Joint Discussion Paper Series in Economics

by the Universities of

Aachen · Gießen · Göttingen
Kassel · Marburg · Siegen

ISSN 1867-3678

No. 26-2017

Max Albert and Hartmut Kliemt

Infinite Idealizations and Approximate Explanations in Economics

This paper can be downloaded from
http://www.uni-marburg.de/fb02/makro/forschung/magkspapers

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

Infinite Idealizations and Approximate Explanations in Economics

Max Albert Hartmut Kliemt

June 30, 2017

Justus Liebig University
Behavioral and Institutional Economics (VWL VI)
Licher Str. 66, 35393 Giessen, Germany

max.albert@wirtschaft.uni-giessen.de

hartmut.kliemt@wirtschaft.uni-giessen.de

Abstract

If we take it that, at least in the social sciences, "realistic" implies "finite", then countless economic models involving infinitary assumptions must obviously be classified as unrealistic—for example, models with infinitely divisible goods, a continuum of traders, consumers optimizing over an infinite time horizon, or players optimizing over an infinite number of interactions. We argue that unrealistic models involving infinities can, in principle, supply explanations in economics. We develop a concept of approximate explanation based on the "method of decreasing abstraction", that is, the practice of approximating complex situations through a sequence of increasingly realistic models. Our account of approximate explanation renders the testing view of economic science compatible with scientific realism. However, compatibility does not extend to the folk theorems for infinitely repeated games, which are used widely in applied economics (e.g., in industrial economics) and beyond (e.g., spontaneous emergence of social order). Explanations based on these theorems are rejected by our criterion.

Keywords: Approximate explanation · Folk theorems of game theory · Infinite idealizations · Method of decreasing abstraction · Methodology of economics · Unrealistic assumptions

JEL Classification: B40, B41, C73

1 Introduction

Among economists, three views concerning the status of economic models prevail. Models are seen as:

1. *Rhetorical devices* (e.g., McCloskey, 1983): Independently of their truth, economic models play only a rhetorical or persuasive role.

2. *Predictive instruments* (e.g., Friedman, 1953): Independently of their truth, economic models are regarded as adequate predictive instruments as long as relevant conclusions are borne out by the facts.

3. *Approximations* (e.g., Gibbard and Varian, 1978): To the extent that they are "approximately true", economic models can play an explanatory role even if they contain unrealistic (that is, false) assumptions.

The first view unduly neglects that economists try to predict events or patterns of events by their theories. The second, instrumentalist view emphasizes prediction but ignores explanation. It has been the most influential position in debates about economic methodology. Here is a succinct summary in Friedman's own words (1953, 15):

> [T]he relevant question to ask about the "assumptions" of a theory is not whether they are descriptively "realistic", because they never are, but whether they are sufficiently good approximations for the purpose at hand. And this question can be answered only by seeing whether the theory works, which means whether it yields sufficiently accurate predictions.

Methodologically, Friedmanian instrumentalism had the advantage of fostering empirical testing in economics. The disadvantage was that it neutralized certain relevant empirical criticisms of economic theory, in particular of the rational-choice approach. Whether the popular interpretation of Friedman's position—that the realism of a model's assumptions is completely irrelevant—captures what Friedman really meant is a matter of debate but does not concern us in this paper (cf., e.g., Mäki 2009 and other essays in the same volume).

While sharing Friedman's emphasis on testing, we reject his downplaying of the role of truth claims. Therefore, we focus on improving the third, non-instrumentalist view. Despite the popularity of Friedmanian instrumentalism, this view is implicit in many contributions to mainstream economics. Gibbard and Varian (1978, 669) argue that "practicing economists ... talk of

the assumptions of a successfully applied model as approximations", the prevailing view being that "when an investigator applies a model to a situation, he hypothesizes that the assumptions of the applied model are close enough to the truth for his purposes". They claim that economists consider models with highly unrealistic assumptions as "helpful in understanding a situation" if they yield "conclusions that are robust, in the sense that they do not depend on the details of the assumptions" (Gibbard and Varian, 1978, 674). Robustness in this sense is also required by Pfleiderer (2014, 5) when he argues that a model "can be rejected simply on the basis that a critical assumption is contradicted by what we already know to be true". Concurring with Pfleiderer, Rodrik (2015, 19, 26-27, 94-98) requires that "critical assumptions", which make a difference to the conclusions, "approximate reality". Similarly, Ng (2016, 182) states that simplifying assumptions are acceptable "[i]f they simplify the analysis without changing the conclusions substantially".

Requiring robustness opens up the possibility of reclaiming an explanatory role for economic theories and models. It is, however, certainly not the case that all economic models are good approximations or yield robust conclusions. The problem is to distinguish between a model's critical and uncritical unrealistic assumptions in a systematic way. Pfleiderer (2014, 20) sees no other solution than to invoke "simple common sense (based on knowledge of the world we live in)". In this paper, we try to be more specific, highlighting the interplay between empirical and theoretical research that is involved when economists try to apply the "real-world filter" (Pfleiderer, 2014, 3-4) to economic models.

Robustness considerations require that we consider sets of models, not single models, in order to validate explanatory claims. In this, we follow Sugden (2000), who argued that robustness arguments are based on a set of models or "credible worlds". However, other than Sugden, we avoid any appeal to inductive arguments (cf. also Albert, 2013). Instead, our approach to "approximate explanation" treats robustness as a testable hypothesis. It is embedded in an interpretation of the process of modeling and research in economics that stays within the confines of hypothetico-deductivism (see, e.g., Musgrave, 2011). Yet, we try to do justice to many prevailing practices of economics (or science in general), and to preserve many of the intuitions and suggestions of Friedman (1953), Gibbard and Varian (1978), Sugden (2000), Pfleiderer (2014), Rodrik (2015) and Ng (2016).

Although we sketch a quite general account of approximate explanation on the basis of unrealistic models, we focus in more detail on a special and widely discussed kind of unrealistic assumptions, namely, infinite idealiza-

tions.[1] Specifically, we single out the, arguably, most important case in the social sciences: the "folk theorems" from the theory of infinitely repeated games. The relevance of these theorems ranges from applied economics to social philosophy (see, for the latter, Binmore, 1994, 1998). Since our focus is on economics, we consider, as our paradigm example, a model of Rotemberg and Saloner (1986) that allegedly explains so-called counter-cyclical price movements on the basis of a folk theorem.

After providing a bird's eye view of the argument including our (quite orthodox) understanding of a scientific explanation (section 2), we introduce the model of Rotemberg and Saloner (section 3). We argue that, despite its prominence, the crucial assumption of infinitely many interactions is an illegitimate unrealistic assumption and that, for this reason, the model should not be considered as an acceptable explanation of counter-cyclical price movements.

In the rest of the paper, we develop and defend the account of approximate explanation on which we rely in our discussion of infinitary assumptions. Our starting point is Musgrave's reinterpretation of Lakatosian scientific research programs as the development of a sequence of models leading up to a prediction or explanation (section 4). This provides the context of our development and discussion of approximate explanations and robustness (section 5). We end with summary conclusions (section 6).

2 A bird's eye view of the argument

A satisfactory scientific explanation of some singular fact (described by the explanandum) must, according to "the received view", satisfy at least three requirements (see, e.g., Hempel and Oppenheim, 1948).

1. The explanation must be a valid deductive argument with the explanandum as the conclusion.[2]

2. The premises of the argument (the explanans) must contain at least

[1] Prominent cases of infinite idealizations are the assumption of infinitely many particles in the explanation of phase transitions in physics, and the assumption of an infinite population in the derivation of the Hardy-Weinberg law in biology (cf., e.g., Morrison, 2015, 25-48). The philosophical literature often distinguishes between idealizations and abstractions; cf. Morrison (2015, 20-21) for a discussion. The distinction is usually not made in economics, where the blanket term "unrealistic assumptions" is used to refer to both. Since our robustness criterion cuts across the categories as they are usually defined, we need not dwell on the distinction between them.

[2] In the case of probabilistic laws, we restrict considerations to the (deductive) explanation of probability distributions.

one law-like hypothesis that cannot be eliminated without invalidating the argument.[3]

3. The explanans must be true.

In many cases, it seems reasonable to add further requirements, for instance, that the explanans should give a correct causal account of the explanandum. However, we are not concerned with strengthening the concept of explanation. Instead, we want to relax requirement 3, the truth requirement, to make room for (some) unrealistic assumptions. The resulting concept of approximate explanation might, in turn, give rise to a concept of approximate causal explanation.[4]

With respect to unrealistic assumptions, we focus on infinitary assumptions in economic theorizing. Economic models of markets assuming infinite divisibility of goods, a continuum of buyers and sellers, consumers optimizing over an infinite time horizon, or actors optimizing over an infinite number of interactions are cases in point. If we take it that, at least in the social sciences, "realistic" implies "finite", then economic theories or models relying on infinitary assumptions must, literally speaking, be false and cannot, according to the "received view", have explanatory force.

