In population biology, for instance, populations might be idealized as infinite. The literal content of this idealization is that a population is infinite. However, the explanatory content may be that genetic drift is not causal in an infinite population. Since genetic drift is related to fluctuations in the frequency of certain alleles in a population due to random chance events in small population, if the population is infinite this is able to effectively eliminate this causal influence.

Strevens presents this account of the content of idealizations in relation specifically to explanations, but it is not truly limited specifically in that fashion. It might be that scientists are interested in investigating the causal relation between selective pressures and frequency of certain alleles in the population. To model this, the scientists might want to control the influence of genetic drift by removing its influence altogether, by representing the population as infinite. While there may be no actual populations that completely lack the influence of genetic drift, controlling this influence in our investigations can provide greater insight into other causal relations we are interested in, such as gene flow or mutation. Just like experimental controls remove causal confounders, idealizations might as well be used to reduce causal confounders.

35 Strevens 2008, ch 8., 2013.

Now, given this account of idealizations as a way to control for complicating variables, we can extend the analogy of robustness reasoning in experiments to that of models. Given the construction of a model using one idealization as a means of control, there is a concern that this idealization might be influencing the results in a way that is unknown. Thus, while the model is able to confirm some claim about causal relations, it is unclear if the model actually represents the causal relation of interest. Alternatively, a model with a different means of control (i.e. idealization) may be constructed to show that the initially indicated means of control is not influencing the results, and allows the scientists to confirm whether or not the causal relation of interest is driving the result.

If the analogy between experiments and models holds, this carves out a confirmatory space for robustness analysis when applied to models. Idealizations can play the role of experimental controls, and robustness analysis can then be employed as a means of showing that these controls are effective, providing insight into the causal relation of interest.

4. Objections

There are several concerns to be answered about this account. First, I will make a concession. I think critics of robustness analysis are correct that many cases where robustness analysis has been applied do not carry the confirmatory power assumed. There are cases where robustness analysis has been attempted where the idealizations might not play the role of control (examples). However, that this criticism holds true in many actual cases does not mean that it holds generally.

There are several important concerns relating to the ontological differences between models and experiments.[36] Models being mathematical, or graphical, while experiments are actually physical does provide some important epistemic distinctions. For instance, the causal relation being studied is not part of the mathematical framework the way that it is part of an experimental dynamic. As a result, this question about a model actually specifying a causal relation is better motivated.[37]

While this is an important point, it also needs to be noted that the fact that the causal relation is a physical aspect of the experimental subject does not mean that models cannot capture all that is needed to appropriately represent the causal relation of interest. It does point out an important difference in degree, in that we have more reason to believe that our experimental subjects instantiate the causal relation of interest, but this difference in degree is not insurmountable. We might have good reason to believe that our mathematical framework is up to the task of representing the causal relation of interest.[38]

As noted, this is a difference in degree, as there is an analogous assumption when it comes to experimentation. We need to have some reason to believe that the methods of measurement being employed are up to the task of detecting the causal relation in question, even if this question does not arise in the experimental subjects. Perrin famously used multiple experiments to calculate Avogodro's number to determine the existence of atoms. Perrin examined Brownian motion under many conditions, and measured Avogodro's Number using distinct methods of measurement. It was the superior precision in calculating Avogodro's

36 See Maki 2005, Morgan 2012, Ch. 6 section 6, and Morrison 2015 for more discussion of the epistemic distinctions and similarities between models and experiments.

37 See Humphreys 2002, Alexandrova 2008, and Odenbaugh and Alexandrova 2011 for more on this concern.

38 See Weisberg 2006.

Number afforded by the assumption of molecular theory, and the remarkable agreement across the many experiments, that led Perrin to conclude that atoms exist.[39] (how?) This replication would only be useful if calculating Avogodro's Number was a worthwhile measurement for determining the existence of atoms. So, while it might be that the causal relation is necessarily a part of the physical experiment in a way that it is not part of a mathematical model, this does not go so far as to present an insurmountable problem. The clarification needed is simply because we need some reason to believe that our methods of representing and detecting the causal relation are up to the task, in both models and experiments.

There are two further concerns derived from this ontological difference (also pointed out by Morgan, 2012, chap. 6.6). One difference between models and experiments is that, despite all of our best efforts, we are limited in how much we can engineer experimental conditions. It is beyond our abilities, for instance, to introduce impossible experimental conditions in the way that we can in models. This difference, however, is not a problem as long as we keep in mind that the idealizations we employ are those that control some causally relevant factors that we are interested in controlling. Whether this is done by impossible or possible means is thus a matter of degree; both can be shown to *not* affect the results detected in a similar manner.

Further, this ontological difference carries with it epistemic advantages of sorts for models. Morgan distinguishes between "surprise" and "confoundment". Models are able to surprise, but not confound, scientists with the results produced, while physical experiments can confound. Since models are mentally constructed, all of the elements that produce the result are known, but it is unknown how the combination of variables work together in an actual system.

---

39 See Psillos 2011. For different accounts of the process of multiple experiments, see van Fraassen 2009 and Hudson 2020.

So, scientists might be surprised by a result of a model, but it is never the result of some unknown variable in the model, since all of the variables were put there. Physical experiments, on the other hand, have the potential to confound, because knowledge of all of the factors that might play out in a physical system is limited. As a result, it is possible for the results of an experiment to present something legitimately new that is not explained or part of current theory. This is to the advantage of models in terms of focusing on relevant causal factors. There is no possibility for confoundment, since all parts of the model are known because they were explicitly put in there, if our interest is to focus on some causal relation, models can prove to be a better option in some cases.

Another line of criticism might stem from the distinction between literal and explanatory content, and whether or not it is *ad hoc*. As with many things in the philosophy of science, an account of idealizations needs to answer not only philosophical concerns but also to the practices of science. Part of why idealizations have been so philosophically vexing is that they seem to be epistemically corrupting but are both prevalent and stubborn in scientific practice. Idealizations are commonplace but it seems that de-idealization is not often a goal.[40] An account that can remove some of this tension has some *prima facie* support.

This distinction between literal and explanatory content can provide some answers to these concerns. While a complete defense is beyond the scope of this paper, it can help make sense of the prevalence and stubbornness of idealizations. Importantly, providing controls on causal influences is a common part of scientific investigations, and if this can be extended to an account of modeling (at least in some instances) this is an advantage. Further, idealizations might be introduced as controls does not conflict with other reasons we might use to explain the

40 Longino 2013, Morrison 2015, Rice 2020, 2021.

prevalence of idealizations in models, such as pragmatic reasons. We can introduce idealizations for a variety of epistemic reasons, and it some idealizations might be introduced for pragmatism or simplicity, while other times they are introduced for reasons of control.[41]

Further, this distinction provides one way of making sense of why de-idealization is rarely the end-goal of a model.[42] While models are often traded out or altered, it is not always the case that a model is headed toward a de-idealized state.[43] Rather than iteratively "improving" models by de-idealizing as we learn more, idealized models are generally kept more or less the same. Once again, it might be that idealizations are kept for pragmatic reasons, because the model would just be too mathematically complex if it was de-idealized, but this is not always the case. That idealizations play a positive epistemic role, not just pragmatic, can provide some insight into why de-idealization is not always the goal. This lends some, at least *prima facie*, reason to accept the distinction between literal and explanatory content.[44]

Given the considerations above, we can acknowledge that there are distinctions between models and experiments. However, when it comes to the robustness I have been discussing, these differences often amount to differences in degree and not a difference in degree extreme enough to warrant legitimate concern about the possibility of confirmation by robustness.

