1 Introduction

Scientists and philosophers use the terms “theory”, “model” and “phenomena” in such a variety of ways that it may seem pointless to try to regiment these uses into some more definite structure. At the same time, practitioners repeatedly run into puzzles of one kind or another. This paper considers one puzzle associated with the semantics of theories. The solution to this puzzle is to distinguish the role of one type of representation, dubbed “theories”, from another type of representation, dubbed “models”. I do not intend to argue that this account of theories and models somehow distills the essence underlying the motley of practice. My point is instead that making this distinction between two types of scientific representations will allow practitioners to avoid this semantic puzzle. The kind of philosophical intervention that I propose is to regiment our use of these terms to avoid puzzlement and achieve a more stable, self-reflective picture of the fruits of scientific
investigation.

If theories and models are two types of scientific representation, then it is an urgent matter to clarify what they are representations of. Following Bogen and Woodward, I propose that theories and models are about phenomena, and that these phenomena should be clearly distinguished from the data that provide evidence for the existence and character of phenomena. In section 2 I outline what I take to be the most important aspects of phenomena. As with Bogen and Woodward, I emphasize the independence of phenomena from scientific investigation. Although some phenomena only occur in scientific laboratories, the character of phenomena is taken to be mind-independent and language-independent (except for those phenomena that involve minds or languages). For this reason, phenomena are the preferred target of the scientific realist. But unlike some scientific realists, I do not suppose that theories (or models) are apt to provide a complete and accurate representation of the phenomena in some domain.

In section 3 I propose that scientific theories be identified with a relatively small collection of claims. The statements of a single theory are about some type of phenomena. Some theories will be aimed at just one phenomenon, but most theories aspire to represent a wider range of phenomena that are thought to be naturally grouped together. The primary constraint that I impose on theories is that a theory is something that a scientist may believe. When a theory is believed, the scientist is taking on a commitment to the truth of each of the claims of the theory. This constraint requires that theoretical claims be fully interpreted propositions that exclude any known falsehoods or idealizations. So, as with the approach to phenomena, this conception of theory is well-suited to articulate a form of scientific realism: theories that are believed are thought to provide accurate, though incomplete, representations of their phenomena.
This regimentation of “phenomena” and “theory” serves to localize one urgent semantic problem for scientific investigation. Scientific practice appears to be permeated with idealizations. These are statements that are false, and believed to be false, when taken to be about the phenomena under investigation. This falsity creates a semantic puzzle: if the statements are false, then how do they contribute to any genuine knowledge of the phenomena? Some respond that these statements are not false as they are actually true of something besides the phenomena. Others posit the ultimate dispensability of idealization from scientific activity. On any of these options, a modest form of scientific realism that aspires to the accurate representation of phenomena seems fanciful.

My recommendation is to tie idealizations to the specification of what I will call a scientific model. A scientific model is a structure that is provided with an interpretation. As I will describe in more detail in section 4, this structure may be abstract or concrete. An interpretation serves to link the model’s structure to some phenomenon (or class of phenomena). A model is thus not a collection of claims, but a structure and an interpretation. This combination generates a collection of claims, some of which may be true of the phenomenon in question, while other claims in the collection are false. The idealizations that generate semantic puzzles pertain to the interpretation of the model’s structure. This role is consistent with accurate theoretical claims also being used to interpret the model’s structure. In certain cases, the empirical success of an adequately interpreted model can be used to support the claims of some theory. This support can accrue even though the idealization is unavoidable. A modest form of scientific realism thus remains and the semantic puzzle tied to ineliminable idealizations is defused.
2 Characterizing Phenomena

Many semantic questions about theories take for granted a semantic distinction between observable and theoretical terms and their intended corresponding properties. It is supposed to be easy to refer to colors, but hard to refer to forces or neutrinos. As Bogen and Woodward emphasize in their groundbreaking 1988 paper “Saving the Phenomena”, this supposition imposes a decided distortion on philosophical discussions of scientific investigations. They note that when philosophers restrict themselves to observable phenomena they are likely to take these phenomena to be the primary objects of theorizing, especially as the targets of prediction, testing and explanation. By contrast, they urge that the primary objects of scientific theories are phenomena quite generally:

For our purposes, what matters most about phenomena is the distinctive role they play in connection with explanation and prediction, the general features they possess which suit them to this role, and the way in which they contrast in these respects with data (Bogen and Woodward 1988, 321-322).

Data are collected in experiments or other forms of observation such as instrument readings. While evidence for the existence of phenomena is tied to observable data, there is no requirement that the phenomena themselves be observable. I follow Bogen and Woodward in claiming that the first step away from semantic puzzlement is to get clear on the nature of, and our access to, scientific phenomena.

Two of Bogen and Woodward’s original examples clarify the breadth of what is to count as a phenomenon and also the means by which we gain semantic and epistemic access to

---

1Here I draw on (Bogen and Woodward 1988), (Woodward 1989), (Bogen and Woodward 1992), (Bogen and Woodward 2003), (Bogen 2011), (Woodward 2011).
2This functional account of phenomena is emphasized by (Apel 2011).
them. One phenomenon is the melting point of lead, while another phenomenon is the rate of neutrino emission from the Sun (Bogen and Woodward 1988, 308, 316). In both cases, what convinces scientists that the phenomenon is genuine is the careful collection and analysis of experimental data. A series of controlled measurements using instruments produces a collection of data on the melting point of lead. The mean and variance of this data supports the inference that lead melts at $327.5 \pm 0.1^\circ C$ (Bogen and Woodward 1988, 309). This is a well-supported claim about the phenomenon of the melting point of lead.

As a phenomenon, it is a repeatable type with a wide variety of instances. The inference from data to phenomenon in the neutrino case is more elaborate, but of essentially the same character for Bogen and Woodward. A certain fluid was buried, periodically flushed and then examined using a Geiger counter for the presence of a radioactive isotope of argon. This process generated data that were used to infer the rate of neutrino emission from the Sun (Bogen and Woodward 1988, 316). As with the melting point of lead, the emission of neutrinos from the Sun has a generic, repeatable character. Among other things, it is an open question whether or not the rate of neutrinos detected on Earth reflects all of the neutrinos emitted from the Sun. Contrary claims about the phenomenon would characterize it in some further respects.

