Homo Oeconomicus
https://doi.org/10.1007/s41412-022-00122-x

RESEARCH PAPER

How to Escape from Model Platonism in Economics: Critical Assumptions, Robust Conclusions, and Approximate Explanations

Max Albert[1]

Received: 9 December 2021 / Accepted: 29 April 2022

Abstract
About sixty years ago, Hans Albert criticized economists for their "model platonism", a methodological attitude that immunizes theoretical models against empirical criticism. Since then, economics has taken an empirical turn; yet, model platonism lingers on. The root of the problem is economists' reluctance to distinguish explicitly between the law-like and the situational assumptions of their models. Without this distinction, it is impossible to give a satisfactory account of the interplay between theory and empirical investigations. Based on Hans Albert's critical rationalism, the paper explains how making the distinction allows economists to escape from model platonism. By identifying critical situational assumptions and robust conclusions, economists can, and sometimes do, find approximate explanations even though they cannot completely avoid unrealistic simplifications.

Keywords Approximate explanations · Critical rationalism · Model platonism · Robustness · Critical assumptions · Unrealistic assumptions

JEL Classification B41

About sixty years ago, Hans Albert (1959, 1963) criticized economists for their, as he called it, "model platonism", that is, a methodological attitude that immunizes theoretical models against empirical criticism.[1] Model platonism is opposed to scientific realism in economics, that is, the view that economic theories and models are attempts to capture interesting truths about the world and, therefore, can at least in principle be criticized empirically, by demonstrating that the relevant claims are false.

[1] To avoid ambiguity, I refer to the works of Hans Albert by using his full name.

✉ Max Albert
max.albert@wirtschaft.uni-giessen.de

[1] Department of Economics, Justus Liebig University, Licher Straße 66, 35394 Giessen, Germany

Published online: 01 June 2022

Since the original publication of the model-platonism critique, economics has changed radically. Historians of economics speak of an “empirical turn” in economics or announce the “age of the applied economist” (Backhouse & Cherrier, 2017). The two or three decades after World War II when pure theory dominated economics now appear as an exception.

The self-image of modern economists emphasizes the interplay between theory and experience (see, e.g., Rodrik, 2018, 276). Economic theory comes in the form of theoretical models. A model is a list of assumptions yielding conclusions concerning some phenomena of interest. The mainstream view is that such a model has to be tested empirically, by confronting its conclusions with relevant observations. If the observations contradict the conclusions, the model must be modified or discarded. If the observations bear out the conclusions, the model can be accepted for explanatory and predictive purposes as well as a basis for deriving policy recommendations.

However, there is a hitch. In the discussion of a model, all sides, including the model’s inventor, typically agree that the model is unrealistic: many or even all of its assumptions are considered as false. Indeed, most economists think that this is unavoidable (see, e.g., Sugden, 2000, 24, 28; Pfleiderer, 2020, 81–82): the complexities of the situations under investigation defy realistic description; whether they like it or not, economists must simplify. If, however, many of the assumptions of the model are believed to be false from the outset, what do we learn from an empirical test? What can “empirical criticism” of such a model mean?

One radical attempt to resolve this difficulty draws upon Friedman’s (1953) methodology of positive economics: a model’s unrealistic assumptions do not matter as long as the conclusions of interest are correct, which is what one finds out by testing them. However, if the realism of assumptions were completely irrelevant, any absurd model with an interesting conclusion would have to be taken as seriously as any traditional model: aliens, magic, fairy tales—anything would go. Indeed, it is hard to see why one would need theories and models at all. Why not just invent interesting and testable hypotheses in an ad-hoc fashion, without recourse to any models?

Actually, of course, most economists tend to take their models seriously. But this requires a less permissive solution to the problem of unrealistic assumptions and an explanation of the relevance of empirical investigations to economic theory and model building. Without a clear understanding of the interaction between theory and experience, model building and empirical work tend to remain separate activities, and severe outbreaks of model platonism are to be expected. Pfleiderer (2020), for instance, criticizes methodological attitudes I would subsume under the term “model platonism”, emphasizing the irrelevance of empirical work inspired by models with highly unrealistic assumptions and the noxiousness of policy advice derived from them.

Though heavy reliance on unrealistic assumptions nowadays often meets with criticism, the critics still fail to clarify the interaction between theoretical model building and empirical work (Sect. 1). The reason for this seems to be a failure to distinguish two kinds of assumptions in a model and their different roles in empirical investigations: law-like assumptions, which taken together form the theoretical part of the model (or the theory, for short), and situational assumptions, which describe the real or hypothetical situation to which the theory is applied (Sect. 2).[2] Without taking this distinction into account, it is impossible to understand the logic of empirical investigations and, specifically, to deal with unrealistic assumptions without falling back into model platonism.

The solution to the problem of unrealistic assumptions lies in robustness considerations. Robustness is often invoked but the usual definition of robustness is unconvincing and the details of how robustness requirements might solve the problem of unrealistic assumptions remain hazy. On the basis of the distinction between law-like and situational assumptions, I propose an improved account of robustness that makes robustness a testable conjecture (Sect. 3). This solves the problem of unrealistic situational assumptions. Moreover, it allows for a straightforward analysis of the sense in which models with unrealistic situational assumptions can provide approximate explanations if their law-like assumptions are true.

The problem of unrealistic law-like assumptions—that is, of false theories—is not amenable to the same kind of solution. However, solving the problem of unrealistic situational assumptions is sufficient to restore the logic of empirical investigations. On this basis, it is also possible to specify the role of false but nevertheless useful theories.

The proposed solution results from a straightforward application of Hans Albert's critical rationalism[3] to the process of modelling in economics. It not only shows how to escape from model platonism. It also yields a more convincing description of the research process than the naïve empiricism of the modern self-image—a description reflecting the practice of critical model discussion in those not so rare instances where methodological common sense prevails (Sects. 4 and 5).

[2] One of the causes of the lingering model platonism in economics might very well be that many economists have not even a basic knowledge of philosophy of science, as argued by Arnold and Maier-Rigaud (2012) in their introduction to the English translation (Hans Albert 2012) of Hans Albert's (1963) paper (cf. also Rodrik 2018, 276). Yet, model platonism seems to be at least as widespread in philosophy of science as it is in economics (see Sect. 1.3 below).

[3] For an introduction, see Hans Albert (1985) and Paitlova (2021). Critical rationalism originated from the work of Popper (notably, 1935) as a philosophy of the natural sciences. As such, it came under attack with the "social turn" in the philosophy of science. Most prominently, Kuhn (1962) argued that the history of the natural sciences showed that science as a social institution does not work as envisaged by Popper. In view of this attack on critical rationalism on its home turf, economists might doubt its suitability as a philosophical background for a discussion of economic methodology. However, the arguments of Kuhn, Lakatos and Feyerabend against a critical-rationalist reading of the history of science have been thoroughly discussed and rejected by Andersson (1994). For a contemporary history of the scientific revolution, critical of Kuhn and his relativist followers and quite in line with modern critical rationalism, see Wootton (2015), whose emphasis on the importance of the idea of natural law should be well heeded by economists. On the institutional aspects of critical rationalism and its role as a constitution of science, see Hans Albert (1985, 48–54; 1978, 52–59; 1987, 157–160, 171–177; 2010), Jarvie (2001), and Albert (2010; 2019).

1 How to be a Model Platonist

1.1 Elementary Model Platonism

Model platonism in economics is a methodological attitude that immunizes economic theory against empirical criticism. Technically, platonism (spelled with a small "p" to indicate that this is a modern position rather than Plato's) is "the view that there exist such things as abstract objects—where an abstract object is an object that does not exist in space or time and which is therefore entirely non-physical and non-mental" (Balaguer, 2016).

Model platonism in this sense results if one takes the assumptions of a model to be statements about an "ideal type" while empirical inquiry is concerned with "real types" existing in space and time.[4] Since the model's assumptions define the ideal type, they are necessarily true (unless they are contradictory). If economic theory talks only about ideal types and not about real types, it contains no testable claims and is immune to empirical criticism.

The term "model platonism" is, however, not restricted to a platonism in the preceding sense. It encompasses all methodological positions denying that economic models can contain testable claims. Nevertheless, model platonism leaves room for empirical economics: one can investigate how historical situations differ from models. However, if economic models involve no testable claims, such comparisons yield no evaluations of the models but only implicit classifications of historical situations: a situation under investigation is judged to be more or less similar to a model, where similarities along some dimensions may, of course, be accompanied by dissimilarities along others. "Unrealistic assumptions" just fall under the heading of dissimilarities.

Model platonism, then, turns empirical economics into economic history, but economic history without a lesson. After all, noting similarities and dissimilarities between models and historical situations is just a way to re-describe the empirical data.[5] By themselves, these data yield neither explanations nor predictions nor policy advice. This would require a theoretical model that applies to the situation under investigation—which we cannot have according to model platonism.

One might, of course, select some model for predictions or policy advice anyway. However, the only conclusion from failures would be that, in the situation under investigation, the model was, in an important respect, not similar enough to the

[4] Platonism does not allow us to contrast ideal types with reality because ideal types are considered as real as real types; they just exist somewhere else. "Ideal type" and "real type" are terms that have often been used in the social sciences (and are connected with the work of Max Weber, whose methodology, however, need not concern us here). A modern philosophical terminology would be "target system" for "real type" and "model object" or just "model" for "ideal type". An alternative to platonism is fictionalism, which argues that the platonists' abstract objects do not exist but are fictions (that is, mental objects or states). Methodologically, there is no difference between model platonism and model fictionalism.

[5] According to model platonism, models play a role similar to that of a classification scheme (a "taxonomy"). Taxonomies can be a useful step in developing theories (which in turn typically require corrections of the classification scheme, as, e.g., the case of biology shows). Model platonism, however, leads to the replacement of theories by classifications.

situation, implying that the researcher chose the wrong model. Model platonists can only blame researchers for failures; the models are always innocent.[6]

In order to save empirical investigations from practical irrelevance, model platonists have two options. The first option is inductivism. Roughly speaking, inductive arguments are arguments whose premises describe observed cases—for instance, similarities and dissimilarities between a model and several historical situations—and whose conclusions are about, or extend to, unobserved cases, in particular, future cases. The problem with inductive arguments is that their conclusions do not follow from their premises, that is, one can, without contradiction, accept the premises and still reject the conclusion. Inductivists maintain, nevertheless, that science proceeds by induction: they claim that, at least under certain circumstances—for instance, if the number of observed cases is large enough—it is rational to accept the conclusions of inductive arguments.

