Where Have All the Theories Gone?*

Margaret Morrison†‡

Although the recent emphasis on models in philosophy of science has been an important development, the consequence has been a shift away from more traditional notions of theory. Because the semantic view defines theories as families of models and because much of the literature on "scientific" modeling has emphasized various degrees of independence from theory, little attention has been paid to the role that theory has in articulating scientific knowledge. This paper is the beginning of what I hope will be a redress of the imbalance. I begin with a discussion of some of the difficulties faced by various formulations of the semantic view not only with respect to their account of models but also with their definition of a theory. From there I go on to articulate reasons why a notion of theory is necessary for capturing the structure of scientific knowledge and how one might go about formulating such a notion in terms of different levels of representation and explanation. The context for my discussion is the BCS account of superconductivity, a 'theory' that was, and still is, sometimes referred to as a 'model'. BCS provides a nice focus for the discussion because it illuminates various features of the theory/model relationship that seem to require a robust notion of theory that is not easily captured by the semantic account.

1. Introduction. Partly as a result of the semantic view of theories and responses to it, a good deal of attention has been paid to the role of models in scientific practice. In fact, the semantic view of theories is, in most of its guises, not about theories at all, but about models, because the former are defined solely in terms of the latter (van Fraassen 1980, 1989; Giere 1988; Suppes 2002). What is common in all these cases is that they tend to situate the discussion of models with reference to the model-

theoretic framework developed by Tarski (1935, 1936).[1] In doing so the semanticists claim not only that they have given a clearer account of theory structure but that their view also captures or can at least be assimilated to the scientist's use of models.[2] In other words, we can have our cake and eat it too. Not only does model-theoretic semantics give us the kind of insight into theory structure that other views lack, but in doing so it brings into focus the relationship between theories and models and provides a definition of what a model is, something that is frequently missing from accounts that pay attention only to the "scientific" use of models.

A good deal of recent work (especially Cartwright, Suarez, and Shomar 1995; Morrison 1998, 1999; Cartwright 1999) has defended a "scientific" account of models focusing primarily on how they are used in scientific practice, their role as autonomous agents in the production of knowledge, and why they ought to be seen as occupying a space distinct from theory.[3] The motivation for this, at least in my own work, was not to dismiss the role of theory altogether, but rather to carve out a space for models that would enable us to better understand their construction and function.[4] However, I believe that the time has come to bring theory back into the picture and attempt a reconstruction of the relation between models and theories that emphasizes a distinct role for each. My reasons for doing so stem, in part, from the fact that what we typically call "scientific knowledge" is not completely captured by ignoring the role of theory or by reducing theories to models. Much of the criticism of the semantic view by the "scientific" modelers has centered on its inability to account for the ways models are constructed and function in scientific contexts. Although the semanticists emphasize that their view is primarily a logico-philosophical account of theory structure, they also emphasize its ability to capture scientific cases. Because of this dual role we need to begin by

1. I say 'tend' here because van Fraassen aligns himself with Beth's and Weyl's approaches rather than Tarski's, but he still defines a model in the spirit of Tarski—as any structure in which the sentences of a theory are true.

2. Many defenders of the semantic view suggest that it was never intended to capture the more practical aspects of modeling. But if we look at what those responsible for its formulation actually say, it becomes clear that capturing the scientist's notion of a model is frequently cited as one of the virtues of the approach. I have more to say about this below.

3. The so-called scientific view can be traced back to the work of Hesse (1953–54, 1970) and Kuhn (1972). Although these writers focused on how models are used in scientific practice, most of the analysis was in terms of either analogies or, in Kuhn's case, exemplars.

4. In that sense my view is different from that of Cartwright and Cartwright et al. in that they want to defend a stronger notion of autonomy for models, making them independent of theory in most respects.

first evaluating the model-theoretic features of the account to determine its success in clarifying the nature of theory structure before going on to assess its merits in dealing with the 'scientific' dimensions of modeling. The question then is whether the semantic view has actually lived up to its promises of providing an account of theory structure and models that satisfies both the logical and practical demands placed on it.

In attempting to answer this question I draw attention to the philosophical reconstruction of theories provided by the semantic view, an account that emphasizes not only the model-theoretic semantics aligned with Tarski but a general emphasis on 'structures' as well. I argue below that despite claims to the contrary, Tarski-style semantics cannot be easily mapped onto the way the semanticists define the notion of a model. Nor is the general notion of a structure particularly helpful in characterizing an account of scientific modeling. This is not a downfall in and of itself because Tarski's account was not intended to capture the way models are used in scientific practice. However, because the semantic view claims to represent the structure of scientific theories while capturing elements of the logician's notion of a model, it ends up in a kind of no-man's-land, as neither an application of Tarski's account nor a theory of scientific modeling; it in fact does a disservice to both models and theories.[5]

But there are other more practical difficulties associated with the semantic view. One such problem relates to the specification of content. Models contain a good deal of excess structure such as approximation methods and other mathematical apparatus that we don't normally include as part of a theory. Moreover, they frequently contain highly stylized descriptions of particular phenomena/properties that we know to be false, descriptions that we don't always identify as part of our theory. To that extent then, we want some way of differentiating what a theory is about (i.e., its content) from the various assumptions required for its application in particular contexts.

As a way of remedying this problem I sketch a way of differentiating models and theories that highlights the different notions of representation and explanation appropriate to each. Part of that differentiation will involve the notion of a theoretical core: a set of fundamental assumptions that constitute the basic content of the theory, as in the case of Newton's three laws and universal gravitation. This core constrains not only the behavior but the representation of objects governed by the theory as well

5. The partial structures approach of da Costa and French (2003) is an extension of Suppes's account with the addition of a structure in which only partial relations are defined on the individuals. The advantage of this is that it can purportedly reflect the fact that we may lack knowledge about the domain we wish to model. The partial structures approach also addresses other issues related to modeling that I discuss below.

as the construction of models of the theory.[6] This is not to say that theory determines the way models are constructed, but only that the way we model phenomena is typically constrained by theoretical laws or principles that are part of the larger scientific context. Once this picture is fleshed out, the hope is that it can provide the beginnings of a solution to some of the problems mentioned above. The context of my discussion is the development of the 'theory' of superconductivity. Not only does this case illustrate the tension that sometimes exists between the notion of theory and model, but it also provides a framework for proposing a possible resolution of those tensions.

2. Models, Model Theory, and Scientific Models. In order to talk about models and theories I should first say a little about how each has been characterized in the literature. The two most popular accounts of theory structure are the semantic and syntactic/received views. The former defines theories in terms of models while the later defines models in terms of theories, thereby making models otiose. The history and variations associated with these views are a long and multifaceted story involving much more technical detail than I can rehearse or do justice to here.[7] What these views do have in common, however, is the goal of defining what a theory is, a definition that speaks to, albeit in different ways, the notion of a model and its function.

On the syntactic view where the theory is an uninterpreted axiomatized calculus or system, a model is simply a set of statements that interprets the terms in the calculus. Because the model and the theory correspond in terms of deductive structure, the model and the theory can both be expressed by the same calculus; consequently, the structure of the two will be identical. In other words, a model for a theory T is simply another theory M which corresponds to the theory T in respect of deductive

6. In other words, the laws of motion and universal gravitation allow us to have a kind of physical representation of how certain objects behave in certain circumstances. The basic motions described by the laws are then reconceptualized in more concrete terms when detailed models are constructed, models that allow us to represent certain behaviors as instances of particular kinds of motion. Here I want to distance myself from the view put forward by Cartwright (1999), who also speaks of different levels of generality between theories and models. However, her account emphasizes the difference between general theoretical models such as the harmonic oscillator and more specific models that apply in well-defined contexts. On her account, models do most of the scientific work since theories are too abstract to represent anything concrete. I claim that theory does play a representational role; hence my goal is to formulate a distinction between theory and model, not simply between different kinds of models.

7. For an extended discussion, see da Costa and French (2003).

structure (Braithwaite 1953, 1962).[8] In this context 'corresponds' simply means that there is a one-to-one correlation between the propositions in *T* and those of *M*, and hence the model as an interpretation of the theory's calculus need not be true; it must only correspond in terms of deductive structure. That is, from the initial propositions of the model (which are correlated with the initial hypotheses of the theory), we must be able to deduce the rest of the model propositions. The difficulties associated with axiomatization and the identification of a theory with its linguistic formulation gave rise to the semantic view whose advocates (Suppes 1961, 1967, 2002; Giere 1988) appeal, in a more or less direct way, to the notion of model defined by Tarski. Although van Fraassen (1981) opts for the state-space approach developed by Weyl (1949) and Beth (1949), the underlying similarity is that the model (or structure) supposedly provides an interpretation of the theory's formal structure but is not itself a linguistic entity. Instead of formalizing the theory in first-order logic, one simply defines the intended class of models for a particular theory.