Phenomena, then, are types of naturally occurring processes, events or states. Based on appropriately collected data, scientists are warranted in making claims about phenomena ranging from the existence of the phenomena through to their quantitative features, such as temperatures or rates at which they occur. A semantic puzzle about these characterizations of phenomena is likely to arise when some parameter is invoked that is not definable in terms of patterns manifest in the data itself. Feest provides a helpful overview of these sorts of claims and how they are to be construed (Feest 2011). She contrasts sur-
face phenomena with hidden phenomena. Surface phenomena are potentially found or instantiated in the data that results from an experiment. For our purposes, what is significant about this instantiation is that the very parameters used to describe the data are also used to characterize the phenomenon. In the melting point of lead, the data consist in a series of temperature measurements. The phenomenon of the melting point of lead is a surface phenomenon because the core claim in question is also a temperature claim. This suggests that there is no serious question about how scientists can make this claim about this phenomena once they are in a position to collect the data. The measurement procedure in this case helps us to see how the semantic content of the claim about the surface phenomena is acquired.

This point is not meant to underplay the presuppositions of the judgment that some data instantiate some surface phenomenon. As Feest emphasizes, this is a “conceptual skill” or ability that will be analyzed quite differently in different cases (Feest 2011, 66). At the same time, the difference between surface phenomena and hidden phenomena raises additional problems. A hidden phenomenon is not apt to be instantiated by the data. In the lead case, for example, we suppose an additional phenomenon tied to atomic structure that is responsible for the melting processes measured in the surface phenomenon. In the neutrino case, we suppose a phenomenon involving the fusion reactions in the Sun as the deeper, hidden basis for the Geiger counter detections here on Earth. The rate of neutrino emission is a hidden phenomenon because it is characterized in terms, like “neutrino”, that are not apt to describe the data or the surface phenomenon that the data instantiate.

How, then, do scientists acquire the ability to characterize phenomena in these novel terms? Feest’s proposal is that a relationship of “stabilization” can obtain between some surface phenomenon and an associated hidden phenomenon:
data can only be treated as indicating a hidden phenomenon insofar as they are also viewed as instantiating a surface phenomena . . . scientists will not be inclined to treat an individual data point as indicative of a hidden phenomenon as long as they do not think that the data point is at least in principle replicable, and can therefore be treated as instantiating a data pattern or surface regularity (Feest 2011, 65-66).

The features of the atomic structure of lead that are tied to the melting point are indicated by the character of the stable melting point of the surface phenomena. The more elaborate process of neutrino emission from the Sun is pinned down by the repeatable pattern of isotope detections here on Earth. When the supposed relation between the surface and hidden phenomena is thus appropriately stabilized, the scientist is able to refer to the most esoteric quantities or entities.³

Much more could of course be said about phenomena and their characterization. The core insight that I adopt from Bogen and Woodward is that appropriately collected and analyzed data can serve to license claims about phenomena, even when those claims involve unobservable entities like neutrinos. Following Feest, I have emphasized the connections between surface phenomena and hidden phenomena. In an extended sense, the features of surface phenomena can serve to characterize or “measure” the features of the hidden phenomena.⁴ This provides a semantic basis for the clarification of the differing representational aims of theories and models.

³I suppose here that Bogen and Woodward’s phenomena are the genuine objects of scientific investigation, as opposed to other candidates such as Weisberg’s “target systems” (Weisberg 2013, 90-91) or van Fraassen’s “surface models” (van Fraassen 2008, ch. 7, 11). Both options tie the objects too closely to the activities of scientists.

⁴Harper emphasizes this aspect of Newton. See (Harper 2016) and (Falkenburg 2011).
3 Theoretical Claims

In Bogen and Woodward’s framework, phenomena are distinguished as the primary objects of scientific prediction and explanation. In their 1988 discussion, they are also clear that scientific theories are the things that do the predicting and explaining: “well-developed scientific theories do predict and explain facts about phenomena” (Bogen and Woodward 1988, 306). However, there is little discussion here about what theories are or how they are able to generate predictions and explanations. We are provided with a necessary condition on a certain kind of “systematic” explanation of features of phenomena: “a satisfactory systematic explanation must show how features of the explanandum-phenomena systematically depend upon the factors invoked in the explanans of that explanation” (Bogen and Woodward 1988, 323). But Bogen and Woodward admit that when scientists actually provide these explanations of phenomena, they are “frequently” forced to resort to “idealizations, approximation, and simplifications of various kinds” (Bogen and Woodward 1988, 324). No guidance is offered, though, concerning the origins of these non-theoretical elements or how they might be rendered consistent with a theoretical explanation of some phenomenon.5

In a footnote to their point about idealization and approximation Bogen and Woodward note the work of Nancy Cartwright and Ronald Laymon. Since the 1980s Cartwright, in particular, has pioneered an increasingly popular approach to scientific theories that argues that theories are not able to provide predictions and explanations without significant supplementation by models.6 This “mediating models” approach insists that scientific

5Of course Woodward has provided extensive discussions of causal explanation elsewhere (Woodward 2003). Still, it is only recently that he has explicitly said much about how idealizations affect causal explanation. See (Woodward 2016), (Woodward 2018).

6Laymon’s work is notable in trying to reconcile idealization with some form of scientific realism
models have been ignored by philosophers of science. Once the proper role for models is acknowledged, central questions like the nature of prediction, explanation, evidence and scientific change will be transformed. Cartwright initially took the consideration of models to undermine scientific realism, but in later work she has emphasized the rejection of laws of universal scope. Given the significance ascribed to models, it is not surprising that philosophers influenced by Cartwright have not reached much consensus on the nature of scientific models or their significance for philosophical debates. In this section and the next it will become clear that I am largely sympathetic to the core claims of the mediating-models tradition. However, I will argue that the need for models is primarily epistemic, and not semantic. That is, models are not semantic mediators between theory and phenomena: theories are already about the phenomena in their domain. Models are required to generate evidence that a theory is correct.