There is, of course, nothing wrong with forming conjectures inspired by observations (or, for that matter, by anything else). According to critical rationalism, however, these conjectures need to be tested severely before they can be accepted. Listing supporting cases is no substitute for severe testing. A severe test is an effort to find counterexamples. It makes use of background ideas and, possibly, competing conjectures suggesting where counterexamples might be found. New conjectures should not be accepted unless they have survived a serious search for counterexamples. Indeed, survival of such a search is the best reason we can have for accepting a conjecture.[7]

This principle applies not only to science, where counterexamples are failed predictions, but also to logic and mathematics (Lakatos, 1976). A logical or mathematical proof is nothing but a chain of conjectures about deductive relations between propositions. Each of these conjectures can, in principle, be refuted by a counterexample. If no counterexamples can be found, the proof is tentatively accepted. The subjective feeling of certainty caused by reading a well-presented proof is irrelevant. What is relevant is the fact that a critical discussion among experienced specialists has yielded no counterexamples to any of the steps in the chain of deductions.

Inductivism, then, cannot solve model platonism's problems. Another option is more attractive. Model platonists might develop assumptions about the circumstances under which models with unrealistic assumptions are, in certain respects, sufficiently similar to situations in time and space. This means that there are two levels of theorizing: a first level where unrealistic models are developed, and a second level concerning the relation of the unrealistic models to experience.

[6] See, e.g., Rodrick (2018, 277) agreeing with Keynes that few economists have the rare gift of choosing a good model.

[7] The criticism of inductivism is one of critical rationalism's points of origin; however, the present paper is not the place for reviewing this topic. See Musgrave (1999, 314–350) for a detailed account of critical rationalism starting from the problem of induction. For recent criticisms of induction from the critical-rationalist perspective, see Musgrave (2011) on non-probabilistic inductivism, Gadenne (2013) and Albert and Hildenbrand (2016) on the new inductivism in experimental economics, and Albert (2017) on probabilistic inductivism (i.e., Bayesianism).

Yet, a list of second-level assumptions would be a second-level model. If this second-level model can do without unrealistic assumptions, the same should be possible on the first level. And if the second-level model also features unrealistic assumptions, model platonists face, again, the same problem they were unable to solve on the first level.

Model platonism, then, is hardly convincing once one begins to wonder how it might make sense of empirical investigations. Before turning to the critical-rationalist alternative, I want to discuss two important variants of model platonism: the theory-as-tautology view and structuralism. Both run into the problems just discussed. In each case, the problem of unrealistic assumptions takes center stage.

1.2 The Theory-As-Tautology View

According to the standard view, economic theories come in the form of models, and models are lists of assumptions. Let A_1 to A_n be a model's assumptions. We write $\wedge$ for "and" and A for the conjunction $A_1 \wedge \cdots \wedge A_n$. Let F be a conclusion of interest (or a conjunction of such conclusions) that follows from these assumptions.

As an example, consider the neoclassical model of a competitive exchange economy, which consists of the following assumptions, stated in a non-technical way: There are many agents, each of which is endowed with specific quantities of several consumption goods. Each agent has a complete preference ordering over the set of all conceivable bundles of these goods. There are perfectly competitive markets for the goods where agents trade these goods at market-clearing prices. External effects are absent. From the conjunction of these assumptions (A), it follows that the allocation of goods resulting from market transactions is efficient (F).

Which claim of this model should be tested? Superficially, the logical structure of the model gives no hints. Clearly, it makes no sense to test just F: the model's message is not that, unconditionally, market allocations are always efficient. Any relevant testable claim derived from the model must be a conditional claim.[8]

One might be tempted to consider, as an alternative to just F, the conditional statement "if A, then F", symbolically $A \rightarrow F$. However, since F follows from A, the statement $A \rightarrow F$ is a conceptual truth or "tautology", as it is often called in economics: it is true but contains no factual information. From a logical point of view, $A \rightarrow F$ is equivalent to "All bachelors are unmarried".[9]

The theory-as-tautology view holds that economic theory consists solely of tautologies like $A \rightarrow F$: the only message of a model is that its conclusions must hold under its assumptions.

This view is quite unattractive. Tautologies, after all, do not explain or predict. We cannot explain why Uncle Bill is unmarried by pointing out that he is a bachelor.

[8] For simplicity, it is assumed here that it can be observed whether allocations are efficient, which may be quite difficult outside of economic experiments.

[9] Technically, "All bachelors are unmarried" is an analytic (or conceptual) truth, which turns into a tautology once one replaces "bachelor" with its definition, "unmarried man".

Nor makes it any sense to use the fact that he is a bachelor to predict that he is unmarried.[10]

Moreover, the theory-as-tautology view runs into problems when confronted with the question of how theory and empirical investigations interact. Empirical testing of a tautology like $A \rightarrow F$ is obviously a waste of time since the result is known beforehand, independently of all empirical facts. If we find no cases where A is true, $A \rightarrow F$ is irrelevant but not refuted. If we find cases where A is true, that is, cases where all the assumptions of the model hold, then F must hold—just as every bachelor must turn out to be unmarried.

So what does empirical work achieve? On the basis of the theory-as-tautology view, there is no reasonable answer. For instance, in a book that, at its time, was regarded as an exemplary combination of theory and empirical investigation, the econometrician and international-trade theorist Edward Leamer retreated to the claim that "[a] judgment about the success of an empirical approximation to a tautological theory is ultimately a matter of aesthetics" (Leamer, 1984, xvi). Yet, claiming that a tautology is only approximately true is a contradiction, as when one claims that almost all, but not all bachelors are unmarried.[11]

The only attractive aspect of the theory-as-tautology view is that it absolves theorists from taking unrealistic assumptions seriously. Leamer (1984), for instance, is concerned with a model of international trade—an extension of the neoclassical model of the competitive exchange economy—whose assumptions are highly unrealistic, as he points out in great detail. Not only are most of the assumptions false for the years and countries he considers; for some of them, it is almost inconceivable that they could ever be true of any group of trading countries at any time and place.

While the conjunction A of the model's assumptions is false, one can still empirically check some conclusion F. In Leamer's case, F is a linear relation between trade vectors and factor endowments across countries. However, what could one learn from testing F? Should F turn out to be true in some situation, it would be unclear why—the model, at least, cannot explain such a result since it states conditions for F that did not hold. Should F turn out to be false, this implies that at least one of the model's assumptions must have been false in the situation under investigation—but this was already known before the test.

Unrealistic assumptions, then, make it difficult to say what a test could achieve. This difficulty should matter. But according to the theory-as-tautology view, it does not. As a tautology, the theory cannot and need not be tested. Of course, while the

[10] With respect to advice, the situation is slightly different. If Uncle Bill asks us how he could avoid marriage, it would be correct but obviously not helpful to tell him that he would have to remain a bachelor. However, in the case of economic models, logical relations between assumptions and conclusions are often not obvious. Therefore, knowledge about logical relations might actually be helpful in solving practical problems. Note that the arguments against the theory-as-tautology view are not arguments against the logico-mathematical analysis of models. The point is rather that logico-mathematical analysis of economic models is important because the models' messages are more than just tautologies (see 2.1 below).

[11] For a critical discussion of Leamer (1984), see Albert (1996). Leamer discusses approximations and robustness extensively; yet, he never explains how these discussions can be made consistent with the theory-as-tautology view.

assumptions of the model might be false, elementary logic implies that the conclusion of interest might still be true. Indeed, in Leamer's (1984, 187) opinion, the linear relationship works surprisingly well in the two years, 1958 and 1975, under consideration, despite drastic deviations from the model's assumptions. Because the whole exercise is not considered as a test of a model or a theory, it does not matter that the econometric test is just concerned with the model's conclusion.

Leamer the trade theorist, then, is completely safe from Leamer the econometrician. The trade theorist states only the tautology $A \rightarrow F$ and is silent on the question of whether any of the assumptions or the conclusion might be true. The econometrician shows that A is false but that F, nevertheless, looks quite good from an econometric point of view. This is considered as a surprise and might please the trade theorist. But if F had been rejected, the tautology $A \rightarrow F$ would still be true.

The empirical results themselves are contributions to economic history. Leamer found that, in two years, a certain linear relationship held up quite well among sixty countries. No conclusion for other times and places follows. Any conclusions going beyond the data require as a further premise some non-tautological theory; however, according to the theory-as-tautology view, economics has no such theories on offer.

1.3 Structuralism: the Non-statement View of Theories

A second variant of model platonism is called structuralism. While the theory-as-tautology view appears only in side remarks, structuralism is a movement in philosophy of science.[12] Structuralists view the model of the exchange economy as a purely formal structure involving variables like "agent 1", "good 1" and so on. An assumption like "agent 1 has ten units of good 1" must be considered as a formula which is neither true nor false. In order to apply the formal structure, the variables have to be interpreted in terms of a specific historical situation. Of course, not any interpretation will do. There exists an intended interpretation: "agent 1" has to be interpreted as a person, "good" as a consumption good, and so on. For instance, in the situation under consideration, there might be a person, Adam, and some apples. We stipulate that agent 1 is Adam and that good 1 is apples, with pieces as unit of measurement. With this interpretation, "agent 1 has ten units of good 1" turns into "Adam has ten apples", which is true or false depending on the number of apples owned by Adam.

Given an interpretation of the complete formal structure, we can ask whether the interpretation turns all formulas into true statements. In this case, the set of all the things providing the interpretation (Adam, the apples, and so on) are said to be a

[12] Structuralism is also called the "semantic view of scientific theories", in contrast to the (logical-positivist) "syntactic view". One prominent current version is the model platonism of Giere (1988, 2004), which explicit introduces "second-level theorizing" (see 1.1 above). In methodological debates on experimental economics, similar views are expressed by, e.g., Guala (2005, 205–209) and Bardsley et al., (2010, 204–204, esp. 206). See Winther (2021) for an overview of different views of scientific theories, including the variants of the semantic view. For a critical discussion of structuralism, see Morrison (2016), whose position on theories and models is quite close to the critical-rationalist position of the present paper.

model (in the logico-mathematical sense) of the formal structure. What economists call a model, then, structuralists consider as a formal structure (a set of formulas), and each interpretation that turns the formulas into true statements is called a model. From this point of view, the principal empirical question is whether the formal structure has any models.

Since structuralism considers scientific theories not as statements about the world but only as formal structures, it has also been called the "non-statement view" of scientific theories. If structuralism meets unrealistic assumptions, model platonism results. A formal structure whose intended interpretation yields unrealistic assumptions has no relevant models (in the logico-mathematical sense) in space and time. The way out is to assume that there exists a model of the formal structure not in time and space but as an abstract object. When we then interpret the formal structure in terms of the abstract object, the formulas turn into true statements about the abstract object. This move brings us back to platonism in the strict sense of the term.

