Joint Discussion Paper
Series in Economics

by the Universities of

Aachen · Gießen · Göttingen
Kassel · Marburg · Siegen

ISSN 1867-3678

No. 52-2013

Max Albert

From Unrealistic Assumptions to Economic Explanations. Robustness Analysis from a Deductivist Point of View

Coordination: Bernd Hayo • Philipps-University Marburg
Faculty of Business Administration and Economics • Universitätsstraße 24, D-35032 Marburg
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: hayo@wiwi.uni-marburg.de

Max Albert, Justus Liebig University Giessen

From Unrealistic Assumptions to Economic Explanations. Robustness Analysis from a Deductivist Point of View

Sugden (2000) offers an answer to the question of how unrealistic models can be used to explain real-world phenomena: by considering a set of unrealistic models, one may conclude that a result common to these models also holds for a realistic model that, however, is too complex to be analyzed, or even just stated, explicitly. This is a kind of robustness argument. Sugden argues that the argument is inductive and that the methodological strategy is inconsistent with received methodological views. This paper argues that Sugden's argument is in need of improvement, that the improved version is deductive, and that the methodological strategy, if applied with care, fits well into one of the received views dismissed by Sugden, namely, hypothetico-deductivism, or the the testing view of science.

1. The Problem

In a much-discussed paper, Robert Sugden (2000) tackles an old but still unsolved problem. He sees an "enormous difference in complexity between the real world and any model we can hope to analyse" (Sugden 2000: 24, cf. also 28). Yet, some of our models seem to explain aspects of the real world. How is this possible? "How do unrealistic economic models explain real-world phenomena?" (Sugden 2000: 117) This is, of course, the central problem of economic modeling.[1]

Sugden's examples are Akerlof's (1970) model of the lemon market and Schelling's (1978) checkerboard model of racial self-segregation. Sugden views these models as paradigmatic cases of good theorizing in economics. Both models are highly unrealistic; both seem to explain features of the real world. Sugden discusses several methodological positions, asking whether they can make sense of the claims and arguments of Akerlof and Schelling. Let us first look at the claims in question, focusing on Akerlof's model.[2]

Let me first summarize—indeed, more or less paraphrase—Sugden's account of Akerlof's model. Akerlof claims that his paper has something to say about a very broad range

[1] In two follow-up papers, Sugden changed his position, now regretting his earlier realist interpretations of economic modeling and emphasizing the relations of his ideas to those of Giere (1988) (see Sugden 2009: 5n, 16-19; Sugden 2011: 718). Since I take Sugden (2000) as a starting point for developing my own view, I refrain from commenting on these further developments, which take him still farther from the position defended here, namely, hypothetico-deductivism.

[2] It seems that, pace Sugden (2000: 8), Schelling's and Akerlof's models are quite different from a methodological point of view. Akerlof's model is based on a well-developed theory, namely, the neoclassical theory of human behavior. In contrast, the behavioral hypothesis in Schelling's model is ad hoc and is not taken very seriously. Schelling seems to be more concerned with a logical point (namely, that segregation is possible without preferences for segregation) than with spelling out the consequences of a specific theory of behavior. It may very well be the case that Akerlof's model is explanatory while Schelling's is not.

of phenomena. He is, however, vague about what the paper says exactly, although explanation is one of the aims. Akerlof makes his points with the help of a highly unrealistic model of the market for used cars. The unrealism of the model is justified as a simplification that allows him to focus on those features of real markets he wants to analyze. The first model Akerlof presents is not really fleshed out; nevertheless, it generates the result Akerlof uses as a motivation (whereas the next model presents an even more extreme result).

Before I go on to summarize Sugden's account of Akerlof's paper, let me give a less sketchy version of Akerlof's first model.

> There is a surprisingly large price difference between new and as-good-as-new cars. This price difference can be explained by the asymmetry of information between buyers and sellers. Assume that there are only two kinds of cars, good and bad. The proportion of new good cars among new cars is q; the proportion of used good cars among used cars is also q. The monetary value of a good car to its owner is $a > 0$, that of a bad car is $0 < b < a$. These values are independent of the number of cars the owner already has or the age of the car, that is, used cars are as good as new cars. Buyers cannot distinguish between good and bad cars, only between new and used cars. Therefore, in a market equilibrium, there is a single price p_n for new cars and a single price p_u for used cars. Owners know the quality of their own car. Assume that a certain proportion of used cars $0 < e < 1$ is sold for some exogenous reason; their value drops to 0 for their owners. The rate e is the same for good and bad cars. We have $p_n = qa + (1\text{-}q)b < a$. We have $p_u > b$, since at least some good used cars are sold. However, $p_u > b$ implies that all owners of bad used cars sell them. Hence, the proportion of bad used cars w is higher than the proportion of bad new cars q; we have $p_u = wa + (1\text{-}w)b$ where $w = eq/[eq\text{+}1\text{-}q] < q$. Therefore, $b < p_u < p_n < a$. The price difference $p_n – p_u$ is surprising in view of the fact that the value for a given car for its owner is the same whether the car is new or used.

Let us call the phenomenon predicted by Akerlof's first model "the excessive price difference between new and almost new cars". According to Sugden, Akerlof presents no systematic evidence, neither for the excessive price difference he wants to explain nor for his central explanatory assumption of asymmetric information. He states no hypothesis. Instead, he refers to the result from his model as the Lemons Principle and then discusses further markets where the Lemons Principle is also at work. Moreover, he goes on to discuss market institutions that can be explained by the Lemons Principle in the sense that these institutions exist in order to solve the problem posed by the Lemons Principle. He presents no evidence that these institutions are absent in used-car markets. But even if they existed, it is plausible that their use would not be costless, so that a price difference between new and used cars due to asymmetric information would still exist.

How, then, do standard methodological positions fare when confronted with Akerlof's claims and arguments and, in addition, Sugden's assumption that Akerlof's paper represents economic theory at its best? Again, I summarize Sugden's accounts, refraining from commenting on his interpretation of the different methodological positions.

Popperian methodology: According to Sugden (2000: 4), the first part of Akerlof's paper fits Popperian prescriptions. There is a received theory (standard microeconomics) which predicts that the prices of new and almost-new cars should be not very different. Contrary to the theory, a large difference is observed. The only way to make the received theory consistent with the observation is an ad-hoc modification: a preference for new cars. Such ad-hoc modifications are to be frowned upon; hence, an alternative theory is needed.

However, Sugden (2000: 4) suggests that Akerlof's alternative account does no longer fit Popperian prescriptions. He moreover believes that, from a Popperian perspective, the applications of the Lemons Principle to other problems than the price of used cars must be classified as "pseudo-science": Akerlof seems to present no testable hypothesis but only an empirically ill-defined principle, the Lemons Principle, which is then confronted with confirming evidence. Moreover, although Akerlof argues that the world is often different from the model, in that there are institutions overcoming the Lemons problem, this is seen as a further confirmation instead of a refutation.

Conceptual Exploration: Sugden (2000: 8-10) concedes (to Hausman 1992) that Akerlof's paper can be partially rationalized as "conceptual exploration", as opposed to "empirical theorizing". Conceptual exploration is the investigation of the internal properties of models, while empirical theorizing is concerned with the relationship between the model and the real world. The results of conceptual exploration apply to existing theories and can be simpler formulations, useful theorems, the uncovering of previously unsuspected inconsistencies, and the development of results that turn out to be useful in completely different domains.

