

**MISSing the world.**  
**Models as isolations and credible surrogate systems**

Uskali Mäki  
Academy of Finland  
TINT

October 2008 version

To appear in *Erkenntnis*

**Abstract**

This paper shows how the MISS account of models – as isolations and surrogate systems – accommodates and elaborates Sugden’s account of models as credible worlds and Hausman’s account of models as explorations. Theoretical models typically isolate by means of idealization, and they are representatives of some target system, which prompts issues of resemblance between the two to arise. Models as representations are constrained both ontologically (by their targets) and pragmatically (by the purposes and audiences of the modeller), and these relations are coordinated by a model commentary. Surrogate models are often about single mechanisms. They are distinguishable from substitute models, which are examined without any concern about their connections with the target. Models as credible worlds are surrogate models that are believed to provide access to their targets on account of their credibility (of which a few senses are distinguished).

Earlier versions were presented at the workshop on Models as Isolations and Credible Worlds (Helsinki, 8 March, 2008) and at the SPSP conference (Twente, 23-25 August, 2007). I am grateful to both audiences for their comments and suggestions, and in particular to Emrah Aydinonat, Till Grüne-Yanoff, Tarja Knuuttila, Jaakko Kuorikoski, Robert Sugden, Petri Ylikoski, and the anonymous referees.

## **1. Introduction**

Recent investigations into the use of models in science refer to them in terms such as representation, idealization and surrogate reasoning (e.g. Suarez 2004, Weisberg 2007, Contessa 2007). In my own work, I have examined the role of idealization in effecting isolation (what I call the experimental moment in modelling), have suggested an account of models as pragmatically and ontologically constrained representations, and have explored ways in which modelling may involve truth acquisition even when employing false idealizations (Mäki 1992, 1994, 2001a, 2005, 2006, 2009a 2009b).

In his celebrated paper on economic models as credible worlds, Robert Sugden (2000/2002) contrasts his account with those of Daniel Hausman (modelling as exploration; Hausman 1992) and myself (modelling as isolation), among others. Ever since Sugden's first presentation of his paper at a conference I organized in 1997, I have failed to see the contrast between his and my accounts as a stark one, or even as existing at all. In what follows I will outline my current understanding of models, and show how it accommodates, elaborates and is enriched by many of Sugden's valuable insights, as well as the idea of modelling as exploration. Thus I suggest that the three perspectives – credible worlds, isolation, and exploration – describe complementary aspects of modelling rather than offer rival alternatives.

My account of models depicts them in terms of isolations and idealizations on the one hand, and of representations and surrogate systems on the other. The first step is to give an outline of the account of Models-as-Isolations-and-Surrogate-Systems (MISS).

## **2. The MISS account of models**

### **MI(SS): Isolation by idealization**

Isolation by idealization is the experimental moment in an important and large class of models in science. What this involves is control for noise so as to isolate some important fact, dependency relation, causal factor or mechanism. The controls are implemented causally in material experiments, while in theoretical modelling other things are

neutralized by way of idealizing assumptions and silent omissions. Idealizing assumptions are often false, but this is not necessarily a reason for concern, since many such assumptions are employed and manipulated as strategic falsehoods that serve the purpose of isolation. In assuming that certain potentially efficacious factors are absent, constant, or in normal or otherwise suitable states, the modeller isolates whatever seems worthy of close attention from those other things (Mäki 1992, 1994, 2005, 2009a, 2009b). An astronomical model of the solar system isolates and examines the operation of gravity between the sun and one planet by assuming that the gravitational forces with other planets and moons are nil. An economic model of international trade isolates and examines the functioning of comparative advantage by assuming that trade takes place just between two countries, each producing just one kind of good, and that there is no trade with other countries.

Sugden has been sceptical about the idea that models are based on isolation (see also Knuuttila 2009). Now consider one of the examples he uses for illustrating his ideas of what models are and do, namely Thomas Schelling's checkerboard models of residential segregation (Sugden 2000/2002, Schelling 1979). These models show how residential segregation emerges as an unintended and perhaps unwanted consequence of individual households' residential preferences and decisions in a very simple toy city on a checkerboard: the households may not want ethnic segregation and yet it may emerge. It seems obvious to me that Schelling's checkerboard modelling also applies the method of isolation. These models involve assumptions that serve the function of neutralizing a number of causal factors: no collective action, no big income or price differences, no forceful zoning, identical locations, and so on. These assumptions help to describe a model that isolates a special kind of causal connection between individual preferences and residential patterns (I will discuss this in more detail later).

The idea that models are based on theoretical isolation, partly effected by means of idealizing assumptions, has been essential to the MISS account. Sugden takes issue with this and suggests that models as credible worlds involve more than isolation, namely *construction*. He suggests that model building is a matter of construction *rather than* isolation. The reason why he is sceptical about my stress on isolation is that he interprets

isolation in a special way (which happens not to be my way). In his view, isolation is a matter of beginning with a particular target object then gradually removing or eliminating or peeling off its features and thereby finally getting to the model. One reason why he thinks this is a wrong image of modelling is that many assumptions, instead of removing real-world features from the model world, appear to *add* to it features that are not to be found in the real world. Sugden believes that he must therefore propose an alternative account of model building as construction.

