Appraising Models Nonrepresentationally
Author(s): Till Grüne-Yanoff
Source: *Philosophy of Science*, Vol. 80, No. 5 (December 2013), pp. 850-861
Published by: The University of Chicago Press on behalf of the Philosophy of Science Association
Stable URL: http://www.jstor.org/stable/10.1086/673893
Accessed: 24/01/2014 07:54

Appraising Models Nonrepresentationally

Till Grüne-Yanoff*

Many scientific models lack an established representation relation to actual targets and instead refer to merely possible processes, background conditions, and results. This article shows how such models can be appraised. On the basis of the discussion of how-possibly explanations, five types of learning opportunities are distinguished. For each of these types, an example—from economics, biology, psychology, and sociology—is discussed. Contexts and purposes are identified in which the use of a model offers a genuine opportunity to learn. These learning opportunities offer novel justifications for modeling practices that fall between the cracks of standard representationalist appraisals of models.

1. Introduction. Philosophers' approaches to appraising models have largely been focused on their representational functions. On the basis of widely accepted accounts of what models are—representations—they propose that models should be appraised accordingly: they are good models to the extent that they are good representations. Various criteria for good representations have been proposed, including isomorphism (van Fraasen 1980), similarity (Giere 1988), and partial resemblance (Mäki 2009). The implicit assumption underlying these accounts is that models represent real targets—entities or properties that are found in the real world. Without this assumption, none of the assessment criteria for models would have much bite: they require comparing model properties with properties that can be independently observed, measured, or at least indirectly inferred.

As long as modeling practices satisfy these requirements, such representational accounts of model appraisal might work fine. However, there is convincing evidence that many scientific modeling practices do not satisfy them. The first kind of evidence comes from the way many modelers describe their own work. Instead of seeking to represent aspects of the real world, they claim to be aiming at constructing possible or parallel worlds that may give relevant insights about the real world in more indirect ways

*To contact the author, please write to: Royal Institute of Technology (KTH), Stockholm & TINT, Helsinki, Teknikringen 78B 10044, Stockholm, Sweden; e-mail: gryne@kth.se.

Philosophy of Science, 80 (December 2013) pp. 850–861. 0031-8248/2013/8005-0017$10.00

(Sugden 2000; Rubinstein 2006). In particular, they claim that these model constructions involve reference to possible processes, possible background conditions, and even possible phenomena or properties. Crucially, modelers claim that such models (at least sometimes) offer a genuine contribution to our knowledge about the real world.

The second kind of evidence comes from a growing realization among philosophers that the common realist defenses of model idealization often do not apply. Among others, it has been pointed out that scientists rarely invest work in de-idealizing an existing model (Frigg and Hartmann 2009), that models often are inferentially highly nonrobust with respect to many of their assumptions (Cartwright 2009), and that models often idealize the very factors they purport to isolate (Grüne-Yanoff 2011). Philosophers' arguments that such models adequately represent their targets despite their many idealizations thus fail in these cases. Yet scientists make use of such models regardless, and so the question arises, whether and how these modeling strategies can be justified.

Philosophers, if they treat such cases at all, have by and large appraised such modeling practices as playing merely a heuristic role, for example, in "conceptual exploration" (Hausman 1992), "getting acquainted with mechanisms" (Hartmann 1995), "defin[ing] the extreme of a continuum of cases" (Wimsatt 2007), or facilitating "creative thought" (Holyoak and Thagard 1995). This heuristic justification is weak because success criteria for such functions are unclear in the extreme. Furthermore, it places the use of such models in the same category as taking a walk, reading the newspaper, or whatever else scientists do in order to inspire themselves to further theory development. Bunching important kinds of scientific modeling together with practices that cannot be rationally accounted for seems an unsatisfactory state, which this article seeks to repair.

In particular, I argue that we can learn from models, even if they lack an established representation relation to real-world targets. Section 2 specifies what I mean by learning from a model. Section 3 draws on the literature of how-possibly explanation in order to distinguish different kinds of learning opportunities from such models. Section 4 illustrates each kind with a concrete scientific model and argues that in particular contexts and for specific purposes, one learns from each. Section 5 presents conclusions.