From a conceptual-exploration perspective, Akerlof demonstrates that certain important results, like Pareto efficiency of market equilibrium, are highly sensitive to the particular informational assumptions. These important results are less robust than previously thought. From the conceptual-exploration perspective, Akerlof's paper is not concerned with markets but with the theory of markets.

This interpretation is consistent with some of the remarks by Akerlof. However, Sugden (2000: 11) argues that Akerlof claims more, namely, that economists will be able to use the ideas of his paper to construct theories that actually do explain important economic institutions. According to Sugden, Akerlof proposes not merely a logical possibility but the sketch of an explanation, of which he is confident that it can be extended into a real explanation, even though his paper as yet presents not more than a sketch. “We are being offered potential explanations of real-world phenomena. … We should expect [these models] to provide explanations, however tentative and imperfect, of regularities in the real world. I shall proceed on the assumption that these models are intended to function as such explanations.” (Sugden 2000: 11).

Instrumentalism: From an instrumentalist perspective, models are sets of assumptions that are used to deduce testable hypotheses. Whether the assumptions are realistic or unrealistic (or, rather, true or false) is irrelevant. For instances, if we construct models to predict prices and quantities on markets, only the predictions concerning these variables are relevant; other data, for instance, concerning individual behavior, are irrelevant.

Sugden (2000: 12) rejects an instrumentalist account of Akerlof’s claims and arguments because the paper contains neither a clear distinction between assumptions and predictions nor explicit and testable hypotheses. Moreover, Akerlof defends the realism of some assumptions. He seems to be concerned with the connection between a real cause, asymmetric information, and a real effect, suboptimal volume of trade.

Incomplete Hypotheses: Sugden (2000: 14-16) discusses the views of Hausman (1992) and Mäki (1992, 1994) as an elaboration of Gibbard and Varian’s (1978) view of models as caricatures. Both accounts assume that there is something missing in economic models. These models are based on general hypotheses about the operation of relevant causal variables. However, these general hypotheses leave something out; they are incomplete in that they do not capture all causal influences that may influence real-world events. Hausman and Mäki both have ideas about how to test these incomplete hypotheses. Independently of testability, incomplete hypotheses are explanatory if they are true.

This account fits Akerlof’s model insofar as Akerlof certainly proceeds from general hypotheses that are widely accepted within neoclassical economics (Sugden 2000: 16-17). Moreover, there is an implicit assumption ruling out further influences, namely, the assumption that there are no countervailing institutions. However, Sugden (2000: 17) notes that there are some assumptions that do not fit into the account, for instance, the assumption that all traders are risk neutral, that there are only two types of cars, and so on. Hausman mentions simplifications, and these assumptions are simplifications. Sugden claims that this implies that the implications of the model are conditional on these simplifying assumptions, which means that the implications are rather weak—indeed, the empirical content is removed from the model because the assumptions do not hold in the real world.

Sugden argues, then, that the methodological positions he discusses cannot make sense of Akerlof's approach. Akerlof claims that he has to offer some insight into the workings of markets. However, all he offers are highly unrealistic models. How can Akerlof be right?

In this paper, I argue that an up-to-date version of Popperian methodology—which is also known as critical rationalism, falsificationism, hypothetico-deductivism (cf., e.g., Musgrave 2011) or, less technically, the testing view of science—can easily accommodate Sugden's many poignant observations on Akerlof's model and the methodology of economic modeling in general. Indeed, hypothetico-dedcutivism yields an improved version on Sugden's answer to the question of how to explain with the help of unrealistic models.

2. Theoretical Models and Theories

This section collects, formalizes, and elaborates on Sugden's assumptions about the nature of models and explains the role of modeling in economics from a deductivist perspective. The question is whether Akerlof's model—or any other highly simplified economic model—offers an explanation of real-world phenomena. Sugden is not completely clear on this point. Sometimes, he seems to say that Akerlof's model already provides explanations (Sugden 2000: 2). The problem, then, is to give an account of this form of explanation. Elsewhere, he seems to acknowledge that Akerlof only suggests how one might develop explanations based on the Lemons Principle (Sugden 2000: 10). However, according to Sugden's own view, more developed models will still be unrealistic. The question, then, is whether, in the development of a sequence of unrealistic models, there might come a point where the models become explanatory, and if so, what kind of threshold must be crossed.

Theoretical Models and Model-specific Implications of a Theory

Step by step, in discussing alternative methodological views, Sugden (2000: 2-19) approaches a description of economic theorizing that features theoretical models in the sense of Bunge and others (see, specifically, Bunge 1973: 91-113, H. Albert 1987: 108-111).

Let us first consider the ingredients of models in economics and other empirical sciences from a logical point of view.[3] A theory is a set, possibly a singleton set, of general

[3] This does not mean that we use the definition of a model as used in logic and mathematics. There, one distinguishes between a formal theory, which is neither true nor false, and an interpretation of the variables occurring in the formal theory. Any interpretation that turns the formal theory into a set of true statements is a model of the theory. In contrast to this definition, we assume here that the interpretation of the different symbols stays fixed. In this paper, I will not deal with the relation between a formal theory and its interpretation. In economics, it is often wrongly assumed that we deal with a formal theory because the meaning of the undefined terms is not clear. For instance, in the theory of the consumer, it is unclear what the consumption goods are, which makes different interpretations possible. However, the basic terms of a theory are, by definition, always undefined *within the theory*. This does not make the theory a formal theory. At some point, we must know what

hypotheses. These general hypotheses, as Sugden explains, are law-like. The theory is combined with non-general or singular assumptions, often called initial or boundary conditions. Initial conditions are singular descriptions of typical, historical, or counterfactual situations. These descriptions use (only) the concepts (or predicates) occurring in the general hypotheses because the general hypotheses cannot be applied to other descriptions. I call these initial conditions the descriptive part of a theoretical model.

If, for instance, the general hypotheses are the laws of Newtonian mechanics, the descriptive part uses concepts like mass and velocity but no prices. If, on the other hand, the general hypotheses are those of microeconomics, mass and velocity do not occur (except, possibly, in the units in which quantities of goods are measured) but prices do. Thus, we may consider people falling from trees—and describe the situation in terms of masses, velocity, etc.—or people buying apples—and describe the situation in terms of preferences, budgets, prices, etc. In both cases, we “abstract”, in the description of the situation, from the color of people’s hair or eyes. This kind of abstraction is enforced by the general hypotheses entertained; it is not to be confused with the more problematic instances of abstraction discussed below.[4]

Thus, a theoretical model is a set of assumptions, where we can distinguish between general hypotheses, which jointly are called the theory, and a descriptive part, which specifies a situation in the language of the theory. Let us denote the general hypotheses by H_i, $i = 1,\ldots,n$ and the descriptive statements by C_j, $j = 1,\ldots,m$. The theory is $T \Leftrightarrow H_1 \wedge \ldots \wedge H_n$. The model is $M \Leftrightarrow T \wedge C_1 \wedge \ldots \wedge C_m$.