According to structuralism, what theoreticians have to say is, again, necessarily true and not subject to empirical criticism. Empirical economists are left with the task of finding out which economic model fits which historical situation. Since, however, no economic model fits exactly (because of the unrealistic assumptions), they can only classify the situations they consider as more or less similar to the theoreticians' abstract objects. This is, of course, just a way of re-describing the observational data. In order to achieve more, structuralists need to embrace inductivism or resort to second-level theorizing about the relations between the theorists' abstract objects and the world of experience.

1.4 Approximation and Robustness

Some economists explicitly oppose the unrestricted use of unrealistic assumptions, demanding greater realism in economic models. On the face of it, these economists seem to reject model platonism. Yet, a closer look reveals that demands for more realism alone are insufficient to escape from model platonism.

Let us begin with the intuitively appealing idea that a model with unrealistic assumptions should in some sense be a good approximation to the situation under investigation. This seems to rule out wildly unrealistic assumptions. Alas, Friedman (1953, 15) defines a model to be a good approximation if and only if the conclusion of interest from the model holds in the situation under investigation. With this definition, the requirement that a model should be a good approximation to the situation under investigation just means that the conclusion of interest should hold. Again, it does not matter whether the assumptions are realistic or unrealistic.

Some economists have tried to improve upon this idea of models as approximations by adding robustness requirements, thereby restricting the use of unrealistic assumptions. According to Gibbard and Varian (1978, 674), even models with very unrealistic assumptions may help us to understand a situation if their conclusions are robust, meaning that the conclusions do not depend on the details of the assumptions. Similarly, Ng (2016, 182) considers simplifying assumptions to be acceptable if they simplify the analysis but do not change the conclusions substantially.

Pfleiderer (2020, 84–85) argues in favor of using a „real-world filter“, rejecting models if critical assumptions contradict what is already known. Rodrik (2015, 19, 26–27, 94–98), citing an earlier version of Pfleiderer’s paper, agrees and requires critical assumptions to be close to reality. Actually, Solow (1956, 65) already expresses similar thoughts in an opening paragraph that reads like an implicit criticism of Friedman (1953).

All these considerations come down to the same point: one may use unrealistic assumptions only if this simplifies the analysis without changing the conclusion of interest. This robustness requirement sounds reasonable, but as it stands, it is useless. Let us state the requirement formally. Let A be the conjunction of the assumptions of the unrealistic model and F the conclusion of interest from the model, so that $A \rightarrow F$ is a tautology. “Unrealistic” means: in the situation under investigation, it is known that A is false. Robustness in this sense would require that the unrealistic model A could, in principle if not in practice, be replaced by a perfectly realistic model A^* also implying F.

It is surprisingly trivial to check empirically whether robustness in this sense holds: just check F. If F turns out to be false in the situation under investigation, this implies that A^* is false for any tautology $A^* \rightarrow F$; hence, robustness fails. If F turns out to be true in the situation under investigation, and if we assume that F is no miracle and, therefore, amenable to an explanation in principle (even if we may be unable to find this explanation), then there must exist some perfectly realistic model A^* implying F. Consequently, the robustness requirement holds if and only if F is true.[13]

For this reason, the robustness requirement invoked by the post-Friedmanian proponents of realism adds nothing to Friedman’s approach to approximation. Again, an unrealistic model turns out to be a good approximation if and only if the conclusion of interest from the model holds in the situation under investigation because then and only then the robustness requirement holds.

While the demand for more realism in economic models goes in the right direction, it must be supported by a more detailed analysis of a model’s components, which then allows for a better characterization of robustness and approximate explanation.

2 Theories and Models

2.1 Law-like Assumptions, Situational Assumptions, and the Rationale of Model Building

The problem of unrealistic assumptions changes completely once one acknowledges that there are two kinds of assumptions in a model: law-like assumptions describing

[13] If F is true, $F \rightarrow F$ is a tautology with true if-part. If F itself is accepted as a perfectly realistic model, robustness is trivial. In the text, it is assumed that A^* provides something like an explanation for F in order to avoid this trivialization of the robustness requirement. Actually, it is up to the proponents of the robustness requirement to spell out reasonable restrictions on A^* which avoid triviality.

relationships assumed to hold always and everywhere,[14] and situational assumptions describing a situation to which the law-like assumptions are applied. The point of testing is to find true law-like assumptions (called "laws") and weed out false ones.[15]

Obviously, both kinds of assumptions appear in economic models. Consider the neoclassical model of an exchange economy already discussed above. The assumption that all agents have complete preference orderings on the set of alternatives is law-like: it is made in each neoclassical model and belongs to the core of the neoclassical theory. On the other hand, the assumption that each agent is endowed with some stock of consumption goods is just a description of the situation where the agents act.

A theoretical model can be written as $T \wedge S$, with T as the conjunction of all the law-like assumptions, usually called the theory, and with S as the conjunction of all the situational assumptions. We also refer to S as the description of a situation. We consider a theoretical model where S is generic, that is, given in general terms, without specifying a time or location of the situation or the persons involved.

Again, we consider some consequence of interest F of the model $T \wedge S$. The statement $T \wedge S \rightarrow F$, then, is a tautology. Now, however, the focus is not on the tautology $T \wedge S \rightarrow F$ but on the statement $S \rightarrow F$. This statement is not a tautology but a law-like consequence of the theory T: the theory implies that, whenever and wherever the situation S obtains, F must hold.[16] If T is true, its law-like

[14] Universality of law-like assumptions may appear as too strong a requirement, especially in economics. However, law-like assumptions are if–then statements. Their scope of application (their generality) can be restricted by adding the relevant conditions to the if-part, without giving up universality. Universality only fails if the conditions in the if-part refers to individuals (specific times, places, or persons). An economic example might be the Phillips curve, which described a law-like relation holding in some countries after World War II until the 1970s. Albert (1973) called such relations "quasi laws" and noted that one typically tries to explain them in universal terms. This is exactly what economists tried to do by explaining the Phillips curve and its breakdown in terms of expectation formation. See also the remarks below on how experimental economists deal with equilibrium assumptions.

[15] In the philosophy of science, law-like assumptions are often called "nomological hypotheses", and situational assumptions are often called "initial conditions" or "boundary conditions". On laws, see Swartz (2021), whose account of economic laws is, however, dubious because it assumes that individual choices are not governed by laws. Laws describe an objective connection between events and provide the only conceivable basis for learning from experiences made at one time and place about what to expect at another time or place. If there were no laws (a position often encountered in discussions in economics), "learning from experience" would be a guessing game where feedback about success or failure is irrelevant because there is nothing that can be learned. Critical rationalists accept the necessitarian concept of laws, which holds that laws imply true subjunctive and counterfactual conditionals, and whose formalization requires modal logic (cf. Albert 1998). The same view is taken as a matter of course in the natural sciences; for amusing examples, see Munroe (2014). Counterfactual conditionals are closely connected to the notion of causality: the claim that a caused b is usually taken to imply that b would not have happened if a had been absent. Modern statistics considers probabilistic generalizations of causation where causes determine the probabilities of consequences; see, e.g., Pearl et al. (2016) for a compact introduction which avoids the term "law" although completely specified "causal models" satisfy all the requirements of nomological theories (cf. also Albert 2007).

[16] The conclusion F may be a conjunction $F_1 \wedge F_2 \wedge \cdots \wedge F_n$, so that $S \rightarrow F$ is equivalent to $S \rightarrow F_1 \wedge S \rightarrow F_2 \wedge \cdots \wedge S \rightarrow F_n$. All consequences of T of the form $S \rightarrow F_i$ (of which there may be infinitely many) taken together form a theory T_S for situations described by S. Of course, T_S follows from T; in fact, T_S is equivalent to $S \rightarrow T$. The model $T \wedge S$, then, corresponds to a theory T_S which captures all relevant implications of T for situations described by S. Specifically, testing a model means testing a theory. See also Albert (1996), where these relations are spelled out in terms of first-order predicate logic.

consequence $S \rightarrow F$ is also true, even if the situation described by S never occurs or is impossible.[17]

Deriving this kind of law-like statement from a theory is the rationale of modeling. Tests, explanations, prediction, policy advice—no matter what we want to do, we have to find out what our theories imply for the situations where we want to apply them. In the case of the competitive exchange economy, the situation is hypothetical, although situations coming close to it can be implemented in laboratories. Considering hypothetical situations is not only relevant for model building but also a crucial element in decision making. A rational decision maker selects from the available options one whose causal consequences he believes to be at least as good as the causal consequences of the others. His assumptions about the hypothetical scenarios resulting from different choices—so-called subjunctive conditionals of the form "if I took action *a*, consequence *c* would obtain"—will be true only if they follow from true law-like assumptions.

The concept of a theoretical model as a combination of law-like and situational assumptions captures the main use of the term "model" in economics and in other sciences (see Bunge, 1973, 97–99). It has an important but often overlooked aspect: a theory comes with its own language, that is, a set of terms occurring in the law-like assumptions and denoting the things to which the theory refers (Hans Albert, 1987, 108–111). The description S of a situation contained in a model $T \wedge S$ uses only the language of the theory.

For instance, neoclassical theory speaks, among others, of agents and goods. These basic terms have no explicit definition within the theory but, of course, a meaning: the agents of economic theory, for instance, are humans.[18] These terms leave some room for interpretation since meanings are not perfectly sharp. Does every human being qualify as an economic agent, or are there, for instance, some age qualifications? In special cases, such fine points might matter. More important, however, is the fact that the language of neoclassical economics lacks many of the terms that are used for describing people's personal characteristics or the characteristics of the goods people's preferences refer to.

Characteristics of a situation that cannot be described with the language of the theory must be ignored in any relevant description of a situation. Therefore, a certain level of abstraction is a built-in feature of any theory. Given the usual law-like assumptions of neoclassical theory, different assumptions about the color of agents' eyes among the situational assumptions would make no relevant difference: according to the model of a competitive exchange economy, trade among blue-eyed agents would be as efficient as trade among brown-eyed agents. The often-heard claim (e.g., Roberts, 1987, 838) that a perfectly realistic model would be as useless as a map of scale 1:1, then, is false. The situational assumptions of a perfectly realistic neoclassical model must be stated in the language of neoclassical theory, and this

[17] Impossible situations are routinely considered in economics, as, for instance, in infinite-horizon models with immortal agents. See also Albert and Kliemt (2017) on infinite idealizations.

[18] The occurrence of undefined terms is, of course, unavoidable in any theory. A non-circular definition of a term introduces further terms, and since a chain of definitions has to stop somewhere, any theory contains undefined basic terms.

theory already implies that an enormous amount of details would have to be left out of the model.

The situational assumptions must state not only what is present in the situation under investigation but also what is absent. For instance, an economic model's assumption that there are two agents is to be interpreted as the assumption that there are exactly two agents: no one else is present. In the same spirit, the situational assumptions of an economic model implicitly exclude any feature of the situation which can be described in the language of the relevant theory but which is not explicitly mentioned.

