Summary of: Robert Sugden - *Credible Worlds: the Status of Theoretical Models in Economics*[1].

Robert Sugden's Credible Worlds: the Status of Theoretical Models in Economics *is a very rich and complete paper on economic models, apt for an introduction on the topic in a Philosophy of Economics class. Its downside is that it is a complex paper. The following is a summary of it, intended* ***for students*** *of a course in Philosophy of Economics. My idea is to provide a "map" of the topics, preliminary to the detailed analysis of the original paper. Additionally, at the end of the summary, I outline and present some possible critiques to aid the critical reading of the paper, such lines of critique are not to be taken as research theses.* ***Please do not quote****.*

In this paper I intend to give a summary of Robert Sugden's *Credible Worlds: the Status of Theoretical Models in Economics*. The research problem, in brief, is to give an account of what economic models are and to explain if and how they relate to world phenomena, similarly to how models in physics, as an instance, relate to physical phenomena. In the paper, the author criticizes a number of accounts previously given in the literature on economic models and, after restating the core issues of the debate, he puts forward his personal account.

In order to take theoretical economics seriously, we should be in the position to say something about how economic models represent or explain the external world[2]. If that were not the case, that is, if theoretical models say nothing about the external world, the truths we learn from models would hold only in the "worlds" built by the economists and models would not be much more than mathematical games.

Justification of theoretical modeling in economics comes from the *prima facie* truth that economic models are telling us something about phenomena that occur in the social world. For instance, Thomas Schelling's model of segregation gives an account of how segregation of different communities, sharing the same territory (e.g.: a city), occurs as

[2] In this summary the locution 'external world', is used instead of Sugden's 'real world', the intended meaning will be the same. The change is motivated by the fact that use of 'real world' justifies the linguistic contrast the expression unreal world, or fake world; these latter, by the logic of the use of 'real world' in the paper, would denote the world of the models (the model-world). Clearly, however, there is not reason to call the world of the models (similar for instance, to the world of mathematics, in the sense of a human or social construct) an unreal world. The locution 'external world', used in this paper, marks the distinction between a world that is external to the human mind, or the social dimension (as preferred), and the natural or physical world, remaining neutral on the ontological status of the two.

an effect of the personal attitudes of individuals towards their neighbors[3]. Similar remarks (the details are omitted in this summary) can be made about George Akerlof's model for the market of "lemons".

Despite that *prima facie* observation, by closer looking at both the aforementioned models, it is clear that the link between the model and the external world is not as justified as one might think. Schelling's model contains assumptions that are not tested or even testable, but only supported by selected evidence: is it true that segregation occurs in cities and communities? And if it is, to what extent? Is the degree of segregation higher on average than the degree of integration, so as to justify the presence of a social force that tends to group individuals of the same kind, and separate them from individuals of a different kind?

What is necessary to understand about Schelling's model, as well as about economic models in general, is the epistemic relation between the model, as a construct of the theoretician, and the external world. In other words, we should ask how and why economic models tell us something about social phenomena such as economic exchanges or urbanization processes. Several alternative explanations are at hand in the literature.

At one of the extremes of the spectrum of possible positions, lies the claim that theorizing in economics is not concerned with descriptions of external-world phenomena, nor it purports to be. Theorizing, that is, is a conceptual exploration of the internal properties of the models themselves and the utility of this practice to concrete applications is never direct but rather indirect: models are developed *qua models*, and they can give insights into theories or into fields different from the ones that were originally thought of when the model was developed[4]. Clearly, however, neither Schelling's nor Akerlof's models are proposed as mere conceptual explorations – there is textual evidence that both are making statements about social or economic phenomena, that is, about the external world.

Alternatively, instrumentalism claims that the problem of whether the assumptions are realistic or not does not arise. This is because the assumptions of a theory serve only as means for making predictions, whose accuracy or inaccuracy is the only measure against which a theory should be evaluated. Also this interpretation runs into problems: it is hardly the case that economic models consistently provide testable predictions and Schelling's model is a good example of it: the questions mentioned above, which the model leaves unanswered, point, among other things, at the testability problem.