Of course, all the implications of the theory also follow from the theoretical model since adding descriptive assumptions can never remove conclusions. However, the point of a theoretical model is that we want to focus on those implications of the theory which are relevant for the situation specified by the descriptive part. How do the model-specific

we are speaking of; otherwise, there could be no interpretation of a formal theory because the interpretation also uses basic terms that are undefined *within the interpretation*. To the extent that different scientists differ in their interpretations of the basic terms of a theory, they consider *different theories*.

[4] Bunge (1973: 97-99) identifies theoretical models with “specific theories” (see the remarks on special in contrast to general theories below). The descriptive part is called “model object”. His account begins with model objects that are not yet linked to general law-like hypotheses. This skips over the fact (emphasized by H. Albert 1987: 108) that, in the case of theoretical models, the descriptive part must use the language (the basic terms) of the general hypotheses. When Bunge discusses model objects that are intended to be “embedded” into a general theory, he does not come back to this point, which remains implicit in his account. His initial discussion therefore slightly exaggerates the extent to which theoretical models are unrealistic and subjective. A good part of the unrealism and subjectivity disappears in the case of theoretical models: features of a real situation that cannot be described in the language of the relevant law-like hypotheses are, according to these hypotheses, irrelevant for the phenomena that can be described in this language. Thus, price theory as well as Newton’s laws force us to ignore the color of people’s eyes. This might be a mistake but this problem cannot be dealt with by enriching only the descriptive parts of the models; one would also have to consider different (or additional) general hypotheses that connect eye color with gravitation or preferences.

implications look like? Here is a simple but precise formalization of Sugden's account (Sugden 2000: 17-19; cf. also Albert 1994, 1996).

General law-like hypotheses can be restated as conditionals or if-then statements. Usually, this is not done because it is rather clumsy. However, general hypotheses always refer to things that interact or have certain properties, like the masses of bodies or the preferences of agents. They can always be stated in the form "If *A* and *B* are two bodies with distance *d* and with mass *a* and *b*, respectively, then …" or "If *A* is an agent, *S* is a set of alternatives, and $a,b \in S$, then …". The general hypotheses are general in that they generalize over many cases; the bodies or agents are any bodies or agents, not specific ones. Thus, we should rather say "For all *A*, *B*, *a*, *b*, *d:* If *A* and *B* are bodies with distance *d* and with mass *a* and *b*, respectively, then …" Hence, general hypotheses are universal conditionals of the form $H_i = \forall x(R_i x \rightarrow Q_i x)$, where, of course, *x* might be a vector and where the predicates R_i and Q_i can be arbitrarily complex.[5]

When we combine universal conditionals with the description of a situation, we add restrictions on the things we are talking about. We are talking no longer about any bodies, but about bodies of equal mass, for instance. Or we are talking not about any agents, but about risk neutral agents. We would like to know what our general hypotheses have to say about these more specific cases. The statements $C_{j,}$ $j = 1,...,m$ of the descriptive part, taken together, require that the vector *x* must satisfy some possibly complex condition, which I write as *Cx*.[6] We are interested in the implications of the theory $T \Leftrightarrow H_1 \wedge \ldots \wedge H_n$ for all those *x* that satisfy *Cx*. Trivially, these implications are equal to the logical consequences of the theory $T_M = G_1 \wedge \ldots \wedge G_n$ where $G_i \Leftrightarrow \forall x(R_i x \wedge Cx \rightarrow Q_i x)$. The theory T_M is the restriction of *T* to the descriptive part of the theoretical model *M*. Any *M*-specific implication *I* of *T* can be stated in the form $I \Leftrightarrow \forall x(Cx \rightarrow Px)$, where *P* is a possibly complex predicate.[7,] This is what Sugden means when he writes (2002: 123): "[A]ny hypothesis that is generated by a deductive

[5] In the form given here, *x* is unrestricted; any domain restrictions are incorporated in R_i. On the definition of law (true law-like hypothesis), see, e.g., Swartz (2009). The formalization in first-order logic ignores an important element of law-like hypotheses. We typically assume that we can deduce counterfactual conditionals from law-like hypotheses, and that these conditionals are true if the hypotheses are true. Thus, law-like hypotheses are not just statements of empirical regularities (and, even less, of observed empirical regularities).

[6] It is not necessary to assume that the initial conditions provide a complete description of the relevant situation. If some elements of the situation remain unspecified (as in a model with $n > 1$ agents), we actually consider a set of fully specified models, with one model for every possible value of *n*.

[7] This does not rule out the possibility that, e.g., $R_1 x$ also belongs to the antecedent conditions of *I* since *Px* might, e.g., be equivalent to $R_1 x \rightarrow Sx$. The important point is that *Cx* is always part of the antecedent conditions of model-specific implications.

method must have implicit qualifying clauses corresponding with the assumptions that are used as premises."[8]

If the descriptive part of the theoretical model is a description of a specific real-world situation and refers to (a vector) a (that is, a specific time and place and so on), then the model can immediately be used to generate predictions or explanations. The testing-view of science is based on the so-called deductive-nomological (DN) account of scientific prediction and explanation. According to the DN account, predictions or explananda must be deduced from a theory (that is, a set of law-like hypotheses) and initial conditions (that is, a description of the relevant situation in the language of the theory). If the initial condition is Ca, then we can derive specific predictions or explananda either directly, or, equivalently, in a two-step procedure by, first, deriving a model-specific implication $I \Leftrightarrow \forall x(Cx \rightarrow Px)$, which together with Ca implies Pa. Modeling, then, occurs as an intermediate step in deriving predictions and explananda for specific situations, where one derives consequences of the theory in question by considering generic situations Cx.

However, the DN account of scientific prediction and explanation requires that the descriptive part of the model must be true. Consider a Newtonian model of a leaf falling from a tree in a vacuum. Newton's laws generate a prediction about the time the leaf would need to reach the surface of the earth. It would be absurd to use this model to predict the way a leaf falls in earth's atmosphere under otherwise identical conditions. Likewise, it would be a mistake to reject Newton's theory because the model of free fall in a vacuum fails to predict how leaves fall in earth's atmosphere. The model-specific implications are conditional on free fall in a vacuum; they say nothing about other situations. For this reason, we must also reject the claim that they *explain* observations made in such other situations, even if the predictions computed with the help of the model are very closely matched by the observations, as in the case of a cannonball dropped from a tree.

Sugden's problem is to reconcile the intuition that unrealistic models like Akerlof's are explanatory with the DN account of explanation. The idea that such a reconciliation is possible at least in some cases is widely accepted (see, e.g. Bunge 1973: 91-113). The

[8] Since $T \Rightarrow I$, that is, I follows deductively from T, we can consider the general hypotheses as axioms and the model-specific implications as theorems of T. On this usage of the term, most theorems of a non-tautological theory T are also non-tautological. (Instead of "tautology", it might be more appropriate to speak of "analytical truth" or "mathematical truth", but these subtleties do not matter here). Some economists, among them Sugden, refer to tautologies as theorems. In this terminology, $T \rightarrow I$ would be the theorem since $T \Rightarrow I$, so that $T \rightarrow I$ is a tautology. Of course, $T \rightarrow I$, in contrast to I, is not an empirical claim. However, if the assumptions of the model are all interpreted as descriptive, T is empty, and $I \Leftrightarrow \forall x(Cx \rightarrow Px)$ becomes a tautology because $Cx \Rightarrow Px$. This case is discussed by Sugden (2000:17) when he considers an interpretation of Schelling's model where all assumptions of the model are viewed as descriptive and rightly concludes that the resulting claim " is not an empirical claim at all: it is a theorem."

application of the model of free fall to a cannonball dropped from a tree might be a case in point. It is also assumed that theories can be tested on the basis of descriptively unrealistic models (Bunge 1973: 107-108). However, the details of when and how this is possible remain unclear. The point of Sugden (2000) is that he makes a specific proposal.