41As well, it might be that we want to control a causal confounder for pragmatic reasons.

42 As Levins (1966, pg. 430) points out, it is rare that a model is made completely precise, but often they are simply replaced with a different model that makes its own idealizations.

43 Cartwright 1983.

44 It could be argued that idealizations are kept for pragmatic and reasons of simplicity. However, if idealizations were seen as providing a legitimate obstacle to understanding it seems that they would be de-idealized. Further, idealizations could be dealt with as computational power increases. It is the case that models have gotten more complex with the incorporation of more computational power into scientific research. However, this has not resulted in de-idealized models, but more complex models. For instance, ecology has started incorporating Individual Based Models, which are not de-idealized versions of older models that focused on populations, but a different and more complex type of model.

5. Conclusion

The above discussion has benefit, but also makes some considerable concessions. Most significantly, I agree with critics of robustness analysis that many actual cases of robustness analysis on models do not generate the confirmation attributed. At best, it is often unclear what confirmatory support is provided by robustness analysis. A further concession is that the conditions under which robustness analysis might be applied are somewhat constrained. If the idealizations are not introduced with explanatory content, or if they are not introduced with the intention of controlling some causal influence, then what I have said lends little support.

The upshot is that there may be a role for robustness in confirming models. Furthermore, introducing idealizations as a means of control by employing some understood notion of explanatory content is not an improbable use for idealizations. It seems that this means of focusing in on certain causal influences is not a rare or unheard of use for models, and so this is a plausible and worthwhile use of robustness analysis.

A Defense of Functional Kinds: Multiple Realizability and Explanatory Counterfactuals

Introduction

Recently, Daniel Weiskopf has provided an update on the discussion of functional kinds and explanations (2011a, 2011b, 2017). Older accounts of functional kinds focused on explanatory laws (Putnam 1967, Fodor 1974), but a laws-based account of kinds and explanations has come under scrutiny, particularly in the sciences that are traditionally home to functional kinds, e.g., Biology, Geology, Chemistry. Focus on the role of models in sciences has grown, along with a models-based account of explanation, where more localized explanatory models take center-stage. Weiskopf's update on functional kinds takes the models-based account to heart.

Although his account is novel in this approach, familiar questions remain about the scientific legitimacy of functional kinds and, insofar as a science relies on them, the explanatory adequacy and autonomy of the sciences that employ them (Kim 2008, Piccinini and Craver 2011, Kaplan and Craver 2011, and Buckner 2015). In this paper, I consider one such criticism, according to which functional kinds fail to be legitimate scientific kinds because they offer inferior explanations to mechanistic accounts of kinds (Craver 2007, Piccinini and Craver 2011, Buckner 2015). These criticisms focus on the counterfactual profiles that each type of kind offers, arguing that a mechanistic explanation of the same phenomenon provides a superior counterfactual profile. In response, I argue that these criticisms mistake the explanatory target of functional kinds. In particular, these criticisms fail to properly take into account that functional kinds may be multiply realized.

Getting clear on the role of functional kinds is important for many reasons. One reason, dating back to earlier accounts (Fodor 1974) holds that the special sciences—like psychology or economics—are built out of functional kinds, or at least heavily employ them. If functional kinds have their own explanatory targets, then the special sciences will have their own explanatory realm. Similarly, if the criticism to be considered is valid, and functional kinds are not scientific kinds, then the special(?) sciences may lack their own set of kinds and lose their explanatory autonomy. Whether or not the special sciences function autonomously plays a significant role in further questions, such as integrating explanations across related sciences.

Further, if the defense of functional kinds that I propose is successful, this creates some problems for the general mechanistic framework. In particular, linking explanations specifically to mechanisms is part of a general framework for unifying psychology and neuroscience. If the defense I present below is successful, this attempt at unifying will ultimately fail. This, however, does not mean that coordination between these sciences will be lost, as I briefly discuss below.

Finally, whether or not there are functional kinds is important to many mechanistic accounts of explanation. Something is explanatory, on some such(?) accounts, only insofar as it illuminates a mechanism. If it turns out that functional kinds explain in a way that does not illuminate a mechanism, then this mechanistic view is questionable. Connected to this view of mechanistic explanation are accounts of integrating explanations across sciences, and an overall picture of the relationship between many sciences (Craver 2007).

This paper proceeds as follows. In section 1: outlining Weiskopf's account of functional kinds. In section 2: establishing clarifying points about multiple realizability. In section 3: counterfactual criticisms of Weiskopf's view. In section 4: arguing that these criticisms fail to take into account what multiple realizability provides in terms of explanatory advantage. In section 5: some concluding remarks about the impact that this discussion has on the mechanistic account of science.

1. Weiskopf's Functional Kinds

Weiskopf provides an updated account of functional kinds by focusing on the role that models play in science (2011b). Older accounts of functional kinds derived them from the explanatory laws found in the special sciences (Fodor 1974). However, there has been a move away from understanding sciences as law-based—particularly the special sciences—and a move towards a focus on models instead. This produces a problem for prior accounts because there are no longer any explanatory laws from which to derive the kinds. Weiskopf developed his account of functional kinds from their roles in explanatory models; however, focus on model construction introduced some new wrinkles in providing a coherent account of kinds.

Weiskopf takes functional kinds to be "abstractly defined functional categories [that] earn their credentials by participating in a range of models that are themselves empirically validated" (2011b pg. 251) Important to these functional abstractions is that they should, in general, not be directly reducible to the components or micro-detail of any target system.[45] This provides a clear distinction from other types of kinds, such as mechanisms, where mechanisms are determined by functional capacity (what it does), *and* the particular components that bring about this functional capacity. A functional kind is just determined by a functional capacity, where the actual components of the target are not taken into consideration.

45 Weiskopf allows that some may simply be difficult to reduce, rather than irreducible *in principle*. I will not consider this deviation here as it is unimportant to the arguments that follow.

These functional kinds are meant to be distinctive of the special sciences, with Weiskopf's particular focus being psychology. However, his account accommodates many other sciences. "Predator" and "prey" in ecological and population biological models may qualify as functional kinds. Various kinds in economics may similarly be considered functional on this account, given that they find their way into empirically confirmed models. The kinds of a science factor into its explanations, and so this is meant to give the special sciences autonomy from underlying sciences. If psychology, for instance, relies on functional kinds, it will have a unique explanatory domain that it alone captures, because its functional kinds are distinct from underlying neurobiological kinds and *cannot* be reduced to them.

These functionally defined or abstracted categories can be introduced into models in a variety of ways. However, abstractions are very common in models, and Weiskopf does not want to permit just *any* abstraction to qualify as a potential functional kind. Given this, there is a focus on three particular methods of abstraction employed in constructing a model (2011b, pg. 329). They will be presented in more detail below, but I will provide a quick glance now.

The first modeling practice Weiskopf calls fictionalization. This is where a category is included in a model that incorporates capacities that the target system is known not to possess. Essentially, there is some component in the model that performs a task that the target system *cannot* perform.