Since economists do not distinguish explicitly between law-like and situational assumptions, economic models are often ambiguous in this respect. For instance, when testing the neoclassical theory of behavior in laboratory experiments by letting players play some game, the relevant model assumes that players' strategies are in equilibrium. Is this a situational assumption or a law-like assumption? This is not easy to say.

If it were assumed that experimental subjects always play according to equilibrium strategies, the equilibrium assumption would be law-like. Alternatively, one might consider the equilibrium assumption as situational, which, however, would rob the theory of its empirical content. Experimentalists often consider a version of the theory which assumes that experimental subjects need time to learn about the game and the other players before equilibrium play occurs. This alternative interpretation invokes (typically: not very precise) law-like assumptions about learning, which are used to choose an experimental design under which the equilibrium assumption is predicted to hold. In effect, problematic situational assumptions are replaced by law-like assumptions claiming that, under relatively easy-to check situational assumptions, the problematic situational assumptions will hold. This is one of the ways to "operationalize" a theory.

Ambiguities concerning the distinction between law-like and situational assumptions must be resolved, explicitly or implicitly, in any test of a theory. Different ways to resolve these ambiguities yield similar but nevertheless different theories.[19] Often, the hypotheses introduced at this stage are so-called auxiliaries, that is, law-like hypotheses required to derive predictions for specific contexts but not really in the center of interest. In experimental economics, for instance, an often-used auxiliary is that experimental subjects who correctly answered some test questions about the experiment have understood the instructions.

In practice, there is a large set of law-like assumptions, some of them quite similar to each other, which are subjected to tests in different combinations. Some combinations of law-like assumptions turn out to be successful in empirical tests, others fail. The process of weeding out false law-like assumptions and identifying true ones is complicated; its discussion is beyond the scope of the present paper. However, for the whole process to work at all, it is necessary to forge a connection between theories, that is, combinations of law-like assumptions, and empirical investigations. And this requires a solution to the problem posed by unrealistic assumptions.

[19] See also Albert and Kliemt (2021, 536) on rational choice theory as a family of distinct theories relying on a common mathematical language.

2.2 Basic Methodological Problem Constellations

On the basis of the distinction between law-like and situational assumptions, the interaction between model building and empirical investigations can be clarified. Given a theory T, a description S of a situation and a conclusion of interest F implied by the model $T \wedge S$, the focus is on the law-like statement $S \rightarrow F$: whenever and wherever the situation S obtains, F must hold. Depending on the status of the model's components, we can distinguish several problem constellations in empirical investigations (see Table 1).

With respect to the theory T, we have to acknowledge that tests are never completely conclusive. False theories might yield correct predictions in some situations, and observational errors might lead us to the false conclusion that predictions from a true theory failed. If theories must be tested by statistical methods, both kinds of errors may, in addition, be caused by sampling variation. Therefore, we can never be certain whether T is true or false. Moreover, a small number of tests, no matter what the results might be, will usually be insufficient to support even a tentative judgment. Therefore, we distinguish three cases: T may be well-corroborated and tentatively accepted as true, untested or insufficiently tested, or falsified and tentatively rejected as false (corroborated, untested, or falsified, for short).

Strictly speaking, the same categories apply to the situational assumptions S. We assume that these assumptions can be checked by direct observation, but since observational errors are always possible, such checks are best viewed as tests. However, this problem seems to be less severe than in the case of theories. We therefore simplify and assume that S is correctly classified as realistic (true) or unrealistic (false) in the situation under investigation.

If the situational assumptions cannot all be checked by direct observation, an empirical investigation might be considered as an indirect test of the unobservable situational assumptions or as a means to estimate some variable whose value cannot be measured in a more direct way. The robustness considerations of Sect. 3 below can be adapted to deal with this case but this is beyond the scope of the present paper.

On the basis of these considerations and restrictions, we can distinguish the six different cases of Table 1. Cases I-III are relatively unproblematic textbook cases. The main idea of the paper is to use robustness considerations to reduce cases IV-VI to their relatively unproblematic counterparts I-III.

2.3 Case I: Untested Theory *T* and Realistic Situational Assumptions S

This is the textbook case of theory testing. The derivation of $S \rightarrow F$ for some observable conclusion F allows us to test T by checking F. If F holds, T is corroborated; otherwise, it is falsified. A single corroboration or falsification is usually not enough to determine the status of a theory. Yet, repeated falsifications usually trigger a search for alternatives for at least one of the law-like statements used in the derivation of F.

If T is corroborated in several tests based on different situations S,S'… yielding different conclusions F, F' … and never falsified, T achieves the status of a corroborated theory and can provisionally be accepted as true.[20] Acceptance is always provisional because even a well-corroborated theory might be false, so that future falsification can never be ruled out.

Checking whether the conclusion F from the model $T \wedge S$ holds, then, is a means to test T, the set of law-like assumptions of the model. Learning about law-like assumptions is crucial in science because only these assumptions have implications beyond the situation under investigation. The aim of tests is to weed out false theories and trigger the search for better ones. As the history of science demonstrates, this process can lead to impressive successes even if the new theories it produces are, again, falsified. These successes are due to the fact that false theories may have important law-like consequences that are true, or approximately true in the sense that predictive errors are small for practical purposes.

2.4 Case II: Corroborated Theory T and Realistic Situational Assumptions S

Typically, the scientific community carries on with testing even well-corroborated and provisionally accepted theories. The point is, of course, not to endlessly check the same conclusions in the same kind of situations but to come up with new, hitherto untested conclusions for new situations.

Alternatively, accepted theories may be used for making predictions or for finding explanations. Predictions are formally identical to tests but are made in the hope not of finding something new but of getting the prediction right. In the case of an explanation, the phenomenon F to be explained has already been observed. The challenge is to show that F follows from the theory and the description S of the situation where F occurred. If this turns out to be the case, the model $T \wedge S$ is said to explain F.[21]

Obviously, predictive failures yield falsifications. The same may happen if the search for an explanation of F fails, that is, if it turns out that the realistic model $T \wedge S$ implies that F should not occur. This case is, however, often less straightforward because it may be quite difficult to observe, or reconstruct after the fact, the relevant situation where F occurred. In contrast, testing a theory allows the researcher to seek out easily observable situations or to implement such situations in a laboratory.

[20] In many cases, F is a statistical hypothesis and must be checked with the help of a statistical test. Although the foundations of statistics are highly contentious, statistical practice shows that statistical tests can lead to corroborations or falsifications. For a critical-rationalist account of statistical testing, see also Albert (1992, 2002, 2007).

[21] Usually, it is also required for a satisfactory explanation that S describes the causes of F according to T. In economic models, this requirement is usually fulfilled: the preferences of agents and the situation where they act are considered as the causes of agents' actions, and these actions or some of their causal consequences are described by F.

Table 1 Possible cases in an empirical investigation based on a model $T \wedge S$

		Situational assumptions $\boldsymbol{S}$	
		Realistic	Unrealistic
Theory $\boldsymbol{T}$	Untested	I	IV
	Corroborated	II	V
	Falsified	III	VI

2.5 Case III: Falsified Theory T and Realistic Situational Assumptions S

Even if the theory T is false, its law-like consequence $S \rightarrow F$ might be true and can, therefore, reasonably be tested by checking F. In this way, a theory that, in principle, has been rejected as false can serve as a heuristic for finding new and true law-like hypotheses.

This is especially relevant if T had been successful for a long time, that is, used to be well-corroborated and had been provisionally accepted as true. The most important case in the history of science is classical mechanics, which must be considered as falsified but is still used for many purposes. Using classical mechanics is, of course, made easy because its successor, general relativity theory, is extremely well-corroborated and predicts in which situations which consequences of classical mechanics should hold.

The situation in economics is less fortunate. The neoclassical theory of human behavior has been thoroughly falsified in laboratory experiments but lacks a well-corroborated successor. Yet, the theory is still used. This can be justified if the theory turns out to be a useful heuristic (see Albert 1996). Testing a new law-like assumption $S \rightarrow F$ derived from a falsified theory T is not only interesting because $S \rightarrow F$ might be true; it can also be considered as a test of the heuristic quality of T. While the logic of testing remains the same, judgments about a heuristic are more lenient: a heuristic may be considered as useful even if its rate of failures is quite high.

Of course, heuristic successes of a falsified theory do not speak against attempts to come up with better theories. In the search for better theories, a falsified theory may serve as a benchmark: one may try to find out in new empirical investigations which consequences of the falsified theory are more or less in agreement with reality. Knowing exactly where and how a theory fails may yield important information for finding a successor—or, less desirable but possibly relevant in the case of behavioral economics, a set of successors, with each new theory covering only a subset of the domain of its falsified predecessor.

2.6 Case IV-VI: Unrealistic Situational Assumptions S

The discussion of cases I-III shows that, with realistic situational assumptions, the logic of empirical investigations of a model based on a given theory T is always the same, independently of T's status. Even in case III, an empirical investigation may make sense and, if undertaken, must proceed according to the same principles as in cases I and II.

With unrealistic situational assumptions, the logic of empirical investigations seems to break down. If S is false in the situation under investigation, the law-like consequence $S \rightarrow F$ following from T is irrelevant: it predicts F for some other situation while implying nothing at all for the situation at hand. Hence, the observation of $\neg F$ would provide no argument against T, implying that the empirical investigation is not a test of T. This also implies that observing F is no argument in favor of T: a corroboration requires a severe test. For the same reasons, the empirical investigation would not contribute to the evaluation of T's heuristic potential if T was already falsified.

However, at this point, it is possible to come back to the ideas of robustness and approximation considered before, although with important modifications made possible by the distinction between law-like and situational assumptions.

3 Robustness as a Testable Conjecture

3.1 The Method of Decreasing Abstraction

Let us consider some theory T and a set of several different models $T \wedge S_1, T \wedge S_2, \ldots$ based on T, with the different situational-assumption parts S_k of the models collected in a (possibly infinite) set $\Sigma := \left\{S_1, S_2, \ldots\right\}$. According to a standard definition, a conclusion F is robust in this set of models if F follows from each model in the set. With a given theory T, we just write that F is robust in Σ.

Let us further consider an empirical investigation, and let S^* be the true description of the situation in the language of T. Thus, S^* is perfectly realistic, while T might be false. As we have seen, the status of T—untested, corroborated, or even falsified—makes no difference for the logic of empirical investigations. Moreover, even if many of the descriptions in $\sum$ were unrealistic, the problem of unrealistic situational assumptions would be absent if it were known that F is robust in Σ and that $S^* \in \Sigma$.