Research Programs and the "Method of Decreasing Abstraction"

However, before we come to this proposal, we have to consider some further details of the research process. Although we construct theoretical models in order to predict or explain real-world phenomena, it would be a mistake to believe that each and every model is intended to provide a prediction or explanation. This has been recognized for a long time. Consider Lakatos' (1970: 135-136) account of Newton's research program:

> Newton first worked out his programme for a planetary system with a fixed-point like sun and one single point-like planet. ... But this model was forbidden by Newton's own third law of dynamics, therefore the model had to be replaced by a new one in which both sun and planet revolved round their common center of gravity. ... Then he worked out the programme for more planets as if there were only heliocentric but no interplanetary forces. Then he worked out the case where the sun and planets were not mass-points but mass-*balls*. ... Having solved this 'puzzle', he started work on *spinning balls* and their wobbles. Then he admitted interplanetary forces and started to work on *perturbations*. At this point he started to look more anxiously at the facts. Many of them were beautifully explained (qualitatively) by his model, many were not. It was then that he started to work on *bulging* planets, rather than round planets, etc.

Lakatos' own interpretation of Newton's research program is inconsistent with central tenets of hypothetico-dedcutivism. According to Lakatos (1970), the research program consists of a "hard core" and a "negative heuristic", which says that the hard core must remain unchanged. Empirical refutations only hit the "protective belt". The "positive heuristic" of the research program determines in advance what should be changed in the protective belt if empirical refutations occur. This makes the theoretical developments largely autonomous. Empirical refutations are taken seriously only if the positive heuristic runs out of steam, that is, if all pre-planned changes in the protective belt have been made and the predictions of the theory are still not consistent with the facts.

A closer look reveals, however, that there exists a more reasonable interpretation of Newton's program. As Musgrave (1978: 189-190) explains:

> The successive 'Newtonian models' which Lakatos describes are the result of trying to find out what Newton's theory predicts about the solar system by a method of successive approximation. ... The autonomy of theoretical science simply reflects how much activity is devoted to logico-mathematical problems of deriving specific predictions. No anti-empiricist lessons can be drawn from it: predictions cannot be tested until they have been derived. ... What he [*sc.*, Lakatos] has done ... is to give us

> a falsificationist account of what it is to develop a theory and defend it against criticism.

In terms of our terminology, the hard core is the basic theory, that is, Newton's laws. The protective belt, which changes from one model to the next, contains the descriptive parts (and some auxiliary assumptions like the absence of interplanetary forces, which are a bit more complicated but which need not concern us here). The aim of modeling is to derive the predictions of the theory for a situation that at least comes close to what is known about the solar system. A reasonable modeling strategy begins with models that are extremely simple and, therefore, descriptively unrealistic in the sense that, at least outside a laboratory, we find no real-world situations fitting the description. For this reason, it makes no sense to insist that all the predictions from these highly unrealistic models are borne out by the known facts about real-world situations. What we know of the solar system cannot contradict Newton's first model because this model is a model of a quite different situation. Of course, drastic failures might motivate scientists to give up a research program very early. For instance, Newton might have given up if he would not have found a simple first model generating closed orbits of the planet around the sun. Some facts should play a role from the very beginning of such a research program. But this is a far cry from empirical testing.

The modeling strategy is to follow a method of successive approximation (or "decreasing abstraction", as it is sometimes called): one tries to make the models more and more realistic in their descriptive parts.[9] Koopmans (1957: 154), developing a similar account of economic modeling, aptly describes the very limited role of empirical observation and the dominance of mathematical problem-solving in this context:

> One may conclude ... that ... theoretical analysis still has not yet absorbed and digested the simplest facts establishable by the most casual observation. This is a situation ready-made for armchair theorists willing to make a search for mathematical tools appropriate to the problems indicated. Since the mathematical difficulties have so far been the main obstacle, it may be desirable in initial attempts to select postulates mainly from the point of view of facilitating the analysis, in prudent disregard of the widespread scorn for such a procedure.

Obviously, this account of modeling covers "conceptual exploration". It would be a mistake, however, to view conceptual exploration in this sense as necessarily distinct from

[9] The term "method of decreasing abstraction" might go back to the first pages of Wieser (1914); cf. also H. Albert (1987: 109). Essentially the same procedure is recommended by Varian (1997) to aspiring economists, with a twist: he recommends to move from one's initial sketch of a model to an even simpler model, the idea being that the simplest model generating some interesting result should be the starting point. Then one should "generalize", say, from a two-goods-two-agents model to an two-goods-*n*-agents model. This kind of generalization is a popular version of the method of decreasing abstraction: if you have no idea how many agents interact in a real-world situation, consider a model (actually: a set of models) with any number of agents. Judgments on whether one model is descriptively more realistic than another may depend on the presentation or language one uses (see Oddie 2008 on Popper's idea of verisimilitude); this does not matter for the present discussion as long as scientists' judgments agree sufficiently often, which seems to be the case.

"empirical theorizing". Unless modeling degenerates into a logico-mathematical exercise undertaken for its own sake, into "model Platonism", it serves the purpose of exploring the empirical content of the theory and, specifically, of deriving explanations of real-world phenomena or testable implications of the basic theory. Model Platonism is one extreme; the other is giving up too early on a research program. Lakatos' problem was how to steer a rational course between the two extremes.[10]

Today, science involves more division of intellectual labor than in Newton's days. Sugden's (2000) paradigm cases of good economic models are actually cases corresponding, at best, to Newton's first models: mere sketches of a basic idea and a research program. In order to publish such a first sketch and to get the research program going, the author must convince his readers that the research program is worth following, by indicating how the descriptive part of the model might be developed and what kind of phenomena one might be able to explain with a more realistic model. However, no serious predictions or explanations are involved; instead, the author engages in "casual empiricism". Popper stressed this point by speaking of "metaphysical research programs", which Lakatos changed to "scientific research program". Both terms make sense. Initially, many theories are not testable, or metaphysical, because their creators are unable to deduce predictions for real-world situations. Finding the predictions requires mathematical work: modeling. However, for the mathematical work, it does not really matter whether the theory at the core of the sequence of theoretical models is already accepted or still viewed as a "mere conjecture". Whenever a theory, new or old, is applied to a new situation that is not easily analyzed, following the method of decreasing abstraction makes sense.