To these suspicions I have two responses that help me to elaborate the MISS account with regard to the nature and role of isolation. First, it is often the case that what may appear not to be a matter of removing a real-world feature is so after all. Let us consider two representative examples. In assuming perfect information on the part of economic agents, a model appears to *add* a feature that does not obtain in the real world: an excessively powerful mental capacity seems to be attributed to the model agents. However, it seems to me that the correct reading of the function of this idealizing assumption is that it is used to *remove* certain real-world features from the model world: the search, acquisition and processing of information. Likewise, when performing equilibrium modelling in a comparative static mode, one typically assumes the instant adjustment (time needed = zero) of certain variables, or their infinitely fast adjustment (speed = infinity). It may appear that a fantastically unreal feature is added to the model world, but again, what happens is that one thereby removes a real-world feature from the model world, namely the process of adjustment. Indeed, appearances may mask what is really the case, also in modelling.

My second response is even more direct. I take theoretical isolation to be a central characteristic of an important class of models, akin to isolation in material experiments (Mäki 1992, 2005; see also Morgan 2005). What these two procedures share is the goal or function of closing a system by neutralizing a number of factors that are not included in the isolated system. This outcome is essential for isolative modelling, while the precise way in which isolations are implemented is inessential. Just as in a chemical experiment one may purify a substance from disturbing ingredients by adding a purifier or cleaning device, theoretical isolation may require adding ingredients to the model that are not

found in the real world. Isolation is not just a matter of “peeling off”, but involves whatever distortion is needed for accomplishing it, such as exaggerations of the included features.

One source of misunderstanding may be the process-product ambiguity that characterizes the term ‘isolation’ in many of its uses. The stress in the MISS account of models is on isolation as the product. Another way of putting my response to Sugden is to say that isolation is the product in model building, whatever the details of the process. Sugden seems to have interpreted isolation as the way in which models are being built (the process), whereas I see it in terms of what models accomplish (the product). What I suggest combines the ideas of construction and isolation. In building a model one imagines a surrogate world, which involves an active process of construction whereby isolations are accomplished. It is thus not construction *rather than* isolation, but *both* construction *and* isolation that are involved in modelling.

### **M(I)SS: representation by surrogate systems**

The general view that models are representations and surrogate systems is now widely held among philosophers, but many advocates appear to adhere to a special version of the idea. I am not an exception (see e.g., Mäki 2001, 2005, 2006, 2009a, 2009b). My account has a distinct realist flavour, but its realism is qualified in two ways. First, it distinguishes between the representative and resemblance aspects of representation, but does not require successful resemblance to be constitutive of representation. Secondly, my account embeds models in a pragmatic context that includes the modeller’s purposes, audiences and commentary. Models are viewed as pragmatically and ontologically constrained representations. Here is my current formulation:

**Agent *A***  
**uses object *M* as**  
**a representative of some target system *R***  
**for purpose *P*,**  
**addressing audience *E*,**  
**prompting genuine issues of resemblance to arise;**  
**and applies commentary *C* to identify and align these components.**

Models represent in two ways, both of which require the model to be a model *of* something else. First, to say that a model is a *representative* of some target is to say that it stands for that target as its surrogate. A model  $M$  represents a target system  $R$  in the sense that  $M$  is a representative of  $R$ : in other words,  $M$  is  $R$ 's surrogate system. The epistemic function of models as representatives or surrogates is that one does not seek to acquire information about target  $R$  by examining  $R$  directly, one rather examines  $M$  directly, engaging in an inquiry into its properties and behaviour, with a view to indirectly acquiring information about  $R$ . Secondly, one could only hope to learn about target  $R$  by examining model  $M$  if  $M$  represented  $R$  in the second sense:  $M$  *resembles*, or corresponds to, the target system  $R$  in suitable respects and sufficient degrees. This second aspect of representation enables models to serve a useful purpose as representatives: by examining them as surrogate systems one can learn about the systems they represent.

However, one might also fail to learn about the target by examining the model, but this should not be taken to imply that there is no representation. Representation does not require resemblance: it only requires *issues of resemblance* to potentially arise. Such issues must be *genuine* in two ways: the model must have a likely capacity to resemble, so resemblance must not be utopian, beyond the reach of the modeller; and irrelevant resemblances do not count (any two things resemble one another in any number of ways, so some limitations are needed).

