Journal of Economic Methodology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rjec20

Economic theories and their Dueling interpretations

Itzhak Gilboa, Andrew Postlewaite, Larry Samuelson & David Schmeidler

To cite this article: Itzhak Gilboa, Andrew Postlewaite, Larry Samuelson & David Schmeidler (2022): Economic theories and their Dueling interpretations, Journal of Economic Methodology, DOI: 10.1080/1350178X.2022.2142270

To link to this article: https://doi.org/10.1080/1350178X.2022.2142270

Published online: 14 Nov 2022.

JOURNAL OF ECONOMIC METHODOLOGY
https://doi.org/10.1080/1350178X.2022.2142270

Economic theories and their Dueling interpretations

Itzhak Gilboa[a,b], Andrew Postlewaite[c], Larry Samuelson[d] and David Schmeidler[a]

[a]HEC, Paris-Saclay, Gif-sur-Yvette, France; [b]Tel-Aviv University, Tel Aviv, Israel; [c]Department of Economics, University of Pennsylvania, Philadelphia, PA, USA; [d]Department of Economics, Yale University, New Haven, CT, USA

ABSTRACT
The interpretation of economic theories varies along several dimensions. First, models can describe reality, illustrate a recommended state of affairs, or analyze the structure and implications of a theory. Second, theories can be used for prediction or for explanation. Third, theories can relate to reality in a rule-based or case-based manner. Fourth, theories can be statements about economic reality or about the act of economic reasoning itself. Fifth, theories can offer predictions or merely critique reasoning. We argue that theories are often open to multiple interpretations which can shift depending on the context in which the theory is applied, the surrounding economic literature, and the argument made by the interpreter.

KEYWORDS
Economic theory; economic methodology; models; analogies; criticism; explanation

JEL CLASSIFICATIONS
A1; B4

1. Introduction

1.1. Motivation

Among the social sciences, economic theory is arguably unique in embracing a single, unifying conceptual framework. Most research in economic theory assumes that each agent maximizes an objective function subject to constraints, and the analyst then focuses on the equilibria defined by the interaction of such agents. Not only is this conceptual framework a fair description of the state of affairs in economic research, but many economists espouse it as a normative stance or even an ideological position.[1]

The adoption of a unifying set of modeling guidelines does not single out economics among *all* sciences. But typically such a commitment is supported by confirmation of the assumptions on which their models are built and on successful predictions. Economics, however, has been under persistent attack for failing to produce accurate predictions comparable to those of the natural sciences. Moreover, over the past decades the fundamental assumptions of the conceptual framework employed by economic theory have been subjected to direct criticism, mostly by psychologists.[2] Some economists and many scholars from other disciplines argue it is high time that economics quit the game of making implausible assumptions and deriving unrealistic conclusions.

Of course, scientific disciplines typically do not totally discard a theory as soon as it is refuted. Instead, a theory can be refined or restricted to exclude the offending observations. Newtonian physics can be refined to incorporate relativity, and restricted to avoid quantum effects. And even when a theory is plainly refuted, the conceptual framework within which it was developed can be used to come up with other related theories that fit the data, allowing the central paradigm to be defended by a 'protective belt' (Lakatos, 1970). Perhaps this is what is happening in economics. For example, experimental findings such as those for the ultimatum game (Guth et al., 1982) might suggest that the theory of subgame perfect equilibria with purely-material payoffs fails. But

CONTACT Itzhak Gilboa tzachigilboa@gmail.com

the introduction of non-material payoffs as determinants of utility can accommodate the experimental data within the conceptual framework (say, subgame perfect equilibria, utility maximizing agents with correct hierarchies of beliefs, and so on). Indeed, to a large extent this has been the direction taken by behavioral economics in recent decades.

And yet, some puzzles remain. First, it is not obvious that theory based on behavioral economics provides more accurate predictions than theory based on the classic assumptions of rationality. Partly, this might be due to the paucity of data – behavioral economics typically relies more heavily on mental phenomena that are not directly observable than does standard economics. Concepts such as reference points, mental accounts, and emotional payoffs are hard to pin down by observable data.[3] Partly, the problem may be due to the psychological theories that are often not very robust.[4] Partly, it may be that much of behavioral economics has focused on accommodating a wider range of behavior, which inevitably blunts its ability to predict. Whatever the reason, many seem to concur that behavioral economics can exhibit the implications of certain effects, but that it falls short of providing the predictions that classical economics was seeking.