The Limits of the "Method of Decreasing Abstraction"

One problem with Sugden's analysis of economic modeling is that he restricts considerations to the first models. These models are, of course, important contributions to economics. However, whether they can ever be developed into a satisfactory explanation of some phenomenon cannot be decided by considering these models and the suggestions of the

[10] Model Platonism is the perpetuation of armchair economics (see H. Albert 1963, H. Albert, Arnold and Maier-Rigaud 2012). It uses the method of decreasing abstraction as an immunizing strategy: empirical and theoretical criticisms of the basic theory are rejected; specifically, any predictive failure is blamed on the unrealism of the descriptive parts of the relevant model. In mainstream economics, model Platonism is still a relevant methodological attitude (see Arnold and Maier-Rigaud 2012), although the emergence of institutional and behavioral economics demonstrates that progress results if this attitude is abandoned and criticism is taken seriously. Note that the model Platonism critique is not directed against the method of decreasing abstraction as such. Indeed, Akerlof's model can be viewed as an example of how to make progress in this way. Akerlof did not introduce new general hypotheses but replaced a descriptive assumption of the standard model of the market, namely, the assumption of symmetric information, by the assumption of asymmetric information, which is more realistic in many real-world situations. This move initiated a new research program based on a received theory.

authors. At this stage of development, the relevant research program offers neither empirically testable predictions nor explanations of real-world phenomena. In Akerlof’s paper, all that is offered is a conjecture that, by following a research program whose outlines are indicated, testable predictions or explanations may be found. This conjecture is not implausible, but we should distinguish between a plausible claim that a problem can be solved and a solution.

Sugden might concede the point but insist that his problem would not be solved by following the proposed research program. Consider, for instance, the research program of neoclassical trade theory, from Heckscher’s and Ohlin’s first sketches (see Flam and Flanders 1991) up to Leamer’s (1984) theoretical summary and empirical investigation.[11] The problem of this research program is that even the best theoretical models are descriptively so inaccurate that they cannot be taken seriously as explanations of the pattern of international trade. Specifically, Leamer’s empirical study provides no empirical test of the theory since it relies, as Leamer himself shows extensively, on dramatic simplifications and unrealistic assumptions that are essential for the predictions of the model. Thus, it is quite dubious at this stage whether this approach can indeed be worked out into an explanation of the pattern of international trade. But you cannot come to this conclusion by considering just the early work of Heckscher and Ohlin. Similarly, we cannot determine on the basis of Akerlof’s paper whether his suggestions can be developed into successful explanations of market prices or institutions.

In principle, then, Sugden’s account of modeling is consistent with hypothetico-dedcutivism. What is missing is the dynamic aspect: the attempt to get closer to a realistic model by adding one complication after another. Economics shows examples of this strategy, for instance, the development of neoclassical trade theory. The problem that plagues neoclassical trade theory is that, although economists tried hard, they never could approach a realistic model. Experiences like this might explain Sugden’s belief—shared, it seems, by many economists—that economic models will always be unrealistic. On the basis of the present discussion, we can say that the main problem is the unrealism of the descriptive part of the models: unrealistic descriptive assumptions rob the theory of its empirical content—because the special theory contains these assumptions in the if-clause, implying that this

[11] See Albert (1994, 1996) for an analysis of the program and its methodological aspects. Leamer’s (1984) book is not the endpoint of the research program. One may discuss the question of whether there has been substantial progress after Leamer (1984) on the issues discussed here. Nevertheless, the first sixty-five years of a research program constitute a better example than a first paper like Akerlof’s.

special theory does not speak about real situations. This fact makes empirical investigations like Leamer's quite worthless.[12]

Sugden (2000) does not consider the question whether the first sketches he discusses have given rise to a sequence of theoretical models of increasing realism. Instead, he sometimes suggests that the initial sketches already offer explanations of real-world phenomena. I find this highly implausible. On the other hand, much he says indicates that he, and the authors he discusses, actually think of the *potential* of these sketches to be developed into explanations. In order to find out whether the potential is fulfilled, it would be necessary to analyze the literature originating from these sketches.

However, if Sugden is right in his belief that the descriptive part of even the best model is still unrealistic, the method of decreasing abstraction can never yield a model that delivers predictions or explanations that satisfy the truth requirement of hypothetico-deductivism. Thus, while the scope of Sugden's discussion of modeling, with the emphasis on first sketches instead of research programs, is, in my view, too narrow, it can be argued that his central problem would remain unresolved even under a wider perspective.

Robustness Checks and the Stability Conjecture

How, then, can we deal with the problem that the descriptive parts of even our best models are still unrealistic (that is, descriptively false)?

		descriptive part of the model	
		realistic	unrealistic
basic theory	corroborated	I	II
	not yet corroborated	III	IV

Table 1: Four different methodological situations.

Let us distinguish between four different methodological situations (see table 1 above). The basic theory might be well-confirmed—or, as Popperians would rather have it, well-corroborated, that is, well-confirmed in severe tests—and accepted, or it might still be

[12] Many economists (and many critics of economics) would argue that the main problem of neoclassical economics is its unrealistic (that is, false) theory of human behavior. As a behavioral economists, I certainly agree that this is an important problem. However, from a methodological point of view, it is much harder to deal with unrealistic descriptive assumptions than false general hypotheses. Even if our general theory T is false (because it contains false general hypotheses about human behavior), some special theory T_M derived from it may still be true. For instance, even if people are not rational and selfish when dealing with family and friends, they might be rational and selfish when trading with strangers on markets. However, if the descriptive parts of a model of market behavior are unrealistic, the if-part of the hypotheses of the corresponding special theory T_M are not fulfilled, meaning that the theory does not deal with the situation to which we apply it. Economists are, at least in principle, quite rational when they insist that one can still work with a false theory. It is much harder to show how one can reasonably work with models whose descriptive parts are unrealistic.

considered as a mere conjecture in need of testing. The descriptive part of our model could be realistic or unrealistic.

The deductive-nomological (DN) account of prediction and explanation assumes methodological situations I (explanation) or III (testing), which are unproblematic. Sugden assumes that these situations do not occur. If he is right, we are stuck with methodological situations II and IV. Of course, unless we can extend the DN account of testing to situation IV, situation II cannot occur, at least given Sugden's premise that all models must be descriptively unrealistic: if testing were impossible, no theory could ever be corroborated. Nevertheless, we first consider situation II, which is simpler, and then show that the ideas developed for this case can be extended to situation IV.

Explanations and the Stability Condition

Methodological situation II is usually considered unproblematic. The basic theory has survived severe testing and is accepted. The problem is to find a good model for explaining what happens in some situation of interest. If the predictions of the unrealistic model are perfect or at least good enough for our purposes, we usually assume that further improvements of the model's descriptive part would make no or no relevant difference. The model is, for our purposes, close enough to the real-world situation. This amounts to a conjecture, namely: a realistic model would yield perfect predictions and, of course, a perfect explanation for the relevant phenomena. What we actually have before us, however, is an approximate explanation. If all models must be unrealistic, approximate explanations are all we can get.

Of course, the conjecture might be false: further improvements of the model might yield worse predictions. If the basic theory is actually true, we would expect that there exists a sequence of unrealistic models for which predictions converge in some sense to perfect predictions. But this does not necessarily mean that we get improved predictions at each step.

Would we consider a model as an approximate explanation if improvements of the descriptive part yield clearly worse predictions? The answer, it seems, is clearly "no": if a more realistic model yields worse predictions, something important must still be missing in the model. Consider Akerlof's model. If including guaranties (that is, an institution that actually exists) to the model eliminated the excessive price difference between new and almost new cars, we would conclude that the model without guaranties offers no explanation of excessive price differences. As already explained, however, Akerlof takes care to argue that obvious descriptive improvements of his model do not destroy the prediction of excessive price differences.

