2016

Modeling

Michael Weisberg
University of Pennsylvania, weisberg@phil.upenn.edu

Recommended Citation (OVERRIDE)

Weisberg, M. (2016). Modeling. In H. Cappelen, T.S. Gendler & J. Hawthrone (Eds.), The Oxford Handbook of Philosophical Methodology. Oxford: Oxford University Press. [doi: 10.1093/oxfordhb/9780199668779.013.26]

Modeling

Abstract

This article focuses on the methodology of modeling and how it can be applied to philosophical questions. It looks at various traditional views of modeling and defends the idea that modeling is a form of surrogate reasoning involving two distinct steps: indirect representation of a target system using a model and analysis of that model. The article considers different accounts of model/target representational relations, defending an account of similarity. It concludes by presenting several examples of the use of models in philosophy, suggestions for philosophers new to modeling, and an assessment of the relationship between thought experiments and models.

Keywords

philosophical methodology, modeling, surrogate reasoning, models, target systems, philosophy, thought experiments

Disciplines

Philosophy

Oxford Handbooks Online

Modeling

Michael Weisberg

The Oxford Handbook of Philosophical Methodology
Edited by Herman Cappelen, Tamar Szabó Gendler, and John Hawthorne

Print Publication Date: May 2016
Subject: Philosophy, History of Western Philosophy (Post-Classical), Epistemology
Online Publication Date: Aug 2016 DOI: 10.1093/oxfordhb/9780199668779.013.26

Abstract and Keywords

This article focuses on the methodology of modeling and how it can be applied to philosophical questions. It looks at various traditional views of modeling and defends the idea that modeling is a form of surrogate reasoning involving two distinct steps: indirect representation of a target system using a model and analysis of that model. The article considers different accounts of model/target representational relations, defending an account of similarity. It concludes by presenting several examples of the use of models in philosophy, suggestions for philosophers new to modeling, and an assessment of the relationship between thought experiments and models.

Keywords: philosophical methodology, modeling, surrogate reasoning, models, target systems, philosophy, thought experiments

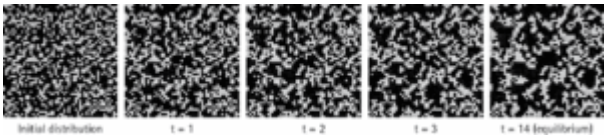
1. Introduction

I live on a racially diverse block in South Philadelphia. A little more than half of my block is Caucasian, a little less than half is African American, and the rest of the people are of Asian or Latino descent. Let’s imagine three things about my block and the city it is in: First, imagine that everyone else on the block values living in a diverse neighborhood. Second, let’s imagine that there is some comfort threshold that everyone on the block has. If, say, less than 30% of the block was African American, the African Americans currently living here might feel uncomfortable and decide to move. Finally, let’s imagine that the whole city has this preference structure.

What will happen to the city in the long run? Will the city gradually move towards more and more integrated blocks like mine (because people value diversity)? Will blocks be relatively integrated, but with some being 30% African American and some 30% Caucasian (because those are the “floor” thresholds)? Will the city become even more integrated? Or will it become more segregated (because people aren’t comfortable being in a very small minority)?

I have a very hard time imagining what would happen in this scenario. When I have asked friends about it, many of them see the 30% threshold as especially salient and think we will find pockets of 30% Caucasians and 30% African Americans. But almost no one who comes up with the right answer, which economist Thomas Schelling discovered (1978) by constructing a model.

Schelling’s original model was concrete, consisting of a chessboard, dimes, and nickels. The squares of the chessboard represented addresses in a city, dimes and nickels represented households consisting of people from two racial groups, which I will call *A* and *B*. The dimes and nickels were distributed randomly throughout the board.



Click to view larger

Figure 15.1 Computer simulation of Schelling’s segregation model. On the left is shown a random distribution of the agent times. As time moves forward, large clusters of the two agent types form.

Besides the individuals and their initially random spatial layout, the model also contained a utility function and a movement rule. The utility function said that each individual prefers that at least 30% of its neighbors be of the same type. So the *A*s want at least 30% of their neighbors to be *A*s and likewise for the *B*s. Schelling’s neighborhoods were defined (p. 263) as standard Moore neighborhoods, a set of nine adjacent grid elements. An agent standing on some grid element *e* can have anywhere from zero to eight neighbors in the adjoining elements.

The model is made dynamic by a simple movement rule. In each cycle of the model, its agents choose to either remain in place or move to a new location. When it is an agent’s turn to make a decision, it determines whether its utility function is satisfied. If if is satisfied, the agent remains where it is. If it is not satisfied, then the agent moves to the nearest empty location. This sequence of decisions continues until all of the agents’ utility functions are satisfied.

When the movement rule and utility function are implemented in Schelling’s physical model, something very surprising happens: a cascade is observed which leads from integrated neighborhoods to highly segregated neighborhoods. In a modern computer implementation of this model on a 51 x 51 grid (shown in Figure 15.1), a preference for 30% like neighbors usually leads to agents having 70% like neighbors.

Schelling urged his readers to actually take out a chessboard and implement his model so they could see the model's dynamics unfold. What one sees in doing this, or reimplementing the model in a computer, is a cascade: agents that start out satisfied become unsatisfied when a neighbor leaves or a new one moves in. This leads to movement, which leads to more agents becoming unsatisfied. A small patch of dissatisfaction can result in widespread movement, and ultimately, segregation. While there are a few agent configurations that are integrated where every agent is satisfied, these states are very rare and nearly impossible to generate from random agent movement. Thus, Schelling's major result is that small preferences for similarity can lead to massive segregation. This result is quite robust across many changes to the model including different utility functions, different rules for updating, differing neighborhood sizes, and different spatial configurations (Muldoon, Smith, and Weisberg, 2012). Schelling's model does what my imagination couldn't do. It predicts that my integrated block in South Philadelphia is unstable in the long run. Although the model is idealized in many ways, it says that over time, my block is likely to become homogenous if the movement rules and utility function of the model are anything like the ones that real people have.

Although Schelling's model is a simple one, I think it very nicely illustrates how an idea can be sharpened using a model. This sharpening forces us to examine our explicit and implicit assumptions, and extends the reach of our imaginative capacities, allowing (p. 264) us to explain and predict complex phenomena that are difficult or impossible to gain a complete cognitive grasp of. This chapter explores the methodology of modeling, showing how it has been applied to philosophical questions and can continue to do so in the future.

2. What is Modeling?

Modeling is a form of surrogate reasoning, a practice in which one constructs and analyzes a model in order to learn, indirectly, about something else. Most commonly in both scientific and philosophical contexts, models are simpler than the real world target systems they represent and they are idealized relative to these targets.

Surrogate reasoning involves two steps: indirect representation of a target with a model and analysis of that model. One first constructs or acquires a model, and specifies the intended target of that model. This step does not involve extensive empirical or conceptual interrogation of a target and construction on the basis of inference from the properties of that target. Schelling didn't derive his model from a detailed description of Philadelphia or some other city. Instead, he asked himself about some of the essential properties of a city, and used those to create his model. So a model shouldn't be thought of as simply as a representation of a target, but rather an intermediary between the target and an analysis. This is why I call model-based reasoning "surrogate reasoning."

After constructing or acquiring the model, one subjects it to analysis. Techniques of analysis vary widely, and depend on both the type of the model and the question of interest. But typically one is interested in understanding the properties of various features of the model, and especially how some mechanistic features give rise to other behavioral features. Sometimes we try to give complete analyses, uncovering everything there is to know about the model. More often, analyzes are goal-directed, trying to answer specific questions. For example, Schelling's model can be used to answer questions about the tipping points or thresholds of segregation.