Second, theoretical behavioral economics, for the most part, retains the conceptual framework of maximization-and-equilibrium and differs from standard theories either by using non-material determinants of utility or by allowing errors and biases in reasoning. While some classical assumptions are relaxed in such models, the fundamental methodology remains unchanged. From this point of view, behavioral economics has not produced a 'paradigm shift,' and very little attention has been devoted to formal models that are of a completely different nature. For example, Simon's (1955) 'satisficing behavior' is rarely used as the basis of behavioral economics. Simon, who coined the term 'bounded rationality,' suggested this model when constrained optimization and equilibrium analysis were still in their infancy. He received the Nobel prize for this contribution, but, despite the attention and the relative primacy, his alternative framework has not changed the field. Many subsequent models have used the term 'bounded rationality,' but applied it to relatively minor modifications, such as bounded memory.

A critic of economic theory might therefore argue that not much has changed in recent decades. Economic theory still clings to the same conceptual framework, despite the fact that many assumptions are unrealistic and many predictions are inaccurate. Why are economists so content with their theories? The question seems particularly pertinent because, since Simon's contributions, other sciences have offered several alternative approaches that economic theory could have adopted to make a total switch of method. Why didn't advances in computing power allow computable general equilibrium models to supplant macroeconomics? Why hasn't economics traded mathematical theorems for neurological studies?[5] Or given up theorizing for prediction based on machine learning techniques? Or replaced maximizing agents with computer simulations of 'agent-based' models?

We believe that economists find value in theories and in specific mathematical results in ways that go beyond their use for prediction or recommendation. Making predictions and offering recommendations are certainly important to most academic economists. Moreover, some predictions are quite accurate and some recommendations are accepted as useful. But if prediction and recommendation were the only goals of economic theory, it would be hard to understand why economists are so blasé about observations that their assumptions are far-fetched, and so unperturbed by refutations of their theories. This paper explores other ways in which economists find their theories and results valuable, in the hope of better understanding these puzzles.

Before proceeding, a few comments are in order. The nature of our analysis is positive, rather than normative. Our goal is to model the way that economists think about their models, and how it differs from the way other academics view these models. Naturally, each of the authors, as any working economist, has views on normative issues relating to the way economic research should be conducted. Yet, these views are not the focus of this paper (nor are they necessarily shared among the authors). Rather, we take a sociology-of-science viewpoint in trying to understand what are the views of many in the field.

It goes without saying that a complete consensus in the field is a rarity. Thus, we view our task as mapping qualitatively distinct ways in which economists can conceptualize models, theories, and results. When pointing out a specific conceptualization, we make the implicit claim that a non-negligible portion of economists would espouse that view (be it positive or normative) – but not necessarily that a majority would. In a few cases we will make bolder statements, involving generally accepted views, but such statements will be the exception.

Claims about the way economists think have empirical content. They can and should be tested. Unfortunately, we are not in a position to report results of such empirical studies. We only suggest the theoretical framework(s) within which these questions can be analyzed. Our theorizing is evidently not conducted in a void: to the contrary, every claim made below (or above) about the way some (or most) economists think is based on observations, gleaned from informal discussions the authors have had with fellow economists over decades of research. Yet, these claims cannot be taken to be empirical facts without further research.

Finally, none of the five distinctions discussed in this paper is novel. All have appeared in previous research, receiving varying degrees of attention. We have discussed three of the five distinctions in previous papers. We accordingly offer only brief descriptions of these distinctions, in the process finding it convenient to focus on the conceptualizations offered in our previous papers, which are clearly separated in our minds. However, we make no claim regarding the precedence or superiority of our specific ways of formulating the distinctions. The main goal of the present contribution is to provide a concise catalog of these distinctions and to highlight their interactions.

1.2. Terminology

This paper refers to the terms 'models,' 'theories,' 'results,' 'conceptual frameworks' and even the word 'paradigm' has already made its appearance above. Some of these terms are fraught with different meanings, and have been defined in various ways in different contexts. We illustrate our usage with an example.