One of the main targets of the mediating models tradition has been the so-called semantic view of theories. The semantic view claims that theories should be identified with collections of models, although the relationship between the models of the semantic view and the models of the mediating-models tradition is often far from clear. Recent work by Halvorson and others has done a great deal to clarify what is at stake in this debate, and also how the semantic view compares to its predecessor, the so-called syntactic view of theories. The syntactic view identifies theories with a deductively-closed set of sentences of some formal language such as the language of first-order predicate logic. I follow Halvorson and Lutz in maintaining that an adequate version of the syntactic view turns out to be “formally dual” to an adequate version of the semantic view. In both

(Laymon 1985). While I have been greatly influenced by his work, I cannot engage with it here.

7Here I draw on (Halvorson 2012), (Glymour 2013), (Halvorson 2013), (van Fraassen 2014), (Barrett 2015), (Lutz 2015), (Weatherall 2016), (Barrett and Halvorson 2016) and (Halvorson 2016).
cases, a formal analysis of the structure of a scientific theory involves the specification and investigation of a network of relationships between formally specified entities. The syntactic approach emphasizes the sentences of some formal language and the network to be studied is the inferential relationships between these sentences. For this task, the models of model-theory are indispensable. By contrast, the semantic approach considers the structural relationships between a family of models, where these are construed as set-theoretic entities or abstract structures like trajectories through some state space. But for these models to be appropriately identified, and for their structural relationships to be investigated, aspects of these models must be designated or labelled in linguistic terms. On either approach, then, some combination of linguistic and non-linguistic entities are invoked.

On this reconstruction, the primary divide in discussions of scientific theories is between the formal approach that unites the syntactic and semantic views and the decidedly informal account of theories offered by the mediating-models tradition. In my view, the best option for the mediating-models tradition is to identify a scientific theory with a small number of propositions. These are not to be thought of as sentences of some formal language equipped with some model-theoretic interpretation. Instead, they are fully interpreted statements that have determinate truth-conditions. I will call this view of theories the propositional view. It has a number of explicit defenders and also seems implicit in many discussions of scientific theories. For example, Bogen and Woodward maintain that theories predict and explain. I suppose that both prediction and explanation require propositions. A prediction is a definite claim about what will happen in the future. An explanation is a definite claim about why something has happened. In addition, it seems

---

8See esp. (Morrison 2016), which is discussed below.
clear that some theories are believed. A belief is a relationship between a person and a proposition. So, if theories are believed, then theories are propositions or, by slight extension of usage, small collections of propositions.

In his recent article “Scientific Theories” Halvorson rejects this belief argument, at least as a criticism of the semantic view: “a classic criticism of the semantic view is that whereas one can believe a theory, one cannot believe a collection of models. But semanticists have been very clear that they see belief as involving the postulation of some notion of similarity or resemblance between one of the models and the intended domain of study” (Halvorson 2016, 597). If a viable syntactic approach is formally dual to a viable semantic approach, then the same strategy is available for a syntactic approach. One will tie belief in the theory to belief in some appropriate relationship between elements of the theory and its intended domain. Perhaps this is what Halvorson has in mind when he says earlier that “we know what it means to believe a collection of sentences” (Halvorson 2016, 595). Presumably for a sentence to be believed is to believe the proposition expressed by that sentence under some appropriate interpretation. Using his updated version of the syntactic view, Halvorson must suppose that these propositions can be identified once the model-theoretic interpretations of his formal sentences are provided. The upshot is that neither the syntactic nor the semantic view has a problem for belief in theory. By extension, neither approach has problems with prediction or explanation. There is thus no reason to abandon these formal approaches to scientific theories in favor of the propositional view of theories.

My response is that considerations of generality advise in favor of the propositional view of theories and against any formal account of the sort discussed by Halvorson. The limited scope of formal approaches is helpfully emphasized at the conclusion of Halvorson’s
If a formal approach to scientific theories has utility, it has a limited sort of utility—precisely the same way that mathematical physics has a limited sort of utility. Mathematical physics represents a sort of limiting case of scientific inquiry, where it is hoped that pure mathematical reasoning can provide insight into the workings of nature. In the same way, formal approaches to scientific theories might be considered as limiting cases of philosophy of science, where it is hoped that pure mathematical reasoning can provide insight into our implicit ideals of scientific reasoning, the relations between theories and the like (Halvorson 2016, 605).

For any scientific theory, we can subject it to some formal analysis. Analogously, for any concrete system, we can study it using the tools of mathematical physics. The appropriateness and fruitfulness of either investigative strategy is partly determined by our aims as well as the character of the theory or system. Mathematical physics is not able to help me to determine when I should water my lawn. In a similar way, a formal approach to scientific theories often does little to illuminate questions about those theories, especially when the issue is to evaluate evidence and determine the rationality of one’s beliefs. By contrast, a propositional view of theories has a much more general application. Every theory that can be analyzed in formal terms can also be identified with a collection of propositions, but there are many collections of propositions that resist analysis into the forms prized by the syntactic or semantic views. As I see it, the central semantic question for theories is how they get to be about phenomena and the primary epistemic question for theories is when evidence indicates that we have got some theory right. I am happy to admit that a
structural or formal analysis of theories is often fruitful, and that it can contribute to the investigation of scientific theories. But when it comes to non-formal “semantic” questions about the rational belief in theories, these analyses seem less helpful.

The propositional view of theories has been recently defended by Morrison in her article “Models and Theories” (Morrison 2016).\(^9\) Morrison, along with Cartwright, is one of the most prominent developers of the mediating-models tradition that emphasizes the independence of models from theories. One of Morrison’s arguments ties in directly to our belief argument:

If we think it is the job of theories to tell us what the world is possibly like, then we need some way of differentiating what the theory is about (i.e., its content) from the various assumptions required for its application in particular contexts. One way to remedy this problem is to differentiate models and theories based on different notions of representation and explanation appropriate to each. My own view is that part of that differentiation will involve the notion of a theoretical core – a set of a fundamental assumptions that constitute the basic content of the theory, as in the case of Newton’s three laws and universal gravitation (Morrison 2016, 380).