Suppes's version of the semantic view includes a set-theoretic axiomatization that involves defining a set-theoretical predicate, (i.e., a predicate such as 'is a classical particle system' that is definable in terms of the notions of set theory), with a model for the theory being simply an entity that satisfies the predicate. He claims that the set-theoretical model can be related to what we normally take to be a physical or scientific model by simply interpreting the primitives as referring to the objects associated with the physical model.[9] He maintains that while the notion of a physical

8. But, because scientists frequently use imaginary models, we need to distinguish their use of the term from that of the logician, where there is a restriction to a sound interpretation. The latter, Braithwaite claims, is too restrictive for philosophy of science. That said, he is critical of the less restrictive (scientists') use of models as well, claiming that the most we can expect of a model is that it yields suggestions for the extension of a theory. While Carnap (1939, 68) claimed that models contributed nothing more than aesthetic or didactic value, Braithwaite was unwilling to wholeheartedly endorse even this role. Models were used by those who found it difficult to "digest" an uninterpreted calculus and instead preferred to think first in terms of a model for the calculus. However, in some cases models were a positive hindrance to our understanding of physical theory since they had the potential to lead us down the road to falsity. Because the goal was the correct interpretation of the theory, the idea that there could be several contradictory yet equally plausible models was anathema to their program; hence, correspondence rules were seen as the only way to avoid the problem.

9. We can axiomatize classical particle mechanics in terms of the five primitive notions of a set *P* of particles; an interval *T* of real numbers corresponding to elapsed times; a position function *s* defined in the Cartesian product of the set of particles and the time interval; a mass function *m*; and a force function *F* defined on the Cartesian product of the set of particles, the time interval, and the positive integers (the latter enter as a way of naming the forces). A realization or model for these axioms would be an ordered quintuple consisting of the primitives $\mathcal{P} = \langle P, T, s, m, f \rangle$. We can interpret

model is important in physics and engineering, he is concerned with the set-theoretical usage. This is the "fundamental" one needed for an exact statement of any branch of empirical science since it illuminates not only "the exact statement of the theory" but "the exact analysis of data" as well (Suppes 2002, 24). Although he admits that the highly physical or empirically minded scientists may disagree with this, he also claims that there seems to be no point in "arguing about which use of the word model is primary or more appropriate in the physical sense" (22).

What this suggests then is that as philosophers our first concern ought to be with the exact specifications of theoretical structure rather than with thinking about how the models used by scientists are meant to deliver information about physical systems. But I take it that recent interest in models by philosophers of science is motivated by different or at least additional concerns, namely, trying to understand the ways in which models function in scientific contexts and attempting to ascertain what the relation is, in that context, between theories and models. To that end, it isn't immediately clear how this type of logico-philosophical reconstruction is going to prove helpful. I will come back to this point below.

Van Fraassen specifically distances himself from Suppes's account, claiming that he is more concerned with the relation between physical theories and the world rather than the structure of physical theory (1980, 67). To "present a theory is to specify a family of structures, its models" (64); and "any structure which satisfies the axioms of a theory . . . is called a model of that theory" (43). The models here are state-spaces with trajectories and constraints defined in the spaces. Each state-space can be given by specifying a set of variables with the constraints (laws of succession and coexistence) specifying the values of the variables and the trajectories their successive values. The state-spaces themselves are mathematical objects, but they become associated with empirical phenomena by associating a point in the state-space with a state of an empirical system.[10] Giere's (1988) approach to models, while certainly identified as 'semantic', does not specifically emphasize this nonlinguistic aspect but

---

this to be a physical model for the solar system by simply interpreting the set of particles as the set of planetary bodies, or the set of the centers of mass of the planetary bodies.

10. But once this occurs, the state-space models take on a linguistic dimension: they become models of the theory in its linguistic formulation. Similarly with Suppes's account: when it comes to specifying the set-theoretical predicate that defines the class of models for a theory, we do need to appeal to the specific language in which the theory is formulated. And, in that context, which is arguably the one in which models become paramount, they cease to become nonlinguistic entities. But as long as no specific language is given priority at the outset, we can talk about models as nonlinguistic structures.

also claims that his notion of a model, which closely resembles the scientist's notion, "overlaps nicely with the usage of the logicians" (79).[11]

As I noted above, each of the formulations of the semantic view claims, to some extent, to incorporate the notion of a model formulated by Tarski, who defines a model as follows: "A possible realization in which all valid sentences of a theory T are satisfied is called a model of T" (1953, 11). In 1961 and later in 2002 we find Suppes claiming that "the concept of model in the sense of Tarski may be used without distortion and as a fundamental concept" in the disciplines of mathematical and empirical sciences (2002, 21). He claims that the meaning of the concept 'model' is the same in these disciplines with the only difference to be found in their use of the concept. Although it is certainly true that this notion of model provides an interpretation of the relevant theory, the issue I'm interested in here is the extent to which this approach helps in articulating a useful account of theory structure.

In contrast to the semantic view, Tarski defines a theory as a set of sentences, and the role of the model is to provide the conditions under which the theory can be said to be true. Hence, the ultimate goal is defined in terms of truth and satisfaction, which are properties of sentences constituting the theory. Because the importance of the model is defined solely in terms of its relation to the sentences of the theory, in that sense it takes on a linguistic dimension. The point of Tarski's definition is that it rests on a distinction between models and theories, something that the semanticists essentially reject. For them, models are not about the theory as they are for Tarski; the theory is simply defined or identified by its models. For example, Suppes's account lacks a clearly articulated distinction between the primitives used to define particle mechanics and the realization of those axioms in terms of the ordered quintuple.[12] Moreover, if theories are defined as families of models, there is, strictly speaking, nothing for the model to be true of, except all the other models. In other words, the models don't provide an interpretation of any "theoretical" framework, but stand on their own as a way of treating the phenomena in question. While there may be nothing wrong with this in principle, it creates a rather peculiar scenario: it provides no way of identifying what is "fundamental" or specific about a particular theoretical framework since, by definition, all the paraphernalia of the models are automatically included as part of the theory. But surely something like perturbation theory, as a mathematical technique, should not be identified as part of quantum mechanics,

11. At the risk of being unfair to Giere, I want to emphasize here that I am referring to his earlier views on models and the semantic account, some of which I gather he no longer explicitly holds.

12. See note 8 for a summary of Suppes's example.

any more than the differential calculus ought to be part of Newton's theory.

One might claim that I am begging the question here against the semanticists by focusing exclusively on the Tarski approach and hence not only restoring the notion of a theory as a set of sentences but also ignoring later work on models that focuses more generally on structures (e.g., Addison 1965). Because the semanticists take the realizations in which these sentences are satisfied as the theory, one might claim that Tarski's account of a theory is irrelevant for their argument.[13] However, my focus on Tarski's account of models is motivated by the semanticists' (particularly Suppes's) own emphasis on that approach and not by an attempt to link the semantic view with a linguistic formulation. Regardless of whether one focuses on later developments that emphasize the notion of 'structure', the problems associated with defining a theory solely in terms of models remain. If a theory is just a family of models, then what does it mean to say that the model/structure is a realization of the theory?

The model is not a realization of the theory because there is no theory, strictly speaking, for it to be a realization of. In other words, the semantic view has effectively dispensed with theories altogether by redefining them in terms of models. There is no longer anything to specify as 'Newtonian mechanics' except the models used to treat classical systems. While there may be, strictly speaking, nothing wrong with this if one's goal is some kind of logical/model-theoretic reconstruction—in fact it has undoubtedly addressed troublesome issues associated with the syntactic account—but, if the project is to understand various aspects of models from within the 'scientific' context, then reducing theories to models seems largely unhelpful. And, even as a logical reconstruction, it isn't at all clear how it has enhanced our understanding of theory structure—one of its stated goals. Replacing theories with models simply obviates the need for an account of theories at all.

Let me emphasize that the point here is not to present an argument against the semantic view of the kind advanced by Friedman (1982) and Worrall (1984), claiming that the semantic and syntactic views are equivalent if one identifies theory structure with the class of intended models. That objection has been effectively handled by van Fraassen (1985). Instead, my claim is that if we want a broader understanding of the role models play in providing scientific knowledge or even an account of theory structure, then the logical/model-theoretic apparatus of the semantic view(s) doesn't really help us, especially in identifying and differentiating crucial aspects of theories and models. Because the use and construction

13. These objections were raised by an anonymous referee.

of theories/models in scientific contexts bears little, if any, resemblance to model-theoretic structures, if becomes difficult to see how the latter aid in understanding the former. Van Fraassen's focus on models as state-spaces is a notable exception in that it closely mirrors the way many physical systems are actually modeled. However, what we need to flesh out this picture is a more robust notion of "theory," something that each formulation of the semantic view lacks.

One of the problems I mentioned at the outset concerns specifying the content of the theory if it is identified strictly with its models, while the other, related issue concerns the interpretation of that content. In particular, the models of many of our theories typically contain a good deal of excess structure or assumptions that we would not normally want to identify as part of a theory. Although van Fraassen claims that it is the task of theories to provide literal descriptions of the world (1989, 193), he also recognizes that models contain structure for which there is no real-world correlate (225–228). However, the issue isn't simply one of determining the referential features of the models even if we limit ourselves to the "empirical" data. Because of the kinds of assumptions we typically build into our models, we often are unable to disentangle the truly empirical aspects from the stylized descriptions, produced via a high-degree mathematical abstraction. I should mention here that my concern is not the problem of surplus structure discussed by Redhead (2001) and others. Instead I am referring to cases in which models contain a great deal of structure that is used in a number of different theoretical contexts, as in the case of approximation techniques or the use of the renormalization group. Because models are typically used in the application of higher-level laws (that we associate with theory), the methods employed in that application ought to be distinguished from the content of the theory/ model, that is, what it purports to say about physical systems.

