

# Appraising Non-Representational Models

Till Grüne-Yanoff

Royal Institute of Technology, Stockholm

[gryne@kth.se](mailto:gryne@kth.se)

## ABSTRACT

Many scientific models are non-representational in that they refer to merely possible processes, background conditions and results. The paper shows how such non-representational models can be appraised, beyond the weak role that they might play as heuristic tools. Using conceptual distinctions from the discussion of how-possibly explanations, six types of models are distinguished by their modal qualities of their background conditions, model processes and model results. For each of these types, an actual model example – drawn from economics, biology, psychology or sociology – is discussed. For each case, contexts and purposes are identified in which the use of such a model offers a genuine opportunity to learn – i.e. justifies changing one's confidence in a hypothesis about the world. These cases then offer novel justifications for modelling practices that fall between the cracks of standard representational accounts of models.

## 1. Introduction

Philosophers' approaches to appraising models have largely been focused on their representational functions. Models are representations; they are good models to the extent that they are good representations. Various criteria for good representations have been proposed, from isomorphism (van Fraassen 1980) through similarity (Giere 1988) to 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.

This differs notably from the way many modellers 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 (for an elaboration of his view, see Sugden 2000). In particular, they claim that these model constructions involve reference to possible processes, possible background conditions, and even possible phenomena or properties. Let me call such models *non-representational models*. Crucially, modellers claim that non-

representational models (at least sometimes) offer a genuine contribution to our knowledge about the real world.

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

Section 2 offers a characterisation of learning from models, and what kind of hypotheses might be learned from non-representational models. Section 3 employs conceptual distinctions from the discussion of how-possibly explanation, in order to analyse different kinds of possibility claims made with non-representational models. Six kinds of non-representational models will emerge. 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 concludes.

## **2. Learning from Models**

Modelling is a set of reasoning practices for cognitively limited beings (Wimsatt 2007). The inferences one can legitimately draw from scientific models are inferences from information already contained in one’s set of beliefs.<sup>1</sup> An ideal Bayesian agent would have no use for scientific models. Being very much unlike ideal Bayesian agents, humans often have to rely on models to justify some of their beliefs.

It is in this sense that we can learn from models. Models facilitate their users in making inferences from their own background beliefs. If these inferences affect the model user’s beliefs about some other hypothesis, then the model user learned from the model. Learning from a model *M*, I suggest, is constituted by a *change in confidence in certain hypotheses, justified by reference to M*.

---

<sup>1</sup> Including beliefs one accepts only tentatively, e.g. for the purpose of a thought experiment.

We do not learn from models in the same way as we learn from straightforward observation. Although observation (of the model) is often part of modelling, we ultimately do not want to learn about the model artefact, but about the real world. Thus the learning I will focus on in this paper concerns changes in confidence in a hypothesis *about the world*.

With representational models this 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 behaviour 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 scale model and actual plane are sufficiently similar with respect to the proportions of their hull elements, the geometry of their wings, etc. If the model user believes in the truth of the model investigation and the sufficient similarity between model and 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 similarly learn from non-representational models. That is, reference to non-representational models may justify changing one's confidence in some hypothesis about the world. By definition, this cannot be accomplished by a belief in the model being a sufficiently accurate representation of a (real world) target. Instead, the inference from model to hypothesis must be licensed differently. I will argue that typical beliefs that license such inferences are those that consider certain background conditions or certain processes "possible", or "credible" (Sugden 2000). Hypotheses whose confidence change is justified through reference to such models include the following types:

- That an entity or property is possible. A special case of this is the hypothesis that something is impossible in the actual world.
- That a process yields a property. A special case of this is the hypothesis that an actual process does or does not have the capacity (in non-actual circumstances) to bring about a certain property.
- That an entity or property possibly is a cause of an actual phenomenon

Of course, such hypotheses do not make claims about particular actual entities or about properties instantiated in the real world. To justify changes in such hypotheses would require models that represented these entities or properties sufficiently well. Nevertheless, these hypotheses are about the world. Reflecting on the impact of such hypotheses on explanation or control supports this claim. Consider for example:

- 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 preferences for reciprocity are adaptive under certain possible conditions might change her evaluation of certain institutional regulations.