Modelling as representation involves both a subject and an object: the agent and the target, the modeller and the modelled. The representative aspect of representation is a product of voluntary cognitive activity, such as imagination and selective choice, while resemblance involves its involuntary dimension. The first is up to the modeller to decide, the latter is for the real-world target to “decide” (although in fact the modeller also decides what the real world has “decided”). A qualification is needed here, however, and this is where the pragmatic context begins to bite. Resemblance is constrained not only ontologically, but also pragmatically. The ontological constraints are due to the objective properties of the target, while the pragmatic constraints derive from the purposes and audiences of modelling. The pragmatic context shapes the respects and degrees of

resemblance that are sought and judged relevant in any given act of modelling (as indicated by the Commentary).

The *purposes* of representation may be epistemic – such as answering some limited explanatory questions or isolating an important mechanism – or non-epistemic – such as solving some practical problem and aiding in policy making. The *audiences* may be academic or non-academic, like-minded or non-like-minded, experts or non-experts, and so on. With respect to various audiences, models may be used for communicating, delivering information, persuading, impressing, excluding (from the expert conversation), and educating (thereby seeking to include the audience in the conversation), for example. Such purposes and audiences have varying consequences as far as the models that are built and used are concerned. Note that regarding the pragmatics of modelling my account is an enrichment of Ronald Giere's account that links the notion of the modeller's purposes with the notion of model (e.g., Giere 1999).

The addition of the idea of *Commentary* is a further major enrichment. Given the complex structure and contextual dependence of representation, the role of the Commentary becomes pronounced. Its task is to identify the various components of representation and to align them with one another. The ways in which the modeller (and others using or discussing the model) provides a Commentary vary from situation to situation, from modeller to modeller, and from audience to audience. The Commentary may be attached to the model, say included in the publication introducing the model, or it may appear only informally at seminar discussions. It might be fairly complete and explicit, or partial and largely implicit. As I will document below, Schelling included a partial Commentary about his checkerboard models in the published papers in which he presented them, and Sugden made the Commentary more complete and explicit (taken still further by this and other articles in the present volume).

Returning now to the idea of models as surrogate systems, I suggest that it is advisable to distinguish between the *model* and the *model description*. Thus we may take many models to be imagined abstract objects rather than anything concrete (such as mathematical symbols, narratives, boxes and arrows, or checkerboards). We could then see these concrete items serving as model descriptions. One advantage of this is that it

gives us a natural way of speaking about one and the same model (as an imagined abstract system) being described variously in terms of different concrete manners and materials.

Another consequence of the distinction between a model and a model description is that it helps us see the way in which model descriptions play an important role in examining the properties of surrogate systems in order to acquire information about target systems. The properties of models are examined by performing inferences among model descriptions, such as making and relaxing assumptions and checking what implications follow (for goals such as checking the robustness of segregated outcomes in variously specified model worlds; on this, see the article by Kuorikoski and Lehtinen in this issue). This is what I believe Hausman and Sugden would call *exploration*. In this sense, exploration deals only with models, and says nothing about their targets or how they relate to their targets (Aydinonat 2007 seems to use ‘exploration’ more broadly).

When models are treated as surrogate systems the exploratory stage is not the final phase of inquiry. Surrogate systems are examined directly in order to indirectly acquire information about some target systems. Nevertheless, the MISS account of models does accommodate the idea of exploration as an integral moment of modelling – rather than providing a separate and rival perspective on it.

### **3. Models as credible worlds are MISS**

Sugden calls models such as Schelling’s checkerboard models “credible worlds”. I have more to say about the credibility of credible worlds in a later section and focus here on the obvious idea that credible worlds are surrogate systems and representations. The representative aspect is in place: a checkerboard city as a credible world is a surrogate system in the sense of standing for real cities and serving as the direct subject of examination by the modeller. Surrogate systems are constructed and manipulated by concrete means (such as ruled paper and pencil, a checkerboard and coins, computers) and idealizing assumptions and silent omissions (needed to neutralize the effects of other things -- such as, in Schelling’s checkerboard cities, collective action and income and

price differences – so as to isolate a social mechanism). The resemblance aspect is also in place: genuine issues of resemblance arise as to whether what is the case in the surrogate systems of checkerboard cities is also the case in the target systems of real cities. I make some observations about the pragmatics of purposes, audiences and commentary in a later section, which further fortify the suggestion that credible worlds are MISS.

I have not yet been very precise about the identity of Schelling's models – nor has Schelling himself, or Sugden. What exactly is the model here? One observation is that it *can be presented and manipulated through* different material media, such as words written by Schelling, a computer, a piece of ruled paper, or a concrete checkerboard plus a set of dimes and pennies on it and a player moving them around. But would we also say that a checkerboard with coins on it – a physical object - *is* a model?