Modeling can be contrasted with *direct representation and analysis*. In this style of theorizing, one begins by representing a target system using what one knows about the target to generate an accurate representation. Although approximations and idealizations may enter the representation for pragmatic or epistemic reasons, the goal is to depict the features of some target system. If we wanted to study segregation by direct representation, we would look carefully at a time series of demographic information, such as data about each census tract. From this data, one could construct a representation of city migration patterns. One might also try to infer likely future patterns, or even the psychological motivations underlying them. It is possible that this would generate something like the Schelling model, but the procedure by which it was constructed would be very different. In direct representation, we analyze the system itself. In modeling, we study a constructed intermediary. The difference is one of practice and procedure, not necessarily the end product (Weisberg, 2007).

One of the virtues of modeling is that models are extremely flexible tools. They can be used to study a single target, a cluster of targets, a generalized target, or even targets known (p. 265) not to exist. In philosophical contexts, models are rarely used to study a single target. Instead, they are most often used to answer *how possibly questions* (Dray, 1968; Resnik, 1991; Forber, 2010) or *what would happen if questions*, and hence usually have generalized systems as their targets. And sometimes, the targets of philosophical models are themselves hypothetical systems which do not exist, such as a perfectly just society, or a universe with only two particles.

3. Models

What kinds of things are models? This basic and central question has remained surprisingly controversial in the philosophical literature. Some philosophers, especially those who defend the semantic view of theories (e.g., Suppes, 1960; Suppe, 1989) argue that scientific models are the same kinds of things as logician's models. The motivating idea for this view is that theories should be language independent. Although we may describe theories with words, equations, and diagrams, they should not be tied to any of these descriptions. Proponents of the semantic view argue that the theory itself is a

structure which satisfies such a description. A true theory, then, is a structure which is isomorphic to structures in nature. Models are thus a kind of mathematical structure.

More recent proponents of the semantic view, especially Bas van Fraassen (1980) and Elizabeth Lloyd (1994), argue that models are sets of trajectories in a state space. A state space is a set of points corresponding to the properties of a system. They are organized in such a way that each dimension of this space is an independent way that the state can vary. Trajectories through the space are time ordered sets of states that describe the temporal evolution of a model system. When there is some kind of match between the trajectories in the state space and trajectories corresponding to the target system, we have a model of the target system.

Another traditional view about models sees them as complements to mathematical theories, and hence not mathematical themselves. In this view, models are material analogies (Campbell, 1957; Hesse, 1966; for a dissenting view, see Duhem, 1906) they allow a scientist to develop an intuitive picture of a complex mathematical principle by comparison to something well-understood and concrete.

Hesse and others have emphasized that many important theoretical advances have been made when theorists understood that some property of one system was materially analogous to that of another. For example, the mathematics describing the propagation of light might be accompanied by an analogy comparing the propagation of light waves to the propagation of water waves. While we no longer think it is necessary for light to propagate though a physical medium, the analogy between light and water waves allowed James Clerk Maxwell to develop the equations describing light propagation.

Although most of the philosophers writing about the nature of theories today do not emphasize material analogies, this view has remained influential in the modeling literature in two ways. First, almost all philosophers of science accept that concrete models can do important scientific work. Watson and Crick used a material model of their proposed DNA structure to make inferences about base-pair hydrogen bonding (Watson, 2011). Walter (p. 266) Newlyn and Bill Phillips constructed a hydraulic model to study how tax rates affect the British economy (Morgan, 2012). And the United States Army Corps of Engineers constructed a working tidal model of the San Francisco Bay and Delta Region in order to study what would happen if the Bay was dammed up (Weisberg, 2013).

A more controversial appeal to concrete systems can be found in a literature which asserts that all scientific models are fictional scenarios. Philosophers defending this view see all models, including mathematical models, as fictional scenarios that would be concrete if they were real. So on this view, Schelling’s model isn’t an abstract configuration of states and set of transition rules, but is actually an imaginary world: a neighborhood with people, preferences, and movement rules (Godfrey-Smith, 2006; see Weisberg, 2013 for a critique).

My own view of models is that they are composed of two parts: structure and interpretation. Like the critics of the semantic view, I think models cannot simply be mathematical objects. Bare structures stand in relations to target systems, but they often have too many of the wrong kinds of relations, and not enough of the right kind. However, mathematical, computational, or concrete structures, suitably interpreted, can stand in the right kinds of relations to represent features of targets. I call the relevant interpretations modelers' construals.

Construals provide an interpretation for the model's structure, set up relations of denotation between the model and real-world targets, and give criteria for evaluating the goodness-of-fit between a model and a target. They are composed of four parts: the *assignment*, the modeler's *intended scope*, *dynamical fidelity criteria*, and *representational fidelity criteria*. The assignment and scope determine the relationship between parts of the model and parts of the target system. The fidelity criteria are the standards theorists use to evaluate a model's ability to represent real phenomena.

Assignments are explicit specifications of how parts of real or imagined target systems are to be mapped onto parts of the model. This explicit coordination is especially important because although the parts of some models seem naturally to coordinate with parts of real-world phenomena, such as grid locations and addresses in Schelling's model, this is often not the case. For example, in a simple model of population growth, a population's growth is described by an exponential function. Nothing about this function suggests population size—it could just as easily signify a nuclear chain reaction. The theorist's assignment is what gives this function its meaning.

Assignments are often not made explicit in discussions of models, because communities of modelers have standard reading conventions for model descriptions. Where conventions are not explicit, are being violated, or where the modeler needs to be especially explicit, he or she will be forced to make the assignment explicit in discussions about the model.

Models inevitably have structure not present in the real-world phenomena they are being used to study. For example, Schelling's model has a perfectly regular grid and perfectly squared off edges. No actual city has these features. So are these features of the model intended to represent something about the target, or are they merely artifacts of the idealizations that went into constructing the model? A model's intended scope specifies the answer to this question, telling the theorist what parts of the model should be taken seriously.

The other aspects of a modeler's construal are fidelity criteria. While the assignment and scope describe how the target system is intended to be represented with the model, fidelity (p. 267) criteria describe how similar the model must be to the world in order to be considered an adequate representation. I divide these criteria into two types: Dynamical fidelity criteria tell us how close the output of the model—the predictions it makes about the values of dependent variables given some independent variables—must be to the output of the real-world phenomenon. Representational fidelity criteria are

more complex and give us standards for evaluating whether the structure of the model maps well onto the target system of interest. Typically, these criteria specify how closely the model's internal structure must match the causal structure of the real-world phenomenon to be considered an adequate representation.

For example, say that Schelling's model of segregation was targeted at the city of Philadelphia. One way to evaluate the model is with very high-fidelity criteria. If we did this, then the model's predicted equilibrium state, as well as the dynamics leading to that state, the utility functions of the agents, the movement rules, and so forth would be compared with the city's distribution of racial groups, looking for a very close match. Another way to evaluate the model is with a qualitative, not quantitative criterion. Yet another kind of fidelity criterion says that the model should be regarded as a how-possibly model, qualitatively matching the segregation patterns of the city, but with no expectation that the movement rules and utility functions were realistic.

Fully describing fidelity criteria requires an account of the model/target relation. If one thinks of models as (ideally) true descriptions of targets, fidelity criteria simply become an error tolerance, specifying how far one can deviate from truth. But when one has an account of the model/target relation that takes into account the highly idealized nature of many contemporary models, the situation is more complex.

Along with a concrete, mathematical, or computational structures, theorists' construals generate models. To say that a model is structure plus interpretation means that models are structures whose parts are interpreted via their assignments. They can potentially denote parts of a target as specified by the theorists' intended scope. And they are evaluated by the theorists' fidelity criteria. These four components of the construal constitute the theorists' interpretation of the model.

Whatever view about the nature of models is adopted, it is important to distinguish between models and their descriptions. Model descriptions specify models, and stand in many-many relationships to them. A single model might be described by words, equations, or diagrams. And any imprecision in a model description, including parameters left as dummy variables, will specify multiple models. Scientists often refer to equations as "models," but I think it is important to see equations as descriptions. Models' structures should be seen as independent of the way they are described.