2. Learning from Models. Learning from a model *M*, I suggest, is constituted by a change in confidence in certain hypotheses, justified by reference to *M* (Grüne-Yanoff 2009). Under the standard, representational account, such learning is accomplished by (i) investigating certain properties of the model and (ii) establishing that the model is a sufficiently accurate representation of a (real-world) target in order to license an inference from model to target. Aerodynamic behavior of a scale model of a new type of airplane, for example, is investigated in a wind tunnel. It is then concluded that an actual airplane of that type has similar properties, given that the scale model and the actual plane are sufficiently similar with respect to the proportions of their hull elements, the geometry of their wings, and so forth. If the model user believes in the validity of the model investigation and the sufficient similarity between the model and the target, and her prior beliefs about the plane's aerodynamics are not identical to the model result, then she has learned from the model about the world.

I claim that one can learn from a model even without establishing its representational adequacy. That is, reference to mere model investigations—with no established representational link to any real-world target—may justify changing one's confidence in some hypotheses about the world.

Let me clarify this claim. First, it is important to stress that reference to such a model should justify changing one's confidence in some hypothesis about the world. It is trivial that investigating a model might result in some justified changes of our beliefs in the model artifact. That is not what I mean here. Instead, I claim that investigating the model can teach us about something different from the model itself.

Second, I do not claim that such models justifiably affect our beliefs about particular actual entities or about properties instantiated in the real world. That would require models that represented these entities or properties sufficiently well. Instead, I claim that they might change claims about possible entities, properties, or processes. Such possibility claims are about the world: they are about what is possible in this world, not about facts in some other, possible, world. Furthermore, many such possibility claims are highly relevant for theoretical and practical objectives. Consider, for example, the following points:

- A policy maker seeking to reduce urban segregation might change her policies upon learning that racist preferences are not a necessary cause of segregation.
- A scientist seeking to explain a population dynamic might change his explanatory strategy when learning that this dynamic cannot be produced from actual background conditions with a set of plausible migration decision rules alone.
- A policy maker who learns that altruistic preferences are adaptive under certain possible conditions might change her evaluation of certain institutional regulations.

Thus, changes in the confidence of hypotheses of the above kind affect the ways we seek to explain and control the actual world. If models would justify changes in the confidence of such hypotheses, one would learn from such models about the world.

Third, such possible entities, properties, or processes are conceptually linked to the world. For example, a possible migration rule in an agent-based model falls under the same concept as actual agents' decision rules concerning migration. But this conceptual link does not require a resemblance between the model rule and any real-world target. Thus, the possibility that such a model presents is relevant for the discussion of actual population dynamics, even if the model rule does not adequately resemble (or is not established as adequately resembling) actual migration rules.

Finally, the most relevant kind of possibility the models I discuss here deal with is epistemic possibility, that is, the possibility of entities, processes, or properties not ruled out by what the agent knows or believes. Acknowledging this makes clear that we do not learn from the model in isolation but rather from the model in conjunction with background knowledge and beliefs. This knowledge at least partly consists in claims about possible, nonactual states, which does not (or cannot) refer to actual targets. In this conjunction, the model performs two epistemic functions: it conceptualizes existing beliefs in a new way, and it draws inferences from the thus established possibility, given all relevant beliefs.

Models thus help only to present existing knowledge and infer from it. Cognitively unlimited beings would have no need for them. But this is just the same as with the representational account: we learn from those models only because they help us make inferences from beliefs we gleaned from the target systems. In the cases discussed here, however, the knowledge processed by the model is different: it contains beliefs about possible entities, processes, or properties, which cannot be obtained by establishing an adequate representation of the model to actual target systems. Thus, adequate representation is not a useful appraisal criterion for such models. Instead, I propose learning as the appropriate criterion.

3. How-Possibly Explanations. How can we learn from models that do not have an established representation relation to a real-world target? To answer this question, the extant literature on *how-possibly explanations* is instructive. This literature controversially discusses what characterizes how-possibly explanations; what distinguishes them from how-actually, potential, or how-possible explanations; and whether how-possibly explanations are explanations at all. In this article, I eschew these controversies. Instead I use the conceptual distinctions offered by this debate to categorize different learning opportunities that models without an established representation relation to a real-world target might provide.

i) The debate commences with Dray's (1957) claim that how-possibly explanations have a different aim and a different structure from how-actually explanations. How-possibly explanations aim at giving an account how events that are considered impossible could have happened. How-actually explanations, in contrast, aim at accounting for how actual events have happened. Furthermore, Dray argues that how-possibly explanations rebut the impossibility of the explanandum by giving a necessary condition for its occurrence. He contrasts this with actual explanations offering sufficient conditions for their explananda.[1]