This consideration implies that even a true basic theory and perfect predictions cannot ensure that an unrealistic model is an approximate explanation. An approximate explanation requires (a) a true basic theory, (b) descriptive assumptions that yield approximately or qualitatively good predictions, and (c) stability, or even improvement, of predictive quality under further improvements of the model's realism.

Condition (c) is crucial for a solution to our problem and should be explained in more detail. I first define stability; then, I state the stability condition.[13]

We consider a general theory *T*, a specific historical situation *a*, an observation *Pa* that is to be explained (an explanandum), and different models $T \wedge C_k a$, $k = 1,2,...$ We write $C_i a \succ C_j a$ iff $C_i a$ is more realistic as a description of situation a than $C_j a$. We then define *stability under descriptive improvements:* The explanandum *Pa* is stable under descriptive improvements of an unrealistic model $T \wedge C_1 a$ iff

(a) $T \wedge C_1 a \Rightarrow Pa$ and

(b) $T \wedge C_k a \Rightarrow Pa$ whenever $C_k a \succ C_1 a$.

Note that the definition implies $T \wedge C_s a \Rightarrow Pa$ if $C_s a$ is true, that is, the explanandum *Pa* follows from the theory if applied to a perfectly realistic description of the situation.

We can now define what we understand by an approximate explanation: Let *T* be a true theory, *Pa* an explanandum, and $T \wedge Ca$ an unrealistic model of situation *a* with $T \wedge Ca \Rightarrow Pa$. Then the model approximately explains *Pa* iff *Pa* is stable under descriptive improvements of the model.

This definition of an approximate explanation retains the truth requirement for explanations[14] and, in effect, requires that an explanation can always be worked out, at least in principle, into a standard explanation. It moreover satisfies the intuitively reasonable requirement that the existence of a predictive improvement leading to a model that no longer implies *Pa* indicates that an important explanatory factor must be missing from the model $T \wedge Ca$—even if the realistic model still implies *Pa*.

These definitions imply that we might reasonably search for simpler explanations once we have found a theoretical model that explains some observation. We might *decrease* the

[13] Similar definitions might be used to cover the case of a false basic theory yielding good predictions for certain situations, which are stable under improvements of the basic theory. Newton's laws and Einstein's relativity theory in situations where masses are not big and relative velocities are not too large provide illustrations.

[14] Many economists are prepared to give up the truth requirement for explanations, probably because they assume that truth is not to be had in any case. This is much too pessimistic in my opinion. While Newton's theory is false according to modern physics, many interesting and practically relevant results derived from it are considered as true—just consider situations where masses and velocities are not too high and replace point predictions like $x = a$ by interval predictions like $x \in [a\text{-}\epsilon, a+\epsilon]$ for suitable values of ϵ. Generally, many interesting consequences of our theories, including law-like consequences, are likely to be true. For this reason, the truth requirement in many explanations might actually be fulfilled.

realism of a theoretical model and find that it still explains the observation in question because this less realistic model as well as all its descriptively improved versions imply that the observed event must occur.

Testing and Accepting the Stability Conjecture

It is, of course, only a conjecture that some unrealistic theoretical model approximately explains an observation. The traditional reason is that the general theory as well as the descriptive part of the model might be false. The extended definition introduces a further conjectural element, namely, the conjecture that the explanandum is a stable conclusion under predictive improvements of the model. In many cases, we may not be able to explain without invoking the stability conjecture.

One problem might be that we never can find a realistic description of the situation. Consider, for instance, the solar system. Obviously, a description that takes all potentially relevant features into account is impossible: too much rubbish flies around, and even the big bodies are so irregular that they defy exact descriptions. We might try to circumvent this problem by some kind of perturbation theory that, in fact, allows us to derive properties that hold for large sets of models that differ only in the fine details we are unable to observe. However, such a program might fail. The same goes for trade theory: the complexity of the situation defies observation (not to speak of description) of all the details. The true model may remain unknown for these reasons.

But even if we were able to observe and state a complete description of all the details, we might be unable to analyze the resulting model. Typically, we would have to resort to simulations of large models. This in itself is not a problem, but if the models are very big, we might have insufficient computing power.

Still, in both cases, predictive improvements of the model might be feasible. And this means that we can test the stability conjecture. In a typical research process, some researchers will take the position that stability holds, meaning that the present model already yields an approximate explanation. Other researchers will criticize this position.

A criticism of the stability assumption can resort to models that are descriptively less realistic than the best current model. All that is needed is a plausible conjecture that a descriptive improvement (that is, the introduction of a hitherto neglected known feature of the relevant situation) destroys the prediction of the best current model. For instance, we do not need a high-dimensional model of international trade in order to show that the linear relation between net trade vectors and factor-abundance vectors predicted by the Heckscher-Ohlin-Vanek model (used by Leamer 1984) does not hold if tariffs and costs of transport cause international price differences in traded goods, as they actually do. Thus, it may be possible to criticize the stability assumption without actually analyzing a model that is more complicated than the best current model.

The stability conjecture is a combination of two conjectures, one mathematical and one empirical.[15] It would be only a mathematical conjecture if the realistic model were known but just too complicated to be analyzed. But even this purely mathematical conjecture needs an empirical input. After all, the stability conjecture does not say that it is impossible to destroy the relevant prediction by changing the descriptive part of the model. The stability conjecture refers to descriptive *improvements*, meaning that changes of the descriptive part are only relevant if they yield a better approximation to the relevant real-world situation.

But due to observation problems, not all potential descriptive improvements of a model may be known at a given time. In such a case, there is also an empirical conjecture involved: the conjecture that we will not discover an aspect of the situation whose introduction into the model would destroy the prediction.

Hence, opponents of the stability conjecture can use empirical and theoretical investigations in order to attack the conjecture. If they are not able, despite serious efforts, to raise such a criticism against the stability assumption, the stability conjecture may become accepted—tentatively, as any other hypothesis. This means that the model is accepted as an approximate explanation of the observation in question. Actually, a very simple model might, after all, become accepted as an approximate explanation, namely, the simplest model for which the stability conjecture is believed to hold.

Sugden's Inductivist Account

As I understand Sugden, he would rely on induction in order to verify, probabilify, or otherwise justify the conjecture that an unavailable realistic model would explain some observation. In his view, the premises of the inductive argument are not, as usual in inductive accounts of science, observational statements but different models. If we find many unrealistic but reasonable models that all deliver the same prediction of an observed phenomenon on the basis of the same theory, we inductively conclude that a realistic model would also yield the same prediction. Hence, we conclude that our theory can explain the

[15] Mathematical results might in general be viewed as conjectures, even if there exists an accepted proof. The classical exposition of this view is Lakatos (1976); for an accessible introduction, see Davis and Hersh (1981: esp. ch. 7). Moreover, even mathematicians who do not accept this view work, of course, with conjectures, proving consequences of such conjectures and deriving relations between them. Thus, mathematical conjectures are nothing new.

phenomenon in question, although we are unable to deduce the phenomenon on the basis of a realistic model.