Consider the following example (discussed in greater detail in Morrison [1999]). Suppose that we want to model the physical pendulum, an object that is certainly characterized as empirical. How should we proceed when describing its features? If we want to focus on the period, we need to account for the different ways in which it can be affected by air, one of which is the damping correction. This results from air resistance acting on the pendulum ball and the wire, causing the amplitude to decrease with time while increasing the period of oscillation. The damping force is a combination of linear and quadratic damping. In the former case the equation of motion has an exact solution, but not in the latter case since the sign of the force must be adjusted each half period to correspond to a retarding force. The problem is solved using a perturbation expansion applied to an associated analytic problem in which the sign of the force is not changed. In this case the first half period is positively damped and

the second is negatively damped, with the resulting motion being periodic. Although only the first half period corresponds to the damped pendulum problem, the solution can be reapplied for subsequent half periods. But only the first few terms in the expansion converge and give good approximations: the series diverges asymptotically, yielding no solution.

All the information and application techniques above are contained in the model, yet we certainly do not want to identify the totality as part of the theory of Newtonian mechanics. Moreover, because our treatment of the damping forces requires a highly idealized description, it is difficult to differentiate the empirical aspects of the representation from the more mathematically abstract ones that are employed as calculational devices. The point here is not just that the so-called empirical aspects of the model are idealized since all models, and indeed theories, involve idealization. Rather, the way in which the empirical features are interconnected with the nonempirical makes it difficult to isolate what Newtonian mechanics characterizes as basic forces.[14] The point here is that we need to differentiate methods of application from the simple fact that theories are expressed in mathematical form.

Why do we want to identify these forces? The essence of Newtonian mechanics is that the motion of an object is analyzed in terms of the forces exerted on it which are described in terms of the laws of motion. These core features are then represented in the models, as in the case of the linear harmonic oscillator which is derived from the second law. The core features not only are common to the models but constrain the kind of behavior described by those models and provide (along with other information) the basis for the model's construction. Moreover, these Newtonian models embody different kinds of assumptions about how a physical system is constituted than, say, the same problem treated by Lagrange's equations. In that sense we identify these different core features as belonging to different 'theories' of mechanics.

But, as we saw above, when we model the physical pendulum, the calculation of the forces involved becomes very complex indeed. The application of the theory to physical phenomena requires special assumptions about how these systems/phenomena are constituted as well as calculational methods for dealing with those assumptions. In other kinds of situations we frequently need to incorporate rigid body mechanics into our models in order to deal with rigid bodies defined as systems of par-

14. Giere (1988, Chapter 3) has also discussed the pendulum model as a way of illustrating the role that idealizations play in models and of showing the importance of models in applying laws to particular problem situations. I have no disagreement with Giere's description of models per se, but only with the idea that theories should be construed as families of models.

ticles. In these cases we assume that the summation $\sum \mathbf{F} = \sum m\mathbf{a}$ and $\sum \mathbf{r} \times \mathbf{F} = \sum \mathbf{r} \times m\mathbf{a}$ over a system of particles follows directly from $\mathbf{F} = m\mathbf{a}$ for the motion of a single particle. That is, we simply assume that this is valid despite the fact that there are limiting processes involved in the mathematics when modeling rigid bodies as continuous distributions of matter. Another problem is that the internal electromagnetic forces cannot be adequately modeled as equal and opposite pairs acting along the same line, so we simply assume that the equations above are valid with the summation confined to external forces. Although the models require rigid body mechanics, no one would suggest that this is an essential feature of Newtonian theory, nor in the case of the pendulum example the specific methods for perturbation expansions. Even if we take account of Nancy Cartwright's point that theories or theoretical laws don't literally describe anything, we would still want to distinguish between what I want to call the fundamental core of Newton's theory (laws of motion and universal gravitation) and the models and techniques used in the application of those laws.[15]

If we identify a theory with a core set of laws/equations, no such difficulties ensue. For example, when asked for the basic structure of classical electrodynamics, one would immediately cite Maxwell's equations. These form a theoretical core from which a number of models can be specified that assist in the application of these laws to specific problem situations. Similarly, an undisputed part of the theoretical core of relativistic quantum mechanics is the Dirac equation. Admittedly there may be cases in which it is not obvious that such a theoretical core exists. Population genetics is a good example. But even here one can point to the theory of gene frequencies as the defining feature on which many of the models are based. My point is simply that by defining a theory solely in terms of its many models, one loses sight of the theoretical coherence provided by core laws, laws that may not determine features of the models but certainly constrain the kind of behaviors that the models describe. Indeed it is the identification of a theoretical core rather than all the features contained in the models that enables us to claim that a set of models belongs to Newtonian mechanics. Moreover, nothing about this way of identifying theories requires that they be formalized or axiomatized. As we shall see below, where a core can be isolated, it enables us to differentiate between theories and models in a way that speaks to many of the problematic issues encountered by the semantic view.

Although the semanticists claim that their account extends to scientific models, questions arise regarding its ability to incorporate the many dif-

15. Cartwright's view on theories is not one that I endorse; in fact as we see below, much of the argument in the remainder of the paper opposes her position.

ferent roles models play and the various ways they are constructed within that broader context. Even if we ignore the difficulties with the theory-model relationship mentioned above, identifying a model with a structure in the logician's sense, as a structure in which the sentences of the theory are true, restricts its function in ways that make it overly theory dependent. Its relation to theory is constrained by the laws of logic and model theory, which typically fail to capture the rather loose connections between theories and models that are more indicative of scientific practice (see Morrison 1990). One of the virtues claimed by the partial structures view of da Costa and French (2003) is its ability to overcome this difficulty; but as we shall see, not only does their formulation fail to solve the problem(s), it introduces ambiguities that further complicate the issue.

3. Partial Structures: A Partial Solution? One of the motivations for introducing the partial structures account of the semantic view was to capture the idea that scientific representations (theories or models) should not be considered true in the correspondence sense but only partially or approximately true. In order to spell out this notion a formalism is required—hence the concept of a partial structure consisting of a set of individuals (observables and unobservables) and partial relations defined on those individuals. For a given domain in which there are gaps in our knowledge, we simply model that domain in terms of a partial structure. Just as Tarski's account attempted to capture the intention of the correspondence view of da Costa and French of truth, the partial structures view attempts to represent the intentions of the pragmatists (specifically Pierce and James) whose notion of truth is much less rigorous, reflecting the epistemic gaps in our knowledge of the physical world (2003, 4). These structures capture the "essential incompleteness and partial nature of scientific theories" (5). So a particular claim or sentence S is partially true if it is true in the partial structure A.

So far so good. But, as da Costa and French quite rightly point out, how does one reconcile an account of theories as families of models with the idea that theories are partially true in a model? To answer this question, they appeal to Suppes's notion of an extrinsic and intrinsic characterization of a theory. In the former case, da Costa and French claim that whatever theories are ontologically, they are represented from the extrinsic perspective in terms of models or classes of models, and with respect to their intrinsic characterization as objects of epistemic attitudes (2003, 34). The extrinsic standpoint, which deals with the description of models, treats models as objects or structures; but when we are interested in the truth or empirical adequacy of the model, we switch to the intrinsic characterization (which focuses on what they do, for lack of a better term) and construe them as "possible realizations" that satisfy the sentences of a

belief report that describes our epistemic attitude (34–35). The upshot here is that it is the belief report, not the theory or the model, that is composed of sentences.

How has this clarified things? From within the constraints of the model-theoretic view, the da Costa/French account seems to provide a useful way of expressing the idea that the knowledge reflected in our models of physical phenomena is incomplete. But in order to specify which relations defined on the individuals are partially true, we must have already designated or know the particular features that have empirical support, features that are then expressed formally in the model. Da Costa and French claim that "insofar as theories have empirical support they can be regarded as quasi-true" (2003, 59). Fair enough; but has this helped us in sorting out which characteristics of the pendulum model we want to isolate as being quasi-true? No, because to do that we must have already chosen which features of the "scientific" model are quasi-true before representing them in the "structure" model(s). Moreover, how much empirical support is required for something to be termed quasi-true? The problems that beset attempts to characterize approximate truth threaten to arise here as well. In short, this "structural" reconstruction hasn't told us anything about the pendulum model that we don't already know, nor has it clarified which structural features we are entitled to call quasi-true. In other words, it allows us to construct a model that reflects an epistemic attitude of partial truth, but when the task of the scientist is to construct a model of a physical system, it is often not clear which relations/features are partially true. In this practical context the partial structures approach fails to capture a crucial epistemic dimension of model construction.

A further problem that besets the partial structures account involves an ambiguity about the role of sentences. At several places da Costa and French refer to the sentences that are partially true in A, a structure that represents a "portion of reality." In fact sentence "S can be said to 'point' to the world by means of a model" (2003, 17).[16] Although the sentences are what are said to be quasi-true, it is the structures that partially represent the world and hence are the "primary locus of epistemic activity" (20). However, when models are used as representational devices, we sometimes speak of them as quasi-true; but this, they claim, is just a *façon de parler* (34). The fact that the models satisfy the sentences of belief reports (where the models are the objects of epistemic attitudes) is not to say that the structures themselves contain sentences. So, when models are considered intrinsically, as the objects of belief reports, we can say that they are

16. Their definition of a partial structure on p. 18 includes a set P of sentences while the one on p. 28 doesn't.

quasi-true because we have switched from the extrinsic to the intrinsic characterization.