Another alternative account[5] on economic modeling focuses on the power of explanation rather than that of prediction: models are not meant to give testable predictions, nor are they to systematically describe some phenomena in the world. On

---

[3] For a slightly more detailed explanation of Schelling's model, see the following webpage: URL <http://web.mit.edu/rajsingh/www/lab/alife/schelling.html.>

[4] For further analysis of this point see: Hausman, D.M. (1992) The Inexact and Separate Science of Economics, Cambridge: Cambridge University Press.

[5] See, for instance: Gibbard, Allan and Hal Varian (1978) "Economic Models", Journal of Philosophy 75:664-677.

the contrary, models are a theoretician's construction, accompanied by some "story", namely, an account of how the model fits the world, but without any claims of a strong connection between the story itself and the external world. The link between model and world is said to be "casual" rather than a strong link. A similar approach, or perhaps an extension of this latter interpretation, is the one that takes models as *metaphors* of phenomena occurring in the world – models are loosely related to reality (normally by some relevant property) in the same way as the related terms stand in a metaphor. As a consequence of this approach, the questions used to evaluate models should be questions like "is it illuminating?", "is it satisfying?", etc. The shortcomings of this approach are that its tenets are vague and little more than metaphors themselves, unable to provide a precise characterization of the link between a model and the external world.

Another and perhaps better-developed approach to economic modeling contends that models operate in "isolation"[6]. Proponents of the "isolation thesis" argue that models point to some tendency in the world (for instance, the formation of segregated areas in American megalopoleis) – a tendency, in turn, is the product of a certain number of proximate causal factors. The modeler has the task of highlighting such tendency, by isolating the relevant causal factors from the ones that are not a component of that tendency. According to this approach, economics makes use of ceteris paribus generalizations, called "inexact generalizations", which are arrived at by use of a 4-step inexact deductive method. Such method, in brief, consists in the following: inexact generalizations are formed by means of isolation of causal factors in the external world; predictions are then deduced from generalizations and these will be tested; if the tests are positive, the theory will be said to be confirmed.

The problem of this interpretation is that it is too restrictive. For instance, Schelling's model contains generalizations that cannot be interpreted as ceteris paribus generalizations of external-world-phenomena (Schelling assumes that individuals in his model are identical among one another, except for the color). The same holds true not only for Schelling's segregation model but also for a wide range of economic models. Under close investigation, there seems to be little in Schelling's model that resembles the external world. Neutral (except for color) agents moving on a checkerboard according to a unique preference function look nothing like people moving from home to home in US cities, apart from the vague resemblance that Schelling himself suggests.

An alternative might be to rely on a counterfactual schema: if assumptions x and y and z and… etc., were true, then the conclusions derived from the model would be true (e.g. cities would be racially segregated). If this schema was taken literally, however, then all the empirical content of a model would simply disappear: clearly the external world is not like the "if-world" of the model.

What then is the function of models in economics? How do models relate to the external world or can at least tell something about the phenomena that take place in it?

---

[6] See: Hausman, D.M. (1992) The Inexact and Separate Science of Economics, Cambridge: Cambridge University Press

A positive account for the interpretation of models can be developed by observing the presence of a regularity in economic models (and, by means of exemplification, in Schelling's and Akerlof's models); what holds for both models is the fact that they deduce the effects of changing the values of a certain variable, when all other variables are held constant. Schelling's segregation model shows that when the preferences of individuals for having a same-colored neighbor is stronger or weaker, this affects how rapidly collections of same-color individuals cluster in segregated subareas of the same territory, whereas if the preference is null (that is, indifference dominates) individuals do not cluster by color.

In general, what Schelling's model intends to show is that "if the values of v_2, …, v_n are held constant at any given value, then the relationship between v_1 and x is…". Again by means of exemplification, Schelling's model predicts that if we keep all other variables constant (for instance, a certain degree of elasticity in the real estate market, the city's geographic configuration, etc.) and change one crucial variable, to wit, the degree of people's preference for same-color-neighbors, assumed that this value is not null, same color individuals will cluster, though at different speeds, in the same areas of the city.