The purely theoretical exploration of a set of models in order to find out in what range of models a given prediction can be deduced is called robustness analysis. Sugden argues that such a robustness analysis is the basis for an inductive argument. This argument justifies the conjecture that a realistic model, which we in fact cannot analyze, would yield the same prediction.

The inductive element in Sugden's account is, in my view, unconvincing. Surely, it is not the sheer number of different models yielding the same prediction that makes robustness analysis convincing. Consider trade theory and Heckscher-Ohlin-Vanek (HOV) theorem, which predicts a linear relation between a country's net trade vector and the country's factor abundance vector (see Leamer 1984). Let us assume that we have derived this result for a two-goods-two-factors-two-countries (2×2×2) model. That is, we have derived a model-specific conclusion $\forall x(Cx \rightarrow Ex)$, where Cx summarizes the descriptive assumptions of the 2×2×2 model and Ex states the linear HOV relationship. Let us assume (counterfactually) that we had no general proof that the same relation holds for the $n \times n \times k$ model. Therefore, we might first prove the result for the 3×3×2 model, then for the 3×3×3 model, and so on, going through, say, some dozen models but stop before we have reached realistic dimensions. Would we then argue that we can, by some inductive argument, conclude that the result holds for a realistic model? Not at all. We would conclude, possibly, that the dimension is not critical for the result (which is, actually, correct), but this leaves other aspects that may still be critical.

What is essential for the conclusion is that we engage in a critical discussion. We must ask which features of the real situation might destroy the result. And, in this case, we do not have to look far. The result relies, for instance, on the equalization of goods prices through trade. In real-world situations, specifically the situations considered by Leamer (1984), goods prices differ between countries, as documented by Leamer himself in his critical discussion of the HOV model. Leamer can find no model whose descriptive part $C'x$ includes international price differences and which yields the conclusion $\forall x(C'x \rightarrow Ex)$. Actually, we may be able to find such a model by making very specific assumptions about the technology, transport costs, consumer preferences, etc. However, once we introduce these specific assumptions, we would have to show that they are descriptive improvements. It is insufficient to argue that specifications of the descriptive assumptions preserving the HOV relationship *exist*. On the present account, we must always argue that these specifications are better descriptions of the situation in question. So far at least, such a result is missing.

Thus, I agree with Sugden that robustness analysis can indeed justify the conjecture that a realistic model that we cannot analyze would yield the same prediction as several unrealistic models that we have analyzed. However, the argument is not inductive, at least not in the usual sense of the word,[16] and completely in line with hypothetico-dedcutivism.

If a theory has survived severe tests, it is rational to accept it, if tentatively. Severe testing requires that we use our background knowledge to devise tests in which the theory can founder or is even expected to founder. A similar argument is invoked in the case of robustness analysis. We must severely test the conjecture that a realistic model would make the same prediction that we have derived from our unrealistic model. A severe test requires that we do our best to find a model that is more realistic than the one we started with and incorporates a feature of the situation that destroys the relevant prediction. In this way, we can severely test the conjecture that the prediction is stable under increased descriptive realism. If we fail to find, despite our best efforts, such a counterexample to our conjecture, it is rational to accept the conjecture, if tentatively. Under these conditions, it would be rational to believe that we have found an approximate explanation for the observed result.

There are three requirements, then, for a successful robustness analysis.

1. The analysis should look for models that are descriptive improvements, which means taking relevant empirical facts into account.
2. The improvements should pick out features of the relevant situation that, given our background knowledge, have a potential to destroy the prediction under consideration.
3. The result of the theoretical analysis must be that the prediction survives.

[16] If one identifies “induction” with “learning from experience”, the testing-view of science is also a brand of inductivism. Usually, however, “induction” means something more specific. Inductive arguments (including abduction, inference to the best explanation, etc.) are deductively invalid, which is why Sugden (2000: 20) speaks of an “inductive leap”. Hypothetico-dedcuvism and inductivism are lucidly explained by Musgrave (2011): Hypothetico-deductivists reject invalid arguments. They consider them as incomplete and insist on adding the missing premises that make the argument deductively valid. If the missing premises cannot be accepted as true, the argument is rejected as unsound. The completion of inductive arguments is always unsound because the arguments’ conclusions are scientific hypotheses that are not (yet) accepted as true (cf., e.g., Sugden 2000: 19-20); hence, on pain of inconsistency, the premises deductively implying them are not (yet) accepted as true. For this reason, hypothetico-deductivists replace inductive arguments with arguments whose conclusions are not scientific hypotheses but *evaluations* of scientific hypotheses (like “it is reasonable to accept, tentatively, hypothesis *H*”). The argument then requires *epistemic* (or, as one might also say, *methodological*) premises specifying the conditions under which it is reasonable to accept hypotheses. These premises can be true independently of the truth or falsity of the scientific hypothesis under consideration. The present paper, then, introduces methodological premises that specify the conditions under which it is reasonable to accept the stability condition. Inductivists could also make use of the stability conjecture but would have to justify it inductively.

Of course, step 3 might be difficult if it involves analyzing a more realistic model. However, as already explained, this is not always necessary: we can, for instance, analyze the role of price differences in the HOV model with the help of low-dimensional versions. If we must consider very complicated models, simulations may often help. If systematic simulations with reasonable parameter values are unable to identify situations where the prediction does not hold, we may conclude, tentatively, that the prediction survives the robustness analysis.

If we have checked all empirically relevant factors that, according to our background knowledge, have the potential to destroy the prediction, and the prediction, upon closer analysis, still survived, it is rational to accept, tentatively, the conclusion that the prediction also follows from the intractable realistic model.

The problem in the case of trade theory is that it is trivial to find aspects of the real-world situation that destroy the linear HOV relation. The stability conjecture must be rejected. Therefore, even if we accept Leamer's (1984) claim that the linear relation survives a statistical test quite well (a claim that is debatable), we must conclude that, as far as we know, the factor-proportions theory cannot explain this empirical result.

Testing Theories Using Unrealistic Models

Under Sugden's assumption that all models are unrealistic, methodological situation II becomes possible only if we solve the problem of dealing with methodological situation IV: it must be possible to test and corroborate a theory although we cannot use a realistic model of the test situation. After all, the theories we consider as well-corroborated today were mere conjectures in their early days. In order to become well-corroborated, it was necessary to test them, which presupposes that we can derive predictions of the theory for the test situation. Again, the standard account of testing requires that the predictions are derived from realistic models, which according to Sugden are not to be had.

However, we can solve this problem along the lines of our analysis of methodological situation II. We now have to derive a prediction from an unrealistic model for the purpose of testing the basic theory. What is required for a test is that the prediction follows also from a realistic model, which, however, we cannot analyze. We conjecture that the prediction is stable under improvements of the model's realism, and use robustness analysis in order to test this conjecture. In order to do so, we incorporate additional features of the real-world situation into the model that, in the light of our background knowledge, might destroy the prediction. If we are unable to find such features, we tentatively accept the stability conjecture. Under these conditions, it is reasonable to consider a predictive failure as a falsification of our basic theory. Of course, further theoretical work can resurrect the theory by showing that there is, after all, a feature of the test situations that destroys the prediction.