4. Target Systems

Models are not compared directly to real phenomena, but to target systems, which are abstractions over these phenomena. The reason for this is that phenomena have many more properties than are represented in even the most realistic models. So when a modeler is ready to start comparing her model to the world, she constructs a target. She does so by **(p. 268)** identifying a spatio-temporal region of interest and the contents of

that region of interest. In scientific cases, the choice of target is driven by the research question of the scientist, specifically, which part of the empirical world is under investigation. Philosophical cases allow somewhat more latitude. Sometimes philosophers are interested in actual extant practices. Other times, they are interested in ideal scenarios such as conditions of perfect justice, or universes with minimal structure. Still other times, the goal of philosophical modeling can be the investigation of concepts, and there are no real or imagined targets for models.

Whatever the case, when a model is targeted at a real or imagined system, it represents only some parts of that system. Theorists must abstract away from the full richness of phenomena and aim their models at a set of features of a real-world phenomenon. For example, say I was interested in modeling a communication system. We might start by identifying the real-world phenomenon of people speaking English to one another. But this phenomenon is far too complex to capture in a model, so the modeler must decide which features to focus on. If the model was being constructed in philosophy of language, perhaps in order to investigate questions about intentionality, we might work with a very abstract target consisting of a set of symbols, states of the word, and transmission channels. However, a linguist would want to include many more details about the nature of language in her target, a communications engineer would include more about the transmission system, and so forth.

This example shows that the relationship between real-world phenomena and targets is one-to-many, which opens the door for a massive proliferation of target systems. Since there are so many different targets that can be generated from the same phenomenon, does anything go? Are there standards that govern the kinds of abstractions that theorists make?

Alkistis Elliott-Graves (ms) has argued that the answer to this question is no. Although many targets can be generated from one phenomenon, there are general norms for constructing appropriate targets. She argues that target system generation should be thought of as consisting of two conceptually distinct stages. Modelers partition the phenomenon into sets of features and then they abstract from these features in order to generate the target. Partitioning, she argues, is guided by the pragmatic norm of usefulness. The modeler should ask whether the relevant features for the topic of investigation get captured by the partition. Abstraction is more highly constrained by the norm of *aptness*, limited to what one can omit without distortion. Whether or not one accepts Elliott-Graves' account, it seems right to say that the enormous latitude of targets is not limitless. The flexibility it affords is positive, but the pragmatics of modeling impose limits.

Philosophical contexts, and, to be sure, some scientific ones, do not always require targets that are abstractions over real-world phenomena. More specifically, constructing and analyzing models of targets known not to exist (e.g. perpetual motion machines, time traveling bricks, or single particles alone in the universe) have played important roles in scientific and philosophical modeling. Sometimes, models are studied simply for their

own sakes, without any target at all in mind. A good example of the latter category is Conway's Game of Life cellular automaton (Gardner, 1970). This model consists of an array of cells, which can each be in an alive state or a dead state. Transition rules determine how the states change, and these rules typically depend on the states of neighboring states.

(p. 269) One of the reasons we study models without targets is in order to help to sensitize our imagination so that we learn how to notice things we might have missed otherwise when looking at real targets. For example, Dennett discusses the interesting fact that when we begin thinking about the Game of Life, we start by describing a grid, cells, and the rules for each cell. But fairly soon we are talking about the patterns and apparent motion in the game.

> Note that there has been a distinct ontological shift as we move between levels; whereas at the physical level there is no motion, and the only individuals, cells, are defined by their fixed spatial location, at this design level we have the motion of persisting objects; it is one and the same glider that has moved southeast … Here is a warming-up exercise for what is to follow: should we say that there is real motion in the Life world, or only apparent motion? The flashing pixels on the computer screen are a paradigm case, after all, of what a psychologist would call apparent motion. Are there really gliders that move, or are there just patterns of cell state that move? And if we opt for the latter, should we say at least that these moving patterns are real?
>
> (Dennett, 1991)

Nevertheless, many of our most important cases of modeling are target-directed. A suitable target is chosen and the model is coordinated to that target by the modelers' construal. When that happens, what kinds of relations must a model stand in to its target?

5. Model/Target Relations

There are two types of accounts of model/target relations in the literature: *model*-theoretic accounts and *similarity* accounts. Model-theoretic accounts are the dominant view. Like other aspects of the modeling literature, they find their original home in discussions of the semantic view of theories. Such accounts typically posit that models must be *isomorphic* to their targets, although some proponents of the semantic view have weakened the requirement to *homomorphism* (Lloyd, 1994), or *partial isomorphism* (da Costa and French, 2003).

Isomorphism is a mapping between two sets that preserves structure and relations. Formally, an isomorphism is a bijective map between two sets such that the mapping function f and its inverse are both homomorphisms, structure-preserving maps between

these two structures. This account of the model/target relation remains influential, but many philosophers have argued that it cannot appropriately deal with the relationship between idealized models and their targets (e.g. Hendry and Psillos, 2007).

As an alternative, Steven French and colleagues have offered the partial isomorphism account. Proponents of this account say that the model/target relation is tripartite, corresponding to the part of the model that is isomorphic to the target, the part that is not isomorphic to the target, and the part that is "left open" with respect to the target. A model is partially isomorphic to its target when a substructure of the model is isomorphic to a substructure of the target. Such an account can deal with some kinds of idealized models. For example, consider the idealization of elastic collisions that is associated with the ideal gas model. This idealization says that when two particles of the gas collide, the pair (p. 270) maintains their combined kinetic energy after the collision. This is not true for molecular gases (hydrogen gas, oxygen gas, water vapor, etc.) because when they collide, some kinetic energy is transferred to the molecules' internal degrees of freedom (internal rotations and oscillations). However, the truth of this idealization is not required for many ideal gas model-based explanations and, in these cases, the idealized model can be confined to the non-isomorphic substructure without any loss.

But this will only handle idealization up to a point. In many cases, it is the idealized features of models themselves that are supposed to be representations of targets' features. For example, the Schelling model's idealized features, such as agents' utility functions and spatial distribution, are the very things that represent properties of real people and do the model's explanatory work. This has led some philosophers, including me, to look elsewhere for an account of model/target relations.

Similarity accounts posit that the model/target relation is one of similarity: a good model is similar to its target in certain respects and degrees (Hesse, 1966; Giere, 1988, Godfrey-Smith, 2006). Proponents of this type of account tend to defend their account along two lines. First, they argue that model-theoretic accounts do not have the resources to account for the relationship between the most common type of model and its target: idealized models relating to realistic targets. Second, they argue that modelers often talk and think about models as if they resemble their targets. This is taken as evidence for the nature of the relation.

There is a long tradition of skepticism about similarity. In "Natural Kinds," W. V. O. Quine argued that similarity was "logically repugnant" (Quine, 1969) because it couldn't be analyzed in terms of more basic notions. He also thought that mature sciences would dispense with similarity all together. In a more detailed discussion, Nelson Goodman (1972) agrees with Quine and adds another challenge. He argues that similarity is too promiscuous a relation to do any philosophical work. For any three objects, there will always be some respect in which two of the objects resemble each other more than the third. This, Goodman argues, shows that there can be no context-free similarity metric.

For many philosophers, this was the end of the matter. Positing similarity as the model/target relation was a dead end. Others took the criticism to be a constraint on a reasonable account of similarity; it must be a context-relative relation. For example, on Giere's (1988) account, a model must resemble its target in certain "respects and degrees." Cartwright (1983) argues that the relevant similarity between models and their targets is "behavioral similarity," meaning the similarity of the model's and the target's causal structures.

While this criticism of Goodman's was simply taken on board, Quine and Goodman also challenged proponents of similarity to give a reductive analysis, showing how some particular model and some particular target could be more similar to each other than other random models and targets. Much less has been written on this question, but some of my own work attempts to give such an analysis.