This unproblematic case of robustness often holds in economics with respect to the "dimensionality" of models. For instance, consider the model of the competitive exchange economy with an unspecified but finite number of goods. From a logical point of view, this model is actually an infinite set of models, each with a different number of goods. The conclusion that trade leads to an efficient allocation is robust in this set of models, that is, it holds independently of the number of goods. To apply the model to some situation, then, one need not know the number of goods in this situation in order to conclude that the theory predicts efficiency.[22]

[22] This also means that a correct description of the situation need not be unique: in a situation with, say, two goods, a description assuming an unspecified finite number of goods would also be correct and sufficient for predicting efficiency. In order to avoid tedious qualifications, however, we ignore this type of non-uniqueness and assume that all descriptions are fully specified. While this makes descriptions of given situations unique, it turns out that the notion of a given situation is not as straightforward as it seems (see below).

The question is whether robustness considerations can be extended to the case where the realistic description S^* of the situation is not available and/or $T \wedge S^*$, if available, cannot be analyzed, so that it is unknown whether F follows from $T \wedge S^*$.

The simplest extension of the robustness argument relies on induction: one argues that one's confidence that $T \wedge S^*$ implies F increases with the size of the set Σ of unrealistic descriptions where F is robust. In this crude form, the inductive argument is obviously not acceptable: it is often easy to come up with many models whose conclusion is $\neg F$. While F might be robust in Σ, $\neg F$ might be robust in Σ'. For the inductive argument to make any sense, the set Σ must be relevant to the situation under investigation.

Sugden (2000) considers an improved version of the argument where the elements of Σ form a sequence $S_0, S_1, \ldots$ of increasingly realistic but still unrealistic descriptions of the situation under investigation. While the completely realistic description S^* is not in Σ, he argues that it may be possible to conclude by induction that, if F is robust in Σ, it is also the case that F follows from $T \wedge S^*$.[23]

In economics, the idea of constructing a sequence of increasingly realistic models is known as the "method of decreasing (or diminishing) abstraction". As already noted by Hans Albert, the method can be misused by model platonists to immunize their theory against empirical criticism. After all, theoreticians could blame any failures of their models on the fact that their assumptions were not yet realistic enough, thereby postponing severe testing of their theory indefinitely. Obviously, an acceptable argument to the effect that F actually follows from the perfectly realistic model $T \wedge S^*$ would block this immunization strategy since it would imply that $\neg F$ speaks against T.[24]

However, inductive arguments are not acceptable. As already explained, the best reason we can have for tentatively accepting a conjecture, even a mathematical or logical conjecture, is that it survived severe tests, that is, a serious search for counterexamples. The challenge, then, is to come up with an improved testable version of robustness.

3.2 Robustness, Approximate Explanations, and Critical Assumptions

The improved definition of robustness[25] involves four elements: a given theory T, a situation under investigation where T is to be applied, an unrealistic description S_0 of this situation in the language of T, and some interesting consequence F of the model

[23] Sugden (2000) defends a realist position and distinguishes between law-like and situational assumptions. However, the form of his inductive argument is not entirely clear. Albert (2013) assumes that he adopts it in its crude form. Yet, some passages (Sugden 2000, 23–24) suggest the improved version. In two follow-up papers, Sugden changed his position, now regretting his earlier realist interpretation of economic modeling and emphasizing the relation of his ideas to those of Giere (see Sugden 2009, 5n, 16–19; Sugden 2011, 718).

[24] On the method of decreasing abstraction see Hans Albert (1959, 6n, 1987, 109) and Albert (2013, 9–11). See also Carrier (2004, 1) on the de-idealization of models.

[25] See also Albert (2013) and Albert and Kliemt (2017) for earlier versions using a different terminology. The robustness conjecture is a generalized version of Musgrave's (1981) "negligibility assumptions".

$T \wedge S_0$. We call F a robust consequence of T for the situation under investigation if and only if F follows from all models that are more realistic than $T \wedge S_0$, that is, all models combining T with a description of the situation that is more realistic than S_0. The set of these more realistic descriptions is denoted by Σ_0.

The realistic description S^* of the situation under investigation belongs to Σ_0 by definition. If F is actually robust in Σ_0, this implies that F follows from $T \wedge S^*$. Therefore, the conjecture that F is a robust consequence of T in the situation under investigation implies that the problematic cases IV-VI of Table 1 can be treated like the relatively unproblematic cases I-III.

This concept of robustness includes an important special case where increasingly realistic models $T \wedge S_0, T \wedge S_1$ etc. lead to increasingly precise predictions F_0, F_1 etc. (see, e.g., Betz, 2011: 657). "Increasingly precise" means that the predictions are numerical intervals, with each interval predicted by a more realistic model being a proper subset of the interval predicted by previous models. This is a special case of robustness because F_{k+1} implies F_k in this sequence: the less precise prediction is correct if the more precise prediction is correct. Hence, F_0 follows from all models more realistic than $T \wedge S_0$.

While the definition of robustness is a generalization of this special case, it is still very strong, which means that the corresponding robustness conjecture is also very strong. A definition, moreover, cannot solve the problem of unrealistic situational assumptions. However, the definition simplifies the presentation of the solution.

In presenting the solution, we focus on the case of explaining an observed phenomenon described by F using a corroborated theory T (case V). Accordingly, we supplement the definition of robustness by a definition of an approximate explanation: in the situation under investigation, the unrealistic model $T \wedge S_0$ approximately explains F if and only if T is true and F is a robust consequence of T.

By definition, an approximate explanation could in principle be extended into a perfect explanation, which is given by the perfectly realistic model $T \wedge S^*$. As before, we assume that this is not possible in practice because S^* is unavailable and/or $T \wedge S^*$ cannot be analyzed. By the definition of an approximate explanation, however, T ist true and F is robust, meaning that F follows from all models more realistic than $T \wedge S_0$, including $T \wedge S^*$. Hence, by definition, an approximate explanation of F implies the existence of a perfect explanation, even if this perfect explanation is not available. Yet, taken by itself, the unrealistic model $T \wedge S_0$ does not explain F because an explanation needs to be true while S_0 is false in the situation under investigation. Hence, the label "approximate explanation" is justified.

Robustness implies that none of the unrealistic assumptions in S_0 is critical: no improvement of the realism of the situational assumptions would lead to a model not implying F. Of course, other conclusions than F also deriving from $T \wedge S_0$ may be false. Moreover, $T \wedge S_0$ may not be the simplest model approximately explaining F. Nevertheless, it seems that an approximate explanation of F by $T \wedge S_0$ would be completely satisfactory.

Consider, in contrast, the case where F is not robust, that is, the case where a more realistic model $T \wedge S_1$ not implying F—or, more drastically, implying $\neg F$—exists. Under these circumstances, $T \wedge S_0$ could not be considered as an explanation of F, even if it were known that $T \wedge S^*$ implied F. Moving from $T \wedge S_0$ to $T \wedge S_1$

improves the realism of the situational assumptions, that is, it introduces some features of the situation under investigation that are represented by S_1 but not by S_0. If $T \wedge S^*$ actually implies F, $T \wedge S_0$ is insufficient as an explanation because the situation under investigation must contain additional features not represented by S_1—features counteracting the effect of the features newly introduced in S_1, thereby restoring F. A satisfactory approximate explanation of F would have to include these countervailing features of the situation under investigation.

A trivial example illustrates the point (see Fig. 1). A ball rolls down a sloping plane. The fact F to be explained is that the ball reaches the floor. The plane is bumpy and crossed by a curved solid ridge. The ridge has a hole at its lowest point where the ball can pass through. Without the hole, the ball would by caught by the ridge.

A first model assumes that the plane is perfectly flat (S_0); together with the theory of gravity T, the model implies F. A more realistic model includes the ridge (S_1); according to this model, the ball is caught at the lowest point of the ridge ($\neg F$). A further, even more realistic model includes the hole in the ridge (S_2). This last model again predicts F but still leaves out the bumps on the plane that have an effect on the exact movement of the ball (not to speak of air resistance, friction, etc.). However, all more realistic models imply that the ball eventually finishes its journey down to the floor.

In the example, the initial assumption that the plane is perfectly flat is false and critical. The assumption that the plane is perfectly flat except for a ridge without holes is also false and critical. The assumption that the plane is perfectly flat except for a ridge with a hole at the lowest point is false but uncritical since accounting for the bumps on the plane would not change the conclusion. The conjecture that the conclusion is robust in the set of all models more realistic than the third model is true.

Given the situation under investigation, the initial model $T \wedge S_0$ cannot satisfactorily explain how the ball could reach the floor, although it predicts this event. A satisfactory explanation needs to mention the ridge and the hole in the ridge. This is achieved by the last model, $T \wedge S_2$. Nevertheless, $T \wedge S_2$ is not a perfect explanation since S_2 is still false; hence, the standard definition of an explanation, which requires that all components of an explanation be true, does not apply. Nor is it possible to achieve a standard explanation by taking in some way or other the robustness conjecture into account. While the robustness conjecture is true in the example, it is neither a law-like nor a situational assumption; it cannot be a part of the model but says something about the model in relation to the situation under investigation. We are therefore stuck with an approximate explanation provided by $T \wedge S_2$.

There is a further sense in which the explanation is approximate. Typically, many conclusions from an unrealistic model will be false. In the example, the model $T \wedge S_2$ is an approximate explanation for the observation that the ball reaches the floor. Let us assume that there are further observations, for instance, the time it takes the ball to reach the floor. The sequence of models just considered focuses on the explanation of just one of the known facts—that the ball reached the floor—at the expense of the other known facts.[26] The last model $T \wedge S_2$ approximately explains the selected fact; however, its conclusion with respect to, for instance, the ball's travelling time might be quite off the mark since this is influenced by a lot of factors not included in the model. The situational assumptions of the model, then, are an approximation chosen for a purpose, namely, explaining a specific fact rather than all the facts.

3.3 The Critical Discussion of Models

Even if the theory T is well-corroborated and tentatively accepted as true, the question of whether an unrealistic model $T \wedge S_0$ can be accepted as an approximate explanation is difficult to answer. If the realistic description S^* is unavailable and/or unanalyzable, the same goes for many elements of the set Σ_0 of more realistic descriptions. Therefore, an argument to the effect that $T \wedge S_0$ can be accepted as an approximate explanation of F involves two problems, which often come together. Identifying members of Σ_0, that is, descriptions of the situation under investigation more realistic than S_0, is an empirical problem. Given some more realistic description $S_1 \in \Sigma_0$, it needs to be shown that $T \wedge S_1$ implies F, which is a theoretical (that is, logical or mathematical) problem.