The semantic view must build the assumptions used to apply the theory into the theory itself. Morrison instead isolates the “core” of the theory as what is found in each of the theory’s applications. In this way, she is able to provide a compelling account of what it is that scientist’s believe when they believe a theory. They believe each of the propositions that make up the “theoretical core”. In my view, we should identify the theory with

\(^9\)See also (Morrison 2007).
Morrison’s theoretical core.

Morrison’s proposal should be distinguished from the claim that talk of theories should be eliminated in favor of talk of propositions. Vickers has developed this form of eliminativism for theories, largely in connection with his own work on allegedly inconsistent scientific theories (Vickers 2013, Vickers 2014). Vickers’ argument comes in two parts. First, it is difficult to identify the propositions that make up the theory, and any attempt to do so generates fruitless philosophical controversies: “Why, exactly, are the selected analysanda important to consider together, as a set, in the given context?” (Vickers 2014, 119). Second, any purpose served by talk of theories can be fulfilled by talk of propositions. For example, Vickers argues that the debate about the consistency of classical electrodynamics turns on the propositions that are included in this theory. One side includes “the equations that are used by scientists” while the other side restricts the theory to “the equations that are/were believed by scientists” (Vickers 2014, 122). Rather than puzzle over which set is identified with the real theory, Vickers urges us to clarify the set at issue and consider its consistency. It does not matter which set is “the” theory.

As we have seen, Morrison’s theoretical core is the set of propositions that we can use to determine what is believed when the theory is believed. It is the set of propositions that is implicated in every application or use of the theory. This might seem to provide a principled reason to side with one account of what a theory is, but I suspect that Vickers would argue that this route really marks no genuine progress. Suppose we identify the theory of universal gravitation with the set of propositions that are involved in each application of the theory of universal gravitation. How are we meant to determine which

---

10See also (French and Vickers 2011), although it is not clear how the proposal in this paper is connected to Vickers’ other work.
activities count as an application of that very theory? To identify the set of propositions, it seems we need some independent access to the content of the theory. There is thus a circular aspect to Morrison’s approach that will leave a skeptic about theories unsatisfied. My response to this problem is to draw on the phenomena from section 2. Morrison claims that “one of the roles of theory ... is to provide a general representation of an entire class of phenomena” (Morrison 2016, 392), but does not say much about what a phenomenon is or what might make some phenomena into a single class. I have drawn on Bogen and Woodward to clarify what a single phenomenon amounts to. In certain cases, scientists will believe that some phenomena are appropriately grouped together due to their natural connections. This commits the scientists to the existence of a theory that covers this class of phenomena in the sense that the theory is able to make predictions and explanations about each of the members of that class. We can break out of Morrison’s circle, then, by supposing that a class of phenomena is selected as the target for some single theory. The theory is then identified with the set of propositions that are deployed in every attempt to predict or explain a claim about these phenomena.

This identification does not prove that a theory really is this set of propositions, and Vickers’ is certainly right that a quest to distill the essence of theory is somewhat pointless. Still, for the purpose we have identified, it is possible and fruitful to identify a theory with this set of propositions. Analysis of particular cases would be needed to determine which propositions count as Newton’s theory of universal gravitation or Maxwell’s theory of electromagnetism. For these cases, I am optimistic that a close tie to a class of phenomena would be revealed, and that it is the connections between the theory and this class that

\[11\] There is no requirement that the theory be able to predict or explain everything about the members of the class.
is central to the activities of the scientists who developed and refined the theory.

One challenge to identifying a theory with a set of propositions is the status of propositions themselves. For example, Halvorson claims that it is a mistake to identify theories with propositions because “we have no direct access to propositions – we only have access to sentences that express these propositions. Thus, we only know how to talk about sets of propositions by using sets of sentences” (Halvorson 2016, 604). If by “we” here Halvorson means human beings, then I do not agree: humans may grasp fully determinate statements about the world independently of the grasp of sentences. Propositions are of course not observable entities or “data” in the sense emphasized by Bogen and Woodward. But just as we have seen the value of invoking unobservable phenomena, there is every reason to posit unmediated cognitive access to propositions. Of course, by “we” Halvorson may mean those that adopt a formal approach to scientific theories. The only formal way to analyze propositions is as sets of sentences under some appropriate interpretation. Again, I am not denying the value of this sort of analysis for certain purposes. But advocates of the propositional view like Morrison are obviously not restricting themselves to formal analyses, and it is no argument against the propositional view to point out that formal analyses cannot take propositions for granted.

Another challenge to the propositional view of theories is central to Cartwright’s account of how theories are applied via mediating models. For Cartwright, the elements

---

12Halvorson here is actually objecting to Vickers’ proposal that we eliminate talk of theories in favor of talk of propositions, but his worry would naturally extend to my proposal as well.

13Halvorson continues: “why should we think that when a scientist puts forward a theory, she’s only putting forward a set of propositions? Why not think that she means to articulate a structured set of propositions (i.e., a set of propositions with inferential relations between them)?” (Halvorson 2016, 604). I am supposing that a proposition is a fully determinate statement, and so the inferential relations between propositions are given along with that set. I am thus in agreement with Halvorson on the importance of what he calls a “structured” view of theories.

14Here I draw on (Cartwright, Suárez, and Shomar 1995), (Cartwright 1999), (Suárez and Cartwright
of a theory are not about phenomena because they are not determinate statements. This means that we need to first formulate a “model” that will be tailored to represent some phenomenon. In some presentations Cartwright emphasizes the abstract character of the claims and concepts found in theories:

the fundamental principles of theories in physics do not represent what happens; rather, the theory gives purely abstract relations between abstract concepts. For the most part, it tells us the capacities or natures of systems that fall under the concepts. ... no specific behaviour is fixed until those systems are located in very specific kinds of situations. When we want to represent what happens in these situations we will need to go beyond theory and build a model (Cartwright 1999, 180).

If theories include only abstract statements, where these statements lack determinate content, then theories are not able to generate predictions or explanations in isolation. Models play the role of a semantic mediator between the indeterminate claims of theory and the concrete target phenomena.