In *Simulation and Similarity: Using Models to Understand the World*, I argue that we can analyze model/target similarity in terms of weighted feature matching, and idea that has its origin in Amos Tversky's *contrast* account of similarity (Tversky, 1977; Tversky and Gati, 1978). The basic idea is that a model's similarity to its a target is a function of the features it shares and the features it doesn't share. Because Goodman is correct and there is (p. 271) no general, context-free account of similarity, some of a model's and a target's features are weighted more heavily than others.

The account can be developed as follows: First, we begin a set of features Δ. This feature set can contain quantitative or qualitative predicated, including "is purple," "is to the left of ξ," "will rain with probability 0.9," and so forth. Further, for model M and target T, m is the set of features in Δ possessed by M and t is the set of features in Δ possessed by T. Modelers' fidelity criteria also implicitly provide a weighting function $f(\bullet)$, which is defined over the power set of Δ. The overall similarity of the model to the target is given by an equation of this form[1]:

$$s(m, t) = \frac{f(M \cap T)}{f(M \cap T) + f(M - T) + f(T - M)}$$

When modeling, it is customary to distinguish between the overall properties and patterns of a system (often called the "output" in computational and mathematical modeling) from the underlying mechanisms generating these properties. I will call the first set of properties *attributes*, and the second set *mechanisms*. It is important for many kinds of modeling, including philosophical modeling, that they be distinguished. The reason for this is that in some instances of modeling, we care far more about feature matching between one or the other type of feature.

As an example, we can return once again to Schelling's model. When the model comes to equilibrium, it contains racially segregated clusters driven by agents' utility functions and rules for movement. Attributes such as degrees of clustering are states of the model and mechanisms such as agents' movement rules are the transition rules of the model. Insofar as Schelling's model explains segregation in actual cities, then there has to be some

relation between the model's attributes and the city's attributes. And there has to be some relation between the model's transition rules and the actual mechanisms that drive segregation in the city.

Now consider two other uses of Schelling's model. If it is used to ask a "what would happen if?" type of question, such as I opened this chapter with, all that is required is that the mechanisms of the model match the mechanisms in the scenario I wished to investigate. It might also be used to answer a how possibly question. Most American cities are highly segregated, even if their overall racial breakdown is mixed. What could possibly cause a racially mixed city to be segregated? Schelling's model offers one answer to this question. In evaluating a how possibly question, all that is required is that the attributes of the target (racially segregated neighborhoods in this case) are shared by the model.

Summing up the last few sections, we can say that most cases of modeling involve indirect representation, where a model is studied in order to learn about some real-world target. Models are interpreted structures which stand in relationships of similarity to target systems. Target systems are, in turn, parts of real-world phenomena.

(p. 272) 6. Asking Philosophical Questions with Models

Thus far, I have primarily spoken about modeling in a scientific context. I have done so both because modeling is most at home in the sciences, and because the philosophical literature about modeling is mostly about scientific modeling. In this section, I turn to several examples of the use of models in philosophy.

6.1 Fairness and the Social Contract

One well-known instance of philosophical modeling involves the application of game theory to foundational issues in political philosophy. Philosophers have long been interested in the question of why rational, egoistic agents would develop the sense of fairness and justice that seems to be the heart of stable political arrangements. Looking to the early modern tradition, Jean Hampton (1988) and David Gauthier (1986) showed that some of Hobbes' arguments could be reformulated as game-theoretic models. More recently, Brian Skyrms (1996) and his students have further developed these ideas and applied evolutionary game-theoretic models to question about the origins of our sense of fairness.

To take just one example from this rich literature, let's consider the game called Divide the Cake. In this game, two players are given a chocolate cake and they have to figure out

a strategy to divide it before it spoils. Each player asks for a certain fraction, and as long as those fractions add up to 1 or less, then both get some cake.

As Skyrms points out, the intuitively correct answer is also the fairest one: both should ask for half the cake. But there is nothing special about this solution. All divisions that add up to a whole cake are *Nash equilibria*, meaning that neither player could be better off changing her strategy unilaterally if that division is employed.

Despite the infinite number of equilibria, we have a very strong sense that the fair division is a 50/50 split. How could that be? What Skyrms was able to show is that when a population of dividers repeatedly plays the game and modifies their strategies according to the ones that lead to the biggest payoffs in cake (formally, employing the replicator dynamics), then the fair strategy can evolve. Skyrms estimates that with nothing else added to the model, the 50/50 division evolves 62% of the time (1996). However, once correlation—interacting with some players more frequently than others—is added to the model, then the fair split evolves most of the time. Skyrms writes:

> In a finite population, in a finite time, where there is some random element in evolution, some reasonable amount of divisibility of the good and some correlation, we can say that it is likely that something close to share and share alike should evolve in dividing-the-cake situations. This is, perhaps, a beginning of an explanation of the origin of our concept of justice.

Thus, Skyrms is able to show that in repeated interactions with correlation under an evolutionary dynamic, which might be instantiated in cultural evolution as much as in biological evolution, fairness norms begin to establish themselves in the population. Skyrms himself primarily offers this as an explanation of the origin of these norms, a sort of genealogy of (p. 273) morals. But some philosophers take this kind of modeling to have normative conclusions. Assuming that we are primarily self-interested, norms of reciprocity and fairness are justified by their good outcomes, as judged by an egoist.

6.2 Meaning and Signaling

A second case of philosophical modeling concerns the origin of meaning. In Lewis' *Convention* (1969), he introduces a game-theoretic analysis of convention and uses this analysis to investigate how communication can arise without a prior shared language or communication system.

In the two-agent version of the model, we imagine that the first agent (the sender) observes the world is in some state S_i for which the second agent (the receiver) ought to perform action R_i . The agents are cooperative such that for each i, if the world is in S_i the sender wants the receiver to perform R_i . In order to achieve this, for each observation, the sender sends a signal σ_j which is received by the receiver. The receiver's contingency

plan specifies the action R_i that will be performed for each signal σ_j . Both sender and receiver aim to achieve a signaling system with the following structure:

$$\begin{aligned} &S_1 \Rightarrow \sigma_1 \Rightarrow R_1 \\ &S_2 \Rightarrow \sigma_2 \Rightarrow R_2 \\ &S_3 \Rightarrow \sigma_3 \Rightarrow R_3 \\ &\quad\vdots \end{aligned}$$

To illustrate this kind of signaling system, Lewis asks us to consider a signaling system that may have been established between the sexton of the Old North Church and Paul Revere. If the sexton saw the British army staying at home, he would hang no lanterns in the belfry. If they set out to attack by land, he would hang one lantern. And if they set out to attack by sea, then he would hang two lanterns. Revere would stay home if he saw no lantern, warn of a land attack if he saw one lantern, and warn of sea attack if he saw two lanterns. The connection between number of lanterns and type of attack is, of course, purely arbitrary. One lantern might have caused Revere to warn of a sea attack. All that matters is that there is coordination between the sexton and Revere such that the signal leads to the right outcome.

In the actual historical case, Revere and the sexton agreed on the code. But what if they hadn't? Could this signaling system evolve by itself? More generally, when two agents (or organisms) have a common interest in communicating, but no shared language, can such a system evolve? Skyrms (2010) investigated this question by adding learning (change within an organism's lifetime) and the replicator dynamic (evolution between the lifetime of organisms) to the Lewis signaling model.

To investigate learning, Skyrms coupled a Lewis signaling model and Herrnestein's matching law. On this probabilistic model of learning, probabilities for taking actions get updated according to the reward accumulated at previous times. When these dynamics are applied to the simplest case involving two signals and two actions, the signaling system (signal 1 leads to action 1, signal 2 leads to action 2, …) is very quickly learned. Skyrms also investigated this process from an evolutionary point of view. When the same type of setup (p. 274) is allowed to evolve under the replicator dynamics, signaling systems are the only stable equilibria. Things get more complicated with greater numbers of signals and actions, but the overall lesson is the same.