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

### **3. How-Possibly Explanations**

Schematically, a model consists of a set of initial conditions  $Q$ , a model process  $P$  and a model result  $R$ , derived from this process and the initial conditions. One learns from such a model if  $R$  affects one's confidence in a hypothesis about the world. In the case of representational models, this may be because  $Q$  and  $P$  are sufficiently similar to a target to consider  $R$  relevant for that target, and hence information contained in  $R$  relevant for the confidence one has in hypotheses about the target. In the case of non-representational models, this may be because  $Q$  and  $P$  are at least considered possible, plausible or credible enough to consider  $R$  a relevant possibility. Considering  $R$  a relevant possibility then may affect one's confidence in certain hypotheses.

What part of a model is considered merely possible (rather than actual) and what kind of possibility is meant here will crucially influence whether the model result is considered a relevant possibility. It is therefore helpful to analyse different model types by the different possibility claims they contain. Here the extant literature on *how-possibly explanations* is very instructive. This literature controversially discusses what characterises 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 paper, I eschew these controversies. Instead I use the conceptual distinctions offered by this debate to categorise different kinds of models, and to elicit the purposes and contexts in which the respective model types might offer learning opportunities.

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. Reiner (1993) has criticised 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.

This distinction is relevant for the present analysis. Actual explanation requires the identification of true (sufficient parts of) causes that brought about the explanandum. Representational models are one mode of identifying and representing these causes. How-possibly explanations, in contrast, identify elements of *possible* causes for an explanandum. Models can represent such possible causes – and hence contribute to how-possible explanations – without representing real-world targets. How-possibly explanations, in Dray and Reiner's sense, give non-representational models a purpose.

More recently, 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 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 that the account offers a mere possible, partial or potential explanation.

One may disagree whether Resnik's type should fall into the category "how-possibly explanation" (for a negative view, see Forber 2010). What is clear, though, is that non-representational models often serve the purpose that Resnik describes, and that this purpose is different from the one Dray and Reiner identify. First, models serving Resnik's type of how-possibly explanation will yield a result that represents a real-world target – otherwise, the similarity to how-actually explanations would not even arise. Models serving Dray-Reiner type how-possibly explanations, on the other hand, may yield results that do not represent real-world targets. Second, models for Resnik-type how-possibly explanations must be "reasonably complete" in order to be turned into how-actual explanations when empirical evidence for their similarity to some real-world target is forthcoming. No such requirement is imposed on models for Dray-Reiner type how-possibly explanations. They may serve their purpose of rebutting impossibilities with a rather sketchy structure, singling out only certain possible processes or background conditions.

Dray type how-possibly explanations focus on identifying some *conditions* that show the possibility of the explanandum. Another kind of how-possible explanation instead focus on indicating the sort of *process* through which the explanandum took place (Reiner 1993). Consecutive authors point out that this may consist in a mere proposal of a possible mechanism, or alternatively in providing a partial mechanism that in fact had the explanandum as outcome. In the latter case, the actual mechanism that produced the explanandum is identified, but in a way insufficient “to see more how the explanandum phenomenon was produced” (Persson 2011). Both purposes are served by non-representational models – the first by a model presenting a possible process, the second by a model presenting an actual process without sufficient causal detail, under possible background conditions.

Finally, Forber (2010) distinguishes between *global* and *local* how-possibly explanation. Global how-possibly explanations account for the possibility that an idealised object has a certain property, produced by a possible process from possible background conditions. Their purpose is to investigate the capabilities of general model processes (Forber 2010, 33). Local how-possibly explanations, in contrast, account for the possibility of a real target object having a certain property, produced by a possible process from actual background conditions. Their purpose is to guide speculation on how a particular model process can produce actual target properties. Forber’s distinction thus points to a difference between non-representational models with an abstract result, and those with a concrete result.