The second is reification. This is where capacities (?) carried out by distinct components of the target system are treated as capacities a single component in the model, or where the capacities of a single component in the target system are treated as capacities of two distinct components in the model. Essentially, multiple actual components of the target system are treated as a single component in the model, or a single component of the target system is treated as multiple, distinct components in the model.

The third modeling practice is functional abstraction. Functional abstraction is most similar to classic functional kinds, where details of a particular component in the target are abstracted away, and only the function it performs remains in the model.

In what follows, I limit my focus to fictionalization and some cases of reification. There are some important distinctions in these modeling techniques, and I argue that functional abstraction and some cases of reification result in what are known as mechanism sketches (Buckner 2015), because, these modeling practices do not introduce functional kinds at all, but are simply part of a mechanistic explanation. However, I accept that fictionalization and some cases of reification result in abstracted categories that *cannot* be equated to mechanism sketches, or any other type of scientific kind (Buckner 2015, pg. 3923). The fact that they are not part of a mechanistic explanation leaves fictionalization and reification open to the criticism that they are explanatorily inferior to mechanistic explanations, therefore not scientific kinds.

With this background, we can see Weiskopf's account as resting on four basic tenets (Buckner, 2015, pg. 3921). These are:

1. The autonomy of the special sciences

2. The multiple realizability of special science kinds

3. A model-based approach to induction and explanation

4. A model-based criterion of kindhood

The first two tenets are standard to accounts of functional kinds, while the last two distinguish Weiskopf's account from law-based accounts of functional kinds, such as those discussed in the work of Putnam (1967) and Fodor (1974).[46]

To understand the recent criticisms of this account, it is best to begin by considering the multiple realizability of special science kinds (tenet 2) in greater detail.

2. Multiple Realizability

The multiple realizability of special science kinds (tenet 2) has often been an impetus for both their defense and criticism, as well as a point of attack or defense for the autonomy of the special sciences (Putnam 1967, Fodor 1974, Kim 1998, 2008).[47] Part of this back-and-forth stems from differing takes on a possible explanatory role of functional kinds.[48] I follow Ross (2020) in highlighting what explanatory advantage multiply realizable kinds might have. Getting this explanatory advantage clear can help set the appropriate explanatory target for multiply realizable kinds.

Ross (2020) has recently discussed some criticisms of multiple realizability (Sober 1999), particularly in relation to their ability to provide causal explanations. Sober argues that multiple

46 Weiskopf does not commit himself to *all* functional kinds being multiply realizable. In fact, Weiskopf takes some of his examples to be models that pick out a single mechanism, in which case they would not be multiply realizable (2011a pg. 323). However, Weiskopf is committed to their being some multiply realizable psychological kinds (2011a), which is all that is needed to get this defense off the ground.

47 Fodor and Putnam present classic defenses of these kinds and explanations while Kim provides a criticism.

48 Some debate also stems from what counts as multiple realizability, as I will briefly discuss below.

realizability does not present a strong anti-reductionist case for scientific explanation arguing that reductive explanations are superior.[49]

Sober's argument is that explanations that include more of the gory details are "objectively" stronger than those that ignore them (Sober, 1999, pg. 549).[50] Sober's basic point is that, while we might accept a less detailed explanation, in terms of some multiply realizable kind that does not specify particular components, this will turn on pragmatic concerns and matters of taste (pg. 551). These higher level kinds are selected because the scientist is not interested in *expressing* the details in this particular context. For instance, when discussing the cause of some disease with their patient, a doctor will leave out many details, as the patient does not need to know them. This does not mean that the details may not be explanatory in this context, however. The downfall of more detailed or lower-level kinds is that they "explain too much" (pg. 547). Thus, we might want to exclude these details for simplicity in some situations. These pragmatic concerns are to be distinguished, Sober contends, from the increase in *objective* explanatory power that the gory details provide, which would include better description of the causal process and the ability to capture more counterfactuals.

Sober presents an example of carcinogens in cigarettes; contending that, in explaining why cigarette smoke causes cancer, it is the micro-structure that is relevant saying, "if smoking

---

49 Sober presents the criticism in terms of causal explanations, and Ross, in kind, presents a defense in terms of causal explanations. This is different from previous discussions where the focus was on a deductive-nomological account of explanation, rather than on explicitly causal account. This is important considering that model-based sciences, like mechanistic accounts, focus on causal and counterfactual aspects of the target rather than the laws of a science.

50 Sober distinguishes between a context of justification and a context of explanation. Under a context of justification, there are several pragmatic "matters of taste" for why we might accept a certain explanation, but a context of explanation does not turn on these matters of taste. He points out that this distinction is similar to the difference between context of discovery and context of justification (1999, pg. 551).

causes cancer, this is presumably because the micro-configuration of cigarette smoke is doing the work" (1999 pg. 548). This is true, even if it turns out that there might be several carcinogens in cigarette smoke, and different cigarettes might carry different carcinogens;

> "the fact that P is multiply realizable does not mean that P's realizations fail to explain the singular occurrences that P explains. A smoker may not want to hear the gory details, but that does not mean that they are not explanatory" (pg. 548-549).

While it might be enough for a doctor to simply give an explanation that "smoking caused your cancer" to a patient, in any particular instance, it is the micro-detail of the particular carcinogen that caused the cancer that gives the superior causal explanation.

Ross' response is that Sober focuses on the wrong explanatory target for multiply realizable kinds and explanations. Multiply realizable kinds capture causal heterogeneity, which is when "distinct instances of the same effect have completely different (or heterogeneous) causes" (Ross 2020, pg. 648). When looking at an individual instance of such effects, e.g., an individual carcinogen as the cause of a particular case of lung cancer—there is no causal heterogeneity, and so a multiply realizable kind is inappropriate. The focus should be at a "population level", where there can be many distinct causes for an effect.

But why insist on this change of explanatory target? Ross considers explanations for lung cancer at the population level. By stipulation of the example, there are multiple possible carcinogens found in cigarettes that cause lung cancer. As a result, it may be that explanations at the population level indexed to a particular realizer will miss out on causal facts related to the *other* realizers.[51] The specific details of a single realizer will only capture the explanations

---

51 I am assuming here that this is enough to count as multiple realization. There are different accounts of multiple realization, some of which would not consider this multiple realization. See Shapiro 2004 and Sullivan 2008 for a discussion of this.

relevant to that realizer, leaving all of the other cases in the population unexplained. As a quick example, let us assume that there are two distinct carcinogens, *A* and *B*, at work in cigarettes. Let us also assume that cigarette consumption in general has risen, and so consumption of both carcinogens has also increased. Given the question of what caused a rise in lung cancer among a population, the details concerning a single carcinogen will be insufficient, because it will provide no account of the possible role that the other carcinogen played. Faced with the counterfactual, "If intake of carcinogen *A* had not increased, lung cancer in the population would not have increased," to answer affirmatively is to give a false answer, since there has also been an increase in carcinogen *B*. Since the effect may be causally heterogeneous the explanation and counterfactual profile needs to take these heterogeneous causes into account. This is why a single realizer, even in all of its gory details, is not sufficient in this case.