[3]Law-like assumptions are laws if and only if they are true. On the definition of law, see, e.g., Swartz (2016). Law-like assumptions imply counterfactual conditionals, which are true if they follow from true law-like assumptions. On this account, law-like assumptions are not just statements of empirical regularities (and, even less, of observed empirical regularities).

[4]See, e.g., Woodward (2014) for a survey on explanation. Woodward's (2003) own account combines causality with counterfactual dependence of the explanandum on initial conditions. As far as we are aware, our account of approximate explanation could be combined with Woodward's (2003) account; at least, in the examples we discuss, the requirement of counterfactual causal dependence is satisfied. However, we are not sure that all explanations are causal. Saatsi and Pexton (2013) argue convincingly that, at least in the explanation of laws, only counterfactual dependence might be relevant. What about non-causal explanations of singular facts? Consider the case where a singular fact can be subsumed under a non-causal law that rules out other conceivable explanations. Here is an example. "All ravens are black" is false because albinism is possible in ravens. Albinism is a rare genetic condition; the modern explanation for a raven's whiteness seems to fulfill Woodward's (2003) requirements of a causal explanation. Assume, now, that Adam sees a swan for the first time and conjectures that its whiteness is caused by albinism. Eve, however, tells him that all swans are white (which is false, but never mind). This seems to offer a different, non-causal explanation of the swan's whiteness—a very superficial and unsatisfactory explanation, to be sure, yet a competitor to the albinism explanation, or so it seems. In contrast, the law connecting the position of the sun and the length of the shadow of a tower with the height of the tower seems compatible with any causal explanation of the height of the tower. We are not aware of an account of explanation covering our example—with the possible exception of pragmatic accounts, which seem problematic for other reasons.

We argue that, nevertheless, infinitary assumptions—and, of course, other unrealistic assumptions—can be part of an adequate (approximate) explanation provided a robustness condition is satisfied. Consider a model of applied economics that is proposed as an explanation of some singular fact F observed in a specific historical situation. A theoretical model in the sense of Bunge (1973) and Albert (1987) is a conjunction $T \wedge D$, where T is a theory and D is the description of a situation. The theory T consists of at least one law-like assumption or hypothesis. The description D describes the situation to which the theory is applied; it consists of at least one singular-descriptive statement or situational assumption.[5]

The descriptive part D must, of course, use only the basic terms of the law-like hypotheses (Albert, 1987, 108).[6] Features of a historical situation that cannot be described in the language of the relevant law-like hypotheses are, according to these hypotheses, irrelevant for the phenomena that can be described in this language. Thus, price theory as well as Newton's laws force us to ignore, or abstract from, the eye color of people dealing in markets or, respectively, falling from ladders. If abstracting from eye color were a mistake, it could not be fixed by enriching only the descriptive parts of

[5]In logic and mathematics, a model is an interpretation of a formal system that turns statement forms into true statements. This concept appeals to economists, who often wrongly assume that leaving open the interpretation of basic terms (like, e.g., "consumer goods") renders a theory formal. However, the basic terms of a theory are, by definition, undefined within the theory (and the basic terms of an interpretation of a formal theory are undefined within the interpretation). Moreover, in the presence of idealizations, the standard interpretations of formal systems (like, e.g., the mathematical conditions describing a competitive equilibrium) yield false statements about the "target systems" (i.e., markets). To avoid this problem, philosophers sometimes assume that these statements are not false statements about the target system but true statements about an abstract object or a hypothetical concrete object, which is then called "model". For a criticism of models as abstracta or hypothetical concreta, see Levy (2015). In contrast, we consider economic theories and models as sets of statements (see also Gibbard and Varian, 1978, 666-667); as far as they are formalized, the meaning of the symbols is given and stays fixed. Changes of meaning are equivalent to a transition to another theory or model. In order to ascertain the meanings of technical terms, one has, of course, to be aware of the tradition and practice of economics. The distinction between law-like and situational assumptions is missing in most comments on Friedman (1953), although already Cyert and Grunberg (1963) recognized its relevance.

[6]Bunge (1973, 97-99) identifies theoretical models with "specific theories", see Albert (1994) for the logical details. His account, however, begins with descriptive parts (called "model objects") that are not yet linked to law-like hypotheses. When he discusses model objects that are intended to be "embedded" into a nomological theory, he does not come back to the point that the theory determines largely how the situation must be described. Bunge's initial discussion therefore exaggerates the extent to which theoretical models are subjective.

the models; one would also have to consider different (or additional) law-like hypotheses that connect eye color with preferences or gravitation, respectively. If, on the other hand, the theories were true, abstracting from eye color in the description D of the situation would not make D "unrealistic" or "incomplete" in any problematic sense.

Now, let T be true, and let D be unrealistic, that is, false. This unrealistic model $T \wedge D$ nevertheless approximately explains F if F follows from the model $(T \wedge D \Rightarrow F)$ and also from all models $T \wedge D'$ resulting from more realistic descriptions D' of the relevant situation. In this case, we say that the conclusion F is "robust under improvements of the singular-descriptive part of the model" (or just "robust").

An approximate explanation is approximate in the sense that, loosely speaking, the description D approximates the exact description D^* of the historical situation closely enough for the purpose of explaining F. However, we do not define degrees of approximation. If F is robust under improvements of D, D is close enough to D^*; otherwise, it is not.

If robustness prevails, it should be possible in principle to replace the approximate explanation by a perfect explanation. After all, the exact description D^* is in the set of all descriptive improvements of D. If we could prove T and D^* together imply F, we would not need the approximate explanation any longer. Yet, typically, we are unable to provide D^*, and even if we might provide it in principle, D^* would, as a rule, be too complex to prove that F follows from T and D^*.

Robustness, then, must remain a conjecture in all interesting cases. Yet, to the extent that the robustness conjecture survives severe tests, the claim that the initial unrealistic model approximately explains F should be accepted.[7]

This argument must also apply to approximate economic explanations containing unrealistic infinitary assumptions in the singular-descriptive parts of the models. Replacing an infinitary assumption by the appropriate finite counterpart should yield a more realistic description of the relevant situation. The crucial question, then, is whether, after making the substitution (along, possibly, with some additional adjustments), F still follows from the model. If this is not the case, we must reject the claim that the original model already provides an approximate explanation.

In economics, replacing infinitary assumptions by their finite counterparts leads, more often than not, to no changes, or just marginal changes,

[7]If we give up the assumption that T is true, accepting the robustness conjecture means that F is accepted as a prediction of T in the situation under consideration and can, therefore, be used for testing T. Apart from some remarks in our conclusion, however, we only discuss explanations in this paper.

of the relevant implications of the models. This is, for instance, true for many models of consumer choice, where the assumption of infinite or perfect divisibility of goods merely facilitates mathematical analysis but could be replaced, without changing the relevant conclusions, by the assumption that there are small indivisible units.[8] There are, however, prominent economic models with infinitary assumptions where the relevant conclusions do not survive such a substitution. To such an example we turn next.

3 The folk theorem in industrial economics

In the important field of industrial economics, infinitely repeated games have been used for a long time. The aforementioned, widely cited paper by Rotemberg and Saloner (1986) is a case in point. The models put forward in the paper feature profit-maximizing firms interacting in an infinite sequence of oligopoly games. For present purposes, we consider a simplified version focusing on Bertrand competition (Mailath and Samuelson, 2006, 201-203).[9]

The simplest Bertrand game assumes that all firms are identical, with no capacity constraints and with identical unit costs c fixed by technology and wages and other factor prices. If the lowest price set by any firm is p, all firms setting price p are allocated the same share of demand while all other firms sell nothing. Collective profits would be maximal if all firms set the price equal to the monopoly price $p_{\max} > c$.

Setting one's price above unit costs is a cooperative move. However, if the lowest price were above unit costs, some firm could increase its profits by slightly underbidding this price. For this reason, the single symmetric equilibrium of the one-off interaction is the competitive, zero-profit solution where all firms set their price equal to unit costs. There are also asymmetric equilibria, which, however, do not differ in any relevant aspect.

Consider now a Bertrand supergame. In a finite Bertrand supergame, firms play the oligopoly game n times in sequence. At each stage, firms maximize the present value of profits, that is, the discounted sum of profits from

[8] On harmless and problematic perfect-divisibility assumptions in economics, see Ng (2016, 187-188). "Indivisibility" means, of course, only that further division leads to the loss of characteristics of the good that are important for its valuation by the consumer. Units that are indivisible in this sense could easily be divisible from a physical point of view. Thus, cars are economically, but not physically, indivisible because you cannot drive half a car (while a car's services are, of course, economically divisible but only up to the precision of time measurement).