The problem here is that we have moved from da Costa and French's claim that sentences (generally construed) are true in a partial structure to their characterization of sentences as belief reports. If we define a partial structure as they do (p. 18),

$$A = \langle A,\ R_k,\ P \rangle_{k \in K},$$

where P is a set of sentences of a language L that is interpreted in A, and R is a partial relation defined on A (a nonempty set), then the claim that P is true in A means that A is a realization of P in virtue of the fact that it satisfies the sentence of the belief report. In that sense the structure (construed intrinsically) is limited to expressing only epistemic attitudes rather than aspects of the world. That is, we can't think of the model as representing physical features of the world because there is no room to accommodate the differences between sentences as the linguistic entities that express propositions about the world and sentences that express belief reports. Indeed, the only characterization of sentences that is acknowledged here is the expression of belief reports. That peculiarity aside, the necessity of moving between the extrinsic and intrinsic accounts in order to capture intuitive ideas about scientists' attitudes illustrates exactly the cumbersome (rather than clarificatory) nature of this approach as an account of scientific modeling.

Before I move on, there is one additional point about the partial structures account that is relevant for my discussion. Da Costa and French claim that their presentation of the model-theoretic account is not tied to a deductive relation between theories and models and can account for the way that models are constructed in actual practice (2003, 56).[17] They focus on the various complicated relations that exist between different types of models (phenomenological and data) and their "high-level" theoretical counterparts (59). The connections between them can be represented by partial isomorphisms that hold between families of partial relations in the structures constituting the models. The partial relations reflect the fact that the representation of a physical phenomenon in any one of these models is incomplete and hence only quasi- or partially true. In that sense,

17. Part of their discussion focuses on nuclear models and the ways they relate to certain aspects of theory. They claim that my discussion of nuclear models (Morrison 1998), as a counterexample to the semantic view, fails to take account of the ways in which these models do appeal to certain features of theory and hence are not theory independent as I claimed. Nowhere did I say that these models were completely independent of theory. My point was that there was no overarching theory from which these models could be derived.

the extrinsic characterization can be seen to support the intrinsic one (59). They claim that this is also able to capture the open-endedness of further theoretical development because the model can be extended in the light of new knowledge. And a significant feature of this characterization is that "any epistemic cut-off point between phenomenological and theoretical models, whether the latter are regarded as merely empirically adequate or outright false, is simply unwarranted" (59). Not only does the distinction between models and theories dissolve (60), but the differences between models are, at best, differences in "degree of partiality."

The difficulty here is that differentiating phenomenological and theoretical models in terms of their epistemic status fails to capture important differences between the two, differences that are crucial for their scientific employment. Although "theoretical" models typically contain deeper microscopic relations than phenomenological models and hence would be considered "less partial" in that respect, they may lack empirical support and consequently have less "partial truth" than a phenomenological model. Moreover, phenomenological models may involve a great deal of structure, as in the case of some sophisticated nuclear models such as the rotation-vibration collective models. Instead of dealing with degrees of partiality as da Costa and French do, we need to focus on ways of differentiating types of structure rather than amount. The emphasis on an epistemic cutoff point here seems moot in that there will often be a great deal of empirical support for phenomenological models and perhaps very little for any given theoretical one.

Ironically a statement by da Costa and French themselves nicely captures what I see as the central problem with the account: "in scientific practice [there] are a variety of structures . . . some . . . get described as theories and others as models" (60), and I would add to that: some are described as phenomenological models. In light of this we surely need some way of differentiating the kind of content embodied in each of these structures, something that the partial structures account seems unable to do. There are good "scientific" reasons for distinguishing these different levels, reasons that relate to what we expect from different types of models (and theories) in terms of knowledge of physical systems (e.g., causal vs. noncausal) together with the way each is used in practical contexts. To say, as they do, that the distinction between phenomenological and theoretical models is simply a difference in partiality not only ignores those reasons but fails to capture a fundamental part of scientific practice.

As I mentioned at the outset, much of the current interest in models has extended beyond logico-philosophical accounts of theory structure into more empirically based analyses of scientific modeling. While the semantic view addresses many issues and problems that arose with the syntactic view, it is, in essence, a philosopher's reconstruction that em-

phasizes the importance of set theory and logic in articulating an account of theory structure. However, if our goal is to achieve a greater understanding of the scientific practices involved in modeling and theory construction, then we need to look elsewhere in achieving that goal. Because the aim of logical reconstruction is not to emphasize practical issues and problems, it provides us with little in the way of resources for uncovering the aspects of models and theories that are crucial in extending our knowledge into new domains.

I have argued elsewhere (Morrison 1998, 1999) that an important aspect of that knowledge 'extension' is the autonomy that models have. I want to emphasize here how that autonomy needn't and in fact shouldn't imply that we can ignore the role that theories play in scientific contexts. The autonomy of models points to the fact that they function in ways that are different from and hence independent of theory, their construction is not necessarily theory dependent (i.e., that all models are models of a particular theory), and they represent physical phenomena/systems in ways that theory does not and cannot.[18] These differences, however, need to be more fully explored. By highlighting the way in which theory functions, we can also begin to have a deeper appreciation for the role it plays in understanding the construction and function of models.[19]

When one is isolating some of the ways theories might be distinguished from models, representation becomes particularly important. In some of the recent literature on modeling, Cartwright (1999) claims that in virtue of their abstractness, theories simply do not represent anything. In what follows I argue the contrary: that both theories and models represent, and the different ways they accomplish this will help to differentiate one from the other. In order to deal with that issue, let me turn to some of the ways in which representation functions in the context of both models and theories.

4. Models and Representations. At an intuitive level we can think of a model as embodying some type of representation of the phenomena under investigation. This notion can be understood in many different ways, but

18. This of course does not mean that models aren't often constructed from an amalgam of different theoretical features. My emphasis on autonomy from theory is meant only to point out that the theory-model relation is frequently not determined by theory.

19. Although van Fraassen claims to be more interested in the relation between theories and the world, his account of models also seems unable to capture many of the features displayed in the context of scientific modeling. Giere's account tends to deal more with the scientist's notion of a model; but in defining theories as families of models, it suffers from the same kinds of difficulties as the other semantic interpretations.

regardless how one interprets it, perhaps the most important feature of a model is that it contains a certain degree of representational inaccuracy.[20] In other words, it is a model because it fails to accurately represent nature. Sometimes we know specifically the type of inaccuracy the model contains because we have constructed it precisely in this way for a particular reason. Alternatively, we may be simply unsure of the kind and degree of inaccuracy because we do not have access to the system that is being modeled and thus have no basis for comparison.

One way of thinking about the representational features of a model is to think of it as incorporating a type of picture or likeness of the phenomena in question, a notion that has had a long and distinguished history in the development of the physical sciences. But, the idea of likeness here can be construed in a variety of ways. For example, some scale models of the solar system were constructed before and after Copernicus attempted to demonstrate that a planetary conjunction would not result in a planetary collision. One can think of these as representing, in the physical sense, both the orbits and the relation of the planets to each other in the way that a scale model of a building represents certain relevant features of the building itself.

Nineteenth-century physics is replete with both the construction of and demand for models as a way of developing and legitimating physical theory. In the initial stages of the development of electrodynamics, Maxwell relied heavily on different pictorial mechanical models of the ether as an aid to formulating the field equations, models that bore no relation to what he thought the ether could possibly be like. What these models did was represent a mechanical system of rotating vortices whose movements would set up electric currents that satisfied certain equations. While no one thought that the ether consisted of vortices, the model represented the way in which electric currents could arise in a mechanical system. Once the field equations were in place, he abandoned these pictorial mechanical models and chose to formulate the theory in the abstract formalism of Lagrangian mechanics, which itself functions as a kind of mathematical model for different types of physical systems, both classical and quantum.

Maxwell's work was severely criticized by Lord Kelvin, who maintained that the proper understanding of nature required mechanical models that could be physically manipulated as a way of simulating experiments. In other words, Kelvin's notion of a mechanical model was something that could be built, not simply drawn on paper; and both he and FitzGerald

20. I should mention here that the notion of representation used by the semanticists—isomorphism—does not really incorporate this notion of inaccuracy. I will say more about isomorphism below.

constructed such ether models as a way of representing the propagation of electromagnetic waves.[21] But here again the important point is that no one really believed that the ether itself bore a similarity to these models; rather they were useful because they represented a mechanical system that behaved according to the electromagnetic equations and because they led to modifications of some of Maxwell's mathematics. Add to this the Rutherford and Bohr models of the atom, the shell, and liquid drop models of the nucleus and we begin to see how models, functioning as both mathematical and physical representations, have emerged as important sources of knowledge.[22] What is important in these examples is that each of the models represents the target system in some particular way, by approximating it either as a likeness/similarity or as a system that obeys the same equations. But, in each case the notion of a likeness/similarity is different.

In keeping with the more logically oriented definition of a model, semanticists have another way of thinking about representation, one grounded in the notion of a structural mapping (e.g., an isomorphism). Since the semanticists' models are defined in terms of nonlinguistic structures, how should we understand their representational capacity? The response given by van Fraassen is that the empirical substructures of the model are candidates for "the *direct representation* of the observable phenomena" (1980, 64; italics added). But how does this "direct representation" take place since one can't have an isomorphism between phenomena and structures?