Ross presents several responses to the possibility of capturing causal heterogeneity in terms of a disjunction of realizers. This was a common response to older accounts of multiple realizability as well, where the multiply realizable kind is simply considered to result from the disjunction of realizers. There are two points to consider here. First, it is not completely clear how to understand a disjunction of causes (Ross pg. 653). Given, say, an interventionist account of causal understanding, it is not clear how to intervene on a disjunction of causes or how that would capture the causal heterogeneity of the multiply realized kind.

A second point that Ross makes is that there is an important question of *why* this causal heterogeneity leads to the same effect (pg. 654). As Ross puts it, "there is an interest in knowing why different factors all produce the same effect and citing a disjunctive set of causes fails to answer this question" (pg. 654). Grouping carcinogens *A* and *B* together into "cigarette smoking" provides a peek into what the heterogenous causes have in common, more so than a disjunction of the realizers. Ultimately, it is this grouping together at a population level that provides some coherence to the scientific explanation rather than simply a list of causes.

This serves to highlight two important explanatory powers that multiply realizable kinds can carry. First, given the appropriate explanatory target, multiply realizable kinds are likely to provide a better counterfactual profile than any single individual realizer. Second, they provide a useful grouping for these population level explanations. What is important to highlight from this discussion of multiple realizability is that keeping the target requiring explanation clear is important. When it comes to explaining why a particular instantiation came about, then the specific details provide a better explanation. However, once we are dealing with causal heterogeneity, a singular explanation may very well be insufficient. Kinds that capture the heterogeneity of these causes capture more of the counterfactuals at the appropriate explanatory level.

3. The Explanatory Criticism

The criticism I focus on is that functional kinds fail to qualify as scientific kinds because they are explanatorily weak, at least compared to mechanistic kinds. The basic criticism works by comparing the counterfactual profile of a functional explanation when compared to a mechanistic one. We can see this kind of criticism presented in Craver (2007), Piccinini and Craver (2011), as well as Buckner (2015). I focus on the version presented by Buckner because it directly focuses on the position presented by Weiskopf, but my argument should apply to all such criticisms.

A further point to make is that the criticism, and my response, focus on fictionalized and some reified kinds, largely for simplicity, since the structure of the criticisms for fictionalized and reified kinds is the same. It can be argued that functional abstraction is actually part of providing a mechanistic explanation, and so does not introduce a purely functional kind (Picinnini and Craver 2011, Buckner 2015). Providing a defense of functional abstraction in terms of multiple realizability requires grappling with these arguments; thus, I start by taking what is given. No one in the debate thinks that fictionalized and reified kinds are parts of a mechanistic kind or explanation, therefore, I can focus on the particular criticism about the counterfactual profile.[52]

We can now start with a bit more focus on the criticism. Buckner provides the basic motivation for his approach by stating, "evidence that it [a category] will not be conserved or does not explain as well as alternatives should weaken our belief that it is a natural kind" (Buckner 2015, pg. 3916). Buckner sees both the possibility of replacement, combined with being comparatively explanatorily weak as related problems for functional kinds. Given that some category provides a weak explanation of some phenomenon, this is an indication that it might be replaced in the future. This judgment of weak explanation is driven by considerations of counterfactual profiles, and given that Buckner takes fictionalized and reified categories to have comparatively poor counterfactual profiles, this is the reason that we are to reject them as kinds. The basic structure is to compare the counterfactual profiles of explanatory models developed with fictionalized and reified kinds to other explanatory models, show that using these kinds produces models with relatively weak counterfactual profiles, and this gives us reason to reject them as kinds.

52 I believe that defense of functional kinds along the lines of multiple realizability can cause trouble for the arguments against functional abstraction being part of a mechanistic kind, but I leave that aside for the reasons discussed.

To understand this better, consider these two modeling practices—fictionalization and reification—in more detail. Fictionalization involves "putting components into a model that are known not to correspond to any element of the modeled system, but which serve an essential role in getting the model to operate correctly" (Weiskopf 2011a pg. 331). These are importantly distinguished from, say, black-boxes in models because fictionalized components are not intended to be replaced or filled in as subsequent models are developed, whereas black-boxes are often treated as placeholders. Further, while black-boxes incorporate a basic functional description of what is going on in the target being modeled, fictionalized categories include distinct causal powers not shown by the target (Weiskopf 2011a, pg. 331). It is these extra properties of the fictionalized category that allow it to play the role it does in the proposed explanatory model.[53]

Weiskopf presents the example of Fast Enabling Links (FELs) as a fictionalized category, while Buckner considers the case of backpropagation.[54] Both have found application in cognitive models— object recognition for FELs[55] and learning in connectionist networks for backpropagation[56]— however both incorporate biologically improbable or impossible properties.

---

53 Note that to deny that these fictionalized categories can be explanatory is to go against granting Weiskopf the point that his models at least qualify as explanatory. As we see, however, Buckner focuses on defenses for why these are explanatory. As well, whether or not fictions that introduce false properties (as opposed to something like an abstraction that simplifies or ignores properties) are explanatory is an open question. See Bokulich (2011) and Weisberg (2009).

54 Buckner makes the change to backpropagation because he does not think that FELs meet the requirement of being applied to several, distinct, targets.

55 Very briefly, FELs are used in some models of object recognition. An FEL is a link between several nodes in the model and functions by sending a signal to all nodes included on the link instantaneously.

56 Very briefly, backpropagation refers to the method of learning modeled in some connectionist networks, where the neural network is given a training task, and then the changes made to the network are calculated by how far off the output of the network is from the actual answer.

FELs, for instance, assume an infinitely fast transfer of information, while backpropagation, similarly, assumes unrealistic transfer of information, including the requirement of individualized error signals to adjust the weight of each node in the network. Without these questionable properties, however, the models that incorporate FELs or backpropagation would not be applicable in ways they are currently used.

Reification is the second modeling technique to be discussed, which involves "a division between model components that does not correspond to a structural division in an underlying mechanism" (Buckner 2015, pg. 3929). Ultimately, the functional and causal capacities included in the model are possessed by the actual system, unlike in fictionalization, but how those properties are divided up in the model does not reflect how the components of the target instantiate these properties. The functional properties of several components of the target system might be combined and treated as a single component in the model, or the causal properties of one object might be divided and treated as separate components in the model. Although all of the causal properties of the target system are accurately represented, components instantiating those properties are separated in a way that does not reflect the target system, and misrepresents the target in a way that cannot be brought in line with the actual components.

Buckner points out that combining or dividing capacities should be treated differently. The first strategy of representing the causal properties of several components in the target system as a single component in the model, is called fusional reification. Fusional reification is when "we introduce a component [into a model] whose causal capacities are actually distributed amongst distinct parts of the system" (Buckner 2015, pg. 3929). Buckner argues that fusional reification reduces to the modeling practice of functional abstraction, which reduces to mechanistic explanation. As a result, I put fusional reification aside for the rest of this paper. The other kind of reification is fissional reification. Fissional reification is when, "we introduce two or more distinct components [into the model] whose causal capacities are actually possessed by the same underlying part of the system (or the system as a whole)" (Buckner pg. 3929). Fissional reification, then, is when the causal properties of a single component in the actual system are treated as capacities of distinct components in the model.[57]

The example of connectionist networks is brought up again in the context of reification. A trained connectionist network is often treated as carrying 'representations' and 'inferences'. One common point, however, is that it is misleading to treat these as distinct entities in a connectionist network. The representations and inferences are the same activations through the neural network; thus, any change to one is a change to the other. It is not quite right to treat the representations as something distinct from the inferences and so this is a case of reification.