This distinction is relevant for the present analysis. Actual explanation requires the identification of true (sufficient parts of) causes that brought about the explanandum. Models that adequately represent relevant targets contribute to such explanations. How-possibly explanations, in contrast, identify elements of possible causes for an explanandum. Models can identify such possible causes—and hence contribute to how-possible explanations—without representing real-world entities or properties. Thus, by affecting our belief in impossibility claims, we can learn from such models.

ii) The Dray-Reiner type of how-possibly explanations focuses on identifying some initial conditions, which show the possibility of the explanandum. Another kind of how-possible explanation instead focuses on indicating the sort of process through which the explanandum could possibly have been produced. Here the purpose lies in giving an account of how certain (possible) initial conditions might yield the explanandum through a possible model process—a question that is of particular interest in those cases in which the explanandum and the initial conditions in question are prima facie incompatible.

iii) Another purpose is to account for the possibility of a real target object having a certain property, produced by a possible process from actual background conditions (what Forber [2010] calls "local" how-possibly explanations). Here both initial conditions and explanandum are given, and the how-possibly explanation either seeks to identify any process that produces the latter from the former or investigates how a particular model process can—or cannot—produce the actual target properties from the given conditions.

iv) Forber (2010) distinguishes such local from what he calls "global" how-possibly explanations. Global how-possibly explanations account for the possibility that an idealized object has a certain property, produced by a possible process from possible background conditions. Forber suggests that the purpose of these how-possibly explanations is investigating the capabilities of general model processes.

v) Finally, how-possibly explanations have been interpreted not in contrast to how-actual explanations but rather as their precursors. According to this view, how-possibly explanations are similar to how-actually explanations in that they satisfy most explanatory virtues, but they are inferior in

1. Reiner (1993) has criticized Dray's account, pointing out that how-possibly explanations do not really identify necessary conditions of the explanandum, but rather necessary parts of a sufficient condition for the explanandum.

that they lack adequate empirical support (Resnik 1991, 143). In particular, they are reasonably complete, showing how the explanandum was generated through a process from initial and background conditions. But process and background conditions are not well supported empirically, so the account offers a mere possible or potential explanation.

Let me summarize. Models that lack an established representation relation to actual targets have a number of distinct purposes, which have been discussed in the philosophical literature under the heading of "how-possibly explanations." As the analysis of some of the key controversies in this literature showed, this notion contains a number of disparate scientific objectives—some of them explanatory, some offering other forms of epistemic gain. In the next section, I discuss each of these five kinds of learning purposes at the hand of a scientific modeling example, showing how in particular situations and for particular purposes, one can learn from each.

4. Five Cases of Learning. In the following subsections, I give five examples of models that lack an established representation relation to actual targets. Each example corresponds to a type of how-possibly explanation discussed in the previous section. For each case, I identify contexts and purposes in which these respective models offer an opportunity to learn about the world.

i. Affecting Impossibility Claims. Schelling's (1971) *checkerboard model* produces an abstract pattern of spatial segregation that he claims can be found in many cities but is not associated with any concrete settlement or even type of settlement. Schelling produces this abstract result with two sets of tokens, initially distributed randomly over a checkerboard. Tokens move according to an iterated rule until no more movements occur. The rule is this. For a given token, if more than half of the tokens on (Moore-) neighboring fields are of a different type, then this token will move to another vacant field, with less than half of the neighboring fields occupied with tokens of the other type. Schelling claims neither this process to represent an actual migration process nor the checkerboard to represent an actual neighborhood. Nor does he justify these model choices with reference to a real-world phenomenon. But he claims that the process is started by an actual initial condition, namely, the (nonracist) preference of individuals not to be in the minority. It is the one aspect of his model that he seeks to connect with the actual world, citing behavioral examples from restaurants, clubs, and classrooms.

Thus, the model system that produces the segregation pattern is not established as an adequate representation of any real-world system. The specific shape of the preferences—the only part of the model that is justified as an adequate representation—alone does not produce the segregation pattern. Appraised from a representational perspective, therefore, such a model should be regarded with suspicion.