The answer involves the notion of "appearances," which include the "structures described in experimental and measurement reports" (van Fraassen 1980, 64). A theory is said to be empirically adequate if it has some model such that all appearances are isomorphic to empirical substructures of that model. For Suppes the situation is similar in that isomorphism enters in a central way. He makes use of representation theorems as a way of characterizing the models of a theory (56). For example, a representation theorem for a theory means that we can pick out a certain class of models of the theory that exemplifies, up to isomorphism, every model of the theory. So if M is the set of all models of some theory T and S is a subset of M, then a representation theorem for M with respect

---

21. For more on Maxwell's use of models, see Morrison (2000). For criticisms of Maxwell's models by Kelvin, see Smith and Wise (1989).

22. My intention here is not to draw a sharp distinction between mathematical and physical models per se because in many cases a physical model, like FitzGerald's ether models, has a mathematical interpretation as well. I don't deny that there may be important ways in which mathematical and physical models differ, but those differences are not relevant for my discussion here.

to S would be the claim that for every model m in M there is a model in S isomorphic to m.[23] I am not going to discuss the relation between this type of representation and the more pictorial type mentioned above except to say that it is analogous to the relation discussed by Suppes between his use of the term model (i.e., the set-theoretic one) and the more general notion of a physical model. While it may be possible to capture the pictorial notion in terms of the more formal one, doing so will undoubtedly add an extra layer of 'structure' that seems unnecessary if our goal is to represent a physical system for the purposes of understanding, say, its causal connections.

Regardless of these possibilities, scientific representation characterized in terms of isomorphism is not without its critics. Suarez (2003) in particular has argued against both isomorphism and similarity as the constituents of scientific representation (see also Frigg 2002). He sees the emphasis on similarity and isomorphism as indicative of a reductive naturalistic approach that ignores scientists' purposes and intentions, thereby relegating the latter as nonessential features of representation. While I agree with many of Suarez's points, my own view is that the poverty of similarity and isomorphism as characterizations of representation stems, ultimately, from adopting the semantic or model-theoretic approach to theories. If one chooses to interpret theories as families of models/structures, then one is all but forced to rely on isomorphism as "the" way to flesh out the notion of representation.

On Giere's (1988) account of the semantic view, which also characterizes models as structures, the model-world relation is explicated via theoretical hypotheses; however, those hypotheses are useful only insofar as they tell us the extent to which the model is similar to the physical system that is modeled. But, if we dissociate models from the notion of a formal structure as I am suggesting, then we are free to talk about similarity as a basis for representation in some contexts but not in others. The advantage here is that in some cases models will be similar to the systems they model, making the concept useful but without committing us to similarity as the only type of representation that is possible. This becomes important when we realize that often a motivating aspect of model construction is a lack of knowledge of the system/phenomena being modeled (as in the Maxwell case), and in those situations the use of similarity as a representational

23. Suppes also uses the notion of definition as a way to think about the representation of concepts of the theory. The first definition in a theory will be a sentence of a certain form that establishes the meaning "of a new symbol of the theory in terms of the primitive concepts of the theory" (2002, 53). The point of introducing new symbols is to facilitate deductive investigation of the theory's structure, not to add to the structure. In that sense then, the definitions are not a crucial aspect of the models.

tool is somewhat circumscribed. My point then is that the difficulties with isomorphism and similarity as accounts of representation arise partly from problems inherent to the notions themselves (especially in the case of similarity), but primarily from the structural account of models that necessitates their use. Consequently, an abandonment of the semantic view will also liberate us from reliance on isomorphism as the way of representing scientific phenomena.

As I mentioned above, the goal in promoting a notion of theory that extends beyond a "collection of models" requires that we be able to capture certain 'general' features used to classify physical systems, features that aren't always obvious or easily extractable from collections of models. To revisit our previous example: If we take Newtonian theory to encompass all the models used in its application, then how do we determine which features of this collection have the appropriate representational status; that is, what features of the models should be singled out as essential for representing physical systems as basically Newtonian? If we can pick out the common elements that make the models Newtonian (e.g., laws of mechanics), then aren't we just isolating certain features that, taken together, specify a core notion of theory from within the family of models—something the semanticists are at pains to avoid and something they claim isn't needed?

One of the roles of theory, one that is not successfully captured by a family of models, is to provide a general representation of an entire class of phenomena. Here "representation" simply means describing the phenomena as exhibiting certain kinds of behavior (e.g., quantum, classical, relativistic). Part of the description involves the constraints placed on that behavior by the laws of the theory. The Schrodinger equation, for example, describes the time evolution of quantum systems, and together with other principles such as the uncertainty and exclusion principles and Planck's constant, form the core of our theory of quantum mechanics. The models of nonrelativistic quantum mechanics all obey these constraints in their role of filling in the details of specific situations in order to apply these laws in models such as the infinite square well potential. In these kinds of cases theory functions as primary in defining a class of models. This is not to say that this is the only role for models; indeed I have spent a good deal of time arguing for their autonomy in various circumstances. Rather, my point is that theory too plays an important role in the way we understand and represent physical systems; consequently, we need some account of how the generality expressed in theories translates into representational power.

Some of the literature on models suggests that only they can function

as the vehicle for scientific representation.[24] Cartwright (1999, 180), for instance, has claimed that "the fundamental principles of theories in physics do not represent what happens"; only models represent in this way, "and the models that do so are not already part of any theory."[25] The reason for this is that theories use abstract concepts, and although these concepts are made more concrete by the use of interpretive models (e.g., the harmonic oscillator, the Coulomb potential, etc.), they are still incapable of describing what happens in actual situations. For that we need representative models that go beyond theory. These latter models, which account for regular and repeatable situations, function like "blueprints for nomological machines" (180).[26]

However, this distinction between interpretive abstract models (and theories) and representative models contains two important caveats: (1) the abstractness of interpretive models prohibits them from representing physical phenomena only if we take a strict notion of similarity as our criterion for representation, and (2) the concrete details allegedly present in representative models don't guarantee that they represent actual situations. Cartwright's argument trades on the assumption that the abstractness inherent in the description of the harmonic oscillator, for example, renders it incapable of describing any real physical phenomena.[27] While theories undoubtedly contain what Cartwright calls abstract concepts, it is important to recognize that these concepts and their interpretive models are part of the way theories classify particular kinds of behavior—by representing it as an instance of, say, harmonic motion. No actual system looks like the model of the harmonic oscillator, but this doesn't mean that the model doesn't represent basic features of harmonic motion or tell us important things. For example, when we use the model to analyze the motion of an object attached to a spring that obeys Hook's law, we find that the period of the oscillator is completely independent of the energy or amplitude of the motion—a result that is not true of periodic oscillations under any other force law. Of course if the frequency varied

24. Although Suppes's use of representation theorems refers to theories, it refers strictly to the set of models of the theory.

25. Not being an advocate of the semantic view, Cartwright typically speaks of models in the sense used by physicists, physical models that are either theoretical or phenomenological, or a mixture of both. Although she doesn't develop a full account of representation, it seems clear that what she has in mind is more closely aligned with the kind of representation afforded by, e.g., the Bohr model of the atom.

26. For a discussion of this notion, see Cartwright (1999).

27. Cartwright (1989) makes a distinction between abstract and idealized models: The latter involves an idealized description of a physical object, such as a frictionless plane, that can be made more concrete by adding back certain features; the former describes an object that cannot be made more realistic simply by adding correction factors.

significantly with the amplitude, the situation would become much more complicated; but most vibrating systems behave, at least to some approximation, as harmonic oscillators with the properties described by the model.

The important point here is how to think about representation. We use models to represent physical phenomena in more or less abstract ways. Some models build in features that describe the system in more detail than others; but regardless of the degree of detail, we typically intend the model to refer to a physical system/phenomenon understood in a certain way.[28] When we say that a diatomic molecule can be considered as two point masses connected by a spring and can undergo quantized oscillations, we are making a point about how quantum mechanics (and the model) predicts that the energy levels of a harmonic oscillator are equally spaced with an interval of h times the classical frequency and have a minimum value (the zero-point energy). It also follows from this that photons absorbed and emitted by the molecule have frequencies that are multiples of the classical frequency of the oscillator. While the model may be highly abstract, it nevertheless refers to basic concrete features of a physical system and in that sense functions in a representational capacity.[29]

The Bardeen-Cooper-Schreiffer (BCS) theory of superconductivity illustrates how theory can 'represent' by isolating the causal mechanism (Cooper pairing) responsible for the production of superconductivity. This type of representation—of a fundamental characteristic of physical systems/phenomena—can be captured only in virtue of the generality that theories embody. While these representations may be abstract in the sense that they don't provide many details about the behavior of these basic features, they nevertheless give us information about concrete physical systems. Moreover, attempts to construct what Cartwright calls "representational" models—more specific accounts of concrete systems—may not necessarily embody realistic features. In other words, we may be unsure of the details of the system we are modeling or may simply choose to model the situation in a way that bears little if any relation to reality.

28. How it does this is a different question, one that may or may not involve matters related to cognitive psychology. My claim here is that theories and abstract models do represent, contrary to Cartwright's point that they don't.

29. Another important concern here is the role of mathematics. Since most models in physics, biology, economics, and many other disciplines are mathematical, the question arises as to the representational relation between mathematics and physical phenomena. This, to my mind, is a separate topic that requires an investigation of the applicability of mathematics, an issue that has been dealt with by Steiner (1999), among others.