It is important to note how fictionalizations and reifications count as abstractions. What is key for Weiskopf is that they do not directly pick out any components, and are produced by a removal or avoidance of the specifics of some particular target to which they might be applied. In a manner similar to how "smoking" washes out the details of the particular carcinogens, the fictionalizations and reifications, are abstractions of the actual goings-on in a particular target, in that they do not pick out real components.[58]

Both modeling techniques receive, generally, the same criticism focusing on the counterfactuals implied by these functional kinds, because they are supposed to negatively impact the counterfactual profile of models that employ them in two ways. One is that they will

---

57 Unless indicated otherwise, when I use the term "reification" I mean fissional reification below.

58 Just to be clear, I am not implying that "smoking" is a category that is fictionalized or fissionally reified. I am simply drawing the comparison that details are washed out.

imply counterfactuals known to be false. The other is that they will obscure important explanatory counterfactuals about their target.

While every model implies false counterfactuals and obscures others, what distinguishes the counterfactuals implied by fictionalization and reification is that they stem from a central aspect of the proposed explanatory category (FEL, backpropagation, etc.). Without these categories, the models would not be explanatory. Models that employ FELs, for instance, only work *because* of the FELs, which only do their job due to the incorporation of instantaneous information transfer. Simplifications and other kinds of abstractions can, in theory, be made increasingly precise or removed. However, in the case of fictionalized or reified kinds, precisification or removal of the questionable properties will remove what allows the kind to capture important explanatory counterfactuals. The false counterfactuals are explicitly implied by what makes the category useful in the first place.

Buckner presents this criticism clearly against fissional reification in his "A without B" challenge.

> "for any two subcapacities A and B, if the system cannot perform A without engaging the very same mechanism that performs B, then an explanation that construes A and B as distinct subcapacities will have less counterfactual power than an otherwise identical model that depicts them as two aspects of the same capacity" (Buckner 2015, pg. 3929).

In a connectionist network, 'representations' and 'inferences' are the exact same activations across the nodes. To reify them as distinct is to imply that we could "prime an inference rule without simultaneously priming a set of associated representations, or that we could add representations to the network without subtly altering generalization patterns for the networks" (Buckner 2015, pg. 3930). Reifying implies false counterfactuals about the ability to manipulate inferences without manipulating representations, as well as obscuring true counterfactuals about the connection between inferences and representations in the connectionist network.

A similar point is made for a model developed with a fictionalized category. FELs, for instance, are meant to capture the fact that distinct neurons work in synchrony. It is not clear how this is done in every case, but it is known that it is *not* accomplished by instantaneous transfer of information. Thus, FELs imply various counterfactuals about synchrony that will not be borne out by the target system, for instance, about the speed of information transfer. Further, Buckner points out that, whatever advantages FELs might bring, they do so “only at the cost of a diminished ability to predict and explain another—namely, the aspect that is fictionalized” (Buckner 2015, pg. 3926). FELs imply false counterfactuals about synchrony and, in doing so, obscure some true counterfactuals about how neurons actually function synchronously.

Both types of kinds are compared unfavorably to a mechanistic account. Although it is unknown exactly how synchrony may be achieved, for instance, a mechanistic model that may be produced will require that it accurately represent the components and their capacities. So, whatever the mechanism of synchrony may be, it will have a stronger counterfactual profile than FELs, because it will not imply the false counterfactuals about information transfer, nor will it obscure by fictionalizing some aspect of the explanation. A similar point is made about reification, where a mechanistic explanation of inferences and representations will require that they be tied to the same components, and the same capacities of the components, so that they cannot be treated differently. In both cases, we have reason to suspect that the explanations are counterfactually weak and, even though we may not have the real mechanistic explanation identified, we know that once we do the functional kind will not be needed (outside of, perhaps, pragmatic reasons or concerns for simplicity).

The general criticism here is that models with functional kinds will provide inferior explanations relative to other models that do not incorporate them. Buckner, in particular, compares functional kinds to mechanisms to show the advantage of incorporating details about components and avoiding such abstractions. Given this, our belief in functional kinds as being *bona fide* scientific kinds should be weakened.

4. Multiple Realizability and Counterfactual Disadvantage

We can now pull out an important tension in Buckner's criticism of functional kinds regarding their explanatory target. Buckner's general criticism is that "the common currency in arbitrating between functionalist and mechanistic interpretations...is counterfactual power, with the interpretation that supports more genuine counterfactuals being preferable, ceteris paribus" (pg. 3928). I begin by looking at the point that functional kinds obscure some important explanatory counterfactuals.

We can see this point in regards to fictionalized models. When discussing FELs, Buckner says that;

> "… it is unclear why modelers should be uninterested in the way that real cognitive systems achieve synchrony. The true explanation for synchronization will be of value not only because it provides additional detail at lower levels of description, but also because it will support more counterfactual knowledge at the psychological level of description" (2015, pg. 3928).

Everything Buckner says here is correct. Modelers should be interested in how real cognitive systems achieve synchrony, and true explanations for synchronization will be of value across various sciences. But nothing about FELs stands in the way of us doing so—if we take FELs to be multiply realized. As with all cases of multiply realized kinds, FEL's cannot provide an explanation of how any *particular* mechanism instantiates synchrony. However, they can provide a way of discussing *types* of systems that use synchrony.

A similar point can be made for reified models. That inferences and representations are the same activation vectors in a connectionist model is important when trying to understand that particular system. But the advantage of talking about inferences and representations is that they are not relevant *only* to connectionist models. They can be instantiated by non-connectionist cognitive systems, and provide a point of similarity between various types of systems.

What's more is that this is not just a way of excusing the obscured counterfactuals of any particular realizer. There is an important counterfactual *gain* in accepting the distinction between the explanatory goal of multiply realized kinds and the kinds that make up its realizers. Barrett makes this point about psychological explanations of working memory by saying;

> "It is not obvious at all what the addition of neuroscience [to psychological explanations] accomplishes...if working memory is multiply realized, then those details will not be capable of telling us the general story about working memory (since the explanation will be indexed to only one of many neurological realizers)" (2014, pg. 2708).

The advantage of a multiply realized functional kind will be that it captures a type of functional capacity that is carried out in several ways. What this means is that, if we were to compare explanations at the "higher-level"—the "population level" that multiply realizable kinds can capture—we can reverse Buckner's criticism. The model of a particular realizer misses important counterfactuals implied by causal heterogeneity that a functional kind model would capture, and to criticize the model of a particular realizer for missing these counterfactuals is to mistake the explanatory target.