[9] See appendix A for basic game-theoretic concepts and more details on Bertrand competition. See Vega-Redondo (2003, 78-83,334-336) for a textbook treatment of Bertrand competition in the finitely and infinitely repeated case.

the current stage on. Backward induction shows that equilibrium behavior will be the same as in the one-off game. Cooperation, that is, not underbidding a price above unit costs, means giving up an immediate profit. This is only rational for a profit-maximizing firm if it leads to higher profits in the future. For this reason, firms set prices equal to unit costs at stage n, after which the game ends. This, in turn, means that nothing can be gained by cooperation at stage $n-1$, implying that prices at stage $n-1$ will be equal to unit costs so that cooperation at stage $n-2$ is not rational. Repeating the argument, we find that prices must be equal to unit costs at all stages.

However, results for infinite Bertrand supergames are dramatically different. Backward induction is not possible because there is no last stage. The theory of infinite supergames played by perfectly rational actors gives rise to so-called "folk theorems" (see, e.g., Vega-Redondo, 2003, 281-323 or Mailath and Samuelson, 2006, 69-104). These theorems are routinely invoked to demonstrate how mutually advantageous behavior that is out of equilibrium in a one-off game can arise from perfectly rational equilibrium strategies in the infinite supergame.

The relevant folk theorem for infinite supergames implies that many patterns of cooperation become possible. In the simplest cooperative equilibrium, all firms plan and play according to a "trigger strategy": Begin with price $p^* > c$ and stick to this price as long as all firms do the same; if, however, any firm underbids p^*, set the price to c in all eternity. If p^* is not too high, no firm can increase the present value of profits by deviating unilaterally from the trigger strategy: the immediate gain from underbidding p^* is offset by the loss of profits at later stages. If firms value future profits highly enough, even $p^* = p_{max}$ may be possible.

Overt behavior in such an equilibrium (technically: behavior on the equilibrium path) will be cooperative throughout. The logic of the folk theorems, then, makes rational cooperation on the equilibrium path of an infinite Bertrand supergame possible. Firms can realize positive profits at a price $p^* > c$ by choosing p^* on the equilibrium path and "threatening" to set their price equal to c—that is, behave competitively—off the equilibrium path. Importantly, the threat of reverting to competitive behavior once underbidding occurs is credible because no firm can gain anything by not executing the threat given that it expects all other firms to do the same. This means that, technically, the equilibrium is subgame perfect.[10]

[10]Cf. also Vega-Redondo (2003, 334-341). It is not sufficient for this result that the number of repetitions is "indefinite" in the sense of "not known exactly". If the number of repetitions is uncertain, cooperation on the equilibrium path is possible only if the possibility of infinitely many repetitions is in the support of the relevant probability distribution (Güth et al., 1991).

However, a wide range of firm behaviors can be "rationalized" as resulting from equilibrium plans. For instance, the trigger strategy explained above can be adjusted to support many different implicit agreements on prices. Prices could be lower or higher and could vary depending on time (that is, the stage of the game) and exogenous events. To yield some clear implication for overt firm behavior, one from the plethora of equilibria needs to be selected. Rotemberg and Saloner (1986) assume that firms coordinate on the efficient equilibrium, where profits achieve their maximum. Efficiency is routinely used as a criterion for equilibrium selection.

Now let there be randomly occurring booms and recessions, that is, periods where the same total quantity sells on the market for a higher (boom) or lower (recession) price, implying that the rewards from cooperation differ between periods. Assume furthermore that the monopoly price (which is high in booms and low in recessions) is out of reach in equilibrium because future profits are discounted strongly enough. In boom periods, the profit from underbidding other firms is higher than in recession periods. The maximal punishment for defectors, however, is always the same: zero profits forever, enforced by perpetual competitive behavior of all firms. Under these conditions, prices in booms must be lower than in recessions in order to reduce the potential gains from underbidding.

Referring to sequences of booms and recessions as business cycles, one could say that the model predicts counter-cyclical price movements. This implication of the model is in line with certain empirical observations but contradicts widely shared intuitions and a preceding literature that had predominantly argued that what is conventionally called a price war should break out in recessions while firms should tend to raise prices during booms.

There are two types of issues here, empirical and methodological. It is an empirical issue whether or not prices indeed move counter-cyclically. It is a methodological issue whether Rotemberg and Saloner's (1986) model provides an acceptable explanation for the alleged fact of counter-cyclical price movements. We are only concerned with the methodological issue.

From the point of view of neoclassical economics, the challenge is to explain counter-cyclical price movements within the constraints of the joint assumptions of rationality and profit maximization. Within this theoretical framework, modeling cooperative behavior is often difficult or impossible unless one invokes the folk theorems of infinite supergames.[11] For this reason, the use of these theorems is widely accepted. Specifically, Rotemberg and

[11]Once one changes the theoretical framework and considers hypotheses from behavioral economics, this difficulty often vanishes. To illustrate this point, appendix B of this paper contains a trivial behavioral variation of the one-off Bertrand model that also yields the conclusion that prices move counter-cyclically.

Saloner's (1986) model is taken seriously by economists as a potential explanation of counter-cyclical price movements even though the rationality of the equilibrium strategies crucially depends on the assumption of an infinite planning horizon.

According to the robustness requirement stated in section 2, however, it would follow that the model must be rejected as a potential explanation. The central conclusion that rational cooperation becomes possible in repeated interaction does not survive the transition to a more realistic finite-horizon model—it is not robust under an essential improvement of the model's realism. In this context, it does not help that we do not know how long a realistic time horizon might be. Any finite upper limit for the number of interactions between agents implies that, according to the theory, cooperation is impossible. For this reason, a model implying cooperative behavior on the basis of a folk theorem for infinite supergames cannot be considered as a potential explanation of cooperative behavior. This is a case of an infinite idealization that should, for methodological reasons, be deemed illegitimate in explanatory uses of rational choice theory.

Of course, the infinite time horizon is not the only infinitary assumption used by Rotemberg and Saloner (1986). Their model also assumes that goods and money are infinitely divisible so that each nonnegative real number can stand for some quantity of a good produced by a firm or for some price paid by a customer. However, other than replacing the infinite-horizon assumption, replacing infinite divisibility by the realistic assumption that goods as well as money come in economically indivisible units does not change the relevant conclusions.

The question of the robustness of a model's conclusions under improvements of the model's realism is important for large parts of applied economics. The robustness criterion is the centerpiece of our definition, or explication, of "approximate explanation" in section 5. The role of approximate explanations in economics is, however, intimately connected to the use of unrealistic assumptions in the process of economic research. We therefore turn to the latter next.

4 The method of decreasing abstraction

4.1 Models and unrealistic assumptions in economics

Economists have often been criticized for their use of "unrealistic assumptions". This criticism had already been raised by the German historical school against the adherents of the classical "political economy" of Adam Smith and,

especially, David Ricardo. Today, the most prominent position with regard to the unrealism of economics is a pop version of Friedman (1953), according to which the realism of assumptions does not matter as long as certain interesting predictions of the model are borne out by the facts.

Putting aside the issue of clarifying what Friedman meant precisely, we shall try to spell out what he, in our opinion, should have said (and meant). In doing so, we present a realist view on how to work with unrealistic assumptions in economic models.

Some of the assumptions of an economic model are clearly law-like, as, for instance, the assumption of rational behavior in neoclassical economics. We refer to the set of law-like assumptions as the relevant theory. In a model, the theory is combined with the description of a situation. This description must, of course, use the concepts, or the "language", of the theory; otherwise, the theory would not be applicable to the description. With the conceptual distinction of two classes of assumptions in hand we can state succinctly: The combination of law-like and situational assumptions is what we understand by a model.

When discussing Friedman's position, defenders and detractors alike tend to speak about assumptions without distinguishing between law-like and situational assumptions. In a given model, both kinds of assumptions may be unrealistic. However, at least for realists, it makes a difference whether the theory or the situational assumptions are false.

Let us first consider the theory. Many applied economists use the combined assumptions of rationality and equilibrium as law-like statements yielding predictions, policy recommendations and, possibly, explanations.[12] For instance, in industrial economics, where the focus is on the behavior of firms, rationality mostly means profit maximization. For the purpose of this paper, we refer to these law-like assumptions (rationality or profit maximization, respectively, and equilibrium) as "applied game theory" (AGT). As the word "applied" is intended to indicate, AGT is not just a body of mathematical assumptions and theorems, or a conceptual analysis of ideal rationality in interactive decision making, but a theory claiming to provide approximate explanations of real phenomena.[13]

[12]See also the prominent methodological defense by Becker (1976, 5) of the assumptions of "maximizing behavior, market equilibrium, and stable preferences", which implies that these assumptions should be treated as law-like (although Becker does not use this term).

[13]Reinhard Selten tended to think of pure game theory merely as inspiration for experiments yet not as approximately true; this is the point of his "methodological dualism" (see, e.g., Selten, 1999). Cf. Kliemt (2009) on related "dualisms" in economic philosophy more generally.