One possibility is to state that the checkerboard is a concrete material model of an imagined abstract model city, and furthermore that the abstract system could be used as a model of various concrete social systems. According to this view we would have layers of model systems (models of models of ...) with real systems at the end, or on the bottom. As I understand the concept of a model, I see no major difficulty with this way of speaking about models. So there would be a hierarchy of representations: a representation (say words and figures in Schelling's *Micromotives and Macrobehavior*) of a representation (checkerboard on Schelling's desktop) of a representation (imagined city) of a real target system (real-world cities). Given this sort of hierarchy, it is also possible to cut across it by talking about the checkerboard (as a concrete object) as a model of real cities (as concrete objects). This would be so by virtue of the supposition that representation is transitive: if A represents B, and B represents C, then A represents C.

It may be illuminating to compare the layers in this hierarchy in terms of their fragility, or the fragility of the mechanism and the outcome in each of these worlds. In the imagined model world this depends on the modeller's will, on the idealizations and initial conditions that he or she chooses to employ to accomplish the isolation. If the modeller chooses to put the mechanism in place and to shield it appropriately by means of idealizing assumptions, the outcome will be segregation. By contrast, in the concrete real world they are fragile – especially the outcome is – given that various kinds of possible

disturbances cannot be blocked, or at least not as effectively as in the model world. Finally, in the world of concrete models of abstract models, both are fragile again: the modeller's small child may muddle the formulas on paper or may upset the checkerboard with dimes and pennies on top, and thereby may destroy the concrete mechanism and the outcome; alternatively, a short circuit may stop the simulation program on the computer. This comparison suggests that isolation is particularly effective in creating and manipulating the imagined systems of credible worlds. Tighter isolation results in less fragility.

#### **4. Ways of MISSing the world: surrogate models and substitute models**

'Miss' is a suitably ambiguous term to be used for highlighting different ways in which one may miss the world in modelling. Good models function as *surrogate* systems, but sometimes a model becomes what I call a *substitute* system. The MISS account (and the shorthand 'MISS') covers both situations (Mäki 2009a). These notions could also be used for characterizing kinds of *success* and *failure* in modelling.

One may *miss* the world in the sense of longing for it (as in "I miss you so much!"). The modeller is craving for contact with the target, hankering after epistemic access to it, and builds a model to fulfil the wish. The model functions as a surrogate system: it is construed and examined with a desire to learn about the secrets of the real world. One yearns for such learning and sets out to build a model in an attempt to satisfy the desire. Surrogate models are intended, or can be employed to serve, as bridges to the world. The issue of resemblance arises and is recognized as a genuine issue. We could say that in this case the model is a *weak success*.

If the issue of resemblance is resolved by establishing that the model resembles the target while meeting the pragmatic constraints, one gets what one wanted, and the feeling of missing<sup>1</sup> evaporates. This is a case of *miss-and-hit*. We could say that it is *strong success* in representing the world: the modeller has successfully used a surrogate system to access facts about a real system.

The issue of resemblance could also be settled negatively: the model turns out not to resemble the target in appropriate ways. This gives us *weak failure* (which presupposes weak success). In this case one misses<sup>1</sup> the world by longing for epistemic access to it, but fails to establish such access. In this second sense one may *miss*<sup>2</sup> the target (as in “I wanted to attend your seminar, but missed it”). This is a case of *miss1-and-miss2*. Here one misses<sup>2</sup> the world by failing to access it; and one’s failure is a failure in one’s attempt to access the world; and one’s attempt is shaped by one’s longing for such access, by one’s missing<sup>1</sup> the world. So in the case of surrogate models, the two senses of ‘miss’ go together.

One may also miss the world in the sense of ignoring the real world, failing to get epistemic access to it simply because one does not even try (as in “I missed the class, but never mind, I’ve decided to stop taking that course”). In this case one does not miss<sup>1</sup> the world in the first sense, but does miss<sup>2</sup> it in the second sense. Here failed access is not a failed attempt. The modeller (and his or her supportive audiences) will not raise issues of resemblance at all since he or she lacks interest in accessing the world. This may happen in consequence of focusing one’s attention and efforts merely on examining the properties of the model without any interest in the resemblance aspect of representation. The exploratory stage of modelling becomes its final stage. The model becomes a *substitute system*, a freely floating subject of inquiry, unconstrained by any concern as to how it might be connected to real-world facts. It substitutes for the real system rather than serves as its surrogate. Substitute systems are *strong failures* of representation. There is a long tradition in economics of blaming economists for failing in just this way: giving all their attention to the properties of models and paying none to the relations of the model worlds to the real world.

One advantage of the MISS account is that it provides a framework for dealing with all these possible situations: surrogate models and substitute models, as well as weak and strong successes and failures in representing the world. The history of economics is filled with debates that could be described in just these terms.

I hasten to add that attributions of ‘surrogate’ and ‘substitute’, as well as of ‘success’ and ‘failure’, may be contestable. If an economist is charged with holding a substitute model only, he could respond by arguing that the model is a surrogate model after all, but its connections to the target are just very indirect; or that the connections will become evident as the model evolves in the long run; or that the model functions as an ideal type that gives rise to explanatory challenges that will be met by building other models that are surrogate models; or that examination of the substitute model promises to produce tools and insights that may be useful for developing surrogate models (Mäki 2009a). No easy and straightforward judgments may be forthcoming.