6.3 The Division of Cognitive Labor

The final example that I will discuss comes from philosophy of science. There was a long tradition in philosophy of science that saw ideal scientists as impartial, cooperative, motivated by the truth, and always striving to do highly significant work. Although philosophers knew that real scientists could fall short of these ideals, they took this to be a more-or-less accurate description of scientists most of the time and the ideals that

scientists should strive for. Scientific communities function better when they cooperated in order to find out the truth.

This picture has been called in to question by historians, sociologists, and philosophers. The classic source for such doubts is Thomas Kuhn's *The Structure of Scientific Revolutions* (1962). Kuhn argued that much of scientific inquiry took the form of "articulating the paradigm," a kind of incremental, puzzle-solving activity. He also argued that in times of scientific revolution, resolution of theoretical controversy had more in common with religious conversion experiences then with rational discourse. Others following in this tradition emphasized all the non-rational and even irrational qualities that characterize scientific behavior. Instead of priests in lab coats, powerful scientists look like mafia bosses, and their underlings like foot soldiers.

Let's say that the historians and sociologists are right, and that much about science seems less than rational. What are the epistemic consequences for the scientific enterprise? Does lack of communication, lack of epistemically pure motivation, and a focus on small-scale puzzles mean that we have to change our views about the authority and productivity and science? A number of philosophers have tried to address these questions by modeling, and the resultant literature is about the division of cognitive labor.

Among the best known philosophical models in this area are those offered by Philip Kitcher (1990) and Michael Strevens (2003). Kitcher and Strevens focus on the question of motivation: What happens when scientists have motives other than the truth? Specifically, what happens when individual scientists are motivated by getting credit for important discoveries (presumably because this leads to fame, higher ranked positions, and money), rather than simply learning about the truth.

Imagine that the scientific community has a particular scientific goal in mind—in his original scenario, Kitcher describes something like the race to find the structure of DNA. In order to reach that goal, there are *n* research approaches that might be taken. Each research approach has a success function, whose input is the number of scientists taking that approach and whose output is the probability that the problem will be solved using that approach.

With this type of model, we can ask two key questions: What is the optimal distribution of cognitive labor? And how will scientists' motivations, including non-epistemic motivations such as those for prestige and credit, lead to different allocations.

On the basis of such models, Kitcher and Strevens argue that classical epistemic norms will lead scientists to misallocate their cognitive labor. If a classically rational, **(p. 275)** truth-seeking agent followed the procedure above, then he or she would join the project with the highest probability of success. But this isn't always what the scientific community as a whole wants to see happen. Maximizing the chance at success might involve distributing scientists across projects, not just to the projects with the best chance of success.

What if scientists were motivated by the accumulation of credit, the recognition that comes from being the first to make a discovery? To answer this question, let's assume that the scientific community adopts the *Priority Rule* (Merton, 1957), that most or all of the credit for a discovery goes to the scientist who makes the discovery first. In this case, scientists will want to take into account both the probability of success of the project and the probability that they will be the first one to complete the project. The first consideration pushes scientists towards the project with the overall highest probability of success, but the second consideration pushes scientists towards projects that have fewer scientists working on them.

To investigate this question more fully, Strevens (2003) models a representative agent who is poised to enter the field for the first time. If this agent can choose between n projects, and knows the current distribution of scientists to projects and the success functions of these projects, which one would she choose? Strevens shows that if the scientific community allocates credit according to the Priority Rule, then the community as a whole achieves the optimum division of cognitive labor. This, he argues, explains why the scientific community has adopted the Priority Rule, the rule that whoever discovers something first gets all the credit. So even if scientists strive for credit instead of truth, the scientific community is well functioning. Further, it might actually function better if scientists strive for credit than if they are only interested in the truth.

Another line of research in this area has looked at cooperation in science. Famous laboratories such as the wartime Los Alamos nuclear weapons laboratory (Rhodes, 1987) and MIT RADAR laboratory (Galison, 1997) planned ways for scientists to continuously integrate their findings and share ideas with one another. Technological innovations such as the Internet and rapid forms of electronic publication make this possible on a much wider and geographically distributed scale. According to classical norms of scientific inquiry, this is unambiguously good and should be encouraged. But it this really true? Might such communication lead to the propagation and fixation of errors as well as knowledge?

Epistemic network models allows us to investigate these questions (Zollman, 2007; see also Grimm et al., forthcoming). In such models, lines of communication between scientists are represented by network graphs, such the ones seen in Figure 15.2. Each node of these graphs represents a scientist and each edge a communication channel. By altering the connectivity of the graph, from the minimally connected cycle to the maximally connected complete graph, we can represent different types of scientific communication—from maximal to minimal.

Click to view larger

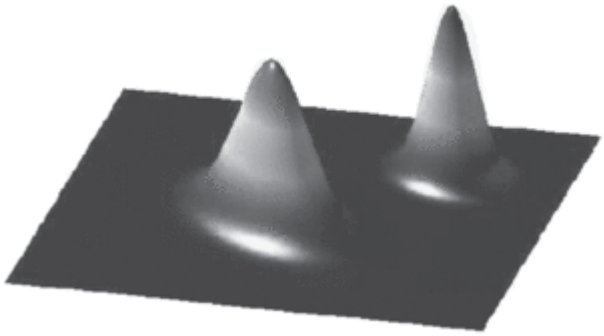
Figure 15.2 Three epistemic networks explored by Zollman. The nodes represent agents, and the edges represent lines of communication.

Images courtesy of Professor Zollman

How can this be used to investigate norms about communication? In a recent paper, Zollman considers networked scientists trying to decide what propositions are true of the world. Imagine that scientists are trying to determine whether the world is in state $s1$ or $s2$. They get information from their own experimentation and also from the other scientists they are connected to. On the basis of this information, they update their beliefs in a standard Bayesian way. When their beliefs reach some threshold (i.e. the probability is high enough), then they decide what to believe. The main independent variable in the model is connectedness of scientists. **(p. 276)**

Zollman's model generates two especially interesting results. First, scientists connected in a cycle converged to the truth more often than scientists connected in a wheel or in a fully connected graph. This suggests that careful limiting of information available to scientists may have certain advantages. Or to put it another way, less well-informed scientists might have an advantage over more well-informed ones, if the goal is to minimize error. However, when more communicative communities converge to the truth (or a falsehood), they do so more rapidly. For the ten-scientist communities Zollman studied, those on complete networks converged about five times faster than those in the cycle network. So one might conclude from this that too much communication is a bad thing. Scientific communities may be better off when they partially limit communication, to ensure that the wrong answers aren't locked in too quickly.

A final line of philosophical modeling in this area considers how scientific communities discover significant significant scientific truths. Epistemic landscape models (Weisberg and Muldoon, 2009) investigate the ways that scientists choose what kinds of problems to work on and the approaches they take in order to do so. They begin by postulating a set of approaches, narrow specifications of how a research topic is investigated. These approaches are then organized by their mutually independent properties. Each one of these properties is represented as a dimension in an epistemic landscape, whose points correspond to approaches. An additional dimension corresponds to the epistemic significance of the approach, giving a topography to the landscape where peaks correspond to the most highly significant approaches, as in Figure 15.3. However individual scientists are motivated, a socially optimal outcome would be one where the peaks are found and so are the many highly significant regions that are not peaks.