The law-like assumptions of AGT have been criticized for a long time and for several reasons, empirical and methodological. Specifically, the rational-choice approach as a general theory of human behavior seems to be false, as has been shown in many carefully controlled laboratory experiments. Nevertheless, researchers in applied economics often decide to stick to AGT. Up to a point, this may be defensible, even from a realist point of view. After all, even if AGT is false, it can have true law-like consequences for many relevant situations.[14] Moreover, the assumption that the quite large organizations typically considered in industrial economics maximize profits is not falsified if individuals in laboratories are unable or unwilling to maximize profits (Albert and Hildenbrand, 2016). Therefore, even realists who accept that the rationality assumption has been falsified might be able to make a reasonable case that AGT as used in industrial economics could still be true.

"Could be true" is, of course, not the same as "well-confirmed". Nevertheless, let us, for the sake of the argument and possibly contrary to fact, assume that AGT, at least when applied to large firms, were a well-confirmed theory which we had tentatively accepted as true. Whether or not this would render explanations based on AGT acceptable then depends on the situational assumptions of AGT models, where different AGT models share the same law-like assumptions and differ exclusively by their situational assumptions.

The situational assumptions employed in AGT are typically false. For instance, the model of Bertrand competition used in Rotemberg and Saloner (1986) derives its conclusions from the law-like assumption of profit maximization under the situational assumptions that firms have no relevant capacity constraints and produce qualitatively identical outputs using the same technology. However, there are always capacity constraints, and the outputs of different firms almost always differ in ways that are relevant to their prospective customers. One can, of course, switch to a more complicated oligopoly model, but all these models employ simplifying unrealistic assumptions. For such and related reasons, many economists believe that the use of simplifying assumptions is unavoidable in economics (cf., e.g., Sugden 2000 or Pfleiderer 2014, 1, 31). If, however, unrealistic assumptions cannot be avoided anyway, how can we single out certain unrealistic assumptions—for instance, certain infinitary idealizations, like the assumption of infinitely many interactions—as more problematic than other unrealistic assumptions?

This would, indeed, be difficult if we just considered a given model. In order to distinguish between more and less problematic situational assumptions, we have to consider sequences of models and the dynamics of the research process within the confines of a given theory like AGT. Since AGT

[14]See Albert (1994). This point is also made by Cyert and Grunberg (1963, 306-307).

is, at least among many industrial economists, an accepted theory, one might say that we consider Kuhnian normal science or puzzle-solving (Kuhn, 1962), although we place these activities firmly within a hypothetico-deductivist framework. In fields where theories are formalized and the derivation of their implications poses in the first place mathematical challenges, searching for an explanation of certain facts of interest always requires that a general theoretical core must, step by step, be worked out into ever more complicated specific models. Our focus will mostly be on this modeling process, not on the question of whether the theory in question is considered as established or not. Complying with the usual practice of philosophy of science, we start with a little bowdlerized physics before we turn to modeling in economics.

4.2 Newton's research program

Lakatos' (1970, 135-136) account of Newton's research program as launched by his originator acknowledges the fundamental role of mathematical considerations and modeling:

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

As is well known, Lakatos characterized Newton's pursuit in terms of "hard core", "protective belt", "negative and positive heuristic" etc. In the conventional reading, this characterization is taken to be incompatible with a falsificationist, testing view of science. Starting from the observations made by Lakatos, an alternative rather straightforward reconstruction of Newton's program in hypothetico-deductivist terms seems plausible, however. As Musgrave (1978, 189-190) explains:

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

Newton intends to derive what his general abstract theory implies for a specific situation of celestial mechanics if due respect is paid to what is known about the solar system. In Musgrave's account, Newton's laws, in particular that of mass attraction, form the basic general theory that fulfills the same role as Lakatos' hard core. Lakatos' protective belt, which changes from one model to the next, corresponds to the singular-descriptive part of the models (and some auxiliary assumptions like the absence of interplanetary forces, which are a bit more complicated but need not concern us here).

A reasonable modeling strategy begins with models that are extremely simple and, therefore, descriptively unrealistic. There need to be no known real-world situations even roughly fitting the initial description. Specifically, what is known of the solar system cannot contradict Newton's first model since that model is not about the real world yet. It is merely a start from which Newton thought he would eventually be able to develop a model close to the known facts.

Even Newton might have abandoned his general theory if he would not have found a simple first model generating closed orbits of one planet around the sun. Some empirical facts will presumably play a role from the very beginning of any such research program. Conformity with some important facts assures the researcher of being "on track" towards an (approximate) explanation. Though this is still a far cry from empirical testing of specific hypotheses and from making specific predictions, Newton could reasonably hope to make progress by this method, which Musgrave characterizes as a "method of successive approximations", while economists sometimes speak of the "method of decreasing abstraction".[15] The idea of the method is to

[15]The term "method of decreasing abstraction" might go back to the first pages of Wieser (1914); cf. also Albert (1987, 109). Essentially the same procedure is recommended by Varian (1997) to aspiring economists, with a twist. He recommends to move from one's initial sketch of a model to an even simpler model, the idea being that the simplest model generating some interesting result should be the starting point. Then one should

render the models more and more realistic in their situational assumptions, in the hope of generating more accurate predictions of known facts and, in later stages, also of new facts.

The method of successive approximation seems not only close to commonsense conceptions of scientific progress in celestial mechanics. It seems also representative of how many mathematical economists have interpreted the role of their modeling efforts.

4.3 Koopmans and economic modeling

Whatever economists may say about the central role of Adam Smith in the constitution of their discipline, it took more than 150 years before Paul A. Samuelson and the group of economists associated with the "Cowles Commission" moved modern economics to a level of mathematical sophistication comparable with that of Newtonian mechanics. Tjalling C. Koopmans, a pivotal figure in this development, describes for mid 20th century economics the same very limited role of empirical observation and the same initial dominance of mathematical problem-solving as pointed out by Musgrave for the development of Newtonian celestial mechanics (1957, 154):

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

Though this account of modeling covers purely "conceptual exploration", it acknowledges implicitly that modeling must eventually justify itself by its contribution to "empirical theorizing" (cf. Hausman, 1992, 221). To the extent that the logico-mathematical exercise is taken as explanatory by itself, it degenerates into what has been aptly called "model Platonism".[16] Yet, to

"generalize", say, from a two-goods-two-agents model to a two-goods-n-agents model. This kind of generalization is a popular version of the method of decreasing abstraction: if you have no idea how many agents interact in a real-world situation, consider a model (actually: a set of models) with any number of agents.

[16]Model Platonism is the perpetuation of armchair economics (see Albert 1963, translated as Albert et al. 2012). It uses the method of decreasing abstraction as an immunizing strategy: empirical and theoretical criticisms of the basic theory are rejected; specifically,

the extent that it is serving the purpose of exploring the empirical content of the theory, it may be seen as one of a series of approximating steps ultimately aiming at explanations of real-world phenomena and testable implications of the basic theory.

Although economics has become mathematically fairly sophisticated, it seems that even today it remains farther removed from its explanatory aims than physics even at the time of Newton. Paradigm cases of what economists regard as "good" economic models correspond, at best, to Newton's first models: they provide a mere sketch of a basic idea and research program. Further development of the research program requires reliance on the division of labor in economic research to a much higher degree than was necessary in the early stages of Newtonian physics. To launch a successful program, a suitable empirically-minded audience must be convinced to join in. The hopeful researcher indicates how the descriptive part of the model might be developed and what kind of phenomena might conceivably be explained in a sequence of increasingly realistic models. In the initial phases, no serious predictions or explanations are involved; instead, the author wooing his audience to take part in developing the program typically engages in merely "casual empiricism" and emphasizes the formal merits of the theoretical tools he proposes.[17]

4.4 Models and explanations

From all this, we may conclude that, once theories have reached a certain level of mathematical complexity, even a realist will be forced to develop explanations in a sequence of models employing, at least at the early stages, simplifying situational assumptions that are highly unrealistic. From the perspective of realism, however, and in marked contrast to Friedman's instrumentalism, the aim should be that of increasingly replacing unrealistic by true assumptions. What should worry a realist is not the use of unrealistic assumptions as such but the quite plausible idea that, at least for many explanations, he might not get rid of them. As it stands, the received view of scientific explanation would not allow him to retain even a single one.

any predictive failure is blamed on the unrealism of the descriptive parts of the relevant model. In mainstream economics, model Platonism is still a relevant methodological attitude (cf. also Arnold and Maier-Rigaud, 2012), although the emergence of institutional and behavioral economics demonstrates that progress results if this attitude is abandoned and criticism is taken seriously. Note that the model Platonism critique is not directed against the method of decreasing abstraction as such.

[17] A good example is provided by Akerlof (1970); cf. Albert's (2013) critical response to Sugden's (2000) methodological interpretation.