Yet we learn from Schelling's model. It shows the possible production of an abstract pattern (a segregation of the two types of tokens on the checkerboard) from a possible and one actual initial condition and a possible process. In the context of spatial residential segregation, where the abstract segregation pattern might be realized, this possible production result is of particular importance: until then it was widely believed that racist preferences were a necessary cause of segregation. Schelling's model shows that segregation patterns might be produced by another cause, which is an actual property of agents in many real-world populations: namely, the preference not to be in the minority (it shows only that it might be produced because it does so in a merely possible context—with an environment and a process that our knowledge does not rule out but that we by no means can assume to be the actual environment or process). The model result thus justified changing one's confidence in hypotheses about racist preferences being a necessary cause of segregation. Anyone who had high confidence in such hypotheses learned from Schelling's checkerboard model.

ii. Identifying Potential Mechanisms. Ainslie's (2001) *feedback model of self-control* produces a behavioral phenomenon: the moderate impulsivity of human choices in the absence of precommitment devices. Ainslie sees this phenomenon exemplified, for example, in the considerable number of addicts, most of whom eventually overcome their addiction. His model produces this result from a particular shape of time preferences (a shape inversely proportional to delay of consumption) and a process of recursive self-prediction—prediction that is fed back to the ongoing choice process. The employed preference shape (also known as "hyperbolic discounting") was first proposed in order to account for impulsive choice and hence is considered an actual initial condition by some. Yet the moderate impulsivity of human choice has led many to doubt that humans actually discount future value hyperbolically.

It is exactly the aim of Ainslie's model to show that the hyperbolic description is compatible with moderate impulsive behavior by directly stoking it on the one hand and by indirectly moderating it, on the other hand, through a process of self-prediction that arises from this hyperbolic form itself. Given this objective, Ainslie cannot treat the hyperbolic shape as a fact, but only as a possibility. Furthermore, he readily admits that the process of recursive self-prediction is inaccessible to controlled experiment and hence remains a mere possibility. Thus, the model produces the behavioral phenomenon from two merely possible elements, neither of which is justified as an adequate representation of real-world properties or processes. Additionally, the behavioral phenomenon to be produced is often seen as evidence against the hyperbolic shape of preferences. A representationalist appraisal of Ainslie's model would therefore surely be negative.

Yet one learns from Ainslie's model in two ways. First, the model justifies a change in confidence in the hypothesis that intertemporal behavioral data are incompatible with a hyperbolic shape of discounting. If self-prediction were operational, then hyperbolic discounting would be compatible with moderate impulsiveness. Given that self-prediction is not ruled out by our background beliefs, we learn from the model that hyperbolic discounting is indeed possible, even if we observe moderate impulsiveness.

Ainslie draws a second and stronger lesson from this possibility. If hyperbolic discounting is possible, he argues, then in order to produce moderate impulsiveness, we need a process like recursive self-prediction. Although self-prediction is only a possible process, the paucity of alternative describable processes not ruled out by our background beliefs lets Ainslie conclude that this result indeed should increase one's confidence in self-prediction. In Ainslie's words, "a small number of selected thought experiments yield a valid rejection of the null hypothesis—that contingent self-prediction is unnecessary for volition" (2009, 145). Thus, we learn from the model about a potential mechanism that produces the phenomenon under specified possible conditions. This in turn justified raising our confidence in the possibility of hyperbolic discounting and—under the additional assumption that alternative possible processes with that result are hard to find—also our confidence in the possibility of the potential mechanism. Consequently, those with low confidence in either of these two claims learned from Ainslie's model.

iii. Learning the Impossibility of Producing a Property. Axtell et al.'s (2002) *Anasazi model* seeks to produce a historically documented population dynamic of a Native American settlement in the US Southwest from contemporary soil and meteorological data, through possible migration decision processes of the modeled people. These processes involved rules whether to reproduce, to split up households, or to leave the settlement, given harvest levels. The model thus seeks to produce an actual phenomenon from actual initial conditions through a set of possible model processes.

Unlike the previous examples, this model was built with the help of computer simulations, which allowed the modelers to run through a large number of different but related possible processes in order to choose the one that produced a result with the best fit to the actual population dynamic. What all these processes have in common is that they model the environmental effect on individual decisions directly (the authors call these "push factors"), disregarding interactive social or cultural influences ("pull factors").

These model processes were not identified by behavioral evidence from the historical populations (no relevant records exist) but rather as possibilities constrained by knowledge about today's microsocieties. Appraised from a representational perspective, the modelers' claim that their model explains (i.e., identifies the true causes of the population dynamic) should therefore be regarded with caution. In any case, the model fails to produce the most salient feature of the actual population dynamic, the complete exodus around AD 1400.