Maxwell's ether models are a case in point, as is the BCS "model" of how the Cooper pairing occurs.

So, in contrast to Cartwright's account, the view I want to defend here ascribes a representational function to theories—one that is significantly different from both the concretization attributed to interpretive models and what she terms the more realistic representative models. As I noted above, representative models may be highly idealized and unrealistic, depending on how they purport to represent physical phenomena. Although theory does not account in detail for specific situations, it sometimes provides us with representations that are more concrete than the interpretive or representative models Cartwright mentions. My use of 'concrete' indicates the fact that theories can represent physical systems/phenomena by isolating and highlighting certain basic features, behaviors, or causes from which more specialized applications can then be derived or constructed via models. For example, Cooper pairing in the BCS theory functions as a concrete representation insofar as it specifies exactly what processes give rise to superconductivity. How that process occurs is spelled out in a more detailed model but one that incorporates, ironically, a number of highly abstract and unrealistic assumptions.

As we can see, none of this implies that the type of basic representation we get via theory is necessarily less realistic than representation by models—that it doesn't describe an actual mechanism or cause. Instead it simply lacks the kinds of details specific to how the cause operates in particular types of situations. A representation provided by either a theory or a model needn't incorporate these kinds of details about concrete situations to be considered realistic. A more appropriate way of thinking about whether something is realistic is whether it attempts to account for how things could be constructed or come about and whether it does so in a way that is physically realizable.[30] Consequently, the theory/model distinction isn't captured by the representational power of the latter over the former, nor by the abstract/concrete relationship. Instead, we need to look elsewhere in attempting to distinguish the role that each plays in producing scientific knowledge. The BCS theory of superconductivity discussed below helps to illustrate the distinction between models and theories by showing a way to isolate the basic structure of a theory from its many models. Thinking about the distinction between theories and models in this way also helps us to make sense of the notion that what began as a model can frequently "evolve into" or "become" a theory in its own right.

---

30. Again, it is important to point out here that simply adding more details doesn't necessarily make the model more realistic. The details might well embody abstract or purely fictional assumptions about how the system is constituted.

5. Theories and Models in Practice. In some earlier work (Morrison 1998, 1999) I suggested that one of the important features of models was their explanatory power, their ability to provide more or less detailed explanations of specific phenomena/behaviors. One of the reasons models have explanatory power is that they provide representations of the phenomena that enable us to understand why or how certain processes take place. Although my claim here is that theories can explain/represent the general features of a physical system, the important difference between their representational function and that of models is that the latter, unlike the former, tend not to be generalizable. Below I highlight some of the differences between these two kinds of representation and explanation by showing how features that begin as model-based representation (and explanation) can, in the right context, develop into a more general theoretical representation characteristic of a theory. Although the first presentation of the BCS account of superconductivity (Bardeen et al. 1957) was referred to as both a 'theory' and a 'model' by its authors, many viewed it as only a model. While it is now generally thought to be a theory, some still refer to it as a model because of its inability to account for high-temperature superconductivity. This problem not withstanding, let us look at why it was initially referred to both ways and some of the reasons that were operant in the alleged transition from model to theory. The ability to do this will, I hope, give us some sense of whether my discussion of representation and explanation might prove useful in articulating the differences between the model and the theory.

Prior to 1957 there existed mainly phenomenological models of superconductivity that focused on thermal and electromagnetic properties of superconductors. Although there were suggestions about what a quantum-theoretic approach might look like, no developed microscopic account was put forward until the famous paper of Bardeen et al., which provided the first coherent picture of how superconductivity actually occurs. The phenomenological models that preceded BCS all incorporated an energy gap to describe thermal properties; the task of the microscopic theory was to explain, in the context of general theories of quantum mechanics, why the gap arises.[31] This energy gap is also related to the problem of resistance, a crucial feature in the explanation of superconductivity.

In ordinary metals, one of the basic mechanisms of electrical resistance is the interaction between moving electrons (i.e., electric current) and vibrations of the crystal lattice. However, if there is a gap in the energy spectrum, quantum transitions in the electron fluid will not always be

---

31. An energy gap is simply a gap between the valence and conduction energy bands; metals, however, do not have separate bands, but a single band containing many more states than electrons to occupy them.

possible; the electrons will not be excited when they are moving slowly. This feature is intimately connected with the possibility of movement without friction, the essential feature of superconductivity.[32] The problem, then, was how to account for this movement together with the presence of the gap. Something about the electrons themselves and their interaction with the crystal lattice ought to provide a clue. Earlier work by Frohlich (1950) and later Bardeen (1951a, 1951b) pointed out that an electron moving through a crystal lattice has a self-energy by being 'clothed' in virtual phonons. This distorts the lattice, which then acts on the electron in virtue of the electrostatic forces between them. In fact, one can think of the interaction between the lattice and electron as the constant emission and reabsorption of phonons by the latter. The problem, however, is that the phonon-induced interaction must be strong enough to overcome the repulsive Coulomb interaction; otherwise the former will be swamped and superconductivity would be impossible.[33]

A quantum mechanical account of the electron-phonon interactions and how they gave rise to the gap required an explanation in terms of the Pauli exclusion principle. Further investigation by Cooper (1956) revealed that two electrons with the same velocity moving in opposite directions with opposite spins had an attractive part that was stronger than the normal Coulomb repulsion. This net attractive interaction involved a dynamical pairing of the two electrons, a process that became known as 'Cooper pairing'. As long as the net force is attractive, no matter how weak, the two electrons will form a bound state separated by an energy gap below the continuum states. In short, the phonon-induced interaction gives rise to Cooper pairing, which produces the energy gap required for superconductivity. These pairs of electrons behave more like bosons, which are capable of condensing into the same energy level; and because they have a slightly lower energy, a gap is produced that prohibits the collision interactions that produce resistivity.[34]

The BSC (1957) paper contained an account of the pairing process that provided both a general representation and an explanation of how su-

32. But, at higher temperatures the interaction between the electrons and the lattice is very weak, and hence there is not sufficient interelectron attraction to overcome the Coulomb repulsion. Consequently, there is no transition into the superconducting state.

33. This is what happens in the case of semiconductors, i.e., solids that also have an energy band gap but yet don't show superconducting properties. The key difference between these two types of metals is of course the presence of Cooper pairs.

34. With systems of fermions no gap is produced since all available levels up to the Fermi energy are filled. The energy gap suggested a phase transition that produced a type of Bose-Einstein condensation, something that single electrons are prohibited from because of the exclusion principle.

perconductivity arises. Essentially it is a property not of the atom but of the free electrons in the metal, electrons that do not move independently. The overall picture is this: The superconducting ground state is a highly correlated one where the electrons are bound together in pairs and occupy a thin shell near the Fermi surface, separated by an energy gap above them on the order of .001 eV. The presence of the gap inhibits the kind of collision interactions which lead to ordinary resistivity. The pairing is caused by an attractive force between electrons that results from the exchange of phonons. Initially this picture functioned as a kind of representation or 'representative model' of the mechanism responsible for superconductivity. This 'representational model' further enabled BCS to focus just on those single electron states that had paired states filled, which in turn led to the construction of what they called a 'reduced' Hamiltonian. This allowed for a more simplified mathematical approach that dealt with only the essential aspects of the superconducting state itself.[35] I refer to the description above as a 'representational model' because it purports to represent what BCS assume is the fundamental or basic causal mechanism (Cooper pairing) involved in superconductivity and situates that representation squarely within the constraints of quantum mechanics (specifically, the exclusion principle).[36] Once this 'representative model' was in place, they then focused on constructing the ground state wave equation that formed the mathematical foundation of the BCS account. But more details were needed to get from the representation of Cooper pairing to a full description of the process.

BSC made several other specific assumptions that, taken together, extended the initial 'representational model' into a more detailed account of superconductivity. It is here that one can begin to see a divergence between the core physical ideas together with the mathematical framework that explain superconductivity and the further assumptions that needed to be presupposed in order to flesh out that framework. These latter assumptions are more tentative but also provide a more specific account of how a superconducting system might be constituted, in essence, a story about how the electron pairing might take place. One of the novel features of the BCS ground state was that it did not have a definite number of electrons, a rather odd situation since there clearly could be no creation processes going on in a superconductor. What made this novel was not the notion of an indefinite number of particles itself, but that this con-

---

35. For a more detailed account of this, see Morrison (forthcoming).

36. It is perhaps important to keep in mind here that this is not an interpretive model because although it describes the process of Cooper pairing within the theoretical framework of quantum mechanics, it is not a model of quantum theory in the sense that Cartwright's interpretive models are.

straint was used to describe Fermi as opposed to Bose particles. The form of the wave function was not novel: others had used it in conjunction with Bose particles since these (e.g., phonons, photons, and mesons) clearly could be created. Similarly, in other high-energy contexts involving scattering phenomena, it was common to write down wave functions that had an indefinite number of particles. However, this was not the case for the low-energy phenomena, where it was assumed that the number of particles should be definite.