Let me summarise. Non-representational models have a number of distinct purposes, which have been discussed in the philosophical literature under the heading of “how-possibly explanation”. 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, some merely heuristic. Crucially, these different purposes are served by different kinds of non-representational models. These models kinds can be distinguished by the modalities of the model result, the model process and the initial conditions. Keeping things simple and merely distinguishing between actual and possible (non-actual) processes and initial conditions, and concrete and abstract model results, we get six different kinds of non-representational models.<sup>2</sup> In the next section, I discuss each of these six non-representational models at the hand of an example, showing how in particular situations and for particular purposes, one can learn from each.

---

<sup>2</sup> Excluding both the representational model with actual initial conditions, actual process and concrete model result, as well as the representational model with actual initial conditions, actual process and abstract model result.

#### **4. Six Cases of Learning from Non-Representational Models**

My preceding abstract discussion leaves many ambiguous cases – a model may contain, say, some merely possible initial conditions, and still represent the workings of an actual process producing some abstract actual property (as e.g. Mäki 2009 argues). Whether such a case is to be counted as a representational or non-representational model will depend on the interpretation of the intentions of the modeller and the objectives of the models' users. Instead of debating this in the abstract, it will be perhaps more fruitful to discuss the issue of learning from non-representational models at the hand of concrete examples.

In the following, I give examples for each of these six kinds of non-representational models. For each case, I identify contexts and purposes in which these respective models offer an opportunity to learn about the world.

##### *i. Possible initial conditions, possible process, abstract result*

Axtell and Epstein's (1996) *sugarscape* is a set of models consisting of agents with individual rates of metabolism and fields of vision, a two-dimensional (51x51 cell) grid which contains different amounts of sugar on each cell, and rules governing the interaction of the agents with each other and the environment. In every step agents look around, find the closest cell filled with sugar, move and metabolize. If their sugar level is below their metabolism rate, they die. Harvested cells grow back one unit of sugar per time period. Using this basic set-up, Axtell and Epstein construct a model of migration, where agents' maximum vision is 10, and all agents are initially clustered together in one rectangular block in the southwest the grid. The authors do not claim that either the initial conditions of the model or the processes established by the model rules represent any actual target; they thus propose a non-representational model with merely possible background conditions and process.

Axtell and Epstein's model produces "waves of migration": a group of agents move outward in north-easterly direction from the initial cluster. Only when this group has progressed a considerable distance does the next group follow them. Although they mention wavelike movements in some mammal herds and economic "herding" as target for other models, they do not argue that the result of their model represents any such actual case. Instead, their result is a mere abstract pattern that *might* be instantiated in the real world.

And yet, one might learn from this model. Axtell and Epstein write that the model

produced “a phenomenon we did not expect” (Axtell and Epstein 1996, 42). Then they analyse the waves as produced by the interplay of food search and consumption by agents, and the slow regrowth of sugar; and they analyse the northeast direction of the migration (a direction in which single agents cannot move) as produced by “a complex interweaving of agents” (ibid.). The model thus justifies reducing one’s confidence in the hypothesis that waves of migration cannot arise from mere food dynamics or that they cannot go in directions single agents cannot move. Because such patterns might be instantiated in the real world, such hypotheses are hypotheses about the real world. Anyone who had high confidence in these hypotheses (like apparently the authors themselves) learned from this model.

*ii. Actual initial conditions, possible process, abstract result*

Schelling’s (1971) *checkerboard model* produces an abstract pattern of spatial segregation that he claims can be found in many cities, but which is not associated with any concrete settlement or even type of settlement. Schelling produces this abstract result with two types 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-) neighbouring fields are of a different type, then this token will move to another vacant field with less than half of the neighbouring fields occupied with tokens of the other type. Schelling neither claims this process to represent an actual migration process, nor the checkerboard to represent an actual neighbourhood. But he claims that the process is started by an actual initial condition, namely the (non-racist) 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 behavioural examples from restaurants, clubs and classrooms. Schelling’s checkerboard model thus is a non-representational model with many possible and one actual initial condition, a possible process and an abstract result.