This, however, speaks against the received view. To illustrate, consider a Newtonian model of a body falling in a vacuum (and let us assume for the moment that Newton's theory of gravitation were true). Newton's laws generate a prediction about the time the body would need to reach the surface of the earth. It would obviously be absurd to use this model to predict the way a leaf falls in the earth's atmosphere under otherwise identical conditions. According to the received view, however, we must generally reject the claim that the model explains observations made in situations other than vacuum. This is implausible. If we consider the case of a cannonball dropped within earth's atmosphere and closely to earth's surface, most people would agree that the model of free fall delivers at least an approximate explanation of the time it takes the cannonball to hit the ground.

Obviously, we need to reconcile the intuition that unrealistic models can be explanatory with the truth requirement as an integral part of the traditional account of explanation. The idea that such a reconciliation is possible at least in some cases is widely accepted (see, e.g. Bunge, 1973, 91-113). It is also intuitively appealing that the model of free fall in a vacuum is "sufficiently similar" to the case of a cannonball dropped from little height so that it explains, at least approximately, the cannonball's fall. However, the details of this intuitively appealing concept of approximate explanation need to be worked out.

5 Approximate explanations

5.1 Explanations and the robustness condition

Let us assume that some basic theory, like applied game theory (AGT), has survived initial testing in ways that induce a group of researchers to proceed on the assumption that the theory is true. The problem is to find a good model for explaining some phenomenon in some situation of interest, say, counter-cyclical price movements. Let us assume that such a model will always contain some unrealistic situational assumptions. Under which conditions would we nevertheless feel warranted to claim that the model yields an approximate explanation?

Our example of the cannonball's fall suggests an answer. In this example, we have good reasons to assume that, even if we do not know the exact details of atmospheric pressure, the amount of air pollution, etc., extending the model to take the atmosphere with all its details into account would not change the prediction, at least within the limits of precision of our measurement. In other words: We have reasons to believe that we have already taken

into account everything that is relevant for our prediction; further improvements of the model's descriptive realism would not lead to a relevant change in the prediction.

Consider, on the other hand, a case where some improvement of the realism of the situational assumptions changes the prediction substantially. Even if further improvements of the model's realism restored the prediction, we would say that the first model did obviously leave out two important factors: a factor that would, considered on its own, destroy the prediction, and some countervailing factor that restored it.

This suggests a general definition of approximate explanation. An approximate explanation requires (a) a true basic theory, (b) situational assumptions that yield correct predictions, and (c) robustness of the prediction under further improvements of the model's realism.

With respect to condition (a), we assume, as before, that the basic theory is well-corroborated and provisionally accepted as true. We focus on conditions (b) and (c). Consider a theory T, a specific historical situation S, and an observation P in situation S that is to be explained (an explanandum). Consider the set $\mathcal{D}$ of all descriptions of S using the language of T. Such a description consists of a finite set of singular statements using the predicates of T and referring to S (for instance, a certain time in Germany). Apart from the reference to S, the descriptions in $\mathcal{D}$ may be wide off the mark (by, for instance, claiming that German GDP 2016 was produced by a single firm).

We introduce a relation $\succeq_S$ on $\mathcal{D}$. We write $D_2 \succeq_S D_1$ for $D_1, D_2 \in \mathcal{D}$ if and only if, as a description of S, D_2 is at least as realistic as D_1. The relation $\succeq_S$ is a partial order on $\mathcal{D}$. It satisfies reflexivity ($D \succeq_S D$), antisymmetry ($D_2 \succeq_S D_1 \wedge D_1 \succeq_S D_2$ implies $D_2 = D_1$), and transitivity ($D_3 \succeq_S D_2 \wedge D_2 \succeq_S D_1$ implies $D_3 \succeq_S D_1$). We write $D_2 \succ_S D_1$ if and only if $D_2 \succeq_S D_1$ and $D_2 \neq D_1$, that is, if D_2 is a descriptive improvements on D_1.

Obviously, $\succeq_S$ is incomplete: if $D_1 \succ_S D$ and $D_2 \succ_S D$, where $D_1 \neq D_2$, we can typically not compare the realism of D_1 and D_2. Moreover, there is a completely realistic description D^*, which is formally a unique maximal element since it is an improvement on any other description: for all $D \in \mathcal{D}$, we have $D \neq D^*$ implying $D^* \succ_S D$.

Let $D \in \mathcal{D}$ be unrealistic as a description of S, that is, false, so that $T \wedge D$ is an unrealistic model. Then P is a robust prediction of $T \wedge Ca$ if and only if

i) $T \wedge D \Rightarrow P$ and

ii) $T \wedge D' \Rightarrow P$ for all D' satisfying $D' \succ_S D$.

Note that robustness implies $T \wedge D^* \Rightarrow P$, that is, the explanandum P follows from the perfectly realistic model.

With the preceding definitions in hand, an approximate explanation can be characterized quite succinctly. Let T be a true theory, P an explanandum, and $T \wedge D$ an unrealistic model of situation S. Then the model approximately explains P if and only if P is a robust prediction of $T \wedge D$.

This definition of an approximate explanation retains the truth requirement for explanations and, in effect, requires that an approximate explanation can at least in principle always be extended into an explanation satisfying the requirements of the received view. Moreover, it satisfies the intuitively reasonable requirement that the existence of a descriptive improvement leading to a model that no longer implies P indicates that an important explanatory factor must be missing from the model $T \wedge D$.

5.2 Various difficulties

As we have seen, sequential modeling typically progresses from simple and quite unrealistic models to more realistic and more complicated ones. However, $D_2 \succ_S D_1$ does not necessarily mean that the description D_1 is simpler than D_2. Simplicity is a different and notoriously difficult notion (see Baker 2016 for a survey), which is not captured by the relation $\succeq_S$. Nevertheless, if D_2 is indeed both, more realistic and more complicated than D_1, and if $T \wedge D_1$ and $T \wedge D_2$ both qualify as approximate explanations of P, we are, on our account, free to prefer $T \wedge D_1$ as the simpler approximate explanation for purely pragmatic reasons. In our context, then, there is no need to justify a preference for simplicity.

These considerations may motivate a search for the simplest approximate explanation. There are two caveats. First of all, there may be no unique simplest explanation. In economics at least, the same phenomenon (say, failure of market clearing) can have several causes (say, sticky prices and asymmetric information) which may all be present and contribute to the overall effect described by P. Thus, there can be overdetermination in the sense that there are two different approximate explanations $T \wedge D_1$ and $T \wedge D_2$ of P satisfying neither $D_1 \succ_S D_2$ nor $D_2 \succ_S D_1$. In such a case, explanations $T \wedge D_1$ and $T \wedge D_2$ may be equally simple, but even if they were not, we would not like to choose between them.

Our second caveat is due to the fact that we only discuss necessary conditions for an explanation. It may be the case that, according to our definition, $T \wedge D_1$ and $T \wedge D_2$ are both approximate explanations of P, with D_1 being simpler and less realistic. However, depending on one's full account of explanation, $T \wedge D_1$ might be too simple while $T \wedge D_2$ may be satisfactory.

If these caveats are respected, nothing speaks against searching for approximate explanations that are as simple as possible. This takes into account the—in a wide sense, practically rational—concern with the simplicity of theories and explanations often emphasized in methodological discussions, in economics and elsewhere.

The requirement that the conclusion P must be true might be considered too strong. After all, we might intuitively expect that the predictions we derive from models are as idealized as the models themselves. For instance, the model of Rotemberg and Saloner (1986) assumes that all firms are identical and predicts that they set the same price. This prediction is probably false in most applications. It seems, then, that we cannot require correct predictions but must focus on approximately correct, or qualitatively good, predictions.

However, this is not the case. The explanandum is not dictated by the model. In the case of Rotemberg and Saloner, the explanandum is the alleged fact of counter-cyclical price movements. A statement to this effect follows from their model. Of course, the model has a host of other implications, like the prediction that all firms set the same price. However, an approximate explanation is proposed for a specific explanandum. Robustness is, and needs to be, claimed only concerning that explanandum. A model that approximately explains a select fact may be quite unsatisfactory as an explanation of other facts, yet this does not undermine the explanation of the fact the model is intended to explain. Models, to adapt Friedman's (1953) words, need only be "sufficiently good approximations for the purpose at hand".