A second part of Buckner's criticism is that fictionalized and reified kinds introduce false counterfactuals, but this generates a similar response. In the case of connectionist networks, Buckner argues that reification of representations and inferences introduces problematic counterfactuals by implying that representations and inferences can be manipulated somewhat independently. However, this occurs only if we take this reification to be focused on how the cognitive system that is exactly captured by how such a model instantiates them. Representations and inferences are not meant to represent how these are instantiated in connectionist networks alone, as they are applicable to non-connectionist models as well. Therefore, while connectionist networks instantiate representations and inferences in a way that means they cannot be manipulated in the same way as other systems that do, applying these categories to distinct types of systems—connectionist and non-connectionist cognitive systems—provides insight into a similarity between these types of systems. Various types of systems represent the world, and we have an interest in what representation in general provides; thus, it cannot be connected to how any particular variation instantiates representations. After all, it is by some shared factor that we bother to apply these categories of representation and inference to various systems, and if we expected them to capture the mechanisms of the systems they were applied to then they would lose this generality.[59]

The case with fictionalized models is a bit more complicated, but follows this same pattern. The fictionalized aspect of the model will introduce false counterfactuals given its impossible nature. FELs require a biologically impossible capacity, for instance. However, the impact of this needs to be judged across the various realizers. Buckner allows that fictionalized components can carry explanatory advantages, but sees them as a net negative given the false

59 See the Barrett quote about above.

counterfactuals implied (pg. 3928). The problem is he considers these false counterfactuals in regards to a *single* realizer, and it may be the case that the explanatory advantages are held across the heterogenous set. This would undercut any concern about a superior mechanistic explanation replacing FELs, for instance, as the mechanistic explanation would capture a single realizer, missing out on important counterfactuals itself.

As shown above, we can flip Buckner's argument around. The model of a particular realizer implies false counterfactuals at the level where causal heterogeneity shows up. Synchrony can be instantiated in a variety of substances, such as neurons, silicone chips, or lasers, and FELs can be used to capture synchrony under these various conditions. There are some general points about synchrony that will be captured by FELs that cannot be captured by a model of how a particular realizer instantiates synchrony. So, while FELs might imply false counterfactuals about any particular realizer, the mechanism that instantiates synchrony in any particular realizer will imply false counterfactuals about all of the other systems that achieve synchrony. Once causal heterogeneity is considered, the counterfactual advantage is had by a kind that can capture this causal heterogeneity.

The point of the above arguments is that, once the appropriate explanatory target is kept in mind, the concerns about the counterfactual profiles of multiply realized, functional kinds seem less pressing. There is a clear explanatory realm that is better captured by these multiply realized, functional kinds than any of their realizers.

It is important to note that this causal heterogeneity is not something that can be easily captured in a mechanistic account of kinds. While there is nothing that rules out the possibility of a population-level mechanism, and there is room to allow multiple realization of components and mechanisms (Craver 2009, Rosenberg 2018), mechanisms are limited in the causal heterogeneity they can capture, because they must pick out components actually in the target. However, different sets of components can enact the same functional profile. Once these components are sufficiently distinct, they can no longer be considered the same mechanism; thus, there would be no single mechanism that captures all of the heterogeneous causes under one kind, the way a functional kind can.[60] If this causal heterogeneity is explanatorily useful, as I believe Ross has shown, then it is something that mechanisms will struggle to capture sufficiently.

All things considered, then, arguments that functional kinds are not scientific kinds *because* of their counterfactual weaknesses missed an important explanatory advantage that multiply realized kinds have. Further, mechanistic kinds themselves will have difficulty capturing the causal heterogeneity that functional kinds can capture due to being determined by components.

5. Concluding Remarks

The defense presented here has been partial. As was noted above, Weiskopf is not committed to *all* psychological kinds being multiply realizable. Barrett (2014) and Weiskopf (2017) have mounted other defenses without considering multiple realization. What is important about the argument presented above is that it makes clear that claims of explanatory superiority need to be made in appropriate context.

This point ties into some further aspects that may be found in the mechanistic framework. Part of this framework is that explanations function by illuminating mechanisms (Kaplan and

60 See Sullivan 2008 for more on this point. As well, see Shapiro 2000 and 2004 for discussions of different ways that functional kinds might be realized and the fact that many mechanisms might realize the same functional profile.

Craver 2011 pg. 611).[61] Something is not explanatory, if it does not describe a mechanism. Functional kinds, at least fictionalized and reified ones, do not do this, or at least do this very poorly. The fact that these functional kinds might carry an advantage in explaining some phenomena, by capturing causally heterogenous counterfactual profiles more appropriately, calls the link between explanations and mechanisms into question.

This is problematic because some aspects of the mechanistic view that are enticing stem from this homogenous explanatory goal. Mechanistic positions provide a promising account of integrating explanations across sciences, for instance (Craver 2007, Piccinini and Craver 2011). If all explanations are about mechanisms, and we accept a mosaic picture of mechanisms, then integrating explanations across sciences is fairly easily handled. The explanation of some phenomenon at one level, say the neuroscientific, provides the components that make up mechanisms at another level, e.g., psychology. Explanations in different sciences work together in providing the components and functional profiles of mechanisms, and this relation can be carried out across a range of sciences.[62] However, if there are some explanations that are *not* about elucidating mechanisms, this simple picture of complete integration is lost.[63]

61 Kaplan and Craver limit their focus to cognitive and systems neuroscience specifically, and claim that a successful explanatory model is one that models a mechanism. There have been extensions of the mechanistic framework beyond this. However, other presentations of the mechanistic framework are a bit more cagey, making claims that certain sciences cannot be *understood* without an appropriate understanding of mechanistic explanations. This is different from claiming that all explanations are mechanistic, as the claim can simply be that mechanisms provide one important type of explanation among several, but this would be rather uncontentious. Further, without the stronger mechanistic assumption there seems to be little reason to provide criticisms of functional kinds.

62 This has been a part of arguments against the autonomy of the special sciences since they simply provide part of an explanation of some mechanism along with

63 I am not denying that *some* explanations are about mechanisms, just denying that all are. This account of integration between mechanistic explanations can be maintained even if we accept that there are some functional kinds.

This account of integration has one advantage in favor of the mechanistic world-view. One criticism of the functional kinds is that, when paired with other types of kinds, prospects for integrating explanations across related sciences seem limited (Buckner 2015, pg. 3939).[64] However, this defense of functional kinds via multiple realizability opens up a path for integration of explanations across sciences. If we take it that some functional kinds might be multiply realizable, the question of how these scientific kinds relate to explanations in other sciences becomes approachable by understanding the realized/realizer relation. While this relationship might not be completely clear, it is a starting point.[65]

Further work from this can go into understanding the realized/realizer relationship as it might count as integrating sciences. Further work for the autonomy of the special sciences might focus on showing how the multiply realized functional kinds interact with non-multiply realized functional kinds that are also posited by Weiskopf to produce a full gamut of kinds for any particular science.

64 See Sullivan 2016 for some general concerns about integrating psychology and psychiatry with neuroscience regardless of the account of kinds assumed.

65 Note that this does not run into a problem of evidence that might make it subject to criticisms found in Piccinini and Craver 2011. They argue that the only way to distinguish a how-possibly from a how-actually explanation is to have evidence for the functional profile at the component level. Ultimately, to know if a psychological explanation is a how-actually explanation, there needs to be some evidence at the neuroscientific, component, level. The realizer/realized relation can simply be a way of connecting explanations, not an evidential one.

Conclusion

The chapters in this dissertation deal with a few, loosely related questions in the philosophy of science. In general, there is a theme of identifying realism of some sort through each of these chapters, with a particular focus on how the development the philosophical literature on scientific modeling has impacted our understanding of realism and truth.