The remarks in this section have been put in terms of resemblance in the abstract. This needs to be qualified. Resemblance always comes – or does not come – in respects and degrees (for the same idea with respect to similarity see Giere 1988, and for an elucidation of how my account modifies that of Giere, see Mäki 2009b)). Issues of resemblance, pursued resemblance, and attained or unattained resemblance are always relative to the pragmatic context of representation. The context determines the relevant respects and degrees that should be at issue. We must therefore ask: MISSing and hitting what exactly?

## **5. MISSing and hitting what?**

Models function as epistemic devices in that the modeller examines the properties and behaviour of a model as a surrogate system in order to learn about the target system. But what exactly is being learned about the target? There are many ways of hitting (and missing<sup>2</sup>) it, and most of them are completely irrelevant for the purposes of modelling. For the acquisition of relevant information and understanding to take place, the surrogate system must resemble the target system in suitable ways. These suitable ways are the preferred but limited respects and degrees of resemblance. They are determined by the pragmatic constraints of the model, its purposes and audiences, as indicated by the Commentary.

Suppose one considered a concrete checkerboard a model of real cities. It would soon become obvious that certain of its features were not relevant for resemblance judgements,

such as the material of the checkerboard, and the shapes and weights of the dimes and pennies on top of it. On the other hand, study of the model may reveal that some other details might be relevant after all, such as the absolute and relative numbers of squares on the checkerboard and of the coins occupying those squares and moving between them: these numbers must be sufficiently high, and the coins must have sufficient space in which to move around for segregation to come about (see Aydinonat 2007, 2008). The general lesson the modeller learns is that, in the generation of segregation, the relevant features have to do with the dynamic interrelations between the coins. This involves something essential that the coins as physical objects lack, however: the capacity to make decisions and to act on them so as to move from one slot to another. The modeller does these things for them – but the modeller is not part of the model. This illustrates why the identity of the model must lie elsewhere, in the imagined world of abstract objects rather than in the physical world of concrete objects. It is in the imagined world in which the agents have preferences, make decisions, and act on them. The task of the modeller now is to imagine such a world in which the agents are able to move themselves – instead of the modeller actually moving them. By exercising imagination the modeller empowers the agents to possess and exercise that capacity in interaction with one another. By moving the coins on the concrete checkerboard he or she manifests the exercise of the agents' capacities in the imagined model world.

The agents' capacities in the imagined model world are part of the structures of interaction that constitute a self-reinforcing invisible-hand mechanism that has the capacity to produce segregation given even weakly discriminatory preferences. By studying and manipulating the model the modeller learns that it contains such a mechanism, the functioning of which in this manner is not dependent on many of the specifics of the process -- provided certain threshold levels are met by variables such as the initial spatial distribution of the agents, the order in which they make their moves, the ratio of the two kinds of agents, and the range of their exact preferences (see Aydinonat 2007, 2008). The process is determined by the mechanism, based on the independent decisions of the agents.

One reason why we need these kinds of models in the social sciences is the lack of transparency of many social processes in the sense that the results of modelling are not trivial or obvious beforehand. In other words, “the outcomes [should] not be so transparently related to the parameters of the model that we can proceed from postulates to results without working the model” (Schelling 1978, 170). More specifically, “the systemic consequences of individual behaviors must not be so transparent that we can treat the aggregate as though it were a collective individual, and do without the model. ... In such cases, studies of aggregates will not permit inferences about individual motives, without the help of *a mediating model*.” (Ibid., 182.) Note the idea of mediation here: a model mediates between individual motives and aggregate outcomes. Another word used in the social sciences for a “mediating model” in this sense is ‘mechanism’.

I take these passages in Schelling to be part of his Commentary, and this Commentary to suggest that his model is about a certain kind of invisible-hand mechanism. We might then consider the idea that his model hits its target just in case the mechanism isolated in it resembles a mechanism functioning in social reality. The Commentary identifies the purposes of the model and the desired respects of resemblance that these purposes imply. The model becomes interpreted as a model of a mechanism.

An important qualification is needed. In the surrogate system of the model the mechanism is in operation and it produces segregated outcomes. In some real target system, it is possible that the same mechanism is in operation and produces segregated outcomes. It is also possible that it is not in operation in some other target system, and that a segregated pattern is produced by some other mechanism(s). Schelling’s Commentary does not go far enough beyond such possibilities to allow him to state that the mechanism isolated by his model is actually in operation and responsible for a segregated pattern. The model isolates a possible mechanism (cf. Aydinonat 2007).