We have followed Morrison in identifying Newton’s theory of universal gravitation with his three laws of motion and the law of universal gravitation. These statements deploy concepts like space, time, mass and force. In addition, the law of universal gravitation invokes a specific force, the force of gravitation. It is hard to understand what Cartwright means when she claims that such concepts are abstract. One account of abstractness that she gives is that “a concept that is abstract relative to another more concrete set of descriptions never applies unless one of the more concrete descriptions also applies” (2008). For additional discussion of Cartwright on theories and models, see (Bailer-Jones 2008).
This is reminiscent of the determinable/determinate distinction: for an object to be red, it must also be a more specific shade of red such as scarlet. So if we claim that an object is red, we must also suppose that it is some more specific shade of red. This notion of abstractness fits well with some of the aspects of Newton’s theory. Newton’s three laws of motion, for example, characterize changes of states of motion in abstract terms. For any acceleration, there must be a net force applied. If we claim that some particle has undergone an acceleration, then this requires that it has undergone some determinate acceleration that is tied to some determinate net force.

This notion of abstractness does not license Cartwright’s conclusion that statements that deploy abstract concepts “do not represent what happens”. If I claim that an object is red, then I have made a definite claim about that object even if my claim does not settle what shade of red the object has. Similarly, if I claim that a system obeys Newton’s three laws of motion, then I have made a definite claim about the object even if my claim does not settle how the system will change over time. What Cartwright needs is a notion of abstractness that will render the semantic verdict she requires. One such notion is that abstract concepts are “concepts that need fitting out in more concrete form” in order for claims made using the abstract concept to be truly about some more concrete system (Cartwright 1999, 180). But this notion of abstractness makes it very implausible that Newton’s concepts are abstract. Newton and his contemporaries took his theory to be about a wide range of phenomena, from the orbits of the planets around the Sun to the tides.15 Newton’s theory needed “fitting out” only in the sense that the claims of the theory are very general (“universal”) and hence not able to make specific predictions or explanations by themselves. But the application of Newton’s theory via models is then

15See, e.g., (Smith 2016), 218-219.
naturally analyzed in terms of the addition of other, non-theoretical propositions. This is how I will make sense of the relationship between theory and model in the next section. Against Cartwright, there is no semantic indeterminacy to the claims of theories. They are ordinary propositions that are thus apt to be believed.

There is a significant and difficult series of questions concerning the semantic content associated with the concepts of a theory. What, after all, is required of the world for Newton’s law of universal gravitation to be correct? It is not entirely clear what the force of gravity was meant to be or how we are supposed to acquire that concept. Similarly, it is not clear what concepts are invoked in the phenomena of section 2, such as temperature and neutrino. As I have developed the picture so far, the primary semantic puzzles are resolved by attending to phenomena, especially the relationship between Feest’s surface and hidden phenomena. These activities settle what magnitudes or parameters are in question for a given phenomenon as well as their semantic character. When phenomena are united into a natural class by scientists, a collection of scientific concepts is generated that may be used to make claims about these phenomena. Some of these claims are united into theories. On this regimentation, then, theories are claims about phenomena, and the contents of those claims are clarified by appeal to the procedures used to characterize those phenomena in the first place. However, to fully appreciate this process, we must turn to models and the distinct way that they relate to phenomena.

4 Idealized Models

If a theory for some phenomenon is a small collection of general claims, then it is clear why an additional element is needed for prediction and explanation. Newton’s three laws of
motion and the universal law of gravitation do not entail any testable claims about some specific phenomenon like the orbit of the Moon around the Earth. So this theory must be supplemented. In line with the mediating models tradition, I claim that models serve this function. In Bailer-Jones’ compact summary, “In models, theories are customized with respect to a specific phenomenon” (Bailer-Jones 2009, 19). This section summarizes one way that this customization can occur. My justification for this proposal is that conceiving of models in this way helps us to address the semantic puzzle of idealization.

Suárez’s recent survey “Representation in Science” emphasizes a contrast between description and representation: “a critical difference between description and representation concerns the applicability of semantic notions such as truth, which are built into descriptions but seem prima facie ill-suited for representation” (Suárez 2016, 440).16 In these terms, the propositional view of theories defended in the last section takes theories to be descriptions of phenomena. But models are not descriptions as they are not to be identified with collections of propositions. Models are about phenomena, but relate to phenomena in a different way than how theories relate to phenomena. My proposal is that a model should be identified with a structure that has been provided with an interpretation.17 This combination will then generate a large collection of claims about phenomena.18

Our focus is on how idealized models can be used to support theoretical claims. For models of this type, the role of the model is to link a theory to some phenomenon so as to allow the theory to predict or explain some aspect of that phenomenon. To regiment

---

16 Suárez credits (Giere 1988) with this contrast.
17 Although different authors use different terms to distinguish these aspects of models, I take my proposal to be very much in line with (Weisberg 2013), (Gelfert 2016) and (Frigg and Nguyen 2017). I have also drawn on (Bailer-Jones 2009). The account of idealized models that I develop is somewhat close to (Mäki 2010), (Mäki 2012), although a detailed comparison is not feasible here.
18 Cf. (Thomson-Jones 2012).
this process, I will distinguish the model’s predictions and explanations from the theory’s predictions and models. Theoretical claims are often central to models, but there seem to be principled barriers to treating all of the model’s predictions and explanations as properly theoretical. These barriers are due to our semantic puzzle: scientists present claims that they know to be false in the course of discussing these models. These idealizations often undermine the link that is needed for a successful prediction or explanation to provide evidence for the truth of the claims of the theory. For if idealizations are central to the prediction or explanation, then how can such a success indicate that the theory is true? This challenge is devastating to an approach to theories that treats their application in terms of the addition of so-called “auxiliary hypotheses”. The falsity of such non-theoretical claims blocks any theoretical success from providing evidence of truth. Philosophers wedded to this way of thinking of the application of theory are thus correct in drawing some non-realist conclusions.19 But conversely this problem motivates the scientific realist to conceive of the application of theory in different terms.