While it is impossible to prove the conjecture that F is robust in Σ_0, the conjecture can be tested. Severe testing means searching for features in the situation under investigation which are missing in S_0 and which, when taken into account, lead to a more realistic model $T \wedge S_1$ not implying F (or, even more strongly, implying $\neg F$). Any rebuttal of the robustness conjecture can trigger a search for an even more realistic model $T \wedge S_2$ implying F and a new robustness conjecture: F might be robust in Σ_2, the set of all descriptions more realistic than S_2.

With respect to the simple example of Fig. 1, we can imagine an equally simple story of a critical discussion leading to an approximate explanation. Let us assume that it is a well-established fact F that balls rolling down the plane reach the floor. Furthermore, let us assume that it is difficult to observe the paths taken by the balls and the exact properties of the plane but that it is already known that the plane is bumpy. Nevertheless, a researcher proposes the unrealistic "perfectly flat plane" model $T \wedge S_0$ as an explanation of the fact F that balls rolling down the plane reach the floor. This model is criticized by a second researcher who found evidence of a ridge crossing the plane. Extending the model to account for the ridge leads to the more realistic model $T \wedge S_1$ implying $\neg F$. Hence, in the situation under investigation, the conclusion F is not robust in Σ_0. It is now unclear how the balls can reach the floor—they might, for instance, jump over the ridge for some reason.

However, further empirical investigations by the first researcher lead to the discovery of the hole in the ridge and, consequently, to an even more realistic model $T \wedge S_2$ which again implies F. Against this third model, the second researcher

[26] In economics, this is often called a "stylized fact": a selection from (or, logically, a non-analytic consequence of) the known facts. Since stylized facts in this sense are just facts, the label "stylized" is somewhat misleading.

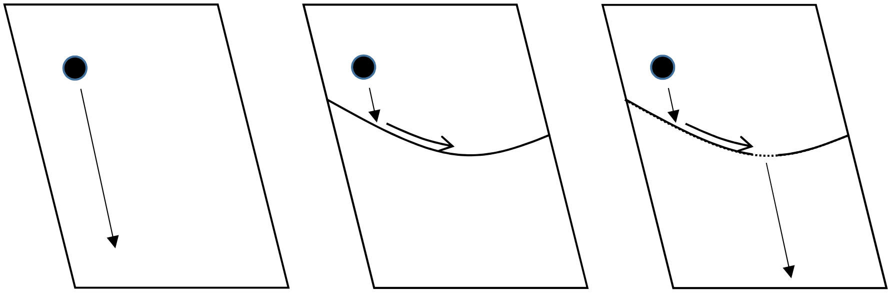

Fig. 1 A ball moves down a sloping and bumpy plane. On the way, it meets a ridge with a hole where it can pass through. The description on the left leaves out the ridge. The description in the middle leaves out the hole. The description on the right, like the others, still ignores the bumps

argues that the bumps on the plane might be so high and lie so dense that the balls cannot pass them. This is conceivable but logical possibility alone is no admissible argument. If despite their best efforts nobody can produce empirical evidence of further impediments that could stop the balls, the conjecture that F is robust in the set of all models more realistic than the third model becomes accepted. This also implies that the third model is accepted as an approximate explanation of F.

This simple story can be extended in several ways. For instance, the researchers may go on to improve the model in order to explain, in addition, the variation in the traveling times of the balls. This may require an extension of the theory used in the explanation: the theory of gravity is insufficient if there is friction and air resistance. If this research program is successful, the aim of model building might become to improve the precision of the model's explanation of traveling times. Yet, given the limited possibilities of observing the features of the plane, it may be impossible to come up with explanations for the distribution of traveling times. Moreover, there may be theoretical problems involved in determining the conclusions of the models.

Counterexamples need not be based on models incorporating all the complexities that have been discussed up to this point. The models starting from some initial model $T \wedge S_0$ may form a treelike structure rather than a line. Along each branch of the tree, the realism of the situational assumptions increases. However, there is no need to compare models from different branches with respect to their degree of realism. In fact, it is completely unnecessary to define the notion of a degree of realism; it suffices to identify increases of realism. This possibility is implicitly admitted by all sides of the debate about unrealistic assumptions: any evidence-based argument to the effect that some situational assumption is unrealistic already indicates what kinds of assumptions would increase a model's realism.

This point is exemplified by, for instance, Leamer's detailed empirical investigations of the situational assumptions of his own model of international trade. For instance, the assumption of no costs of transport is shown to be unrealistic by presenting statistics on these costs, which in the years he considers were quite high. A more realistic model would have to include these costs (but would no longer imply the linear relation tested by Leamer).

Any empirical criticism of a situational assumption, then, points the way to improvements of realism. However, this holds only for situational assumptions. A falsification of the theory-part of the model does not indicate what a more realistic theory would look like. It just throws up a fact contradicting the theory. All we learn from the falsification is that a new theory must avoid to be in conflict with this fact (as with other known facts).

Importantly, the focus on criticizing robustness allows for a broader range of arguments than a focus on proving robustness. Numerical simulations are often considered as insufficient surrogates for mathematical proofs. However, simulations within the empirically relevant range of parameter values provide counterexamples to the robustness conjecture if parameter values are found for which the conclusion of interest F fails. In the search for counterexamples, it is legitimate to focus on extreme values of the parameters—values at which background knowledge suggests that robustness might fail—as long as one stays in the empirically relevant range. Consequently, a failure to find counterexamples by simulations that are, in this way, "rigged" against F is a corroboration of the robustness conjecture.

Laboratory experiments may also provide robustness checks. To see this, let us again assume that the aim of research is to explain some observed fact F on the basis of some accepted theory T. Let there be an unrealistic initial model $T \wedge S_0$ implying F, and some more realistic model $T \wedge S_1$ which, however, is too difficult to analyze so that it is unknown whether it implies F. Yet, it may be possible to implement the situation described by S_1 as an experimental design in the laboratory. If F fails to occur in this experiment, and if T is indeed true, F does not follow from $T \wedge S_1$ and is, therefore, not a robust consequence of T. Hence, laboratory experiments with designs that cannot be analyzed theoretically can provide counterexamples to robustness conjectures. The failure to refute robustness in the laboratory contributes to the corroboration of the robustness conjecture and, consequently, helps to establish the simple model as an approximate explanation of F.

There is no logical endpoint to the search for an approximate explanation. Any conclusion is accepted only tentatively, until somebody comes up with a new argument. In this respect, the search for approximate explanations is not different from any other dispute in science (including logic and mathematics). Different contributors take different positions, and one side wins if the other sides run out of arguments. Although it is never possible to ascertain conclusively what is true and who is right, there are methodological rules regulating this competitive process. They determine which kind of arguments are admissible and which side has, at a given time, the upper hand. While the details of scientific methodologies, which depend on accepted scientific theories and available technologies, change over time, the top level rules regulating the interplay of theoretical argumentation and empirical investigation remain the same. The conjecture that these top-level rules are the best rules for promoting scientific progress can be accepted in view of the success of science and the fact that, despite intense critical discussions, no better rules have been found.

As already explained, searches for an explanation can lead to the conclusion that F cannot, in fact, be explained by T. This can, of course, also happen in the case of approximate explanations, and it seems to me that it happened, for instance, in the case of the movement of the perihelion of the planet Mercury, for which no explanation on the basis of Newton's theory of gravitation could be found. Given the complexity of the solar system and the fact that, according to Newton's theory, all masses in the solar system and, actually, in the whole universe should instantaneously affect the movement of the bodies in the solar system, the search for an explanation in this case must be considered as a search for an approximate explanation—a search that famously failed, while an (it seems: also approximate) explanation on the basis of general relativity theory could be found.

The example of Newton's theory suggests that the notion of a given situation under investigation is not as straightforward as it seems. While the language of the theory T determines what could conceivably be relevant—namely, anything that can be described in this language—, the content of T determines which elements of this description are actually relevant in which way. Without consulting T, it is impossible to say what has to be included and where the limits of the situation under investigation are to be drawn. Depending on T, things that are far away in space and time might be relevant. Moreover, the question is: relevant to what? If we focus just on one conclusion of interest F, some things turn out to be irrelevant that may be relevant to some other conclusion.

Fortunately, it is not necessary to clarify the limits of the situation under investigation in advance. As far as such a clarification is needed, it emerges as a byproduct of the robustness discussion. The discussion begins with the problem of explaining, with the help of T, some observed phenomenon F. The observation of F always comes with a rough-and-ready notion of the situation under investigation. We propose a model $T \wedge S_0$ and call F a robust consequence of T for the (not precisely defined) situation under investigation if and only if F follows from all models that are more realistic than $T \wedge S_0$. Such a more realistic model replaces S_0 by situational assumptions S_1 improving on S_0 in the light of some empirical criticism raised against S_0. If it can be shown that F is not a consequence of $T \wedge S_1$, robustness has been successfully criticized and it has been shown that the phenomena newly included in S_1 belong to the situation under investigation. If, after serious attempts at refuting it, we accept the conjecture that F is a robust consequence of some model $T \wedge S_n$, we thereby also accept that some observable features of the situation under investigation describable in the language of T but not captured in S_n are irrelevant for an approximate explanation of F. Whether these irrelevant features are, intuitively, considered as part of the situation under investigation or not makes no difference.

Robustness conjectures, then, lend a specific structure to a critical discussion of potential approximate explanations of some observed phenomenon F on the basis of a well-corroborated and tentatively accepted theory T. The same kind of interaction between empirical and theoretical considerations is relevant if T is untested or falsified (cases IV and VI in Table 1). If one wants to argue that the result of checking some conclusion F from an unrealistic model $T \wedge S_0$ is relevant for the assessment of the truth or the heuristic potential of T, one must argue that, in the situation under investigation, F is a robust consequence of T. Of course, experimental robustness checks presuppose that T is true; they make no sense if T is not accepted. But apart from this caveat, the robustness conjecture triggers always the same kind of critical discussion.

Independently of the status of the theory T, then, the logical structure of the critical discussion of models is always the same and drives model building in the direction of increasing realism or “decreasing abstraction”. This variant of the “method of decreasing abstraction” is not a pretext for indefinitely postponing empirical criticism. Rather, it is an application of the basic principles of critical rationalism: a way of integrating empirical criticism into the modelling process.

4 Robustness in Economic Research

In the case of Leamer (1984), robustness of the conclusion of interest—the linear relationship between factor endowments and trade—in the situations under investigation fails, which makes it hard to explain what the empirical investigation of this conclusion is meant to achieve. But there are other examples where robustness seems to hold. I want to discuss two of them.

The first example comes from the field of mechanism design. The typical problem in mechanism design is to find institutional arrangements that generate some desirable result or avoid some undesirable result. In the present context, mechanism design is interesting for two reasons. First, designers frequently use numerical simulations and experiments to supplement theoretical considerations. Second, design problems illustrate an important general point: the same considerations relevant in the search for explanations are also relevant in the search for solutions to practical problems.