Milgrom and Stokey's well-known no-trade theorem (Milgrom & Stokey, 1982), introduced here as a convenient but rather arbitrary example, establishes the *result* that (under certain conditions) differences in information alone cannot give rise to trade. This result appears in the context of a *model* of a competitive exchange economy whose expected-utility-maximizing agents share common (or at least concordant) prior beliefs. This model is a special case of a *theory* of agents who maximize utilities defined over their own material rewards subject to market-clearing prices, which in turn is an application of the *conceptual framework* of utility maximizing agents.[6]

There are no bright lines between these various categories, but the meaning is typically clear from context. A model may specify all relevant assumptions, or only provide the template within which assumptions can be imposed. A theory may be associated with a single model, but typically encompasses a collection of models with several results for each model. A conceptual framework may be identified with a single theory that is most commonly investigated, but typically allows various interpretations, resulting in different theories.[7] Adding interpretations for the formal objects and guidelines for how the framework is to be applied yields a theory, which in turn can be specialized to construct models and derive results.

We are concerned with theoretical rather than empirical work in economics, and so we refer to 'economic theory' throughout. However, we typically refer to 'economists' rather than economic theorists, reflecting a view that even if the production of economic theory is primarily the province of economic theorists, other economists are at least complicit in the enterprise. Applied work typically begins with an economic model, and a first-year grounding in economic theory is standard in graduate programs, regardless of the intended field of study. Of course many applied economists are deeply committed to analyzes that make useful predictions, often with the goal of informing the policy debate, and so it is certainly a mischaracterization to claim that economists as a whole are unconcerned about prediction. However, the theoretical models that provide the building blocks for such analyzes are not constructed by theorists with the goal of making accurate predictions, nor are they typically evaluated by applied researchers in terms of the realism of their assumptions.

1.3. Outline

In the next five sections we examine five distinctions that are useful in identifying types and purposes of economic theories, some more relevant to understanding economists' views of counterfactual results and others more relevant to understanding unrealistic assumptions. The distinctions are:

(i) Models can be positive or normative, or can be used to analyze a theory.
(ii) Models can aim to explain rather than predict.
(iii) Models can be used for prediction in a case-based rather than a rule-based manner.
(iv) Theories can be useful when they apply to the work of the economist, rather than to economics proper, and they can be tools that economists use.
(v) Results can be useful tools for critiquing reasoning, even without suggesting predictions.

The dividing lines inducing these distinctions are sharper in some cases than in others. We do not attempt to offer an example of an economic theory falling into each of the seventy-two cells induced by these distinctions, and indeed many cells are empty. Section 7 argues that theories can be open to multiple interpretations, which can shift depending on the context in which the theory is applied, the surrounding economic literature, and the argument made by the interpreter.

2. Positive, normative and analytical models

Our first distinction concerns the relationship of economic models to reality.

2.1. The distinction

It may be useful to first think of physical models such as maquettes constructed by an architect. Assume, for concreteness, that we examine a maquette of a town square. Such a maquette can

(i) describe a square that exists, or that existed in the past;
(ii) describe the architect's proposed design for a square; or
(iii) test the feasibility of a possible square.[8]

The first is a positive model, the second a normative model, and the third is what has sometimes been referred to as an 'analytical' model.

There is nothing in the maquette itself that can unambiguously determine which the intended interpretation is (among the above and possibly others). The same maquette can be constructed to illustrate the square to people who cannot visit it, to propose a reconstruction of a square that looks quite different at present, or to test whether suggested buildings can fit into the space.

Economic models can similarly be interpreted in these three ways. Interpretation (i) gives a descriptive/positive reading of an economic theory. A theory demonstrating that increased educational attainment on the part of women and increased assortativeness in marriages will lead to increased inequality in family incomes is positive. Interpretation (ii) is normative. A paper arguing that we should subsidize the cost of higher education in order to flatten the profile of education attainment and hence, family income, is normative.

The positive and normative categories have garnered the lion's share of attention in introductions to economics and in expositions of economics to outsiders, with interpretation (iii) often not mentioned. Indeed, interpretation (iii) does not have a widely accepted, 'official' title. However, economists often refer to this type of reasoning in explaining the value of models. For example, Aumann's

(1987) model of Bayesian rationality in games initially defied categorization. The paper made no prescriptions, and so was not normative. It was difficult to interpret as a positive model of decision making, since each state specified the actions of each of the decision makers, seemingly leaving no scope for assessing and selecting alternatives. Nonetheless, the paper clearly made a methodological contribution, whether interpreted as providing a motivation for correlated equilibrium or as an argument that a failure to observe correlated equilibrium reflects difficulties in the concept or assumption of rationality. Aumann's response was that the model was analytical, designed to demonstrate that the notion of the common knowledge of rationality could be incorporated within standard techniques.[9]