These observations could be put in terms of explanation. Many economic models are not intended to answer any ‘why’ questions: they are not used to explain actually occurring phenomena by indicating why they came about. They are rather used to answer questions about how something *could* have occurred. Rather than, “*Why did P occur?*” the question

posed is, “*How could P have come about?*” where the explanandum *P* is a general pattern rather than a particular event (such as segregation in urban housing markets). The latter seems a more modest question requiring different resources from those required for answering a straightforward ‘why’ question. To use the popular philosophical expression, the model does not permit inference to the best explanation. It only permits *inference to a possible explanation*: it articulates a possible mechanism that could have produced (or more weakly: could have contributed to the production of) the aggregate outcome that is observed. We could then say that Schelling’s model may at most have hit a possible mechanism. (See also Grüne-Yanoff 2009.)

Schelling’s Commentary at times expresses even weaker ambitions. He explicitly states that the rationale of his models is not the ambitious one of serving as “first approximations that can be elaborated to simulate with higher fidelity the real situations we want to examine” (Schelling 1979, 183). This negative statement is in line with having surrogate systems that represent possible mechanisms, but it is also consistent with aiming at more ambitious representations of actual mechanisms. Nevertheless, Schelling goes further and lowers his level of ambition below either of the above two: he makes the positive statement that his models at most serve the modest purpose of illustrating “the kind of analysis that is needed, some of the phenomena to be anticipated, and some of the questions worth asking” (Schelling 1979, 183). Whether or not he would be justified in making more ambitious claims, this statement, taken in isolation from what else he says and does, prompts the question whether he is wavering between considering his models as surrogates or as substitutes.

A plausible option is that he really views them as surrogate models but rhetorically plays down the intended ambitions in order to minimize the space for criticism. This would exemplify an important general characteristic of model commentary: audience-sensitivity. Model commentaries are sensitive to the background beliefs and expectations of the various audiences for modelling. It is therefore unlikely that there is a single complete and fixed commentary per model, stable across audiences and contexts. Commentaries are contextual.

## 6. Kinds of and grounds for credibility

It is now time to be a little more specific about the credibility of models as credible worlds. Very generally, to be concerned about credibility is to miss the world by longing for epistemic access to it through surrogate systems. Both aspects of representation are involved in credible worlds in this broad sense: representative (an imagined surrogate world) and resemblance (the credibility of the surrogate world). I have suggested that credible worlds are models as surrogate systems, and that a Commentary isolates parts of such surrogate systems that, it is supposed and hoped, resemble some parts or aspects in the target systems of the models.

‘Credibility’ is sufficiently ambiguous to invite a brief discussion of some notions that can be used for generating different versions of it. I have suggested (in Mäki 2002, 16 – this is the introductory essay in the volume in which Sugden’s paper is included) - that Sugden’s credible worlds are credible in that they meet what I call (in Mäki 2001b) the *www constraint* (‘www’ for the ‘way the world works’). The *www constraint* is an ontological constraint on theory and model choice requiring that a model be coherent with a set of ontological convictions. Such constraining convictions may be based on scientific theory and observation, or commonsense beliefs and political ideologies, and they are concerned with things such as human behavioural dispositions and social structure, or generally with the way the world works (this would seem to be in line with the broader idea of built-in justification, see Boumans 1999). I think this suggestion is on the right track, but it is good to note that there are several versions of credibility.

*Imaginability* is a property of models as surrogate systems, but mere imaginability, being a possible product of our powers of imagination, is not sufficient for credibility. To qualify as a credible world, a surrogate system must meet further constraints. *Conceivability* could be viewed as one epistemic form of credibility in which imaginability is constrained by general factual beliefs concerning the sorts of things that populate the target system (X is conceivable given our beliefs about society and the human mind/brain, for example). The mechanism isolated by a model may be just conceivable in that the modeller and his or her audience are able to conceive of it as a

possible mechanism in the target, given their background beliefs. These background beliefs may come up against various further constraints, which give different versions and degrees of conceivability. The ontic correlate of conceivability is *possibility* (X is possible given certain facts about society and the human mind/brain). Thus we may say that it is possible that the mechanism in a model is also there in the target system. The connection between conceivability and possibility is not tight since neither implies the other, but it is often justifiable to take conceivability as a (fallible) guide to possibility. However, the passage from conceivability to actuality is longer.

I presume it is safe to suggest that Schelling's models are credible in the minimal sense of being conceivable, and that this is what Schelling himself thought. If we wish to attribute a stronger epistemic form of credibility to them we could consider *plausibility*. In judging a surrogate system (that isolates mechanism X) to be plausible about the target the modeller believes that it is likely that X is actually in operation in the target system too – in a particular one or in a class of systems. It is not sufficient to have some general ontological convictions (about human behaviour and society) to support such plausibility judgements. More specific information is needed, based on empirical inquiries into specific cases. Schelling himself did not provide this sort of information, but probably thought favourably of taking these further steps. Later research by others has followed this line (for a brief survey, see Aydinonat 2007).