How did BSC justify this rather bold step? Essentially they appealed to a fundamental idea from statistical mechanics. Given that there were so many pairs spread over such a large volume, it made sense to think of them as not being completely correlated with one another, but correlated only in a statistical sense.[37] In other words, the wave function represented a kind of statistical ensemble in which the pairs interacted but were not strongly correlated; they were partly independent, constituting a superposition of states with different numbers of particles. A Hartree-type approximation (which does not conserve the number of particles) was used in which the probability distribution of a particular state does not depend (at the level of description that is given) on the distribution of the others, something that had never been applied to electrons. This was justified by arguing that the occupancy of some one state was basically independent of whether other states were occupied.[38] In addition, they made other idealizing assumptions such as the neglect of anisotropic effects, the impact of which was that superconducting properties appeared to depend only on gross features rather than details of the band structure.

All these more specific assumptions extend the 'representational model' beyond the initial claims about the existence of Cooper pairs and the energy gap. Although the explanation concerning the indefinite number of electrons initially seemed ad hoc, this indefiniteness later emerged as

37. Bardeen later pointed out, however, that the form of the wave function, with the all-important common momentum for paired states, is determined by energetic rather than purely statistical considerations. See his 1957 reply to Dyson (Bardeen 1973).

38. This aspect of the model definitely had the appearance of an ad hoc assumption, and despite its apparent efficacy, BSC were not happy with the idea of an indefinite number of states. They attempted to downplay the idea by claiming that the superconducting wave function could be taken to be the projection of the state that they had introduced onto the space that did have a definite number of particles. Later developments in the theory would reveal that this indefiniteness was an essential feature of the superconducting state since it allowed for the introduction of the quantum phase defined over the whole of the macroscopic superconductor. However, it was still not possible to think of electrons being created or destroyed since one was dealing with low-temperature solid-state phenomena. To that extent Bardeen et al. emphasized that their system really did have a definite number of electrons despite the form of the wave function.

an essential feature of the superconducting state that related to the phase of the wave function and was crucial for understanding interactions taking place in the metal. To that extent the picture BCS presented was a coherent one despite containing some highly speculative elements. The process by which they arrived at the account of superconductivity depended primarily on the representation of Cooper pairing and how it produced the energy gap. The structural features of the representational model were mirrored by the form of the Hamiltonian (it was possible to focus on the 'reduced' problem that considered only those single electron states that had paired states filled). These qualitative ideas about the nature of the superconducting state in turn provided constraints on the formulation of the BCS wave function. However, the form of the wave function presupposed a physical picture that BCS had to then spell out in further detail: the ground state was treated as a linear superposition of the pair states and the pairs were treated as a statistical ensemble that allowed the ground state to have an indefinite number of electrons. These rather bold assumptions extended the physical picture provided by the initial representational model by furnishing a more specific account of how superconductivity is produced.

6. But Is It a Theory? It is easy to see why Bardeen et al. referred to their account as a 'model' of superconductivity. It contained many idealizing assumptions as well as some rather controversial ideas about the specifics of the superconducting state. But, they also referred to it in several places in their paper as a 'theory' without any explicit way of differentiating the two. As I mentioned above, one reason for focusing on this particular case is that it embodies many of the tensions present in the literature on models and theories but also provides fertile ground for examining some possible resolutions. My primary goal here is not to philosophically reconstruct the BCS case, but rather to use it as a way of illustrating how a more robust notion of theory can play a role in capturing the various levels of explanation and representation present in the larger "theoretical" environment. So the question we must ask is whether there is any basis for a theory-model distinction here that can be extrapolated to a broader context, one that is consistent with the idea of a theoretical core introduced earlier.

Recall that the BCS account provided a general representation and explanation of how the superconducting state arises (Cooper pairing), and did so on the basis of quantum mechanical principles. This together with its good quantitative agreement with the equilibrium properties of superconductors are qualities that, taken together, could support its status as a 'theory'. In the conclusion of the paper the authors acknowledge that their calculations are based on a rather idealized 'model' that they

nevertheless feel is "essentially correct" (Bardeen et al. 1957, 1198). They go on to remark that they would like to see an improvement in the general formulation of the 'theory', specifically, the desire for a renormalized interaction in which higher-order terms can be taken into account. But these are details that do not affect the fundamental assumption on which the 'theory' is based, specifically, the net attractive interaction between electrons for transitions in which the energy difference between the electron states involved is less than the phonon energy.

What these remarks reveal is exactly how disparate levels of explanation and representation can provide different types of information about the superconducting state. These can, in turn, provide evidence for a fundamental distinction between models and theories that enables us to differentiate a specific role for each, both in terms of explanatory power and in terms of representational capacity. In this particular case there exists a fundamental idea that is thought to constitute the core of the 'theory', that is, Cooper pairing, which explains and represents how superconductivity can arise. The physical details of that idea and how it relates to specific aspects of superconductivity are then spelled out in a rather idealized model.

Let me flesh out exactly how I see the theory-model distinction at work here. We saw above that the BCS account explains how the electron-phonon interactions associated with superconductivity could give rise to an energy gap, and in particular how one could show this primarily as a result of the exclusion principle. The pairing hypothesis was then incorporated into a more general treatment that involved the construction of a ground state wave function with a reduced Hamiltonian that focused specifically on the pairing. In other words, the basic causal mechanism responsible for superconductivity was isolated and placed in the broader context of a quantum mechanical treatment. To that extent the BCS account provided not only a general conception of how superconductivity arises but also a method (i.e., the mathematical framework) required for arriving at quantitative predictions. It is in this sense that we can think of the account as a 'theory'. But these features might also be characteristic of models: they also provide a conception of how a phenomenon might occur or behave, and they frequently provide the mathematical apparatus necessary for working out problems specific to the context. So, what can we point to that distinguishes the two? The feature that, to my mind, seems important here is the level of generality that is involved in both the conception and the method and how this differs for theories and models.

Even if we ignore the new types of superconductors discovered after 1979, it is possible to distinguish between a theory of superconductivity whose foundation is the formation of Cooper pairs and the more specific

assumptions made in the original BCS paper (and indeed later accounts of Cooper pairing) about the form of the attractive potential and so forth. While there have been several subsequent accounts in the literature about exactly how the pairing hypothesis explains superconductivity, they are all insensitive to the form of the pair wave function and hence to the origin of the attraction responsible for the pairing. In other words, as an explanation and representation of how superconductivity arises, Cooper pairing remains the fundamental cause that accounts for an entire class of superconductors.

In addition to electron-lattice interaction and the formation of Cooper pairs the other important feature of superconductivity is that these coupled electrons can take the character of a boson and condense into the ground state leaving a band gap between them. The BCS account not only deals with one of the possible pairing mechanisms (phonons) but also provides a description of the condensed state in general terms independent of the pairing mechanism. This general conception of pairing is an important feature in establishing BCS as a 'theory'. Moreover, the theory shows that the ratio between the value of the gap at zero temperature and the value of the superconducting transition temperature takes a universal value independent of the material. But, and perhaps most important, there is another feature of the story that has important implications for identifying aspects of BCS as a theory, specifically the relation to higher-order principles such as spontaneous symmetry breaking. If we look at BCS from the point of view of general principles, we find that the core features of the theory could be generalized in a way that would allow for gauge invariance. This relied ultimately on what was found to be the broken symmetry of the BCS state, something that was not available in the original presentation but could be shown without violating the fundamental features of the original account. In that sense, then, the relation to spontaneous symmetry breaking can be seen as the final step in establishing BCS as a theory.[39]

Compare these conditions with the type usually in play when constructing a model. Initially, at least, there is a target level of accuracy that is aimed at. We want to explain a particular phenomenon, such as nuclear fission, but we cannot do so within a mathematical framework that is generalizable to other nuclear phenomena. And in some cases there is no background theoretical framework the different models have in common (see Morrison 1998). As a result we invoke a specific representation that works well for the case at hand. By contrast, the BCS theory not only

---

39. This was also true as a matter of sociological fact. Many objected to the original version of BCS because it was not gauge invariant, but accepted it once this was shown by Philip Anderson to be the case.

tells us the cause of superconductivity but embeds that causal claim into the larger quantum theoretical framework and subsequently relates it to the more general notion of spontaneous symmetry breaking. To that extent we have a general representation and explanation of what is involved in superconductivity. Moreover, despite the lack of details we can call this description concrete insofar as it furnishes a causal story. There are a variety of specific accounts of the pairing process on offer, but they all fall within the domain of models. In other words, the BCS 'theory' has different 'models' in addition to the one initially formulated by Bardeen et al. This way of differentiating theories and models allows us to make sense of the uses of both terms in their 1957 paper. They saw themselves as putting forward a 'theory' of superconductivity because the fundamental idea of Cooper pairing and the form of the wave function were taken to be essentially correct. In addition, however, there were specific features of the account that could only be classified as providing a 'model' of how the pairing process took place.

None of what I have said here undermines the autonomy of models. They still function as the source of specific representations, and we can see how this plays out in the treatment of Cooper pairs within the original BCS account. The initial representational model of how the pairing took place contained a fundamental idea that functioned as the representational centerpiece around which the mathematical treatment was constructed, an idea that then formed the core of the theory of superconductivity. The evolution from model to a more fully developed theory involved the embedding of the basic elements of the pairing phenomenon into the mathematical framework of quantum mechanics, an account that was then shown to be gauge invariant and resilient to changes in more specific explanations of the interactions involved in the pairing. In other words, an idea that originally formed the basis of a model evolved into a theory as a result of its ability to function in a fundamental and general way.