If theories are applied using models, and a model is a combination of a structure and an interpretation, then the challenge of idealizations can be fruitfully addressed. The key move is to take an idealization as a means of specifying a model’s interpretation. The idealization looks to be a claim about either the model or the target phenomenon. But the idealization should instead be viewed as a kind of pseudo-statement. This account of idealized models thus emphasizes the non-representational, expressivist elements of idealization: in offering an idealized model, a scientist is committing herself to the appropriateness of a certain kind of intellectual demand. The commitment is that investigating the phenomenon with a model with this interpretation will aid scientists in finding out

19See e.g. (Duhem 1954), (Elgin 2004), (Elgin 2009).
The first aspect of a model to clarify is the model’s structure. A model’s structure is some entity that has internal articulation. It may be a concrete object or an abstract object. For example, an appropriately shaped piece of wood may be the structure for a concrete, scale-model of a ship. Or a purely abstract mathematical structure may be the structure of some model. The structure of a model may be physically constructed or picked out by some other means, such as via the solution to some system of differential equations. A model’s structure is central to the resulting use of the model to represent some definite phenomena. Typically, many of the features of the model’s structure will be unknown when the model is specified. Investigating the model then largely turns on acquiring new knowledge about the model’s structure.

Clearly a model structure, in isolation from any interpretation, is not yet a representation of some phenomenon: it is not about that phenomenon in any sense. In addition, a model structure is not intrinsically the model of some theory. The interpretation of the model structure thus serves this dual semantic purpose. First, the interpretation renders the model about the phenomenon. Second, when the interpretation is carried out using the claims that comprise the theory, then the resulting model is a model of the theory. It seems essential to distinguish at least two parts of the interpretation. The first part assigns denotations to elements of the model’s structure. The second part clarifies how those elements, with those denotations, are supposed to relate to the phenomenon. The elements of the model are assigned denotations using the parameters or magnitudes that we have already encountered in our discussion of phenomena and theories. In the Newtonian case, the interpretation must make clear what elements of the model’s mathematical structure stand for spatial and temporal quantities. In addition, some elements will stand
for masses and others will stand for forces. In this way, a model structure is labelled or made to stand for a series of magnitudes. But as a number of commentators have emphasized, this sort of interpretation is not yet a full interpretation of the model structure. Fixing the denotations of the elements of the model’s structure, a scientist can vary how that model is meant to relate to the phenomenon. One dimension of variation is the strictness of the relation. A very strict interpretation might require a complete “isomorphism” between the model’s structure and the structure of the phenomenon. In a Newtonian case, for example, a model may be intended to track the exact positions of each particle at each time. This strictness can of course be relaxed, so that the model is only meant to track the positions of the particles within some small boundaries. The interpretation may be “coarse-grained” in this sense. Another dimension of variation is the scope of the relation between the model’s structure and the phenomenon. The scientist may aim to track the long-term or the short-term behavior of the particles found in the phenomenon. Or she may aim to track the small-scale or the large-scale behavior of bodies such as the Earth and the Moon. An adequately interpreted model will make clear how the model is meant to relate to the phenomenon in question, given the denotations assigned to the elements of the model. The crucial indication that a model has been fully interpreted is that it generates a collection of claims about the phenomenon. The members of this collection may not be obvious once the interpretation is given. For concrete models, an empirical investigation of the concrete structure is often needed to figure out what claims are generated by the model. Similarly, in the abstract case, a mathematical analysis may be needed to determine which claims are generated by the model. Some of these claims will amount to potential predictions or explanations. I call these “model predictions” or “model explanations”.

23
The interpretation of a model thus requires a series of statements. Some of these statements are about the model. The denotations of the elements of the model’s structure are assigned by making stipulations about the model: this element stands for $X$. Here a scientist is able to draw on the semantic resources accrued through the investigation of phenomena. For example, the scientist is able to say that some coordinate stands for mass or temperature because they are able to refer to masses and temperatures. Some of the statements used to provide an interpretation of the model may come from a theory. In our main example so far, Newton’s theory of universal gravitation is composed of Newton’s three laws and the law of universal gravitation. If all of the claims of a theory are used to interpret a model’s structure, then the resulting model is a model of the theory. However, for reasons that are now familiar, a model will not be fully interpreted using just the claims that make up the theory. On our way of thinking about the interpretation of the model, this actually turns out to be a trivial claim. For even when we use all the claims from some theory to interpret a model, it still remains to specify the intended relation between the model and the phenomenon. Consider again an attempt to model the orbit of the Moon around the Earth. A scientist may pick out some abstract mathematical structure as the solution to some set of differential equations, and may also interpret this structure using the claims of Newton’s theory. But this still does not settle the strictness of the relation between the model and that phenomenon or the intended spatial or temporal scope of that model-phenomenon relation. There is thus a need for additional, non-theoretical claims in order for the model to be adequately interpreted.

It is precisely at this stage that idealizations often take a central place in modeling. To make the problem concrete, I will bring in a well-known example of mathematical
modeling: Kelvin’s 1863 model of the age of the Earth. Kelvin deployed Fourier’s theory of heat to conclude that the Earth was around 98 million years old. This estimate proved to be incorrect, but it can still serve as a useful instance of a prediction. As we will see, an idealized model plays an essential role in this prediction. In line with our discussion of Newton, we can identify Fourier’s theory with his celebrated heat equation:

\[
\frac{\partial}{\partial t} u = \alpha^2 \frac{\partial^2}{\partial y^2} u, \quad y > 0, \quad t > 0.
\]  (1)

This equation is a definite claim about how the temperature \( u(y, t) \) at point \( y \) at time \( t \) changes over time. The right-hand side of the equation relates these changes to the spatial distribution of temperatures around \( y \) at time \( t \) and the parameter \( \alpha^2 \), known as the rate of heat conduction. Famously, Fourier did not take a stand on the nature of heat, but claimed only that this parameter reflects how readily heat is transferred through a given material. In this way, he sought to accurately describe the variations in the temperatures of points over time. Fourier and Kelvin took Fourier’s theory to be about various phenomena, including the way that temperatures increased as one moved from the surface of the Earth towards the center of the Earth. But if the theory is just this interpreted equation, there is no way to generate any predictions or explanations using the theory by itself.