There are more markedly social versions of credibility. *Persuasiveness* is a rhetorical version in which credibility appears as a function of several things, such as the properties of the model, the beliefs of an audience, the rhetorical skills of the modeller, and the academic institutions within which persuasion takes place. *Credit-worthiness* is another social version that refers to the social supportability of a model, in other words the extent to which it is perceived as worthy of acceptance and further examination. This is also dependent on the academic institutions in which the modelling is embedded as well as on the rhetorical skills of the modeller. Given the continued attention attracted by Schelling's models, it is obvious that they are also credible in terms of being persuasive and credit-worthy.

Now consider such forms of credibility from the perspective of resemblance and the distinction between surrogate and substitute models. Conceivability and plausibility seem fit to take the issue of resemblance seriously, thus to treat the model as a surrogate system. Persuasiveness and credit-worthiness seem more neutral in this respect, compatible with both surrogate and substitute modelling. However, if we were to require nothing but persuasiveness or credit-worthiness of our models, this would be too weak to encourage anything more ambitious than substitute modelling.

## **7. Inference from surrogate systems to target systems**

As I see it, there are two classes of inference that can be exercised in relation to models. First, we may draw inferences *among model descriptions* such as mathematical equations or geometric images. Because they are inferences among model descriptions, they are also inferences *about models*. This class of inferences amounts to examining the properties and behaviour of models as imagined systems, which is what the “exploration” of models means. When Schelling sets out to check what happens in his checkerboard cities under various initial conditions, he is performing just this kind of activity. Exploration or inference among model descriptions about the model’s properties is an honourable scientific activity that an adequate account of modelling must accommodate.

Modellers performing no other inferential activity but model exploration may be dealing with substitute models. They may have no interest in the properties of the target and do not use the model as a surrogate, but if they do, they must engage in performing another kind of inferential activity: inference from surrogate systems to target systems. Naturally, the division of scientific labour may allocate these two inferential tasks to different groups of researchers. In such a (rather typical) case, the research community as a whole treats the model as a surrogate system.

It is one thing to infer from models to conclusions about the real world, which is about as easy as inferring conclusions about the model. It is quite another thing, and far more difficult, to infer to conclusions about the world that are true or are likely to be true about the world. Inference to true conclusions about the target requires the issue of resemblance

to be *settled* favourably – not only the issue to potentially *arise* as when performing inferences to conclusions that may be true or may be false. In stating this we should be precise about the kind of conclusion in which we should be interested, given what is written above in section 5 (entitled ‘MISSING and hitting what?’). The focus here is on inferring from a model mechanism to a real mechanism (rather than from a model to an empirical prediction of the outcome of the functioning of the relevant mechanism).

The challenge now is to understand and justify inference from the model to the world, from an imagined surrogate system to a real target system. There is supposed to be a positive analogy between the two systems (the same mechanism) in the ocean of negative analogies. It is this analogy that permits a move from the surrogate system to the target system, and it is here that credibility judgements enter the picture. Justified model-to-world inference requires the model to be a credible surrogate system in being conceivable and perhaps plausible insofar as what it isolates – the mechanism – is concerned. The justifiability and reliability of this inference are enhanced if the model meets the *www* constraint and further empirical constraints. Credibility therefore plays a key role in using a model as a surrogate of some target.

Consider the possibility that Schelling-type models represent social reality – a purpose that seems more ambitious than Schelling’s own modest one. Let us take checkerboard models to be models of racial segregation in the housing market in many US cities. Examination of them shows that racial segregation emerges in checkerboard cities. The question we need to ask is this: which fact or facts about real cities could checkerboard cities represent? One negative answer is that they do not represent the shape of real cities (suggesting that they consist of squares in a checkerboard-like structure), and a positive answer is that the models represent a causal mechanism that possibly also functions in real cities. It is this causal mechanism that produces racial segregation in the housing markets of model cities, and may produce it in real cities. Checkerboard models help to isolate this causal mechanism as the one possibly responsible for the patterns of racial segregation observed in reality.

Strong racial segregation emerges in Schelling's imaginary checkerboard cities. It is also a fact in many large industrial cities in the North East of the USA. Sugden (2000/2002) suggests an analogy between inductive inference from real cities and from imaginary model cities. From the fact that there is racial segregation in Baltimore, Philadelphia, Detroit and Pittsburgh we can inductively infer that it also occurs in Cleveland due to the presence and efficacy of the same causal mechanisms. Similarly, from the fact that racial segregation emerges in checkerboard cities we can infer that it emerges in all cities in which it occurs, whether imaginary or real, due to the same causal mechanisms. Both the imaginary checkerboard cities and the real cities in the North East of the USA serve as models – theoretical and natural respectively - that represent the responsible causal forces, the same in all cases (Sugden 2000/2002).