One could endeavor to explain examples of this type along the lines of (i), as standard descriptive science, arguing that Aumann's analysis is meant as a model of how rational people will behave. But often economists who suggest such models do not argue that their model is the 'correct' model, let alone the simplest account of the observed phenomenon. Theorists who explain apparent anomalies as outcomes of Bayesian Nash equilibria need not believe that such equilibria, or even common knowledge of rationality, are very plausible. If asked, 'Why do you propose this explanation, then?,' they might say, 'Well, I believe that there is some value in testing whether the standard assumptions are compatible with the phenomenon at hand.' Thus, they resort to motivations along the lines of (iii) above. In a sense, they use the term 'model' as it is used in mathematical logic: an analysis of the structure and implications of a theory. Proving that there exists a formal mathematical structure within which all agents are rational, have commonly known beliefs, and exhibit a certain behavior pattern is an exercise that proves the consistency of the assumptions with some stylized facts.

This way of interpreting economic models is akin to accounts of economic theories by Hausman (1992) and Sugden (2000, 2009). Both relate to the disparity between economic models and observed reality, and highlight the importance attributed to the model as such. Hausman focuses on the plausibility of the model's basic assumptions, while Sugden focuses on the role of logical coherence in the judgment of 'credibility.' While these notions are closely related, the main feature of an analytical model seems to us a formal problem of the type: Is there a model in a given class that can exhibit a given type of observations?[10]

2.2. Some comparisons

A good point of departure in appreciating this third use of economic models is the well-known statement of Box (1979) that 'all models are wrong, but some are useful.' Every field of science must deal with the recognition that its models are not literal descriptions of reality.[11] There will accordingly always be some observations inconsistent with the existing models. However, different fields have different ways of dealing with these contradictory observations.

One might expect that, upon encountering observations that are inconsistent with an existing model, scientists would contract the scope of applicability of the existing model and then either refine it or find new models to deal with the formerly anomalous observation. This preference for 'accuracy' is complimented by a preference for a unifying theory, which, thanks to simplicity and generality, tends to enjoy greater credibility as a basis for prediction in future problems, especially under novel circumstances. In some cases, mostly in the natural sciences, one can enjoy both accuracy and simplicity. In others, as often happens in the social sciences, different theories are developed for different phenomena, and thus unification is sacrificed for accuracy.

For example, psychologists respond to observations that are inconsistent with an existing model by contracting the scope of applicability of the model, and then enriching their set of models to include one encompassing the formerly anomalous observation. Physicists, in contrast, aspire to a single model that will explain all physical phenomena. When encountering contrary observations, the model is refined in order to make more nuanced predictions that accommodate the observation.

Economics is an anomaly. The field's (revealed) preference appears to be for a unifying framework even at the expense of predictive accuracy. Economists work with a common theory, building their analysis up from the foundation of people maximizing relatively stable utility functions in response to possibly shifting constraints, but respond to challenges to the theory neither by modifying or generalizing the theory nor by encompassing alternative theories. Instead, economists sometimes go to great lengths to show that seemingly anomalous results can be encompassed within the standard theory, even if it is difficult to defend the resulting models as realistic. As a result, the third type of models, illustrating the consistency of a theory with a given phenomenon, is probably more common in economics than in other disciplines.

Economists rarely assert they have 'the' correct model of a phenomenon, and routinely refer to a portfolio of models and analysis when making a point.[12] However, they insist that these models be drawn from a common modeling framework. The idea is that the predictions of a suite of models sharing a common conceptual core will be more useful than either a suite of models organized by no common vision or a single model. We can view analytical models as exercises in showing that seemingly anomalous observations are indeed consistent with the common modeling framework. In some cases, this leads to surprisingly simple potential explanations – an argument that familiar models cannot account for the fact that people appear to envy the economic well-being of others gives way to the observation that parents may be concerned about the welfare of their children, who will face a tournament in which desirable mates go to the wealthy. In other cases, accommodating the anomalous observation may require such modeling gymnastics as to suggest a relaxation of the common modeling framework. The standard is then to undertake the least drastic relaxation capable of explaining the phenomenon of interest. Behavioral economists argue that more flexible models of preferences are the most useful way to explain some behavior, while the current rather unsettled place of behavioral economics in the discipline reflects a lack of consensus as to which such models constitute the least drastic relaxation.