Click to view larger

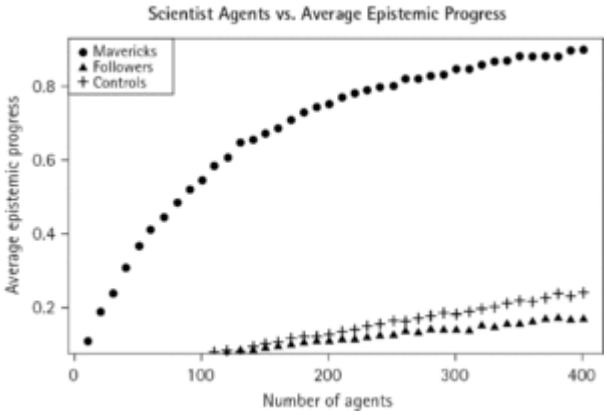
Figure 15.3 A low dimensional epistemic landscape investigated in Weisberg and Muldoon, 2009. The *x* and *y* dimensions correspond to aspects of the research approach and the *z* axis corresponds to degree of epistemic significance.

Weisberg and Muldoon, 2009

Scientists are represented in epistemic landscape models as agents who make strategic choices about what approaches to take. They get feedback from the landscape about the significance of the approaches they have taken, and have the possibility of communicating with other agents about the significance of the approaches that these other agents have taken. So an exploration strategy will be the rules an agent follows in determining which approaches to adopt in each cycle of the model. Should it keep the current approach, or move on? Should it take into account what others are doing? If so, how should this information be taken in to account. (p. 277)

In one model, Muldoon and I investigated the last of these questions. We looked at two extreme ways that scientists can take account of what others are doing in their search for approaches of high significance. Followers think that the best way to find more significant truths about the world is to find the approach which has yielded the highest significance so far, and move in that direction. This is simulated in several steps. At the beginning of each cycle of the model, followers examine the patches in their Moore neighborhood, the eight approaches immediately adjacent to the one on which they are currently located. These patches correspond to the most conceptually similar approaches to the agent's current approach. Model agents then move to the previously explored approach of maximum significance in their Moore neighborhood, if such an approach is available. Like followers, mavericks pay attention to what others are doing, but they use this information differently. Instead of moving towards approaches yielding high significance, mavericks move away from explored territory.

When first working on this project, Muldoon and I expected the followers to do quite well because the strategy is essentially imitative. We predicted that the followers would help each other get to the "frontier" of unexplored knowledge, then they would spread out and discover what there as to discover. Mavericks, we predicted, would do less-well. Since they are always adopting new approaches, they never allow themselves to build on the knowledge of those that came before them, which seems to create a disadvantage if they are trying to find the peak of the epistemic landscape.



Click to view larger

Figure 15.4 Epistemic progress of communities of agents of different types.

Weisberg and Muldoon, 2009

As it turns out, our imagination led us widely astray. When we take these ideas about maverick and follower strategies and implement them as models, we see a very different result. For ease of visualization, let's look at a simple, three-dimensional landscape: two dimensions correspond to ways that approaches can vary, a third corresponds to epistemic significance. At the beginning of a simulation, scientist-agents are placed in random, low-significance regions of the landscape. They are then allowed to implement the follower (p. 278) strategy or the maverick strategy. The results are striking: the mavericks massively outperform the followers.

Figure 15.4 shows the *epistemic progress* (fraction of significant approaches investigated) against the number of scientists of different types. As we can see from this graph, even small populations of mavericks massively outperform followers. So our intuition that followers would be effective was incorrect. In fact, we also were able to show that populations of followers do no better, and sometimes do worse, than populations that don't share any information. This result suggests that while the sharing information can be a good thing, it can also have unforeseen consequences. Sometimes the community is better off "spreading out" in epistemic space, not simply building on the best that came before.

I opened this section by saying that the traditional view of scientists was that they were impartial, cooperative, motivated by the truth, and always striving to do highly significant work. Kuhn and others called this picture into question, and drew radical epistemic conclusions from it. The philosophical modeling described in this section shows some of the ways that these less-than-ideal individual epistemic virtues may nevertheless be beneficial at a societal level.

7. Relationship Between Thought Experiments and Models

Throughout this chapter, I have often introduced the models I was discussing in the way one introduces thought experiments. I might introduce Schelling's model by saying, "Imagine a city where all the houses are arranged on a grid." This certainly sounds very similar to many philosophical thought experiments: "Suppose that I'm locked in a room and given a large batch of Chinese writing" (Searle, 1980). "Suppose you are the driver of a (p. 279) trolley. The trolley rounds a bend, and there come into view ahead five track workmen, who have been repairing the track" (Thomson, 1985).

In this section, I want to think about this apparent similarity between thought experiments and simulations and ask a couple of questions: Are models and thought experiments the same kind of thing? Are models better than thought experiments? Should models replace thought experiments in philosophical theorizing?

7.1 Thought Experiments and Models

Although there is a large philosophical literature about thought experiments (e.g. Gendler, 2000a; Sorensen, 1992), there is little consensus about what kind of thing thought experiments are. Most philosophers accept that thought experiments are imaginary scenarios, and that is how I will understand them in this chapter. If thought experiments are imaginary scenarios contemplated in order to help us learn something about the world, then they look very much like models. The main difference is the role of imagination: I have argued that models are interpreted concrete, mathematical, or computational structures. Thought experiments, on the other hand, are products of imaginations. What is the relationship between these things?

One view is that all models are actually a kind of thought experiment. Godfrey-Smith defends what I call the *fictions' view*, arguing that even mathematical models are best understood as fictional scenarios. In discussing the ways that modelers talk and think about their models, he writes:

> I take at face value the fact that modelers often take themselves to be describing imaginary biological populations, imaginary neural networks, or imaginary economies. An imaginary population is something that, if it was real, would be a concrete flesh-and-blood population, not a mathematical object.
>
> (Godfrey-Smith, 2006)

On a more standard view of models that sees them as structures, not all models are thought experiments. There are many mathematical and computational structures that are difficult or even possible to imagine, so they can't be the same kind of things as thought experiments, unless one thinks that thought experiments do not literally have to be imagined. So on this view, the set of models is clearly much larger than the set of thought experiments. But the converse question remains: are all thought experiments models or proto-models?

It is tempting to simply say that thought experiments are models, where imaginative structures take the place of mathematical, computational, or concrete structures. A more refined way to link thought experiments to models is to say that the narrative parts of thought experiments are model descriptions for concrete or computational models. Without seeing the dynamics of Schelling's model play out on a computer screen, I can't imagine much about the scenario he describes. But here is a very simple Schelling-like thought experiment: There are 10 houses on a block and only four families, half Caucasian and half African American. Each family wants to have 50% of its neighbors be of the same race and, although they are initially placed randomly in houses, they can move freely in the block until they find a configuration that satisfies them. It is easy to imagine the equilibrium state of the model: four houses together alternating in racial makeup.

(p. 280) What was this thought experiment I was engaging in? There are clearly the same kinds of computational elements as Schelling's original model: a configuration of agents, a utility function, and movement rules. The main difference is that this case is sufficiently simple to have a good intuitive grip on, that could be checked with a computer program or chessboard. So in this case, the thought experiment is functioning in exactly the same way as a computational model.

Here is another kind of case. Gendler (2004) asks us to

> [t]hink about your next-door neighbor's living room, and ask yourself the following questions: If you painted its walls bright green, would that clash with the current carpet, or complement it? If you removed all its furniture, could four elephants fit comfortably inside? If you removed all but one of the elephants, would there be enough space to ride a bicycle without tipping as you turned?

What is happening in this case when we contemplate, along with Gendler, whether bright green would in fact clash with the carpet? Gendler argues that our judgments about these questions are made by creating the relevant mental image of green walls, carpets, elephants, and bicycles, and then forming a judgment by examining this image. She says that using a mental image to determine if elephants will fit is analogous to taking "a three-dimensional scale-model of the room, along with four similarly scaled plastic elephants … putting the elephants into the room, and seeing whether they fit" (1158). So in this case, we are using our imagination in exactly the same way that we would use a concrete model.