This is not specific to economic explanations. Consider an approximate explanation of the fact that helicopters can rise from the ground. For this explanation, it suffices to consider, in addition to the helicopter's environment, the mass of the helicopter's body, the size of the main rotor and the form and inclination of its blades, and the angular momentum of the rotor. A simple model along these lines, however, would imply that the body of the helicopter turns in the opposite direction of the rotor. If the explanandum is not only lift-off but also absence of rotation of the body, the model must additionally take into account the existence of tail rotors (or, depending on the type of helicopter, other anti-torque devices). A more specific explanandum as a rule requires an (approximate) explanation in terms of a more realistic model.[18]

Many economists are prepared to give up the truth requirement for explanations, probably because they assume that truth is not to be had in any case. This may be much too pessimistic. While Newton's theory is false according to modern physics, many of its interesting and practically relevant

[18]We owe the helicopter example to Sam Fletcher.

implications are still accepted as true—just consider situations where masses and velocities are not too high and replace point predictions like $x = a$ by interval predictions like $x \in [a-\epsilon, a+\epsilon]$ for suitable values of ϵ. Generally, many interesting consequences of our theories, including law-like consequences, are likely to be true. On the whole, it seems reasonable to assume that the truth requirement in many explanations might actually be fulfilled. Possibly, pessimists assume that "truth" means "certain truth" or "the complete truth" (or both); however, neither is implied by the truth requirement for explanations.

Of course, the definition (or explication) of approximate explanation suggested here presupposes that one can find sequences of models of the same situation whose singular-descriptive parts can be ranked according to their realism. This may seem to invite problems similar to comparisons of the truthlikeness or verisimilitude of different theories, which turned out to depend on the presentation or language one uses (see Oddie, 2008).

However, we think that such problems, if relevant at all, are not relevant in practice. The theory under consideration restricts the set of potential features of the relevant situation that can be taken into account. In the case of Newtonian models of the solar system, only bodies with their masses, positions and instant velocities are relevant; colors of bodies, for instance, play no role because color concepts are not part of the theory. It seems that a model taking one more actually existing body in the solar system into account is more realistic than a model where this body is missing but which is otherwise the same. Similarly in economics: if one tries to model the competition on the European car market, a duopoly model is less realistic than an oligopoly model involving three firms.[19]

Typically, researchers can agree quite easily which features of a situation a perfectly realistic model should take into account. This includes also hypothetical features: if there were a further planet in the solar system, it would have to be taken into account.

[19] A typical modeling strategy in economics is to begin with twoness assumptions (two agents, two moves, two goods, two countries, two periods), proceed to threeness, and then try to analyze the general case of $n > 1$. The n-ness model is actually an infinite set of models covering any situation to which the model may be applied. The strategy characterizes much of the development of the neoclassical theory of international trade (cf. Albert, 1994). Note that a threeness model cannot be derived from a twoness model just by adding to the conjunction of statements making up the model; a twoness model assumes that there are exactly two things, not at least two things. This holds more generally. Realism can rarely be increased by adding statements because models typically assume that the description of the situation is complete in the sense that all possible influences work through changes of exogenous variables that are already part of the model.

5.3 Testing and accepting the robustness conjecture

In all interesting case, robustness of predictions must remain a conjecture. Consider, for instance, the solar system. Obviously, a description that takes all potentially relevant features into account is impossible: too much rubbish flies around, and even the large bodies are so irregular that they defy precise descriptions. We might try to circumvent this problem by some kind of perturbation theory that allows us to derive properties that hold for large sets of models that differ only in the fine details we are unable to observe. However, such a program might fail, too.

In many fields of economics (like, say, trade theory), the complexity of the situation defies observation (not to speak of description) of all theoretically relevant details. Here, as in most real-world economic situations, the perfectly realistic model may remain unknown almost in principle. Moreover, even if we were able to observe and state a complete description of all the details, we might be unable to analyze the resulting model.[20] Like physicists, economists would like to take resort to simulations of large models but might not be in command of sufficient computing power.

When analyzing the perfectly realistic model is impossible, robustness cannot be proved. Nevertheless, selected descriptive improvements of the unrealistic models under discussion might be feasible. This implies that the robustness conjecture becomes testable.

In a typical research process, some researchers will endorse the view that the best current model already yields an approximate explanation, while others try to criticize this conjecture. All that is needed for such criticism is a demonstration that some descriptive improvement (that is, the introduction of a hitherto neglected feature of the relevant situation) destroys the prediction of the best current model. In order to refute the robustness conjecture, even a one-off simulation might suffice.

A severe test of a robustness conjecture amounts to an extended search for features of the relevant situation that might destroy the result. In many cases, scientists have some ideas about countervailing factors that might prevent the occurrence of some event or the realization of some phenomenon. As in other kinds of testing problems, researchers are guided by background knowledge in their critical discussion of explanatory claims. If the search for refutations fails, it is reasonable to provisionally accept the robustness conjecture.

[20] In economics, this point is often made by comparing models to maps: maps with a 1 : 1 scale containing all the details of a landscape would not be useful (see, e.g., Rodrik, 2015, 43-44). The metaphor, however, applies only to the singular-descriptive part of the model. As a general illustration of the concept of a model, the idea of models as maps is quiet misleading.

In our experience, robustness considerations are often the core issue when economists discuss conclusions from unrealistic economic models. Those who defend the conclusions discuss features of real-world situations not contained in the model, arguing that taking them into account would not change the conclusions. Critics often point to realistic complications, arguing that taking them into account would change the conclusions. This amounts to a discussion of the robustness conjecture.

The robustness conjecture is a combination of two conjectures, one mathematical and one empirical.[21] It would be only a mathematical conjecture if the realistic model were known but just too complicated to be analyzed. But even this purely mathematical conjecture is based on an empirical input. After all, the robustness conjecture does not say that it is impossible to destroy the relevant prediction by changing the descriptive part of the model. The robustness conjecture refers to descriptive improvements, meaning that changes of the descriptive part are only relevant if they yield a more realistic description of the relevant real-world situation. The latter can be known empirically only. Moreover, due to observation problems, not all potential descriptive improvements of a model may be known at a given time. In such a case, there is again an empirical conjecture involved: the conjecture that there exists no so far undiscovered element of the situation which would, when taken into account, destroy the relevant prediction.

The fact that we might not be able to observe all relevant features of the situation leads to further complications. In economics at least, it is often the case that different assumptions about (at the time, or in principle) unobservable features make a difference to the relevant predictions. A reasonable strategy, then, is to take into account additional predictions of the model—not just the explanandum that originally motivated the relevant strand of research, but further predictions that depend on unobservable but critical features of the situation (cf. also Rodrik 2015, 83-112 on model selection and calibration). The more a research program matures and approaches approximate explanations, the closer must empirical and theoretical considerations interact in order to make further progress.

In sum, to attack a robustness conjecture, opponents can rely on empirical and theoretical investigations. If, despite their serious efforts, they are unable to raise a decisive criticism against the robustness assumption, the assumption may become accepted—tentatively, as any other hypothesis. This means that the model is provisionally accepted as an approximate explanation of the observation in question.

[21] Mathematical results might in general be viewed as conjectures, even if there exists an accepted proof. The classical exposition of this view is Lakatos (1976).

6 Concluding remarks

In this paper, we have given a definition, or explication, of "approximate explanation" in terms of a robustness criterion: an approximate explanation requires a model based on a true theory that robustly predicts the explanandum. Robust predictions are those that are not destroyed by improvements of the realism of the model's singular-descriptive, or situational, assumptions.

The robustness conjecture can be tested, by checking strategically those descriptive improvements that, according to our background knowledge, have the potential to destroy the relevant prediction. As the preceding discussion shows, there seem to be three requirements for a successful robustness analysis.

1. The analysis should look for models that are descriptive improvements, that is, it should take additional relevant empirical facts into account.

2. The improvements should focus on features of the relevant situation that, given our background knowledge, have the potential to destroy the prediction under consideration.

3. The prediction should, nevertheless, survive the improvements.

Of course, step 3 might be difficult to judge since it requires analyzing more realistic models. If an analytic proof cannot be found, systematic simulations with empirically reasonable parameter values may be used. If these simulations are unable to identify situations where the prediction does not hold, we may conclude, tentatively, that the prediction is robust. That is, simulations are used as tests, not as proofs, of the robustness conjecture.

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

The robustness conjecture adds a further source of identifiable error to scientific explanations. In any purported explanation, the basic theory might be false even if it has survived severe testing; the situational assumptions we believe to be true might nevertheless be false because we made a mistake in our observations; and we might, of course, have made a logical mistake in deriving the explanandum from the model. In an approximate explanation, we must add the possibility that, in fact, the explanandum is not a robust prediction of the model.

We have argued that, by and large, the research process in economics and the accompanying methodological discussions often turn on robustness

in this sense. Yet they typically fail to identify the robustness criterion as such. We have also shown that an application of an explicitly formulated robustness criterion leads to the rejection of prominent explanatory claims in economics that are widely accepted. That models like that of Rotemberg and Saloner (1986) cannot provide approximate explanations of counter-cyclical price movements is a case in point of a general insight extending beyond the important but relatively narrow field of industrial economics. It applies to rational choice theories of social and political order almost across the board. Theorists who believe that infinite-horizon folk theorems offer no less than a solution for the central so-called Hobbesian problem of providing an explanation for the emergence of cooperative order in competitive interactions must think twice. Contrary to what many rational choice theorists claim, some of the most popular results of pure game theory, the folk theorems for infinite supergames, have no legitimate claim to explanatory status, not even in the sense of an approximate explanation.