It is also important to point out that the different accounts of pairing that followed involved the relaxation of some assumptions made by BCS, for instance, that all electrons are bound in pairs with the same molecular wave function. This results in a rather different account of how Bose condensation occurs. Similarly for the case of post-1979 superconductors, especially the heavy-fermion variety, the pairing is thought not to be of the BCS type, but rather 'exotic', involving different crystal symmetries. In that sense then, the different models of superconductivity retain an autonomy that allows for more specific representations of pairing while maintaining the integrity of the fun-

damental representation that forms the core of the theory—Cooper pairing—a concrete causal claim.[40]

7. Conclusions. My primary task was to provide some elementary way of distinguishing the kinds of representation and explanation provided by models and theories and show how this distinction might be used to differentiate models and theories themselves. My motivation resulted from what I perceive as a shortcoming in many of the accounts currently on offer that discuss the model/theory relation. Reflecting on the few details I have given about the BCS theory of superconductivity, it seems that we have no clear way of accommodating its structure or development in the formal framework of the semantic view, nor on accounts of models that emphasize complete independence from theory. The BCS theory refers to a well-defined core that involves the notion of pairing and a connection with the broader theoretical principle of spontaneous symmetry breaking, both of which constrain the way superconducting phenomena can be represented. To say that the theory consists simply of a 'family of models' would be to include the different (and contradictory) accounts of pairing, something that proves unhelpful in understanding the theory's structure and its development. Different models provide various instantiations of the causal mechanism of Cooper pairing, and each exhibits a degree of autonomy to the extent that its specific features are not derived from or dictated by the 'theory' of superconductivity. The contradictory features of the models cannot be adequately accounted for on the partial structures approach because we have no way of delineating which aspects of the models qualify as partially true. Resorting to 'structures' as a way of explicating the theory/model relation simply adds a layer of complexity to the discussion, the benefits of which are far from clear.

Similarly, accounts that stress the complete isolation of models from theory are at a loss to explain the stability of (1) the basic assumptions that each of the different models instantiate and (2) the form of the wave equation that is immune from variations in different accounts of pairing, features that are best identified as forming part of the 'theory'. We have

---

40. Interestingly enough, this situation has parallels with the development of the Glashow-Weinberg-Salam electroweak theory. Weinberg's initial paper on the topic was entitled "A Model of Leptons" (1967). Yet, even after the remarkable experimental confirmation with the discovery of weak neutral currents, he remained skeptical about the $SU(2) \times U(1)$ model, claiming that it was not sufficiently natural or realistic to win general acceptance and that the parameters had to be carefully rigged to achieve even qualitative agreement with the experiment (Weinberg 1974, 258, 262). That pessimism, however, was directed at the specific model that had been constructed and not at the core idea that the weak and electromagnetic interactions were associated with a spontaneously broken gauge symmetry.

seen how theory and models involve different notions of explanation and representation that are appropriate at different levels of generality. The ability to identify a fundamental core or idea that can be framed within a mathematical treatment and accounts for a broad spectrum of phenomena seems like a good place to start in the difficult task of differentiating models and theories.

REFERENCES

Addison, J. W. (1965), "Some Notes on the Theory of Models", in J. W. Addison, L. Henkin, and A. Tarski (eds.), *The Theory of Models*. Amsterdam: North-Holland, 438–441.

Bardeen, J. (1951a), "Electron-Vibration Interactions and Superconductivity", *Review of Modern Physics* 23: 261–270.

——— (1951b), "Relation between Lattice Vibration and London Theories of Superconductivity", *Physical Review* 81: 829–834.

——— (1973), "History of Superconductivity", in B. Kursunoglu and A. Perlmutter (eds.), *Impact of Basic Research on Technology*. New York: Plenum.

Bardeen, J., L. Cooper, and J. R. Schrieffer (1957), "Theory of Superconductivity", *Physical Review* 108: 1175–1204.

Beth, E. (1949), "Towards an Up-to-Date Philosophy of the Natural Sciences", *Methodos* 1: 178–185.

Braithwaite, R. (1953), *Scientific Explanation*. Cambridge: Cambridge University Press.

——— (1962), "Models in the Empirical Science", in E. Nagle, P. Suppes, and A. Tarski (eds.), *Logic, Methodology and Philosophy of Science*. Stanford, CA: Stanford University Press, 224–231.

Carnap, R. (1939), *Foundations of Logic and Mathematics: International Encyclopedia of Unified Science*. Vol. 1, no. 3. Chicago: University of Chicago Press.

Cartwright, N. (1989), *Nature's Capacities and Their Measurement*. Oxford: Oxford University Press.

——— (1999), "Models and the Limits of Theory: Quantum Hamiltonians and the BCS Models of Superconductivity", in M. Morgan and M. Morrison (eds.), *Models as Mediators: Perspectives on Natural and Social Science*. Cambridge: Cambridge University Press, 241–281.

Cartwright, N., M. Suarez, and T. Shomar (1995), "The Tool Box of Science", in W. E. Herfel et al. (eds.), *Theories and Models in Scientific Processes*. Amsterdam: Rodopi, 137–149.

Cooper, L. (1956), "Bound Electron Pairs in a Degenerate Fermi Gas", *Physical Review* 104: 1189–1190.

da Costa, N. C. A., and S. French (2003), *Science and Partial Truth: A Unitary Approach to Models and Scientific Reasoning*. Oxford: Oxford University Press.

Friedman, M. (1982), "Review of van Fraassen *The Scientific Image*", *Journal of Philosophy* 79: 274–283.

Frigg, R. (2002), "Models and Representation: Why Structures Are Not Enough." CPNSS Discussion Paper Series, DP MEAS 25/02.

Frohlich, H. (1950), "Theory of Superconducting State 1: The Ground State at the Absolute Zero of Temperature", *Physical Review* 79: 845–856.

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

Hesse, M. (1953–54), "Models in Physics", *British Journal for the Philosophy of Science* 4: 198–214.

——— (1970), *Models and Analogies in Science*. Oxford: Oxford University Press.

Kuhn, T. (1972), *The Essential Tension*. Chicago: University of Chicago Press.

Morrison, M. (1990), "Unification, Realism and Inference", *British Journal for the Philosophy of Science* 41: 305–332.

——— (1998), "Modelling Nature: Between Physics and the Physical World", *Philosophia Naturalis* 38: 65–85.

——— (1999), "Models as Autonomous Agents", in M. Morgan and M. Morrsion (eds.), *Models as Mediations: Essays in the Philosophy of the Natural and Social Sciences*. Cambridge: Cambridge University Press, 38–65.

——— (2000), *Unifying Scientific Theories: Physical Concepts and Mathematical Structures*. Cambridge: Cambridge University Press.

——— (forthcoming), "Theories as Representational Structures", in Stephan Hartmann, Carl Hoefer, and Luc Bovens (eds.), *Nancy Cartwright's Philosophy of Science*. London: Routledge.

Redhead, M. (2001), "The Intelligibility of the Universe", in A. O'Hear (ed.), *Philosophy and the New Millennium*. Cambridge: Cambridge University Press.

Smith, C., and N. Wise (1989), *Energy and Empire*. Cambridge: Cambridge University Press.

Steiner, M. (1999), *The Applicability of Mathematics as a Philosophical Problem*. Cambridge, MA: Harvard University Press.

Suarez, M. (2003), "Scientific Representation: Against Similarity and Isomorphism", *International Studies in the Philosophy of Science* 17: 225–244.

Suppes, P. (1961), "A Comparison of the Meaning and Use of Models in the Mathematical and Empirical Sciences", in H. Freudenthal (ed.), *The Concept and Role of the Model in Mathematics and Natural and Social Sciences*. Dordrecht: Reidel, 163–177.

——— (1967), "What Is a Scientific Theory?", in S. Morgenbesser (ed.), *Philosophy of Science Today*. New York: Basic Books, 55–67.

——— (1974) "Recent Progress in Gauge Theories of Weak, Strong and Electromagnetic Interactions", *Reviews of Modern Physics* 46: 255–277.

Tarski, A. ([1935] 1956), "The Concept of Truth in Formalized Languages", in A. Tarski, *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*. Oxford: Clarendon Press, 152–278.

——— ([1936] 1956), "The Establishment of Scientific Semantics", in A. Tarski, *Logic, Semantics, Metamathematics: Papers from 1923 to 1938*. Oxford: Clarendon Press, 401–408.

——— (1953), "A General Method in Proofs of Undecidability", in A. Tarksi, A. Mostowski, and R. M. Robinson (eds.), *Undecidable Theories*. Amsterdam: North-Holland.

van Fraassen, B. (1980), *The Scientific Image*. Oxford: Oxford University Press.

——— (1985), "Empiricism and the Philosophy of Science", in P. Churchland and C. Hooker (eds.), *Images of Science*. Chicago: University of Chicago Press, 245–308.

——— (1989), *Laws and Symmetries*. Oxford: Oxford University Press.

——— (2002), *Representation and Invariance of Scientific Structures*. Stanford, CA: CSLI Publications.

Weinberg, S. (1967), "A Model of Leptons", *Physical Review Letters* 19: 1264–1266.

Weyl, H. (1949), *Philosophy of Mathematics and Natural Science*. Princeton, NJ: Princeton University Press.

Worrall, J. (1984), "An Unreal Image, Review of van Fraassen *The Scientific Image*", *British Journal for the Philosophy of Science* 35: 65–80.