Roth (2002) reviews the history of the US American labor market for new doctors seeking a first job at hospitals. After the market came into existence around 1900, intense competition caused it to “unravel”: in an effort to secure an attractive partner before their respective competitors became active, hospitals and medical students entered into contracts earlier and earlier in students’ careers. By the 1940s, students were hired almost two years before graduation, at a time when hospitals still lacked reliable information about the prospective doctors’ qualifications, and students had not yet found their preferred field of specialization. As a consequence, matchings between doctors and hospitals were highly inefficient.

Attempts to reform the matching process led, in the 1950s, to the creation of a centralized clearinghouse. Hospitals interviewed graduates as they saw fit and then provided preference rankings of graduates to the clearinghouse, while graduates provided their preferences rankings of hospital positions. The clearinghouse used these rankings to propose a matching using an algorithm that subsequently was improved several times. Although participation in the clearinghouse was voluntary and the matching provided by the clearinghouse was a non-binding proposal, this solved the problem of unraveling. In the 1990s, however, hospitals and medical students became dissatisfied with the operation of the clearinghouse (for reasons we need not discuss), and Roth was asked to improve the situation. This led to the adoption of a new algorithm. Subsequently, we focus on selected aspects relevant in the design of this algorithm.

Roth (2002, 1348) begins with the question of what explains the successes and failures of clearinghouses in various labor markets. The explanation he seeks is based on the neoclassical theory of human behavior and, specifically, game theory. As the paper shows, Roth accepts this theory in a slightly weakened form. It is not assumed that equilibria are reached without a learning phase; indeed, exploratory behavior may persist to some degree even after the learning phase. Moreover, it is conceded that individuals are unable to solve very complex problems; therefore, equilibrium predictions for some situation are taken to be theoretically relevant also for slightly different situations where, actually, small gains could be achieved by using different but hard-to-find strategies. These deviations from hard-core neoclassicism are not unusual in applied and, specifically, experimental economics.[27]

The starting point for the explanation is what Roth calls a "too simple model" of the labor market. The model assumes that the clearinghouse cannot be circumvented and finds a stable matching between doctors and hospitals. A matching is stable if and only if (a) all participants are assigned partners acceptable to them and (b) there exists no hospital-doctor pair where both prefer each other to their assigned partners. Stability goes beyond efficiency: it ensures that searching for a better match will not be successful.

Stability of the proposed matchings is a plausible requirement for the success of a clearinghouse. With unstable matchings, participants dissatisfied with their assigned partner may find better matches, thereby displacing others who would then have to search for a new partner. Depending on the extent of this effect, using the clearinghouse may become unattractive, leading to the decentralization of the market and unraveling.

In line with this intuition, an empirical investigation of several clearinghouses from different markets suggested that using "stable algorithms" tends to prevent unraveling while using "unstable algorithms" does not (Roth, 2002, 1351). A matching algorithm is called (un)stable if it leads to (un)stable matchings on the basis of stated preferences. Specifically, the algorithm used by the successful clearinghouse on the medical labor market was stable. One important question, then, is whether the initial success of the medical clearinghouse is, indeed, explained by the stability of its algorithm.

Due to the complications of the medical labor market, this question could not be answered on the basis of existing models. Roth's discussion of these complications is an instance of robustness checks in the sense of the present paper. There is a situation, the medical labor market in the 1950s. The theory T is the (slightly weakened version of the) neoclassical theory. The conclusion F of interest is that a stable algorithm prevents unraveling. While this intuitively appealing conclusion does not follow from the "too simple" model—in this model, the clearinghouse cannot be circumvented—, an extended model actually implies F (Roth & Xing, 1994). However, this model still ignores several complications of the medical labor market. We discuss just two of them.

One complication is presented by incentives to manipulate the outcome of the algorithm by lying about one's preferences. In the model of Roth and Xing (1994),

[27] On learning and exploratory behavior after the learning phase, see the discussion of the experimental results in Kagel and Roth (2000). On the practical irrelevance of hard-to-find strategies, see Roth (2002, 1354 n. 16).

such incentives are absent; however, the theory implies that incentives to lie must exist in the medical labor market (Roth and Sotomayor 1992, 525–527). Lying about one's preferences is a problem because a stable algorithm ensures stability of the matching only with respect to stated preferences. If market participants lie about their preferences, the resulting matching can be unstable with respect to their true preferences.[28]

Yet, the incentive problem may be irrelevant. In sufficiently large markets and with a lack of information about the preferences of other market participants, manipulating the algorithm by lying is difficult and the profits tend to be small.[29] The problem is that there exists no clear-cut theoretical result yielding a threshold beyond which a market is large enough. It is therefore unknown whether one can conclude that there is no incentive problem in the medical labor market.

Kagel and Roth (2000) tackle this problem experimentally.[30] The experiment is not a test of the basic theory: Kagel and Roth were unable to derive an equilibrium prediction for the experimental design; moreover, they do not express any doubts that the theory is correct in this domain of application.[31] Nor is the experiment a simulation of any labor market encountered in the field: the experimental design is still much too simple (see Kagel & Roth, 2000, 208). The experiment can only be interpreted as a robustness tests. And, indeed, Kagel and Roth (2000, 202, 229) repeatedly claim that the experiment checks the robustness with respect to market size of the hypothesis that clearinghouses using stable algorithms prevent unraveling. This seems to be correct although the authors do not provide a complete account of this robustness check. The missing step in the argument, which may have been obvious to the authors, is that, in the very small markets implemented in the experiment, incentive problems loom much larger than in large markets. Given this premise, the experiments provide a severe robustness check. The experimental subjects first gained experience with a decentralized market that led to unraveling. Then, a clearing house was introduced which used either a stable or an unstable algorithm. However, subjects could still make early binding contracts instead of waiting for the clearinghouse to open. Moreover, lying was possible as well as potentially profitable

[28] In practice, lying often takes the form of stating one's preferences incompletely: not ranking a potential partner is taken by the algorithm to mean that being matched with this partner is worse than remaining unmatched, which, however, is typically not true.

[29] See Roth and Peranson (1999, 749, 763) and the corrections by Lee (2017), which, however, concern not the result but only its explanation.

[30] We ignore several other problems which, according to Kagel and Roth (2000), are addressed by the experiment like, e.g., the influence of the costs of early matches or the details of the adjustment process triggered by the introduction of the clearinghouse.

[31] As explained above, this position is not inconsistent with considering the theory as falsified. A falsified theory T can still imply a restricted true theory T_S of some situation S. In the medical labor market, agents deal with an important decision, take an effort to construct preferences (by participating in job interviews before contracts are made), and it is known that they think about the strategic aspects of the situation (see, e.g., Roth 2002, 1347 on "gaming the system"). Hence, this may be a domain of application where the (slightly modified) neoclassical theory yields a true theory of how the market works.

and did occur.[32] Yet, the clearinghouse reduced unraveling significantly if it used a stable algorithm; if the algorithm was unstable, unraveling prevailed.

Another complication in the medical labor market, which became increasingly relevant in the 1970s, is the presence of couples of doctors seeking jobs in the same city. Couples pose a problem because they cannot individually state their true preferences and may, therefore, have reasons to circumvent the clearinghouse. While the algorithm of the medical clearinghouse can be, and has been, adjusted to allow for couples seeking positions at the same hospital, it can be shown that, depending on participants' preferences, stable matches may not exist. It is just a conjecture that some modified algorithm is stable. However, numerical simulations with actual stated-preference data from several years indicated that this seems not to be a problem in practice: a suitably modified algorithm always found stable matches and, therefore, seems to be stable within the range of observed stated preferences (Roth, 2002, 1359). Again, it seems that these simulations can only be interpreted as robustness checks in the sense of the present paper since no theory or model is tested.[33]

Thus, the conclusion that the modified algorithm would prevent unraveling in the medical labor market has survived at least two serious robustness checks.

It is perhaps no surprise that model platonism plays no role when economists work out solutions to practical problems. Yet, robustness considerations are also relevant for the development of pure theory, as shown by a further example, Akerlof's (1970) lemon-market model. This is one of the two paradigmatic examples of useful unrealistic models discussed by Sugden (2000).

The lemon-market model is an extremely simple model of the market for used cars. The observation the model intends to explain is the price of almost-new used cars, which often seems to be much lower than the quality of the car would warrant. Akerlof's explanation for this alleged fact is that buyers on the used-car market cannot distinguish between almost new good cars and almost new "lemons" (that, is, cars with manufacturing defects), while sellers know the quality of their cars. This information asymmetry between buyers and sellers prevents an efficient market equilibrium even if all other assumptions of a competitive market hold because the willingness to sell as such signals low quality. In the extreme case, the market could break down, meaning that only the lowest quality is traded at all: the good cars do not command the price at which sellers would sell them, although buyers would pay this price if they could be sure not to get a lemon.

Akerlof's (1970) basic model and the extensions he discusses are quite unrealistic. Nevertheless, one gets the impression that one learns a lot about markets from reading the paper. The reason, however, is not, as Sugden (2000) argued, that one concludes inductively that conclusions from several unrealistic models should also

[32] On the occurrence of lying, see Kagel and Roth (2000, 225). On the incentives for lying, see Roth and Sotomayor (1992, 517, corollary 31), which applies to those treatments in Kagel and Roth (2000) where several stable matches exist.

[33] These simulations are not simulations of a complete equilibrium model. However, they cover the crucial part of such a model. They demonstrate that including couples would not endanger stability and, therefore, would not trigger unraveling in an equilibrium model.

follow from a realistic model of the used-car market. The persuasive power of Akerlof's argumentation is based on his robustness discussion, which shows that obvious counterarguments against his conclusion fail. He extends his model by adding familiar market institutions like warranties which, at first sight, might be able to overcome the inefficiency caused by asymmetric information. Then he goes on to show that these institutions are unable to restore efficiency.

While the discussion of the model extensions is informal, it is quite clear that one might write down a model along these lines which still shows a market failure. At the end of Akerlof's discussion, one is left without objections: there seems to be no conceivable remedy for the problem. Akerlof's arguments also demonstrate that a believed-to-be-robust consequence of neoclassical economics—competitive markets are efficient if there are no externalities—is actually not robust.[34] This forces believers in market efficiency as well as non-believers to consider existing complications on competitive markets and to continue Akerlof's informal discussion with theoretical and empirical arguments.

Akerlof's line of argument illustrates the fact that robustness discussions in the sense of the present paper are an essential element in the evaluation of economic models. Critics of a model point out complications existing in the situation under investigation but missing in the model and try to show that accounting for these complications invalidates the conclusions from the initial model. Defenders of the model do not argue that Friedman taught us that the realism of assumptions is irrelevant. Instead, they try to show that accounting for the complications does not change the conclusions. Such robustness discussions drive model building in the direction of greater realism of the situational assumptions of economic models.