Perhaps the closest parallel to the use of models in economics is provided by evolutionary biology. Evolutionary biologists similarly have a common modeling framework, consisting of selection among phenotypes generated by random mutations. This again gives rise to a collection of particular models (reproduction may be sexual or asexual, fitness may be frequency independent or frequency dependent, and so on). Considerable attention is devoted to analytical models, designed to identify the extent to which some behavior is consistent with the selection framework. Sometimes, the answer may be no – spandrels may be an unselected byproduct of the process (Gould & Lewontin, 1979). However, evolutionary biologists would argue that there is much to be learned from looking for evolutionary explanations (Dawkins, 1976), and much to be learned from a suite of models built on the framework of mutation and selection.

2.3. Identifying interpretations

As in the case of an architectural maquette, the suggested interpretation of an economic model generally cannot be inferred from the model itself. Moreover, the preferred interpretation might vary from one economist to another, and even the same researcher might prefer different interpretations in different contexts or at different times. A celebrated example is Savage's interpretation of his axioms in the face of Allais's (1953) paradox. According to legend, at a conference in Paris in May of 1952, Allais presented Savage with his 'paradox' and the latter gave the typical response that violates expected utility theory. Allais is reported to have asked Savage how he could expect people to satisfy his axioms if he didn't satisfy them himself. The next day Savage came up with a response that became rather influential: he argued that his axioms can help him correct his choices when they were wrong; that is, that when the model does not work well descriptively, it might be used normatively.[13]

The switch from a descriptive interpretation of a model (i) to a consistency test (iii) is more continuous and less dramatic than between descriptive and normative interpretations. The latter, namely, the distinction between the 'is' and the 'ought,' is a qualitatively stark distinction, and, as pointed out above, the 'ought' is most interesting when it differs from the 'is.' By contrast, one may suggest a model with a descriptive interpretation in mind, but, when facing an aggressive audience, one might take a step back and rather than promoting the model as an explanation of a real-life phenomenon, present it as a 'proof of concept' or 'merely an exercise' in testing the scope of the standard paradigm.

3. Prediction vs. explanation

There is a rich literature in the philosophy of science and in statistics about the distinction between prediction and explanation in general, and specifically, as possible goals of modeling (see Shmueli, 2010, and the survey therein). While prediction is defined by the accuracy of fitting yet-unobserved data, explanation is identified with the 'warm feeling inside' that one experiences at 'a-ha' moments. In one approach, explanation can be modeled formally. For example, Kolmogorov complexity or minimal message length models have been used to measure the degree to which we have a sense of 'understanding' or 'explaining'[14] (see Dowe et al., 2007). Alternatively, explanation is often assessed on an 'I know it when I see it' basis.

Economics has often been described as especially prizing explanations. Keynes (1936) wrote

> The theory of economics does not furnish a body of settled conclusions immediately applicable to policy. It is a method, rather than a doctrine. An apparatus of the mind, a technique of thinking, which helps its possessors to draw correct conclusions.

More recently, Aumann (1985) emphasized explanation, as opposed to prediction, as a major goal of game theory. Simon (2001) distinguished between basic and applied science, where the former seeks explanations more than predictions, as compared to the latter. The distinction between explanation and prediction has received renewed attention in recent years, with the rise of machine learning (ML) techniques, which often provide better predictions than do theoretical models, but do not necessarily provide a sense of 'explanation.' These advances raise the questions, why don't the social sciences adopt ML techniques, or even, who needs to theorize if we can predict better using theory-free, general-purpose techniques? In response to these questions, economists often respond that they seek causal relationships, which are particularly important for policy questions, and therefore they cannot abandon their theories in favor of ML techniques. Why is this the case? Why do economists place such emphasis on an often vague sense of understanding in addition to the concrete ability to predict, and why are they often willing to sacrifice predictive ability in quest of understanding?[15]

First, we believe that 'mere understanding' is not precisely what economists have in mind when they defend models that do not predict very well. Most economists would agree that the understanding obtained by explanations might be useful in other contexts. Indeed, one may argue that humans evolved to enjoy the feeling of understanding because such understanding is likely to be useful in future, yet-unspecified decisions. Thus, the need to predict differs from the need to explain in the immediacy of the new problems for which prediction is needed. This distinction tends to be a matter of degree, not of kind.