This suggests further proliferation of types of models: we not only have theoretical models of cities – layers of them – but also “natural models” , which are real cities such as Pittsburgh and Detroit. If we take real cities to serve as models, however, what could they be models of, what could they be taken to represent? An obvious idea is that natural models are taken to represent relevant causal mechanisms.

Since this form of expression may be confusing, allow me to suggest a clarification. Consider real cities as the extreme end of a series (or a continuum) of model cities along which the series is ordered in terms of numbers of attributes, simplicity of structure, and so on. We may not want to say that natural models are real cities *an sich*, or under any description. Natural models are not real cities in their entirety, but are rather real cities under a description that covers a selected set of relevant features plus particular features that distinguish them from other particular cities. What makes the first set of features relevant is that they are related to the causal mechanism that is isolated as the ultimate target of representation.

## **8. Conclusion**

In the foregoing remarks I have sought to show that there is no proper clash between viewing theoretical models in terms of credible worlds (as Sugden does), of exploration

(as Hausman does), or of isolation and surrogate systems (as in my MISS account). These three perspectives complement one another, highlighting different aspects of modelling. The MISS account can be enriched through exploitation of the key insights of the other two, or in other words an enriched MISS account can accommodate their basic features. However, more work is required to remove any remaining doubts on this matter.

## References

- Aydinonat, N. Emrah (2007) "Models, conjectures and explanation: An analysis of Schelling's checkerboard model of residential segregation", *Journal of Economic Methodology*, 14, 429-454.
- Aydinonat, N. Emrah (2008) *The Invisible Hand in Economics. How Economists Explain Unintended Consequences*. London: Routledge.
- Boumans, Marcel (1999) "Built-in justification", in *Models as Mediators*, ed. M.S. Morgan and M. Morrison. Cambridge: Cambridge University Press. Pp. 66-96.
- Contessa, Gabriele (2007) "Scientific Representation, Interpretation, and Surrogative Reasoning", *Philosophy of Science* 74:1, 48-68
- Giere, Ronald (1988) *Explaining Science*. Chicago: University of Chicago Press.
- Giere, Ronald (1999) "Using models to represent reality" in *Model-Based Reasoning in Scientific Discovery*, ed. L. Magnani, N. Nersessian, and P. Thagard. New York: Kluwer. Pp. 41-57.
- Grüne-Yanoff, Till (2009) "Learning from minimal economic models" *Erkenntnis*, this issue.
- Hausman, Daniel (1992) *The Inexact and Separate Science of Economics*. Cambridge: Cambridge University Press.
- Knuuttila, Tarja (2009) "Isolating representations vs. credible fictions? Economic modelling in theory and practice", *Erkenntnis*, this issue.
- Kuorikoski, Jaakko and Lehtinen, Aki (2009) "Incredible worlds, credible results", *Erkenntnis*, this issue
- Mäki, Uskali (1992) "On the method of isolation in economics", *Poznan Studies in the Philosophy of the Sciences and the Humanities*, 26, 319-354.

Mäki, Uskali (1994) "Isolation, idealization and truth in economics", *Poznan Studies in the Philosophy of the Sciences and the Humanities*, 38, 147-168.

Mäki, Uskali (2001a) "Models", in *International Encyclopedia of the Social and Behavioral Sciences*. (Volume 15, pp. 9931-9937). Elsevier.

Mäki, Uskali (2001b) "The way the world works (www): Towards an ontology of theory choice", in *The Economic World View. Studies in the Ontology of Economics*, ed. U. Mäki. Cambridge University Press, 2001. Pp. 369-389.

Mäki, Uskali (2002) "The dismal queen of the social sciences", in *Fact and Fiction in Economics. Realism, Models, and Social Construction*, ed. U. Mäki. Cambridge University Press. Pp. 3-34.

Mäki, Uskali (2005) "Models are experiments, experiments are models", *Journal of Economic Methodology*, 12, 303-315.

Mäki, Uskali (2006) "Remarks on models and their truth", *Storia del Pensiero Economico*, 1/2006, 7-19.

Mäki, Uskali (2009a) "Realistic realism about unrealistic models" in *Oxford Handbook of the Philosophy of Economics*, ed. Harold Kincaid and Don Ross. Oxford: Oxford University Press.

Mäki, Uskali (2009b) "Models and the locus of their truth", forthcoming in *Synthese*

Morgan, Mary (2005) "Experiments versus models: New phenomena, inference and surprise", *Journal of Economic Methodology*, 12, 317-329.

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

Suarez, Mauricio (2004 "An inferential conception of scientific representation", *Philosophy of Science*, 71, 767-779.

Sugden, Robert (2000/2002) "Credible worlds: the status of theoretical models in economics", *Journal of Economic Methodology* 7, 1–31. Also in *Fact and Fiction in Economics*, ed. U. Mäki. Cambridge: Cambridge University Press 2002. Pp. 107-136.

Weisberg, Michael (2007) "Who is a modeler?" *British Journal for the Philosophy of Science*, 58, 207-233.