Yet, one learns from the Anasazi model—specifically from its failure. The model shows that with any available push-type process, the exodus cannot be produced. This should increase one's confidence that it is impossible to produce this property with these kinds of processes. Consequently, the model justifies increasing one's confidence in the belief that another capacity (cultural pull factors) must be included in a model to produce the actual population dynamics from the initial conditions.

iv. Capacities of a Possible System. Güth's (1995) *indirect evolutionary approach* offers a model of preference evolution, which produces preferences for reciprocity. The model starts with a population of agents who have different preferences over objects of choice (e.g., consumption bundles or behavioral strategies). Agents' rational choices then are determined according to their preferences so that different preferences lead to different choices. Depending on their choice (and the environment in which the choice is made), an agent will have greater or lesser reproductive success than other agents with different preferences and hence different choices. Assuming that preferences are inherited, differential reproduction of agents then leads to differential replication of preferences in the population. Clearly, the background conditions of this model, in particular, the distribution of preferences in the population and the differential reproductive success of certain choices, are not rooted in beliefs about actual real-world targets but are mere possibilities. Similarly, the model processes by which preferences are reproduced across generations, and by which preference-based choices yield reproductive advantages or disadvantages, have no established relation to real-world targets. Instead, they are mere possibilities, adapted from selection processes of much simpler (and more uniform) traits. Even the result—preferences for reciprocation—is described only in abstract terms, and Güth makes no attempt to link it to concrete real-world targets. So again, a representationalist appraisal would not find much to praise in this model.

Nevertheless, one can learn from Güth's model. It shows that preferences with certain abstract properties can be produced through natural selection in nonactual circumstances.[2] More concretely, a possible adaptive system,

2. In this case reciprocation, but in related papers, Güth also produces preference for fairness and trust.

free from external intervention or constraints, has the capacity to produce preferences for reciprocation. This is a lesson of potential practical relevance. Anyone who with high confidence believed that reciprocation, fairness, or trust could not possibly be an adaptive trait has good reason to change his belief when confronted with this model.

v. Potential Explanation. Trivers's (1971) *reciprocal behavior model* produces a concrete actual result, the particular behavioral patterns exhibited by cleaner fish (*labroides dimidiatus*) and their hosts. To this end, it employs an actual process, frequency-dependent selection, which is found in many instances of biological and cultural evolution. Cleaner and host, so Trivers argues, are engaged in an indefinitely repeated prisoner's dilemma game, where the gains of cooperation (i.e., the cleaner feeds on the host's parasites and the host does not eat the cleaner) are sufficiently high to ensure differential reproductive success over unilateral defection. However, Trivers's model does not employ actual but rather possible background conditions. In fact, the very purpose of his model is to identify initial conditions that would license a selection explanation of reciprocal behavior between cleaner and host. These include "that hosts suffer from ectoparasites; that finding a new cleaner may be difficult or dangerous; that if one does not eat one's cleaner, the same cleaner can be found and used a second time; that cleaners live long enough to be used repeatedly by the same host; and if possible, that individual hosts do, in fact, reuse the same cleaner" (1971, 41).

That Trivers list these conditions in this way makes clear that his model, at least at the time of writing, lacks established representational links to its potential target system. It does not represent what the initial conditions of the cleaner-host system actually are, but rather what they possibly could be. A representationalist appraisal would—correctly—point out that Trivers's model, because of the nonestablished initial conditions, does not provide a how-actually explanation of the cleaner-host symbiosis.

Yet, contrary to the representationalist appraisal, this does not relegate the model to a merely heuristic function. Rather, one can (or, rather, could in 1971) learn important lessons from the model. It offers a potential explanation of a specific phenomenon that previously was either unexplained or explained differently. Before 1971, many people probably believed that no such adaptive explanation was possible. Or if they did, they probably had low confidence that Trivers's list would be the set of necessary conditions for such an adaptive explanation. His model gave them good reasons to change these beliefs, and it taught many marine biologists what to look for when observing *labroides*.

5. Conclusions. I have argued that one can rationally justify the use of models that do not have an established representation relation to any actual target by showing that one learns from them about the world. To this end, I characterized learning as justifying a change in confidence in certain hypotheses about the world. I then discussed a number of hypotheses relating to possibility claims and argued that changing one's confidence in any of them would affect the way scientists and policy makers seek to explain and control the actual world. These hypotheses, although consisting in possibility claims, thus are about the world.