7.2 Advantages of Models

If thought experiments are models or proto-models, are there any advantages to being more formal and constructing full-blown models to replace thought experiments? Is it sufficient to rely on thought experiments, or should we take the construction of a fully explicit model as some kind of regulative ideal? In order to answer these questions, I

think it is worth considering some of the advantages models have over thought experiments.

In a philosophical context, the main advantages of modeling over thought experiments are explicitness, reduced inter-philosopher variation, and the ability to deal with imaginative resistance. Creating and analyzing a model is a process that involves, among other things, forcing oneself to make all assumptions explicit. Our imaginations are very flexible and we can construct an imagined scenario from a very minimal script. But when one has to derive an equation, build something out of plastic, or write a computer program, this kind of vagueness is not allowable. Programs won't compile, equations cannot be derived, and plastic models will fall apart if details are left unspecified. One of the most common experiences reported by model builders is discovering a missing or hidden assumption in the course of modeling. This involves finding a resolution, or realizing that there are multiple avenues worthy of investigation.

A second major advantage is the reduction of inter-philosopher variation. There will always be philosophical disagreements about how to assess the results of thought experiments. This is especially true in normative domains, but also true in epistemology and metaphysics cases. However, sometimes our imaginations cannot even resolve the (p. 281) phenomenon we are supposed to be judging, or we seem to think different things would happen. For example, here is a thought experiment suggested by philosopher Simon May in personal correspondence: Imagine we allowed and encouraged professors and students to carry firearms on campus. Would there be many fewer or many more campus shooting fatalities? I can imagine this scenario, and I even know what I think would happen. But I have no confidence that my imagination is able to resolve the scenario properly and, therefore, no confidence in my judgment of the case. A major advantage of modeling in such cases is a determinate answer about what would happen in such a case, given such-and-so assumptions.

Finally, there are many scenarios that we are simply incapable of imagining, a phenomenon often called imaginative resistance (Gendler, 2000b; Walton, 2006). Although morally repugnant cases of imaginative resistance have received the most attention in the literature, the complex counterfactuals, high-dimensional spaces, massive aggregation over agents, and atypical mental states that arise in philosophical thought experiments may also induce resistance. While some of these scenarios may also resist analysis by modeling, mathematical and computational models do not face the cognitive and memory limitations of humans. In such cases, a careful description of the setup and initial conditions may yield to computation, even if not to human imagination.

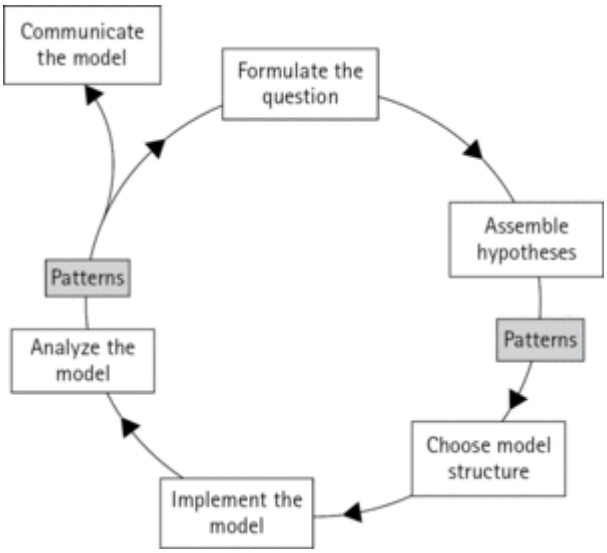
So should explicit modeling replace thought experiments in philosophical analysis? I think such a position is both too strong and premature. There are many cases where we can engage in a thought experiment, but really have no idea how we can create a model for the case. Moreover, if what I have said in this section of the chapter is true, thought experiments already have a modeling-like character, so it isn't obvious how much of an improved understanding we get by modeling. However, there are enough advantages to

modeling that a reasonable norm might be as follows: if one can construct and analyze a philosophical model, then one should attempt it. Short of actually building a model, many of modeling's internal norms such as explicitness, publicness, cycling back and forth between constructing the model, analyzing the model, and revising the model, also make good norms for thought experiment analysis.

8. How to Get Started Modeling

In this final section, I will make a few brief comments about how to get started modeling. Unfortunately, it is hard to make such comments without being so general as to be unhelpful, or so specific as to only be relevant for a particular type of modeling. I will therefore divide these comments into two sections: the first about general principles of modeling and the second about agent-based modeling, a type of computational modeling particularly well suited to many philosophical questions.

8.1 Modeling Cycle



Click to view larger

Figure 15.5 A depiction of the modeling cycle from Railsback and Grimm (2012).

Figure used with permission.

Modeling begins in much the same way thought experiments begin, by formulating a question to be answered and choosing a scenario to be investigated. Rather than the scenario (p. 282) being something imagined, however, the scenario is what the model's interpreted structure will represent. In some of the recent literature on modeling methodology, this scenario is referred to as a hypothesis under investigation. Because models are stripped down versions of real-world scenarios, the modeler must pay special attention both to what gets "put in" a model. Grimm and Railsback (Grimm, et al., 2005; Grimm and Railsback, 2012) recommend that this be done in a pattern-oriented fashion, as depicted in Figure 15.5. One identifies patterns observed in some target systems and uses these to guide the construction of the model.

> Using observed patterns for model design directly ties the model's structure to the internal organization of the real system. We do so by asking: What observed patterns seem to characterize the system and its dynamics, and what variables and processes must be in the model so that these patterns could, in principle, emerge?
>
> (Grimm et al., 2005, 987)

How one goes about translating a scenario to a model depends on the type of model one wishes to build. If one wants to build an agent-based model, then one has to decide, among other things, who or what the agents are, what the agents are trying to accomplish, what their resources are, and what rules guide their decisions. If one wishes to make a game-theoretic model, then one has to identify a game that represents the scenario, the payoff structure of that game, and if a repeated game, the space of possible strategies. And so on for other types of models.

Across many model types, some general questions arise. For example, do the model's variables represent individuals or aggregates? Is time represented? Are the transition rules deterministic, probabilistic, or stochastic? Are these rules discrete or continuous with respect to time? Once answers to these questions and others are settled, one can choose a model structure and develop a construal.

(p. 283) Once the model is constructed, the process of analysis can begin. In general, there are two possible kinds of analysis here. If the modeler wishes to engage in complete analysis of the model, then he or she she will aim to determine:

1. the static and dynamic properties of the model
2. the allowable states of the model
3. the transitions between states allowed by the model
4. what initiates transitions between states
5. the dependence of states and transitions on one another.

I refer to this list as the total state of the model (Weisberg, 2013).

Complete analysis is usually associated with relatively simple mathematical models. In such cases, one can give analytical solutions which describe the models' behavior for every initial condition and every intermediary state. For more complex models, intensive computation, along with some approximation, can generate a complete or near-complete analysis of a model. But in many cases, complete analysis is too difficult to be practical, and not necessary. In such cases, modelers engage in goal-directed analysis, where they are investigating a specific set of properties or patterns of the model.

At the beginning of this chapter, I said that modeling was the process of indirect representation and analysis, and spent some time explaining how a model can represent a target in virtue of being similar to such a target. In order to "transfer" the results of an analysis of a model to a target, we need to know something about this similarity. Since models are almost never truthful representations of their targets, we are not looking for

confirmation that the model is truthful. Rather, we are looking for validation, that the model resembles the target in certain respects, and then confirmation that the analytical results are confirmed in virtue of this validation.

8.2 Agent-Based Modeling

Much recent work in philosophical modeling, including many of the examples I have discussed, take an agent-based approach. Such models explicitly represent individuals, as opposed to aggregates with aggregate-level properties. For philosophers new to modeling, I suggest beginning with models of this type because of the availability of a straightforward, powerful, and free tool called Netlogo (Wilensky, 1999).