The same argumentative structure, exemplified above by Schelling's model, is present in all economic models. Nonetheless, the question remains: why should that generalization be true in the external world? In other words, why should that generalization, deduced in the model, hold outside the "model world"? Deductive justification of the step from model-world to external world is clearly hopeless. It seems however possible to make an inductive step, in the following sense: if we take *R* to be a regularity (a process that occurs repeatedly) and *F* to be a set of causal factors, the schema for the interpretation of economic models would look like the following:

> "E1 – in the model world, R is caused by F.
>
> E2 – F operates in the real world.
>
> E3 – R occurs in the real world.
>
> *Therefore, there is reason to believe:*
>
> E4 – in the real world, R is caused by F."
>
> Sugden (2000:19)

The passage from E3 to E4 is not deductive, nor probabilistic; it is an inductive inference, intended as any mode of reasoning that takes from specific inferences to more general ones.

What justifies the inductive step? Sugden's next (and final) step is to give a justification of this inductive step that takes from the world of models to the external world. After dismissing two alternative accounts, the *separability account* by Hausman, and Gibbard's and Varian's *robustness account*, he intends to justify induction in terms of "credible worlds". Consider the aforementioned schema, by which economic models explain phenomena: when the values of some factors (relevant to a certain phenomenon) are fixed – ceteris paribus condition – then we can find a function, if there is one, that

relates the values of some extra factor v_1 (also relevant to the same phenomenon) with the phenomenon x.

But again, consider Schelling's model as an example, what it show is that the relation holds in a specific model, though there could be similar yet slightly different models that are not explicitly accounted for. However, we are allowed (with a certain degree of reasonableness) to infer that whatever is true of that model is true of a certain class of similar models: cities in Schelling's model are rectangular-shaped, yet in this case we can be fairly sure that in triangular shaped cities the same truths would hold, similarly for other possible changes in the original formulation.

What Sugden claims, is that by using the same inference that we make from the original model-city to other similar yet slightly different model-cities, we can make inferences from model cities to non-model cities, that is cities in the external world, provided that they are part of the relevant class of "similar cities". The criterion for such inference is that the model must represent a *credible* state of affair. In other words, the model presented by Schelling is itself an instance of a set of *possible cities* of which also "real" cities are members.

One important implication of these latter statements is that models are not simplifications, as instead it is commonly held, of the external world, rather instances of a class of objects that cuts across cities in the external world, and model cities as economists study them. All the instances of such class, *could* be real, that is, they describes states of affairs that are *credible* (Sugden 2000: 25).

Sugden's account is innovative in that it tries to bridge the gap between the model world and the external world by denying that there is a qualitative difference between the two. To Sugden, a rectangle-based checkerboard city (in which the "segregation theorems" are proven to hold), reasonably different model-cities (in which we are assured that similar theorems would hold), and cities in the external world, all them are *possible* cities, and we can therefore have a certain degree of confidence that the explanation given in the model holds in the external world as well. In other words, also in cities such as Philadelphia or Boston the following will hold: segregation happens because of individual preferences towards similar neighbors and, as the preferences become stronger or weaker the speed of the segregation process increases or decreases.

Clearly there is a certain obvious objection to this latter section: given this account, what makes the difference then between a realistic model and an unrealistic one? That is, what makes the difference between a model that represents the external world, and one that fails to do so? Here is where Sugden's notion of "credibility" enters the game. In order to make the inference from models that represent and models that fail to do so a model must be "credible", where the notion of credibility is rendered in analogy with the notion of "realisticness" in novels: "In a realistic novel, the characters and locations are imaginary, but the author has to convince us that they are credible – that there could be people and places like those in the novel." (Sugden 2000:25)

Is there a more precise notion of credibility? Sudgen makes it clear that he cannot give a complete notion of credibility, but he hints at some criteria that contribute to the credibility of a model; in particular credibility is increased if the model is coherent in two ways, that is, coherent in the logical sense and coherent with what we know about the causal processes in