These considerations also shed some light on theory testing. As Kuhn (1962) has emphasized in his description of normal science, much of applied science is puzzle-solving: the theory is taken as given; the puzzle is how to explain certain phenomena on the basis of the theory. Following Musgrave (1978), we have placed normal science within a hypothetico-deductivist framework. The aim of normal science is often to provide approximate explanations, and robustness checks are needed to establish explanatory claims. These robustness checks are important drivers of theoretical research. They lead researchers to incorporate further aspects of real-world situations into their models, either in an effort to attack or defend explanatory claims.

In contrast to Kuhn, we would argue that this process can lead to the falsification of a theory. If, despite our best efforts, we cannot solve the puzzle, that is, if we cannot even approximately explain a phenomenon on the basis of a relevant theory, then we should, at least provisionally, accept that the facts are inconsistent with the theory.

Of course, such a falsification can, in principle, be refuted by supplying the missing explanation; any falsification is provisional. Moreover, we cannot specify the exact amount of effort that must have been invested in the search of an approximate explanation before we can speak of a falsification. Yet, the point of methodological rules is not to state hard and fast rules for collectively deciding which theories should be abandoned and which should be retained. Rather, a methodology is part of the incentive system of institutionalized science (Albert, 2010, 2011). In order to fulfill this role, it is sufficient if a methodology points out when a theory is in difficulties and what would be required to resolve the difficulties, thereby shifting the burden of proof among adherents and critics of a theory.

On this account, applied game theory, as it is used in industrial economics, is in severe difficulties because of its failure to provide an approximate explanation for certain cooperative behaviors among firms. Of course, this is especially damaging to the theory if alternative and, possibly, robust explanations on a different theoretical basis (in this case: behavioral economics) seem to be within reach.

As stated before we believe that our argument extends to the fundamental Hobbesian problem of explaining the emergence of a cooperative social order. The models proposed as explanations on the basis of a narrow rational-choice theory—ruling out genuine rule following behavior as well as social, or other-regarding, preferences—all rely on the same unrealistic assumption of infinitely many interactions; without this assumption, the models do not predict cooperation.

One might argue that, actually, people cooperate because they ignore, or do not believe, that only finitely many interactions are possible: they perceive a finite supergame as an infinite supergame. In order to defend the assumption of an infinite horizon, Mailath and Samuelson (2006, 105) write, in the spirit of Friedmanian instrumentalism:

> The key consideration in evaluating a model is not whether it is a literal description of the strategic interaction of interest, nor whether it captures the behavior of perfectly rational players in the actual strategic interaction. Rather, the question is whether the model captures the behavior we observe in the situation of interest.

In this connection, they cite the prominent textbook of Osborne and Rubinstein (1994, 135). Interestingly enough, the passage cited by Mailath and Samuelson is followed by a discussion between Osborne and Rubinstein where Osborne sharply criticizes the position taken by Mailath and Samuelson. With reference to the behavior of experimental subjects in a finitely repeated prisoners' dilemma, he writes (Osborne and Rubinstein, 1994, 135):

> The fact that it may be consistent with some subgame perfect equilibrium of the infinitely repeated game is uninteresting since the range of outcomes that are so-consistent is vast.

He goes on to discuss experimental evidence showing that many features of behavior in other finite supergames are, indeed, nicely explained by the theory of finite supergames, while the theory of infinite supergames offers no insights in these cases.

It might be added that the rational-choice approach assumes (as a law-like hypothesis) that the relevant situation is perceived correctly, implying

that we need not ascertain beliefs in applying the theory. Once false beliefs are admitted, the beliefs held by the players in a game become situational assumptions. Since these situational assumptions make a difference for the predictions, one would have to show that they are true. This is rather unlikely. Indeed, people cooperate in laboratory experiments with a similar structure where it is highly implausible that they falsely believe in the possibility of infinitely many interactions (for a relevant survey, see Cooper and Kagel, 2016). This indicates that their reasons for cooperation might be found elsewhere.

As matters stand, narrow rational-choice models of infinitely many interactions should be rejected as approximate explanations of human cooperation. If no other approximate explanations of observed cooperation on the basis of narrow rational-choice theory can be found, this speaks clearly against the theory.

A Bertrand competition

We first consider a one-off Bertrand game. Let the market demand function be $x(p) = \max\{a - \min p, 0\}$, where x is the quantity demanded by consumers, p is the vector of prices set by the $N \geq 2$ identical firms serving the market (so that $\min p$ is the lowest price at which consumers can buy), and a is a parameter which is high in a boom and low in a recession. Let the firms produce, without relevant capacity constraints, the good in question at unit costs c with $0 \leq c < a$. Mailath and Samuelson (2006, 101) simply assume $c = 0$. The price set by firm i is p_i, and $p = (p_1, \ldots, p_N)$ is the vector of prices. Let $\nu(p)$ be the number of firms setting the lowest price $\min p$. The firms setting $\min p$ share the market equally while firms setting a higher price have no customers. The profit of firm i, then, looks as follows:

$$\pi_i(p) = \begin{cases} (p_i - c)(a - \min p)/\nu(p) & \text{if} \quad p_i = \min p \\ 0 & \text{if} \quad p_i > \min p_i \end{cases} \tag{1}$$

The monopoly price is $p_{\max}(a) = (a + c)/2$.

In general, game theory assumes that each agent's moves are made according to a complete "contingency plan", called a strategy. Strategies and moves must be distinguished because, in general, strategies cover all situations that might conceivably arise in a game. In the special case of a simultaneous one-off game, there is only one relevant situation where one move must be selected, and strategies and moves need not be distinguished.

For the purposes of explanation and prediction, economists characteristically must assume that all observed behavior can be accounted for as

execution of equilibrium strategies. A list containing one strategy for each and every agent is called a strategy profile. An equilibrium is a strategy profile with the property that each player's payoffs are maximal given the other players' strategies.

In a one-off Bertrand game, players are firms, payoffs are profits, and strategies are prices. Prices, quantities and costs are real numbers. It follows that slightly underbidding the lowest price chosen by one's competitors is always the profit-maximizing move unless this lowest price is equal to unit costs. An equilibrium is characterized by $\min p = c$ independently of the demand parameter a. However, there are several such equilibria: since profits in equilibrium are zero anyway, up to $N - 2$ firms may choose $p_i > c$, an option we refer to as "exiting the market". Since all these equilibria yield zero profits for all firms, they need not be distinguished for our purposes. We therefore simplify by restricting considerations to the symmetric equilibrium $p = (c, \ldots, c)$ and $\nu(p) = N$.

In a finitely repeated Bertrand game, the same firms play the Bertrand game, now called the stage game, $n > 1$ times. The game resulting from repeating the stage game is called a supergame. At each stage t, $1 \leq t \leq n$ of a supergame, there exists a history of length $t - 1$, namely, the move combinations and payoffs of the $t - 1$ previous stages. The history of stage 1 is empty and has length 0. Agents are assumed to know the history when called upon to make a move. The rest of the game after some history of length $t < n$ is called a subgame (where the supergame itself is the subgame reached by the empty history).

Stage games at the same stage t must be distinguished according to their history. A firm's strategy for the supergame selects a price in the first stage game and then prices for each stage game. Thus, a firm's strategy can let the firm react in every conceivable way to the history.

At each stage, the firm decides in view of the future causal consequences of its moves, which in this case are the profits on the path determined by its own strategy given the other agent's strategy. At stage $t > 1$, payoffs from earlier stages are no longer relevant for decision making. Let the payoff of firm i at stage s be $\pi_{i,s}$. We assume that the supergame payoff computed at some stage t, $1 \leq t \leq n$ is the present value $\pi_i^t = \sum_{s=t}^{n} \delta^{s-t} \pi_{i,s}$ of the stage profits for the rest of the game, where $\delta \in (0, 1)$ is a time-invariant discount factor.[22]

[22]The assumption $\delta = 0$ is ruled out since then the evaluation of strategies in the repeated game amounts to a sequence of evaluations of disconnected one-shot games. The assumption $\delta = 1$ amounts to ascribing "infinite patience" to the agents since then only the sum of payoffs and not their timing matters to the firms. However, the sum of payoffs is infinite for positive payoffs; for this reason, we assume $\delta < 1$, which implies finite present

In supergames, a stricter definition of equilibrium, called subgame perfect equilibrium, is used. A strategy for the supergame fixes a strategy (a complete contingent plan) for each subgame. A strategy profile is a subgame perfect equilibrium if and only if the strategy profile implied for each subgame is an equilibrium. Each subgame perfect equilibrium is perfect in the sense that it applies the criterion of equilibrium planning consistently to all subgames including the game itself. Or in other words: In equilibrium, plans for all contingencies are such that there would be no reason to change plans if the contingency arose.