The structure of Kelvin’s model is an abstract sphere, where each point on the sphere is labelled with three numerical coordinates. Kelvin assigned the natural denotations to these elements of his structure in terms of distance, time and temperature (with units of feet, years, and degrees Fahrenheit, respectively). In addition, he interpreted the model using a number of non-theoretical claims. First, he offered a boundary condition for what

\(^{20}\)I follow (Tung 2007), ch. 13, but the mathematics is the same in Kelvin’s original paper (Thomson 1890).
occurred at the surface of the Earth:

\[ u(0, t) = u_s, \ t > 0. \] (2)

That is, for all times, the temperature at the surface \((y = 0)\) has some constant value \(u_s\).

Second, Kelvin offered the initial condition

\[ u(y, 0) = u_0, \ y > 0. \] (3)

This means that when the Earth was formed \((t = 0)\), for all points within the Earth (for all positive \(y\)), the temperature \(u\) was some value \(u_0\).

These steps completed the purely mathematical side of Kelvin’s model: for every choice of the three parameters \((u_0, u_s, \alpha^2)\), the system of equations (1)-(3) fixes a unique temperature profile near the surface of the Earth as a function of \(t\). To show this, Kelvin indicated how (1)-(3) required

\[ \frac{\partial}{\partial y}u(y, t) = \frac{(u_0 - u_s)}{\sqrt{\pi \alpha^2 t}} e^{-\frac{y^2}{4\alpha^2 t}} \] (4)

The left-hand side of this equation was then treated as a constant, and the exponential term was replaced by 1, yielding

\[ \Delta = \frac{(u_0 - u_s)}{\sqrt{\pi \alpha^2 t}} \] (5)

By measuring \(\Delta\), Kelvin was able to calculate \(t\) as a function of \((u_0, u_s, \alpha^2)\). This led to his estimate of 98 million years.\(^\text{21}\)

\(^{21}\)Kelvin does not make this precise estimate, but allows for a range of 20 to 400 million years (Thomson 1890, §11).
The genius of Kelvin’s model is that it used Fourier’s heat equation to relate a measurable quantity to a practically unmeasurable quantity. By fixing the rate at which temperature increased with depth near the surface of the Earth, one could determine the otherwise inaccessible age of the Earth. Clearly, though, the interpretation of the model is permeated with idealizations. Some of the most obvious idealizations are (i) (2)’s treatment of the surface temperature of the Earth as a constant, (ii) (3)’s claim that the initial temperature throughout the Earth was some single temperature, (iii) the model’s structure of a sphere as opposed to the actual shape of the Earth and (iv) considering only the one-dimensional heat equation for the dimension of depth as against the three-dimensional character of the Earth. I will focus, though, on a crucial or essential idealization having to do with the rate of heat conduction. The core theoretical equation (1) makes \( \alpha^2 \) a constant, and Kelvin’s model incorporates this claim. But there is every reason to think that the material making up the Earth is not homogeneous in this way. In fact, in the very paper where Kelvin provides his estimate for the age of the Earth, he also claims that the Earth’s interior is heterogeneous.

Our approach to models and their interpretation provides a principled way to handle these idealizations. On the one hand, we have statements that are used to interpret the model’s structure. On the other hand, we have statements that are generated once the model is fully interpreted. Let us suppose that all idealizations are statements that are used to interpret the model’s structure but that they are not included in the statements generated by the fully interpreted model. We have already encountered non-idealizations that exhibit these features. For example, when Kelvin specifies which element of the model structure denotes time in years, he is making a statement about how the model should be interpreted. This specification is not a claim about the age of Earth. But the
specification is an essential step in generating genuine claims about the age of the Earth, including Kelvin’s prediction that the Earth is 98 million years old. My suggestion is that we take the same attitude towards the statement expressed by “$\alpha^2 = 400 \text{ ft}^2/\text{year}$”. This is a statement that is used to specify the model. It is an essential ingredient to the specification of the fully interpreted model. But it is not included in the collection of statements that are generated by the fully interpreted model.

Two difficult questions thus arise for this approach to models. First, how are we supposed to figure out which statements that are used to specify the model fail to be included in the claims generated by the model? Second, what is the function of statements like Kelvin’s statement about the constant rate of heat conduction if it does not appear in the collection of claims generated by the model? I do not think there are any easy, universal answers to these types of questions. We must attend to specific models and the contexts in which the models are presented in order to figure out what is going on. For the cases we are focused on, models are being used to generate predictions or explanations that are not possible using the theory alone. The problematic statements are idealizations, and I have taken these to be statements that are deployed in the practice of modeling, even though they are believed to be false of the phenomenon under investigation. Kelvin says “$\alpha^2 = 400 \text{ ft}^2/\text{year}$” even though he believes that the Earth’s interior is heterogeneous. So for these cases at least there is every reason to block these claims from membership in the claims generated by the model. A predictive model is used to generate claims that stand as predictions, and the users of the model do not intend their idealizations to be included among their predictions. Similarly, when a model of a theory is being used to generate explanations, the users of the model do not intend their idealizations
to be included among their potential explanations. So, when an idealization is used to specify a model, we should grant it a role in interpreting the model, but keep it out of the collection of claims generated by the model.

What, then, is the role of a claim like \( \alpha^2 = 400 \text{ ft}^2/\text{year} \) in interpreting a model if it is not meant as a prediction or explanation? My suggestion is that these claims serve to express the model user’s commitment to a certain feature of the claims generated by the model. In Kelvin’s case, he is expressing his commitment to the accuracy of the predictions generated by his model. Like any other commitment, this commitment can be based on reasons, and so be subjected to rational scrutiny and evaluation. Kelvin’s commitment is justified to the extent that he has reasons to think that the claims generated by his idealized model are accurate. That is, he is supposing that using this false claim to interpret the model will not undermine the truth of the generated claims, especially the claim about the age of the Earth. Sometimes a model will be specified for some phenomenon that has not previously been investigated in these terms. In these cases, a commitment to the accuracy of some of the model’s predictions may be quite speculative and unwarranted. However, once a model or family of models has accrued a track-record of successful predictions, a commitment to the accuracy of novel model predictions can be quite rational. The question is always how the distortions introduced by the idealization affect the accuracy of the intended predictions. In certain cases, the predictions are unaffected even though the idealizations are considerable.