The first chapter tackled this concern head on, focusing on how the prominence of "model-based" science has impacted philosophical approaches to scientific realism. In particular, the main argument for and against scientific realism require an understanding of scientific practice that cannot accommodate "model-based" science. This is problematic for the realism debate, since we now lose the two most prominent arguments for and against realism. I am skeptical of an appropriate account of scientific realism that can accommodate the NMA and PMI *and* a model-based view.

The second chapter moves on to questions of confirmation. This is important in a very general view about scientific realism, since an appropriate understanding of realism is tied to truth in some sense. The concern dealt with here is that the idealizations of models make them poor representations, and therefore undercuts prospects for them conveying truth, and in particular providing confirmation. One method of resolving concerns about idealizations has been robustness analysis. Robustness analysis has been roundly criticized for not actually providing confirmation.

I argued that this criticism overstates some concerns, and by drawing an analogy between idealizations in models and controlled experimental conditions, made the case that models might provide confirmation. I argued that the criticism leveled against robustness analysis made a mistake about the impact of idealizations and the role of robustness reasoning. Idealizations can be used to control causal confounders that are unwanted. However, in controlling causal confounders, it is not clear if that means of control is influencing the results in some unexpected way. Just as experiments are replicated to make sure that experimental conditions are not influencing the results in an unexpected way, so should idealized models be subjected to robustness analysis.

Finally, I tackled questions about the nature of scientific kinds. In particular, it has been argued that functional kinds do not qualify as scientific kinds since they are explanatorily inferior to other explanatory kinds. I argued that these criticisms mistake the explanatory power of multiple realization. In particular, multiply realized kinds capture unique counterfactuals that cannot be captured by their instantiations. Given that functional kinds are multiply realizable, they then capture explanatorily important counterfactuals that are lost by more specific kinds.

I believe that all of these papers point towards further work needed in clarifying various aspects of scientific realism. In general, I believe that many accounts of the realism and anti-realism debate do not take into consideration the generally pragmatic nature of reasoning. This is an oversight, I believe, because it has often tied explanatory power to exact or close representation of a target. This, I believe, is not reflected in actual practice and stems from a conflation of our means of representing the world with how we reason and learn about them. I look to extend this work by focusing on how exactly it is that we might leverage falsehoods to learn about the world.

Works Cited

1. Aarts, A.A, et al. “Estimating the Reproducibility of Psychological Science.” Science (American Association for the Advancement of Science), 349(6251), 2015, 943.

2. Achinstein, P., 2001. *The Book of Evidence*. Oxford University Press.

3. Alexandrova, A., 2008. “Making Models Count.” Philosophy of Science, 75(3), 383–404.

4. Barrett, D., 2014. “Functional Analysis and Mechanistic Explanation”. *Synthese, 191*(12), 2695-2714.

5. Belot, G., 2007. “Is Classical Electrodynamics an Inconsistent Theory?” *Canadian journal of philosophy*, 37(2), 263–282.

6. Bogen, J, and Woodward, J., 1988. “Saving the Phenomena.” The Philosophical Review, 97(3) 303–352.

7. Buckner, C. 2015. “Functional Kinds: A Skeptical Look”. *Synthese, 192*(12), 3915-3942.

8. Cartwright, N. 1983. *How the Laws of Physics Lie*. Clarendon Press ; Oxford University Press,.

9. Cartwright, N. 1991. “Replicability, Reproducibility, and Robustness: Comments on Harry Collins”. *History of Political Economy*, 23(1), 143-155.

10. Cartwright, N. & Jones, Martin R, 2005. *Idealization XII : Correcting The Model : Idealization And Abstraction In The Sciences*. Amsterdam ; New York, NY: Rodopi.

11. Cesario, J., 2014. "Priming, Replication, and the Hardest Science : Behaviorial Priming and Its Replication." *Perspectives on Psychological Science* 9(1), 40-48.

12. Chakravartty, A., 2007. *A Metaphysics For Scientific Realism : Knowing The Unobservable*, Cambridge: Cambridge University Press.

13. Costa, N.C.A.da. & French, S., 2003. *Science And Partial Truth : A Unitary Approach To Models And Scientific Reasoning*, New York: Oxford University Press.

14. Craver, C., 2007. *Explaining the Brain.* Oxford: Oxford University Press.

15. Craver, C., 2009. "Mechanisms and Natural Kinds". *Philosophical Psychology, 22*(5), 575-594.

16. Dawid, R. & Hartmann, S., 2017. "The No Miracles Argument Without The Base Rate Fallacy". *Synthese (Dordrecht)*, 195(9), pp.4063–4079.

17. Devitt, M., 2011. "Are Unconceived Alternatives a Problem for Scientific Realism?" *Journal for general philosophy of science*, 42(2), pp.285–293.

18. Doppelt, G., 2007. "Reconstructing Scientific Realism to Rebut the Pessimistic Meta-induction". *Philosophy of science*, 74(1), pp.96–118.

19. Downes, S.M., 1992. "The Importance of Models in Theorizing: A Deflationary Semantic View". *PSA (East Lansing, Mich.)*, 1992(1), pp.142–153.

20. Elgin, C.Z., 2017. *True enough*, Cambridge, MA: MIT Press.

21. Elliott-Graves, A., and Weisberg, M., 2014. "Idealization." Philosophy Compass, vol. 9(3) 176–185.

22. Fodor, J. A. (1974). "Special Sciences (Or: The Disunity of Science as a Working Hypothesis)". Synthese, 28(2):97–115.

23. Giere, R.N., 1988. *Explaining Science : A Cognitive Approach*, Chicago: University of Chicago Press.

24. Giere, R.N., 2006. *Scientific Perspectivism*, Chicago: University of Chicago Press.

25. Godfrey-Smith, P., 2006. “The Strategy Of Model-Based Science”. *Biology & philosophy*, 21(5), pp.725–740.

26. Guala, F. (2005). *The Methodology of Experimental Economics*. Cambridge University Press.

27. Hempel, C.G., 1965a “Aspects of Scientific Explanation”, in Hempel 1965b: 331–496.

28. Hempel, C.G., 1965b. *Aspects Of Scientific Explanation, And Other Essays In The Philosophy Of Science*, New York: Free Press.

29. Henderson, L., 2017. “The No Miracles Argument And The Base Rate Fallacy”. Synthese, 194(4), 1295–1302.

30. Hohwy, J. & Kallestrup, Jesper, 2008. *Being Reduced : New Essays On Reduction, Explanation, And Causation*, Oxford ; New York: Oxford University Press.

31. Howson, C., 2013. “Exhuming The No-Miracles Argument”. Analysis, 73(2), 205–211.

32. Humphreys, P., 2002. “Computational Models.” Philosophy of Science, vol. 69 (S3), S1–S11.

33. Jones, M.R., 2005. “Idealization And Abstraction: A Framework”. In Cartwright and Jones 2005. Poznań Studies In The Philosophy Of The Sciences And The Humanities. pp. 173–217.

34. Kaplan, D., & Craver, C., 2011. The Explanatory Force of Dynamical and Mathematical Models in Neuroscience: A Mechanistic Perspective. *Philosophy of Science*, *78*(4), 601-627.

35. Kim, J., 2008. Reduction and Reductive Explanation: Is One Possible Without the Other? In Howhy and Jesper 2008. Oxford: Oxford University Press.