5 Conclusion

By acknowledging the distinction between law-like and situational assumptions, then, economists can escape from model platonism. Yet, as the examples above show, it is not necessary to refer to the distinction explicitly in order to come up with reasonable arguments. Nor do economists need to specialize in philosophy of science. Apart from a few basic ideas, all it takes to practise Hans Albert's critical rationalism is a commitment to realism and critical discussion, and some common sense.

Acknowledgements For helpful comments, I am grateful to Friedrich Breyer, Hartmut Kliemt, and three referees.

Funding Open Access funding enabled and organized by Projekt DEAL.

[34] See Stigler (1957) on the concept of perfect competition. His treatment of the crucial homogeneity condition (Stigler 1957, 260) shows that this condition was considered as expressing the idea that suppliers on the market supply the same product, a condition that is fulfilled in the lemon-market model. See also Roberts (1987, 838) on the homogeneity condition in Debreu's general equilibrium model. Akerlof's (1970) paper forced economists also to come up with a more precise concept of perfect competition.

Declarations

Conflict of interest There is no conflict of interest.

References

Akerlof, G. A. (1970). The market for „lemons". Quality uncertainty and the market mechanism. *Quarterly Journal of Economics, 84*, 488–500.

Albert, H. (1959). Der logische Charakter der theoretischen Nationalökonomie. Zur Diskussion um die exakte Wirtschaftstheorie. *Jahrbücher für Nationalökonomie und Statistik, 171*, 1–13.

Albert, H. (1963). Modell-Platonismus. Der neoklassische Stil des ökonomischen Denkens in kritischer Beleuchtung. Repr. In H. Albert (Ed.), *Marktsoziologie und Entscheidungslogik. Zur Kritik der reinen Ökonomie* (pp. 108–137). Mohr Siebeck.

Albert, H. (1973). Macht und ökonomisches Gesetz. Repr. In H. Albert (Ed.), *Aufklärung und Steuerung* (pp. 123–159). Hoffmann und Campe.

Albert, H. (1978). *Traktat über rationale Praxis*. Mohr Siebeck.

Albert, H. (1985). *Treatise on critical reason*. Princeton University Press.

Albert, H. (1987). *Kritik der reinen Erkenntnislehre. Das Erkenntnisproblem in realistischer Perspektive*. Mohr Siebeck.

Albert, H. (2010). The economic tradition and the constitution of science. *Public Choice, 144*, 401–411.

Albert, H. (2012). Model Platonism. Neoclassical economic thought in critical light. *Journal of Institutional Economics, 8*, 295–323.

Albert, M. (1992). Die Falsifikation statistischer Hypothesen. *Journal for General Philosophy of Science, 23*, 1–32.

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

Albert, M. (1998). *The Logic of Risk and Uncertainty.* Konstanz: unpublished. https://www.researchgate.net/publication/289672539_The_Logic_of_Risk_and_Uncertainty

Albert, M. (2002). Resolving Neyman's paradox. *British Journal for the Philosophy of Science, 53*, 69–76.

Albert, M. (2007). The propensity theory. A decision-theoretic restatement. *Synthese, 156*, 587–603.

Albert, M. (2010). Critical rationalism and scientific competition. *Analyse & Kritik, 32*, 247–266.

Albert, M. (2013). From unrealistic assumptions to economic explanations. Robustness analysis from a deductivist point of view. *MAGKS Joint Discussion Paper Series in Economics 52–2013*. https://www.econstor.eu/bitstream/10419/93518/1/773934235.pdf. Accessed 22 Oct 2021.

Albert, M. (2017). How Bayesian rationality fails and critical rationality works. *Homo Oeconomicus, 34*, 313–341.

Albert, M. (2019). Karl Popper und die Verfassung der Wissenschaft. In G. Franco (Ed.), *Handbuch Karl Popper* (pp. 321–336). Springer.

Albert, M., & Hildenbrand, A. (2016). Industrial organization and experimental economics. How to learn from laboratory experiments. *Homo Oeconomicus, 33*, 135–156.

Albert, M., & Kliemt, H. (2017). Infinite idealizations and approximate explanations in economics. *MAGKS Joint Discussion Paper Series in Economics 26–2017*. https://www.econstor.eu/bitstream/10419/174322/1/26_2017_albert.pdf. Accessed 22 Oct 2021.

Albert, M., & Kliemt, H. (2021). Classical game theory. In M. Knauff & W. Spohn (Eds.), *The handbook of rationality* (pp. 529–541). MIT Press.

Andersson, G. (1994). *Criticism and the history of science: Kuhn's, Lakatos's and Feyerabend's criticisms of critical rationalism*. Brill.

Arnold, D., & Maier-Rigaud, F. P. (2012). The enduring relevance of the model platonism critique for economics and public policy. *Journal of Institutional Economics, 8*, 289–294.

Backhouse, R. E., & Cherrier, B. (2017). The age of the applied economist: The transformation of economics since the 1970s. *History of Political Economy, 49* (Supplement), 1–33.

Balaguer, M. (2016). Platonism in metaphysics. In: E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Spring 2016 Edition)*. https://plato.stanford.edu/archives/spr2016/entries/platonism. Accessed 21 Mar 2021.

Bardsley, N., Cubitt, R., Loomes, G., Moffatt, P., Starmer, C., & Sugden, R. (2010). *Experimental economics: Rethinking the rules*. Princeton University Press.

Betz, G. (2011). Prediction. In I. C. Jarvie & J. Zamorra-Bonilla (Eds.), *The SAGE handbook of the philosophy of social sciences* (pp. 647–664). SAGE.

Bunge, M. (1973). *Method, model and matter*. Reidel.

Carrier, M. (2004). Knowledge gain and practical use: Models in pure and applied research. In D. Gillies (Ed.), *Laws and models in science* (pp. 1–17). King's College Publications.

Friedman, M. (1953). The methodology of positive economics. In M. Friedman (Ed.), *Essays in positive economics* (pp. 3–43). University of Chicago Press.

Gadenne, V. (2013). External validity and the new inductivism in economics. *Rationality, Markets and Morals, 4,* 1–19. https://jlupub.ub.uni-giessen.de/bitstream/handle/jlupub/464/04_Article_Gadenne.pdf?sequence=1. Accessed 16 Apr 2022.

Gibbard, A., & Varian, H. R. (1978). Economic models. *Journal of Philosophy, 75*, 664–677.

Giere, R. N. (1988). *Explaining science. A cognitive approach*. University of Chicago Press.

Giere, R. N. (2004). How models are used to represent reality. *Philosophy of Science, 71*, 742–752.

Guala, F. (2005). *The methodology of experimental economics*. Cambridge University Press.

Jarvie, I. C. (2001). *The republic of science. The emergence of Popper's social view of science 1935–1945*. Rodopi.

Kagel, J. H., & Roth, A. E. (2000). The dynamics of reorganization in matching markets: A laboratory experiment motivated by a natural experiment. *Quarterly Journal of Economics, 115*, 201–235.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. University of Chicago Press.

Lakatos, I. (1976). *Proofs and refutations*. Cambridge University Press.

Leamer, E. E. (1984). *Sources of international comparative advantage*. MIT Press.

Lee, S. (2017). Incentive compatibility of large centralized matching markets. *Review of Economic Studies, 84*, 444–463.

Morrisson, M. (2016). Models and theories. In P. Humphreys (Ed.), *The Oxford handbook of the philosophy of science* (pp. 378–396). Oxford University Press.

Munroe, R. (2014). *What if? Serious scientific answers to absurd hypothetical questions*. John Murray.

Musgrave, A. (1981). 'Unreal assumptions' in economic theory: The F-twist untwisted. *Kyklos, 34*, 377–387.

Musgrave, A. (1999). *Essays on realism and rationalism*. Rodopi.

Musgrave, A. (2011). Popper and hypothetico-deductivism. In D. M. Gabbay, S. Hartmann, & J. Woods (Eds.), *Handbook of the history of logic* (Vol. 10, pp. 205–234). North-Holland.

Ng, Y. K. (2016). Are unrealistic assumptions/simplifications acceptable? Some methodological issues in economics. *Pacific Economic Review, 21*, 180–201.

Paitlová, J. (2021). Hans Albert's systematic approach to critical rationalism. *Homo Oeconomicus*. https://doi.org/10.1007/s41412-021-00107-2

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics. A primer*. Wiley.

Pfleiderer, P. (2020). Chameleons. The misuse of theoretical models in finance and economics. *Economica, 87*, 81–107.

Popper, K. R. (1935). *Die Logik der Forschung*. Springer.

Roberts, J. (1987). Perfectly and imperfectly competitive markets. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *The New Palgrave dictionary of economics* (Vol. 3, pp. 837–841). Macmillan.

Rodrik, D. (2015). *Economics rules. The rights and wrongs of the dismal science*. Norton.

Rodrik, D. (2018). Second thoughts on economics rules. *Journal of Economic Methodology, 25*, 276–281.

Roth, A. E. (2002). The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *Econometrica, 70*, 1341–1378.

Roth, A. E., & Peranson, E. (1999). The redesign of the matching market for American physicians: Some engineering aspects of economic design. *American Economic Review, 89*, 748–780.

Roth, A. E., & Sotomayor, M. (1992). Two-sided matching. In R. J. Aumann, & S. Hart (Eds), *Handbook of game theory* (Vol. 1, pp. 485-541). Elsevier.

Roth, A. E., & Xing, X. (1994). Jumping the gun: Imperfections and institutions related to the timing of market transactions. *The American Economic Review, 84*, 992–1044.

Solow RM. (1956). A contribution to the theory of economic growth. *Quarterly Journal of Economics*, *70*, 65–94. https://doi.org/10.2307/1884513

Stigler, G. (1957). Perfect competition, historically contemplated. In G. Stigler (Ed.), *Essays in the history of economics* (pp. 234–267). University of Chicago Press.

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

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

Sugden, R. (2011). Explanations in search of observations. *Biology & Philosophy, 26*, 717–736.

Swartz, N. (2021). Laws of nature. In: J. Fieser, & B. Dowden (Eds.), *The Internet Encyclopedia of Philosophy*. https://iep.utm.edu. Accessed 6 Sept 2021.

Winther, R.G. (2021). The structure of scientific theories. In: E.N. Zalta (Ed.). *The Stanford Encyclopedia of Philosophy (Spring 2021 Edition)*. https://plato.stanford.edu/archives/spr2021/entries/structure-scientific-theories. Accessed 30 Aug 2021.

Wootton, D. (2015). *The invention of science: A new history of the scientific revolution*. Allan Lane.