An idealization can serve to interpret a model even when that model is also specified by the claims of some theory. But when are these idealizations consistent with considering

---

22This is of course a controversial assumption about explanation, but I cannot argue for the requirement that explanations be true here. This requirement is certainly central to most accounts of explanation, but not all.
the predictions of the model to also be predictions of the theory? We can approach this
question by recalling the point of theories predicting features of phenomena. The point
is epistemic. When a theory makes an accurate prediction, and the prediction is verified,
then we have acquired evidence that the theory is correct. So we can use the possibility
of acquiring evidence as a sufficient condition for an appropriate link between a theory
and the claims generated by an idealized model. Our question thus becomes: when could
the verification of a claim generated by an idealized model of theory constitute evidence
that the theory is correct? Additionally, when would the falsification of a claim generated
by an idealized model of a theory constitute evidence that the theory is incorrect? My
suggestion is to bring in the favored mode of inference of scientific realists: inference to
the best explanation. In certain cases, the best explanation for the capacity of an idealized
model of a theory to generate verified predictions is that the claims of the theory are true.
In those cases we should count the predictions of the model as predictions of the theory
and conclude that these successes count as evidence for the correctness of the theory. In
other cases, the best explanation for the capacity of an idealized model of a theory to
generated falsified predictions is that the claims of the theory are false. Then we should
also count the predictions of the model as predictions of the theory, but now conclude that
we have evidence that the claims of the theory are false.

If we apply this framework to Kelvin’s model of the age of the Earth, we get a complex
result. Kelvin’s model counts as a model of Fourier’s theory of heat because all the claims
of Fourier’s theory were used to specify the model. Once all the parameter values are
fixed, Kelvin’s model makes the prediction that the Earth is 98 million years old. At the
time that Kelvin made this prediction, there was no way to independently verify the age
of the Earth, and so this prediction could not be used to support or undermine Fourier’s
theory of heat. However, later developments revealed that the Earth must be much older than Kelvin realized. It thus became an urgent research program to diagnose the failure of Kelvin’s model. It appears that nobody took the failure of the model’s prediction to provide evidence against Fourier’s heat equation. In our terms, then, later investigators did not take this prediction of the model to be a prediction of the theory. Interestingly enough, geologists did not simply dismiss the model based on the idealizations, but aimed to pinpoint which idealization had undermined the accuracy of the prediction. The consensus now appears to be that some combination of additional heat from radioactive materials and mantle convection was missed in Kelvin’s original model, and this accounts for the flawed prediction (Richter 1986).

This brief example is meant to illustrate how idealized models can be rendered consistent with theoretical predictions and explanations. The upshot of this approach to modeling is that idealized models may be both essential to generating theoretical predictions and explanations, and also crucial epistemic mediators between theory and phenomena. The evidence that we acquire that theories are correct (or incorrect) is mediated by models, but this does not stand in the way of a modest, piece-meal realism for theories that generate predictions and explanations in the right way.

5 Conclusion

The semantic puzzle for theories that we have considered is that idealizations occupy a central role in the application of theory. In summary, our solution to this puzzle involved three controversial regimentations of “phenomena”, “theory” and “model”. Phenomena were presented as the proper object of scientific theorizing. They are repeatable types of
processes, events or states. The evidence for the claims that characterize these phenomena are tied to observable data, even though the phenomena themselves may not be observable. Theories are collections of claims that are united by their purported ability to predict and explain features of phenomena. Theories are thus construed as collections of fully interpreted, general propositions. This created the need for a second sort of representation, a scientific model. When a theory is used to interpret a model’s structure, what results is a model of a theory. Typically, idealizations will also be used to specify the interpretation of the model’s structure. But in certain circumstances the predictions or explanations generated by the theoretical model will afford evidence for the truth or falsity of the claims of the theory. This is my suggestion for how to resolve the semantic puzzle of idealization. It requires that we conceive of models as epistemic, but not semantic, mediators between theory and phenomena.

This summary and the structure of this paper might suggest that phenomena must be characterized independently of theory and also that models arise only through theories. Neither commitment is part of my proposal. In line with Bogen and Woodward, it is entirely possible that some data may provide evidence for the existence of some phenomenon for some scientist only because that scientist believes some theory (Woodward 2011). This is clearly what happened in the case of the rate of neutrino emission. Attention to the details of such cases would reveal how the theory of particle physics was used to link the data collected to the claim about the rate of neutrino emission. Unsurprisingly, I would require that an idealized model be used to flesh out the theory and link it to the phenomena in question. There is no escape from theory and idealized models. There is only a wider process of stabilization of the sort envisaged by Feest that includes not only the links between surface and hidden phenomena, but also associated theories and models.
Although the semantic puzzle is most pressing when theories are applied via idealized models, there is no general requirement that every model be a model of a theory. Many idealized models are arrived at prior to the articulation of any theory, or before it is clear how a theoretical treatment of the phenomena in question would work. One celebrated example is Poiseuille’s law for fluid flow through pipes (Sutera and Skalak 1993). Poiseuille established this law through the careful analysis of data generated by fluid flow of various sorts through small tubes. This experimental work supported his law and a corresponding model of such flows, but it took more than twenty years for that model to be tied to a theory of fluid dynamics. The tie between Poiseuille’s model and theory in turn served to clarify various issues for the theory of fluid dynamics. In general, then, some models are developed without theories, while other models are specified in part using theoretical claims.

References


Philosophy of Science 66, 801–828.


Elgin, C. Z. (2009). Exemplification, idealization and scientific understanding. In M. Suárez (Ed.), Fictions in Science: Philosophical Essays on Modeling and Ide-
alization, pp. 77–90. Routledge.


Cambridge University Press.


36