36. Kuorikoski, J, et al., 2010 "Economic Modelling as Robustness Analysis." The British Journal for the Philosophy of Science, 61(3), 541–567.

37.

38. Levins, R., 1966 "The Strategy of Model Building in Population Biology." American Scientist, 54(4), 421–431.

39. Levins, R., 1993 "A Response to Orzack and Sober: Formal Analysis and the Fluidity of Science." The Quarterly Review of Biology, 68(4), 547–555.

40. Maaki, U., 2005, "Models Are Experiments, Experiments Are Models." The Journal of Economic Methodology, 12(2), 303–315.

41. Maxwell, S. E., Lau, M. Y., and Howard, G. S., 2015. "Is Psychology Suffering From a Replication Crisis?" *The American Psychologist* 70(6), 487-98.

42. Morgan, M. S., 2012 *The World in the Model*. Cambridge, UK: Cambridge University Press.

43. Morgan, M.S. & Morrison, M., 1999. *Models As Mediators : Perspectives On Natural And Social Science*. New York: Cambridge University Press.

44. Morrison, M., 2007. “Where Have All the Theories Gone?” *Philosophy of science*, 74(2), pp.195–228.

45. Morrison, M., 2015. *Reconstructing Reality : Models, Mathematics, and Simulations*. New York: Oxford University Press.

46. Nosek, B. A., and Errington, T. M., 2017 "Making Sense of Replications." ELife(6).

47. Odenbaugh, J., 2011. “True Lies: Realism, Robustness, and Models”. Philosophy of science, 78(5), pp.1177–1188.

48. Odenbaugh, J, and Alexandrova, A. 2011. “Buyer Beware: Robustness Analyses in Economics and Biology.” Biology & Philosophy, 26(5), 757–771.

49. Orzack, S H. (2005). “Discussion: What, If Anything, Is ‘The Strategy of Model Building in Population Biology?’ A Comment on Levins (1966) and Odenbaugh (2003).” Philosophy of Science, 72(3), 479–485.

50. Orzack, S. H., and Sober, E., 1993. “A Critical Assessment of Levins's The Strategy of Model Building in Population Biology (1966).” The Quarterly Review of Biology, 68(4) 533–546.

51. Piccinini, G., & Craver, C. 2011. “Integrating Psychology and Neuroscience: Functional Analyses as Mechanism Sketches”. *Synthese*, 183(3), 283–311.

52. Plutynski, A., 2006. “Strategies of Model Building in Population Genetics”. *Philosophy of science*, 73(5), pp.755–764.

53. Psillos, S., 1999. *Scientific Realism : How Science Tracks Truth*, London ; New York: Routledge.

54. Psillos, S., 2011. "Moving Molecules Above the Scientific Horizon: On Perrin's Case for Realism." *Journal for General Philosophy of Science* 42(2), 339-63.

55. Putnam, H. 1967. Psychological Predicates. In *Art, Mind, and Religion.* Harvard University Press.

56. Putnam, H., 1975. *Mathematics, Matter, And Method*. New York: Cambridge University Press.

57. Rice, C., 2019. “Understanding Realism”. *Synthese (Dordrecht)*, 198(5), 4097–4121.

58. Rice, C., 2019. "Models Don't Decompose that Way: A Holistic View of Idealized Models", The British Journal for the Philosophy of Science, vol. 70(1), 179-208.

59. Rice, C., 2021. *Leveraging Distortions : Explanation, Idealization, and Universal Patterns in Science*.

60. Rosenberg, A. 2015. Making Mechanism Interesting. *Synthese. 195*(1), 11-33.

61. Ross, L. 2020. Multiple Realizability from a Causal Perspective. *Philosophy of Science* 87 (4), 640-662.

62. Rowbottom, D.P., 2019. A methodological argument against scientific realism. *Synthese (Dordrecht)*, 198(3), 2153–2167.

63. Schulz, A. W., 2015. "The Heuristic Defense of Scientific Models: An Incentive-Based Assessment." *Perspectives on Science* 23(4), 424-42.

64. Shapiro, L. 2000. “Multiple Realizations”. *The Journal of Philosophy, 97*(12), 635-654.

65. Shapiro, L. 2004. *The Mind Incarnate* (Life and mind). Cambridge, Mass.: MIT Press.

66. Simons, D. J., 2014 "The Value of Direct Replication." *Perspectives on Psychological Science* 9(1), 76-80.

67. Stroebe, W., and Strack, W., 2014. "The Alleged Crisis and the Illusion of Exact Replication." *Perspectives on Psychological Science* 9(1), 59-71.

68. Sober, E., 1999. “The Multiple Realizability Argument Against Reduction”. *Philosophy of Science*, 66(4), 42–564.

69. Soler, *et al.* (eds), 2012. *Characterizing Robustness in Science: After the Practice Turn in the Philosophy of Science.* New York, NY: Springer-Dordrecht.

70. Sugden, R. 2001. “Credible Worlds: the Status of Theoretical Models in Economics.” The Journal of Economic Methodology, vol. 7(1), 1–31.

71. Sullivan, J., 2008 “Memory Consolidation, Multiple Realizations, and Modest Reductions”. *Philosophy of Science*, *75*(5), 501-513.

72. Sullivan, J. 2016. “Construct Stabilization and the Unity of the Mind-Brain Sciences”. *Philosophy of Science, 83*(5), 662-673.

73. Van Fraassen, B. C., 1980. *The Scientific Image*. Oxford, UK: OUP.

74. Van Fraassen, B.C., 2008. *Scientific representation : paradoxes of perspective*, Oxford : New York: Clarendon Press ; Oxford University Press.

75. Van Fraassen, B. C., 2009 "The Perils of Perrin, in the Hands of Philosophers." *Philosophical Studies* 143(1), 5-24.

76. Weisberg, M., 2013. *Simulation and similarity : using models to understand the world*, New York: Oxford University Press.

77. Weiskopf, D. 2011a. The Functional Unity of Special Science Kinds. *The British Journal for the Philosophy of Science*, *62*(2), 233-258.

78. Weiskopf, D. 2011b. Models and Mechanisms in Psychological Explanation. *Synthese, 183*(3), 313-338.

79. Weiskopf, D. 2017. The Explanatory Autonomy of Cognitive Models. In *Explanation and Integration in Mind and Brain Science*. Oxford University Press.

80. Wheeler, B., 2020. Idealization, Scientific Realism, and the Improvement Model of Confirmation. *Science & Philosophy*, 8(2), pp.7–15.

81. Wimsatt, W. C. 2012. “Robustness, Reliability, and Overdetermination.” in Soler *et al.* (eds). 2012, 61-78.

82. Wimsatt, W. C. 2007. *Re-Engineering Philosophy for Limited Beings : Piecewise Approximations to Reality*. Harvard University Press.

83. Woodward, J. 2006. “Some Varieties of Robustness.” The Journal of Economic Methodology, 13(2), 219–240.

84. Worrall, J., 1989. “Structural Realism: The Best of Both Worlds?” *Dialectica*, 43(1-2), pp.99–124.

85. Worrall, J., 2007. “Miracles And Models: Why Reports Of The Death Of Structural Realism May Be Exaggerated”. *Philosophy (London)*, (61), 125-154.

86. Wray, K.B., 2018. *Resisting Scientific Realism*. Cambridge, UK. CUP.