We learn from Schelling’s model because it shows the *possible production* of an abstract pattern (a segregation of the two types of tokens on the checkerboard) from possible and one actual background condition and a possible process. In the context of spatial residential segregation, where the abstract segregation pattern might be realised, 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 condition in many real-world populations: namely the preference no to be in the minority. 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.



*iii. Possible initial conditions, actual process, abstract result*

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 behavioural 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 mere possibilities. The process by which the model result is produced, however, is an actual process, namely natural selection through differential reproduction. It has clear instantiations both in the domains of cultural and biological evolution. The result – preferences for reciprocation – are only described in abstract terms, and Güth makes no attempt to link it to concrete real-world targets. Nevertheless, one can learn from Güth's model. It shows that preferences with certain abstract properties<sup>3</sup> can be produced through selection in non-actual circumstances. That is, anyone who with high confidence believed that reciprocation, fairness or trust cannot be adaptive traits has good reason to change his belief when confronted with this model.

*iv. Possible initial conditions, possible process, concrete result*

Ainslie's (2001) *feedback model of self-control* produces a concrete result: the *moderate* impulsivity of human choices in the absence of precommitment devices, exemplified for example in the considerable number of addicts, most of whom eventually overcome their addiction. Ainslie produces this result with a possible description of delayed human value as an inverse proportion of delay, and a possible process of recursive self-prediction – prediction that is fed back to the on-going choice process. This description of value (also known as hyperbolic discounting) was first developed 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 impulsiveness, by directly stoking it on the one hand, and by indirectly moderating it through a process of self-prediction that arises from this hyperbolic form itself. In the model, Ainslie thus intentionally casts the hyperbolic shape as a mere possibility. Furthermore, Ainslie readily admits that the process of recursive self-prediction is inaccessible to controlled experiment, and hence remains a

---

<sup>3</sup> In this case reciprocation, but in related papers Güth also produces preference for fairness and trust.

mere possibility.

Interestingly, in Ainslie's model, the proposed process of recursive self-prediction arises as a reaction to hyperbolic discounting, and it acts on future choices in the way often described as an effect of "the will" or "volition". Thus, one learns from Ainslie's model in two ways. First, the model justifies a change in confidence in the hypothesis that intertemporal behavioural data is incompatible with a hyperbolic shape of discounting. Second, the model justifies a change in confidence in the hypothesis that self-control can grow "from the bottom up" – from reactions to the hyperbolic shape of discounting. 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" (Ainslie 2009, 145). All those who had low confidence in such a claim learned from the failure of this model.

*v. Actual initial conditions, possible process, concrete result*

Axtell's et al. (2002) *Anasazi model* fails to produce a historically documented population dynamic of a settlement in the US southwest from soil and meteorological data, through any member of a set of possible migration decision processes of the modelled people. These possible decision processes involved rules whether to reproduce, to split up households, or to leave the settlement, given harvest levels. The model thus seeks to produce a concrete, actual phenomenon from actual initial conditions through a set of possible model processes. One can learn from this model by learning from its failure.

In particular, reference to the model justifies changing one's confidence in the hypothesis that the Anasazi's migration decisions based on subsistence considerations was sufficient to produce the exodus of the Anasazi around 1400 AD. Axtell's et al. model shows that with plausible processes, such a result cannot be produced from the actual conditions. Therefore, the model justifies increasing one's confidence in the belief that another capacity (cultural "pull factors" as the authors call it, in contrast to subsistence consideration "push factors") must be included in a model to produce the actual population dynamics from the initial conditions.