Subgame perfect equilibria in finite supergames are found by backward induction, as explained in the text. We consider only the symmetric case. Backward induction implies that the price vector in each stage game is $(c, \ldots, c)$, just as in the one-off game. Again, there exist asymmetric equilibria where up to $N-2$ firms exit in some or all stage games. In all these equilibria, all firms earn zero profits in all stage games. Cooperation is impossible.

This result extends to all finite supergames with only one equilibrium of the one-off game or, as in the present case, several equilibria all yielding the same payoffs to each player. In these games, subgame perfect equilibria allow only for behavior that can also occur in the one-off game.

For the infinite supergame, replace n by ∞. A profile of trigger strategies as discussed in the text is a symmetric subgame perfect equilibrium. If δ is low enough, equilibrium prices in booms and recessions must be lower than the relevant monopoly price $p_{\max}(a)$ (Mailath and Samuelson, 2006, 202).

B A behavioral model of Bertrand competition

This appendix presents a very simple behavioral *ad hoc* variation of the one-off Bertrand model of appendix A replicating the result of Rotemberg and Saloner (1986) that prices are low in booms and high in recessions.

Assume that owners or managers of firms pursue two goals: maintaining friendly personal relations with other managers or owners and generating profits for their firm. Let underbidding put a strain on personal relations within the industry. From a static model of utility optimization, it follows that underbidding will occur only if the additional profits generated by underbidding are higher than some threshold value $s > 0$, the monetary value of maintaining friendly relations.[23]

values for profits since stage game profits have a finite upper limit.

[23] If firms have different threshold values, s must be the lowest such value. All results remain unchanged.

If a firm underbids if and only if this results in a profit higher than $s > 0$, the condition determining the equilibrium price q is $(q-c)(a-q)-(q-c)(a-q)/N \leq s$, that is, the difference between profits from marginal underbidding and market-sharing profits is not above the threshold value s. Assuming that firms choose the best equilibrium satisfying this condition yields an equilibrium price

$$q(a) = \frac{a+c}{2} - \sqrt{\frac{(a-c)^2}{4} - \frac{N}{N-1}s}\,. \tag{2}$$

This equilibrium requires $a \geq c + \sqrt{\frac{4Ns}{N-1}}$ even in recessions; with a smaller value of a, firms would adopt the monopoly price $p_{\max}(a)$. In equilibrium, $p_i = q(a)$ for all $i = 1, \ldots, N$. Since $q'(a) < 0$, prices are low when demand is high and vice versa.

Acknowledgements

For helpful suggestions and discussions, we are grateful to Geoffrey Brennan, Volker Gadenne, Werner Güth, Sebastian Krügel and the participants of the 2016 workshop "Infinite Idealizations in Science" at the *Munich Center for Mathematical Philosophy*, especially Sam Fletcher, Margaret Morrison, Juha Saatsi, Paul Teller and Michael Weisberg.

References

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

Albert, H. (1963). Modell-Platonismus. Der neoklassische Stil des ökonomischen Denkens in kritischer Beleuchtung. In H. Albert and F. Karrenberg (Eds.), *Sozialwissenschaft und Gesellschaftsgestaltung. Festschrift für Gerhard Weisser*, pp. 45–76. Berlin: Duncker und Humblot.

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

Albert, H., D. Arnold, and F. P. Maier-Rigaud (2012). Model platonism: neoclassical economic thought in critical light. *Journal of Institutional Economics 8*, 295–323.

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

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

Albert, M. (2011). Methodology and scientific competition. *Episteme 8*, 165–183.

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).

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

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

Baker, A. (2016). Simplicity. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 ed.). `https://plato.stanford.edu/archives/win2016/entries/simplicity/`.

Becker, G. S. (1976). *The Economic Approach to Human Behavior*. Chicago: University of Chicago Press.

Binmore, K. (1994). *Game theory and the Social Contract, Vol. 1: Playing Fair*. Cambridge, Mass.: MIT Press.

Binmore, K. (1998). *Game Theory and the Social Contract, Vol. 2: Just Playing*. Cambridge, Mass.: MIT Press.

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

Cooper, D. J. and J. H. Kagel (2016). Other-regarding preferences. In J. H. Kagel and A. E. Roth (Eds.), *Handbook of Experimental Economics Vol. 2*, pp. 217–289. Princeton: Princeton University Press.

Cyert, R. M. and E. Grunberg (1963). Assumption, prediction and explanation in economics. In R. M. Cyert and J. G. March (1963), *A Behavioral Theory of the Firm*, pp. 298–311. Englewood Cliffs: Prentice-Hall.

Friedman, M. (1953). The methodology of positive economics. In *Essays in Positive Economics*. University of Chicago Press.

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

Güth, W., W. Leininger, and G. Stephan (1991). On supergames and folk theorems: a conceptual discussion. In R. Selten (Ed.), *Methods, Morals and Markets*, Game Equilibrium Models, pp. 56–70.

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

Hempel, C. G. and P. Oppenheim (1948). Studies in the logic of explanation. *Philosophy of science 15*(2), 135–175.

Kliemt, H. (2009). *Philosophy and Economics I: Methods and Models*. München: Oldenbourg.

Koopmans, T. C. (Ed.) (1957). *Three Essays on the State of Economic Science*. New York.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Lakatos, I. (1970). Falsification and the methodology of scientific research programs. In A. Musgrave and I. Lakatos (Eds.), *Criticism and the Growth of Knowlegde*, pp. 91–196. Cambridge: Cambridge University Press.

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

Levy, A. (2015). Modeling without models. *Philosophical studies 172*(3), 781–798.

Mailath, G. J. and L. Samuelson (2006). *Repeated Games and Reputations. Long-run Relationsships*. Oxford etc.: Oxford University Press.

Mäki, U. (2009). Reading *the* methodological essay in twentieth-century economics: map of multiple perspectives. In U. Mäki (Ed.), *The Methodology of Positive Economics. Reflections on the Milton Friedman Legacy*, pp. 47–67. Cambridge: Cambridge University Press.

McCloskey, D. N. (1983). The rhetoric of economics. *Journal of Economic Literature 21*(2), 481–517.

Morrison, M. (2015). *Reconstructing Reality. Models, Mathematics, and Simulations*. New York: Oxford University Press.

Musgrave, A. (1978). Evidential support, falsification, heuristics, and anarchism. In G. Andersson and G. Radnitzky (Eds.), *Progress and Rationality in Science*, pp. 181–201. Dordrecht: Reidel.

Musgrave, A. (2011). Popper and hypothetico-deductivism. In D. M. Gabbay, S. Hartmann, and J. Woods (Eds.), *Inductive Logic*, Handbook of the History of Logic, pp. 205–234. Amsterdam etc.: North-Holland.

Ng, Y.-K. (2016). Are unrealistic assumptions/simplifications acceptable? some methodological issues in economics. *Pacific Economic Review 21*(2), 180–201.

Oddie, G. (2008). Truthlikeness. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*.

Osborne, M. J. and A. Rubinstein (1994). *A Course in Game Theory*. Cambridge, Mass and London, England: MIT Press.

Pfleiderer, P. (2014, March). Chameleons: The misuse of theoretical models in finance and economics. Working Paper 3020, Stanford University, Graduate School of Business.

Rodrik, D. (2015). *Economics Rules. The Rights and Wrongs of the Dismal Science*. New York and London: W.W. Norton & Company.

Rotemberg, J. J. and G. Saloner (1986). A supergame-theoretic model of price wars during booms. *The American Economic Review 76*(3), 390–407.

Saatsi, J. and M. Pexton (2013, December). Reassessing Woodward's account of explanation: Regularities, counterfactuals, and noncausal explanations. *Philosophy of Science 80*, 613–624.

Selten, R. (1999). Response to Shepsle and Laitin. In J. Alt, M. Levi, and E. Ostrom (Eds.), *Competition and Cooperation: Conversations with Nobelists about Economics and Political Science*, pp. 303–308. New York: Russell Sage Foundation.

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

Swartz, N. (2016, October). Laws of nature. In J. Fieser and B. Dowden (Eds.), *The Internet Encyclopedia of Philosophy*. http://plato.stanford.edu/archives/win2014/entries/scientific-explanation/.

Varian, H. R. (1997). How to build an economic model in your spare time. *The American Economist 41*, 3–10.

Vega-Redondo, F. (2003). *Economics and the Theory of Games*. Cambridge: Cambridge University Press.

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

Woodward, J. (2003). *Making Things Happen. A Theory of Causal Explanation*. Oxford etc.: Oxford University Press.

Woodward, J. (2014). Scientific explanation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2014 ed.). http://www.iep.utm.edu/lawofnat/.