Netlogo is a high-level programming language, especially appropriate for creating simulations of social and natural phenomena that can be broken down into individuals or agents, which are called turtles in this framework. Although Netlogo is a powerful language that has been widely adopted by modelers across the natural and social sciences, it is very straightforward to learn due to its close association with a family of programming languages specifically designed for teaching. The original Logo language was one of the first pieces of educational software written in the late 1960s for the PDP-1.

The best way to get started with Netlogo is twofold: First, one should choose a few of its dozens of example models and explore them. Both the interface side and the programs themselves are well documented and explained. Much can be learned by modifying, (p. 284) breaking, and fixing these existing models. Second, *Agent-Based and Individual-Based Modeling* (Railsback and Grimm, 2012) is an essential textbook for beginners. It combines helpful discussions on all aspects of agent-based modeling methodology with practical advice on programming in Netlogo, and it is accessible to complete beginners.

References

Campbell, N. R. (1957). *Foundations of Science*. New York: Dover.

Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford: Oxford University Press.

Da Costa, N. C. A., and French, S. (2003). *Science and Partial Truth*. Oxford: Oxford University Press.

Dennett, D. C. (1991). Real Patterns. *Journal of Philosophy*, 88(1), 27–51.

Dray, W. H. (1968). On Explaining How-Possibly. *The Monist*, 52(3), 390–407.

Duhem, P. (1906). *The Aim and Structure of Physical Theory*. Princeton: Princeton University Press.

Elliott-Graves, A. (ms) Target Systems and Their Role in Scientific Inquiry. (Unpublished doctoral dissertation). University of Pennsylvania

Forber, P. (2010). Confirmation and Explaining How Possible. *Studies in History and Philosophy of Science Part C*, 41(1), 32–40.

Galison, P. (1997). *Image and Logic: A Material Culture of Microphysics*. Chicago: University of Chicago Press.

Gardner, M. (1970). The Fantastic Combinations of John Conway's New Solitaire Game "Life". *Scientific American*, 223, 120–3.

Gauthier, D. P. (1986). *Morals by Agreement.* New York: Oxford University Press.

Gendler, T. (2000a) *Thought Experiment: On the Powers and Limits of Imaginary Cases*. New York: Garland Press.

Gendler, T. (2000b). The Puzzle of Imaginative Resistance. *The Journal of Philosophy*, 97, 55–81.

Gendler, T. (2004) Thought Experiments Rethought—and Reperceived. *Philosophy of Science*, 71: 1152–64.

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

Godfrey-Smith, P. (2006). The Strategy of Model Based Science. *Biology and Philosophy*, 21, 725–40.

Goodman, N. (1972). Seven Strictures on Similarity. In N. Goodman (ed.) *Problems and Projects*, pp. 23–32. Indianapolis: Bobbs-Merril.

Grim, Singer, Fisher, Bramson, Berger, Reade, Flocken, and Sales (forthcoming). Scientific Networks on Data Landscapes: Question Difficulty, Epistemic Success, and Convergence. *Episteme*.

Grimm, V., and Railsback, S. F. (2012). Pattern-Oriented Modelling: A "Multi-Scope" for Predictive Systems Ecology. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367 (1586), 298–310.

Hampton, J. (1988). *Hobbes and the Social Contract* Tradition. Cambridge: Cambridge University Press.

Hendry, R., and Psillos, S. (2007). How to Do Things with Theories: An Interactive View of Language and Models in Science. In J. Brzeziński, A. Klawiter, T. A. Kuipers, K. Łastowski, (p. 285) K. Paprzycka, and P. Przybysz (eds.), *The Courage of Doing Science: Essays Dedicated to Leszek Nowak*, pp. 59–115. Amsterdam: Rodopi.

Hesse, M. B. (1966). *Models and Analogies in Science*. South Bend: University of Notre Dame Press.

Kitcher, P. (1990). The Division of Cognitive Labor. *The Journal of Philosophy*, 87(1), 5–22.

Kuhn, T. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lewis, David. 1969. *Convention*. Cambridge: Harvard University Press.

Lloyd, E. A. (1994). *The Structure and Confirmation of Evolutionary Theory* (second edn.). Princeton: Princeton University Press.

Merton, R. (1957) Priorities in Scientific Discovery. *American Sociological Review*, 22, 635–59.

Morgan, M. S. (2012). *The World in the Model: How Economists Work and Think*. Cambridge: Cambridge University Press.

Muldoon, R., Smith, T., and Weisberg, M. (2012). Segregation That No One Seeks. *Philosophy of Science*, 79, 38–62.

Quine, W. V. O. (1969). Natural Kinds. In W. V. O. Quine (ed.), *Ontological Relativity and Other Essays*, pp. 26–68. New York: Columbia University Press.

Railsback, S. F., and Grimm, V. (2012). *Agent-Based and Individual-Based Modeling: A Practical Introduction*. Princeton: Princeton University Press.

Resnik, D. B. (1991). How-Possibly Explanations in Biology. *Acta Biotheoretica*, 39(2), 141–9.

Rhodes, R. (1987). *Making of the Atomic Bomb*. New York: Simon and Schuster.

Schelling, T. C. (1978). *Micromotives and Macrobehavior*. New York: Norton.

Searle, J. R. (1980). Minds, Brains, and Programs. *Behavioral and Brain Sciences*, 3(3), 417–57.

Skyrms, B. (1996). *Evolution of the Social Contract*. Cambridge: Cambridge University Press.

Skryms, B. (2010). *Signals: Evolution, Learning, and Information*. Oxford: Oxford Univeristy Press.

Sorensen, R. A. (1992) *Thought Experiments*. Oxford: Oxford University Press.

Strevens, M. (2003). The Role of the Priority Rule in Science. *The Journal of Philosophy*, 100(2), 55–79.

Suppe, F. (1989). *The Semantic Conception of Theories and Scientific Realism*. Chicago: University of Illinois Press.

Suppes, P. (1960). A Comparison of the Meaning and Uses of Models in Mathematics and the Empirical Sciences. *Synthese*, 12(2–3), 287–300.

Thomson, J. J. (1985) The Trolley Problem. *The Yale Law Journal*, 94(6), 1395–415.

Tversky, A. (1977). Features of Similarity. *Psychological Review*, 84, 327–52.

Tversky, A., and Gati, I. (1978). Studies of Similarity. In E. Rosch and B. Lloyd (eds.), *Cognition and Categorization*, pp. 79–98. Hillsdale, N.J.: Erlbaum.

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

Walton, K. (2006). On the (so-called) Puzzle of Imaginative Resistance. In S. Nichols (ed.), *The Architecture of the Imagination*, pp. 137–48. Oxford: Oxford University Press.

Watson, J. D. (2011). *The Double Helix: A Personal Account of the Discovery of the Structure of DNA*. New York: Scribner.

Weisberg, M. (2007) Who is a Modeler? *British Journal for the Philosophy of Science*. 58, 207–33.

Weisberg, M. (2013). *Simulation and Similarity: Using Models to Understand the World*. New York: Oxford University Press.

Weisberg, M., and Muldoon, R. (2009). Epistemic Landscapes and the Division of Cognitive Labor. *Philosophy of Science*, 76(2), 225–52.

(p. 286) Wilensky, U. 1999. NetLogo. **<http://ccl.northwestern.edu/netlogo/>** (accessed September 18, 2015). Center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL.

Zollman, K. J. (2007). The Communication Structure of Epistemic Communities. *Philosophy of Science*, 74(5), 574–87.

Notes:

([1]) This is a much-simplified form of the similarity equations developed in Weisberg, 2013.

Michael Weisberg

Michael Weisberg is Professor and Chair of Philosophy at the University of Pennsylvania. His research focuses on the philosophy of science, especially the role of idealization in modeling. His other research includes social and cultural

evolutionary theory, the nature of the chemical bond, the division of cognitive labor, and the public understanding of evolution and climate change.