Conclusion

The main problem considered by Sugden (2000) is the question of how economists can explain with the help of unrealistic models. Sugden proposes an inductivist version of robustness analysis to answer this question: if we find that many unrealistic theoretical models predict and explain an observation, we can conclude, inductively, that the realistic model of the situation (which is unknown or too complex) also predicts and explains the observation.

I have argued that this idea is unconvincing as it stands. However, it can be improved and extended into a convincing argument—an argument that seems to fit in with good economic theorizing as well as the testing-view of science, also known as Popperian methodology, hypothetico-deductivism, or falsificationism. The crucial element of the improved and extended argument is the stability conjecture: the conjecture that, if the model's descriptive part is made more and more realistic, up to a completely realistic model, the prediction (or the explanandum) still follows deductively. Or in other words: the stability conjecture says that the prediction survives, or is not destroyed by, descriptive improvements.

The stability conjecture can be tested, by checking strategically those descriptive improvements that, according to our background knowledge, have the potential to destroy the prediction. If the prediction survives this strategic version of robustness analysis, it is rational to accept, tentatively, the stability conjecture, which implies that the prediction holds for the situation at hand.

Explanations based on unrealistic theoretical models require that the basic theory is already accepted. The stability conjecture then refers to some observation we wish to explain. If the stability conjecture is accepted, we accept the conjecture that the basic theory is able to explain, on the basis of a realistic model, the observation in question. Since we have accepted the conjecture that descriptive improvements of our unrealistic model do not change the conclusion, it seems reasonable to say that the unrealistic model already offers an (approximate) explanation of the observation.

A theory of scientific research has the same problem as any other theory: it is explained in terms of models with unrealistic assumptions. One unrealistic assumption that we often make when analyzing the process of research is that this process is unreasonably tidy. Actually, the division of intellectual labor and the competitive nature of science lead to all kinds of gambits, which may or may not turn out to be successful. For instance, scientists may skip robustness checks and proceed with empirical testing, leaving the robustness analysis to others. Moreover, the picture emerging from the analysis above is certainly too simple. No room has been given to the fact that we often have competing explanations and competing theories. Since, however, we had only to stretch and extend hypothetico-deductivism a little bit to cover the case of unrealistic assumptions, we should be able to accommodate these complications in the usual way.

Acknowledgments

This paper is based on a comment on Robert Sugden's paper at a workshop at the *Frankfurt School of Finance and Management* in September 2012. For discussions, hints and comments, I am grateful to the participants of this workshop, especially Marlies Ahlert, Hartmut Kliemt, Bernd Lahno, and Robert Sugden, to Volker Gadenne, and to participants of a session of the *Economics Seminar* at the University of Bielefeld in October 2013, especially Herbert Dawid, Bernhard Eckwert, Christoph Kuzmics, and Frank Riedel.

References

Akerlof, George A. (1970), The market for “lemons”: quality uncertainty and the market mechanism, *Quarterly Journal of Economics* 84, 488-500.

Albert, Hans (1963), Modell-Platonismus. Der neoklassische Stil des ökonomischen Denkens in kritischer Beleuchtung, in: Friedrich Karrenberg und Hans Albert (eds), *Sozialwissenschaft und Gesellschaftsgestaltung – Festschrift für Gerhard Weisser*, Berlin: Duncker und Humblot, 45-76. Translated as Albert, Arnold and Maier-Rigaud (2012).

Albert, Hans (1984), Modelldenken und historische Wirklichkeit, in: Hans Albert (ed.), *Ökonomisches Denken und soziale Ordnung*, Tübingen: Mohr Siebeck 1984, 39-61.

Albert, Hans (1987), *Kritik der reinen Erkenntnislehre*, Tübingen: Mohr Siebeck.

Albert, Hans, Darrell Arnold and Frank P. Maier-Rigaud (2012), Model Platonism: neoclassical economic thought in critical light, *Journal of Institutional Economics* 8: 295-323.

Albert, Max (1994), *Das Faktorpreisausgleichstheorem*, Tübingen: Mohr Siebeck.

Albert, Max (1996), "Unrealistische Annahmen" und empirische Prüfung, *Zeitschrift für Wirtschafts- und Sozialwissenschaften* 116, 451-486.

Arnold, Darrell and Frank P. Maier-Rigaud (2012), The enduring relevance of the model Platonism critique for economics and public policy, *Journal of Institutional Economics* 8: 289-294.

Bunge, Mario (1973), *Method, Model and Matter*, Dordrecht and Boston: Reidel.

Davis, Philip J. and Reuben Hersh (1981), *The Mathematical Experience*, Boston: Houghton Mifflin.

Flam, Harry and M. June Flanders (eds) (1991), *Heckscher-Ohlin Trade Theory*, Cambridge/Mass. and London: MIT Press.

Gibbard, Alan and Hal R. Varian (1978), Economic models, *Journal of Philosophy* 75, 664–677.

Giere, Ronald (1988), *Explaining Science*, Chicago: University of Chicago Press.

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

Cambridge University Press.

Koopmans, Tjalling C. (1957), The construction of economic knowledge, in: Koopmans, Tjalling C., *Three essays on the State of Economic Science*, New York 1957.

Lakatos, Imre (1970), Falsification and the methodology of scientific research programs, in: Imre Lakatos and Alan Musgrave (eds), *Criticism and the Growth of Knowlegde*, Cambridge: Cambridge University Press, 91-196.

Lakatos, Imre (1976), *Proofs and Refutations*, Cambridge: Cambridge University Press.

Leamer, Edward E. (1984), *Sources of International Comparative Advantage*, Cambridge, Mass.: MIT Press.

Mäki, Uskali (1992), On the method of isolation in economics, *Poznán Studies in the Philosophy of the Sciences and the Humanities* 26, 316-351.

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

Musgrave, Alan (1978), Evidential support, falsification, heuristics, and anarchism, in: Gerard Radnitzky and Gunnar Anderson (eds), *Progress and Rationality in Science,* Dordrecht: Reidel, 181–201.

Musgrave, Alan (2011), Popper and hypothetico-deductivism, in: Dov M. Gabbay, Stephan Hartmann and John Woods (eds), *Handbook of the History of Logic, Vol. 10: Inductive Logic*, Amsterdam etc.: North-Holland, 205-234.

Oddie, Graham (2008), Truthlikeness, in: Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, http://plato.stanford.edu.

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

Sugden, Robert (2000), Credible worlds: the status of theoretical models in economics, *Journal of Economic Methodology* 7, 1-31.

Sugden, Robert (2009), Credible worlds, capacities and mechanisms, *Erkenntnis* 70, 3-27.

Sugden, Robert (2011), Explanations in search of observations, *Biology and Philosophy* 26, 717-736.

Swartz, Norman (2009), Laws of Nature, in: James Fieser and Bradley Dowden (eds), *The Internet Encyclopedia of Philosophy*, http://www.iep.utm.edu/.

Wieser, Friedrich v. (1914), Theorie der gesellschaftlichen Wirtschaft, in: *Grundriss der Sozialökonomik*, I. Abteilung, Wirtschaft und Wirtschaftswissenschaft, Tübingen: J.C.B. Mohr.

Varian, Hal R. (1997), How to Build an Economic Model in Your Spare Time, *The American Economist* 41, 3-10.