To analyze different kinds of possibility claims made with such models, I employed conceptual distinctions from the discussion of how-possibly explanations. Five kinds of learning opportunities emerged. I illustrated each with a concrete scientific model. In particular contexts and for specific purposes, I argued, one could learn from each of them.

Of course, one cannot learn from every model of this kind: many such models present possibilities that everybody already believes in or impossibilities nobody doubts. For those models, my account does not provide justification. This is just how it should be. It is the delicate art of the modeler to sense the scientific community's conviction that something is necessary or impossible and then to construct a model that convinces them to change their conviction. When this art fails, such models do not teach anything. But where it succeeds, these important functions justify modeling practices that are neglected by a representationalist appraisal of models.

REFERENCES

Ainslie, George. 2001. *Breakdown of the Will*. Cambridge: Cambridge University Press.

———. 2009. "Recursive Self-Prediction in Self-Control and Its Failure." In *Preference Change: Approaches from Philosophy, Economics, and Psychology*, ed. Till Grüne-Yanoff and Sven Ove Hansson, 139–58. Berlin: Springer.

Axtell, Robert L., Joshua M. Epstein, Jeffrey S. Dean, George J. Gumerman, Alan C. Swedlund, Jason Harburger, Shubha Chakravarty, Ross Hammond, Jon Parker, and Miles Parker. 2002. "Population Growth and Collapse in a Multiagent Model of the Kayenta Anasazi in Long House Valley." *Proceedings of the National Academy of Sciences* 99 (3): 7275–79.

Cartwright, Nancy. 2009. "If No Capacities Then No Credible Worlds: But Can Models Reveal Capacities?" *Erkenntnis* 70 (1): 45–58.

Dray, William H. 1957. *Laws and Explanations in History*. Oxford: Oxford University Press.

Forber, Patrick. 2010. "Confirmation and Explaining How Possible." *Studies in History and Philosophy of Biological and Biomedical Sciences* 41 (1): 32–40.

Frigg, Roman, and Stephan Hartmann. 2009. "Models in Science." In *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. Stanford, CA: Stanford University.

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

Grüne-Yanoff, Till. 2009. "Learning from Minimal Economic Models." *Erkenntnis* 70 (1): 81–99.

———. 2011 "Isolation Is Not Characteristic of Models." *International Studies in the Philosophy of Science* 25 (2): 1–19.

Güth, Werner. 1995. "An Evolutionary Approach to Explaining Cooperative Behavior by Reciprocal Incentives." *International Journal of Game Theory* 24:323–44.

Hartmann, Stephan. 1995. "Models as a Tool for Theory Construction: Some Strategies of Preliminary Physics." In *Theories and Models in Scientific Process*, ed. William E. Herfel, Wladyslaw Krajewski, Ilkka Niiniluoto, and Ryszard Wojcicki, 49–67. Amsterdam: Rodopi.

Hausman, Daniel. M. 1992. *The Inexact and Separate Science of Economics*. Cambridge: Cambridge University Press.

Holyoak, Keith, and Paul Thagard. 1995. *Mental Leaps: Analogy in Creative Thought*. Cambridge, MA: Bradford.

Mäki, Uskali. 2009. "MISSing the World: Models as Isolations and Credible Surrogate Systems." *Erkenntnis* 70 (1): 29–43.

Reiner, Richard. 1993. "Necessary Conditions and Explaining How-Possibly." *Philosophical Quarterly* 44 (170): 58–69.

Resnik, David B. 1991. "How-Possibly Explanations in Biology." *Acta Biotheoretica* 39:141–49.

Rubinstein, Ariel. 2006. "Dilemmas of an Economic Theorist." *Econometrica* 74 (4): 865–83.

Schelling, Thomas C. 1971. "Dynamic Models of Segregation." *Journal of Mathematical Sociology* 1 (2): 143–86.

Sugden, Robert. 2000. "Credible Worlds: The Status of Theoretical Models in Economics." *Journal of Economic Methodology* 7 (1): 1–31.

Trivers, Robert L. 1971. "The Evolution of Reciprocal Altruism." *Quarterly Review of Biology* 46:35–57.

van Fraasen, Bas. 1980. *The Scientific Image*. Oxford: Oxford University Press.

Wimsatt, William. 2007. *Re-engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Cambridge, MA: Harvard University Press.