the world[7]. The latter notion of coherence cannot work in this context: if we knew what forces are in the world (individuals move according to their preferences for certain types of neighbors), then we would not need a theory in order to explain what the cause of segregation is. We do know that neighborhoods tend to segregate: in most big cities there's a Chinatown, a Little Italy, a Little India and so on, but do we know that people have preferences for "similar-colored" (a.k.a. same-type) neighbors? Mathematically, this same-colored-neighbors preferences are expressed by the specific functions the model uses: does person *a* move to square *ij* or not? If square *ij* has at least *n* neighbors that are colored like *a*, *a* moves, if not, *a* does not move – . In Schelling's model 'preference' is a neutral term, which could denote what we normally mean by 'prefer', as in "I prefer hot chocolate to hot milk". On the other hand, citizen's preferences, in cities like Philadelphia or New York, may hide cultural pressure on same-race-citizens to live together, or alternatively, easiness of living conditions for people living with same-race-people. Moreover, segregated neighborhoods don't form starting from a randomly distributed checkerboard, rather from a checkerboard with many empty squares, and with new agents entering and leaving the checkerboard at different times (immigration and emigration flows). If we are dealing with racial issues, like the segregation model seems to be, the identification of the correct social forces seems to be an important issue but Schelling's model does not seem to be indentifying the exact social forces that determine segregation, since "preference" is a completely neutral term in Schelling.

So from the former discussion, coherence with the external world cannot be a feature of the model, coherence is the *explanandum* in this context, not the *explanans*. On the other hand, if we are left only with internal coherence of the model as an indicator of its credibility, this clearly won't do either: models can be perfectly coherent yet fail to represent the external world, the problem is well know in the philosophy of science context.

What can we say about the notion of credibility? Sugden does explain such notion in a little more detail than just as in terms of coherence. As mentioned above, credibility is the type of relation that stands between a "realistic novel" and the world that the novel *could* stand for. This type of explanation, however, falls short of the critique that the author himself makes of the metaphorical account of models. To recall, the metaphor-based account takes models to be metaphors and claims that the connection between a model and the external world is casual rather than systematic. Sugden's comparison of his notion of realisticness with the same notion in literature seems casual rather than a systematic (or scientific) one.

A further worry about Sugden's account is that it falsely depicts the difference between the model world and the external (social) world. It should be remembered that one of the key tenets of his account is that the economic theorist is not abstracting from the external world in order to build a model, nor is she approximating, nor postulating if-then conditions: "these models have not been formed merely by abstracting key features from the real world; in important respects they have been constructed by the authors." (Sugden 2000: 28). The inductive inference that takes us from the many-models world to the external world is possible because the two classes share a same type, that is belong to the same class of *possible worlds*.

---

[7] For instance, in Schelling's model, agents move from square to square in the checkerboard according to their preference functions, as it happens in the external world, where people move from home to home according to their preferences for some or other factors.

Ignoring the perhaps pedantic problems that philosophers bring up about inductive reasoning, it is clear that I can infer from many *y*s to all *y*s, only if the *y*s belong to the same class. I can infer that all ravens are black from the fact that $raven_1$ is black and $raven_2$ is black and $raven_n$ is black, etc. but not from the fact that *toy-* $raven_1$ is black and *toy-* $raven_n$ is black and *toy-* $raven_n$ is black to the fact that all *ravens* (viz. the birds) are black. So from a purely inferential point of view it is crucial for Sugden that all cities (model cities, physical cities, perhaps fictional cities too) be same-type cities.

There are at least two issues to the latter requirement: the first is logical the second semantic. Physical cities are, so mostly believe, actual, that is they exist as natural objects, model-cities on the other hand do not exist *actually*, they exist only *potentially* as physical cities (one could actually build a physical model of city exactly like the theoretical model city[8]). Are we allowed to say that *all* cities (actual or potential) are possible?

The second point is that it is seems *ad hoc* to say that model cities and physical cities belong to the same relevant class, semantically, the interpretation is quite stretched. One would like to maintain a certain degree of semantic common sense and in this case commons sense needs to be stretched in order to allow inductive inference from model-cities to non-model cities. Moreover, one could ask whether we are allowed, given Sugden's account to make inferences from the latter type of cities to the former. If we take a logical interpretation of the account, we should be allowed to. This is problematic, as the following example shows: I can infer from looking at the cities in the world that in all of them people reproduce and multiply themselves whereas hardly any even close phenomenon such as reproduction goes on in any model city.

[8] This is actually done in certain branches of technology: the performance of wings in race-cars and aircrafts is done buy testing models in wind tunnels as the mechanics of fluids, such as air, would require too much computational power to be tested exclusively by computer simulations.