*vi. Possible initial conditions, actual process, concrete result*

Trivers (1971) *reciprocal behaviour model* produces a concrete actual result, the particular behavioural 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 a indefinitely repeated Prisoners' Dilemma game, where the gains of cooperation (i.e. the cleaner cleans and the host does not eat the cleaner) are

sufficiently high to ensure differential reproductive success over unilateral defection. However, Trivers' model does not employ actual, but rather possible background conditions. In fact, the very purpose of Trivers' model is to identify initial conditions that would license a selection explanation of reciprocal behaviour 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” (Trivers, 1971, 41).

That Trivers list these conditions in this way makes clear that his model is a non-representational model with merely possible initial conditions. Yet one learns from this model: it gives one good reasons to change one's confidence in hypotheses about what the necessary conditions are for reciprocal behaviour between cleaner and host to be an adaptive trait.

## **5. Conclusions**

I have argued that one might justify non-representational models by showing that one learns from them about the world. I did not claim that one can learn from every non-representational model, and therefore that every non-representational model is justified. Instead, I described a possible way of appraising them, which is stronger than merely justifying them as heuristic tools.

To this end, I characterised 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 relating to possibility claims, thus are about the world.

To analyse different kinds of possibility claims made with non-representational models, I employed conceptual distinctions from the discussion of how-possibly explanation. Six kinds of models emerged, distinguished by the modality of their background conditions, processes and results. Each of these kinds I illustrated with a concrete scientific model. In particular contexts and for specific purposes, I argued, one could learn from each of them. By demonstrating this, I showed that it is possible to justify each type of non-representational models, in particular contexts and for specific purposes. This concludes my argument.

## References

- Ainslie, G. (2001). *Breakdown of the Will*. Cambridge: Cambridge University Press.
- Ainslie, G. (2009). Recursive self-prediction in self-control and its failure. In T. Grüne-Yanoff & S. O. Hansson (Eds) *Preference Change: Approaches from Philosophy, Economics, and Psychology*. Springer, 139-158.
- Axtell, R. L., Epstein, J.M. (1996) *Growing Artificial Societies Social Science From the Bottom Up*. Cambridge, MA: MIT Press.
- Axtell, R. L., Epstein, J. M., Dean, J. S., Gumerman, G. J., Swedlund, A. C., Harburger, J., Chakravarty, S., Hammond, R., Parker, J., & Parker, M. (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), 7275–7279.
- Dray, W. (1957). *Laws and explanations in history*. Oxford: Oxford University Press.
- Forber, P. (2010). Confirmation and Explaining How Possible. *Studies in History and Philosophy of Biological and Biomedical Sciences* 41(1), 32-40.
- Giere, R. N. (1988). *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.
- Güth, W. (1995). An evolutionary approach to explaining cooperative behavior by reciprocal incentives, *International Journal of Game Theory*, Vol. 24: 323 - 344.
- Hartmann, S.(1995). Models as a Tool for Theory Construction: Some Strategies of Preliminary Physics. In Herfel, W., Krajewski, W. Niiniluoto, I. and Wojcicki, R. (Eds.) *Theories and Models in Scientific Process*. Amsterdam: Rodopi, 49-67.
- Hausman, D. M. (1992) *The inexact and separate science of economics*. Cambridge: Cambridge University Press.
- Holyoak, K. and Thagard, P. (1995). *Mental Leaps. Analogy in Creative Thought*. Cambridge, Mass.: Bradford.
- Mäki, U. (2009). MISSING the World. Models as Isolations and Credible Surrogate Systems. *Erkenntnis* 70 (1), 29-43.

Persson, J. (2011). Three conceptions of explaining how possibly - and one reductive account. In de Regt, Okasha, Hartmann (Eds.) *The European Philosophy of Science Association Proceedings*, vol 1. Dordrecht: Springer, pp. 275-286.

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

Resnik, D. B. (1991). How-possibly explanations in biology. *Acta Biotheoretica* 39: 141-149.

Schelling T C (1971) Dynamic Models of Segregation. *Journal of Mathematical Sociology*, 1(2), pp. 143-186

Sugden, R. (2000). Credible worlds: The status of theoretical models in economics. *Journal of Economic Methodology*, 7(1), 1–31.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, 46, 35–57.

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

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