Second, we find that the need to establish causality is correlated with the need for understanding and explaining, but there is no inherent causal relationship between them. Specifically, one can establish causal relationships without any understanding. For example, if we have a database generated by a random-controlled-trials medical experiment, and we find that the vast majority of those who received treatment ended up with no symptoms, whereas the opposite is true of the patients who received no treatment, we can conclude that treatment is the cause of the outcome even if we do not know what mechanism drives it. Indeed, the human mind naturally finds causality in the phenomena it observes, and it can use this practical understanding of causality, as long as potential confounds are noted, without understanding the mechanisms behind it. We suspect economists are hesitant to rely exclusively on ML and similar techniques not because of an absence of understanding *per se* but rather because of the (presumably related) difficulty in exporting knowledge from a given database to making predictions in a very different one. For example, predicting the impact of global warming, or of the Covid-19 pandemic are challenging tasks even without the need to establish causal relationships: it is simply difficult to predict out-of-sample when the sample is not representative. Indeed, new government policies can often generate such non-representative prediction problems, but it is not the causal relationship that is at the crux of the matter, but the novelty of the new problems.

Third, conceptualizing the quest for understanding as an input for yet-unspecified and potentially-novel prediction problems can help us explain economists' affinity for observationally equivalent models. Some of the more celebrated contributions in economics consist of producing models that are observationally equivalent to existing models. One especially sees this in decision theory, with its emphasis on formulating systems of axioms whose implications are equivalent to those of an existing decision rule, but similar exercises occur in game theory, social choice theory, mechanism design, and so on.[16] Of what use is a model whose empirical content precisely duplicates that of an existing model? Axiomatizations and equivalent models can further understanding. For example, positing that a decision maker maximizes a utility function is equivalent to arguing that she makes decisions and can do so in a transitive way.[17] Thus, by dint of the underlying equivalence theorem, any database that is consistent with utility maximization is also consistent with a (complete and) transitive preference order, and vice versa. The axioms that characterize utility maximization provide understanding: economists feel that they understand the reason that certain databases are or are not consistent with utility maximization, and may be more comfortable defending complete and transitive preferences than defending an equivalent maximization procedure. This 'feeling of understanding' has implications for predictions made in different problems. For example, suppose that an economist considers choices between bundles of groceries and finds that they can be explained by utility maximization. Concluding from this finding that utility maximization is also a good model choices of careers or financial portfolios might require a leap of faith. However, if one understands that transitivity of choices is the key feature, one may find the aforementioned leap of faith more manageable. In other words, the feeling of understanding allows us to project from one domain to another. The further are the 'out of sample' predictions one needs to make from the observed sample, the more one needs to resort to understanding, and, correspondingly, the more one may benefit from reformulations of the same theory.

These points lead us to the view that the novelty of the problems for which predictions are needed, relative to the problems one has observed in the past, is of paramount importance in determining the relative emphasis that various differences put on prediction and explanation. This degree of novelty can explain the emphasis of social scientists on explanation, as well as the value they attach to seemingly prediction-free explanations. Moreover, novelty can also explain distinctions within the social sciences in their tastes for models: economists cherish their unified conceptual framework and are often critical of psychology, where different theories seem to be developed for different phenomena, without an obvious unifying principle. It seems that, as compared to psychologists, economic theorists put a higher weight on the role of models as a source of general, unified explanations than as tools for generating predictions. This difference is related to varying degrees of novelty: because psychology, like medicine, focuses on individuals, it can hope to have many observations that are causally independent and that represent the prediction problems. This puts a natural focus on models that provide good predictions. By contrast, when economics deals with phenomena such as financial crises or growth of economies, it is restricted to learning from relatively few observations, which are hopelessly causally intertwined. In such situations, an economist tends to relinquish the hope for accurate quantitative predictions and seek qualitative ones. It is then that the role of models as providing explanation is highlighted.

Economics is perhaps the social science in which there is the greatest variability in the degree to which problems can be isolated and experimented with. Some problems in economics deal with the behavior of individuals and can be studied experimentally. Indeed, certain problems are studied both in experimental economics and social psychology – where one could expect the model to provide predictions. Other problems that are clearly within the realm of economics involve entire economies and long time scales. In these, experimentation and even isolation of a problem is often impossible. Yet, there are economic models that serve both purposes. For example, maximization of discounted utility is used both to explain simple decision problems with a few time periods, as well as the behavior of an entire economy over decades. It follows that the same model can sometimes be interpreted as a tool for predictions, and sometimes as a building block in constructing an explanation.

4. Rules vs. cases: models as analogies

Several authors, including Sugden (2000, 2009), Grune-Yanoff and Schweinzer (2008), and Cartwright (2010), have suggested that models are used for analogical rather than deductive inference. This section draws on Gilboa et al. (2014), who suggest that different models, as well as empirical and experimental evidence, can be combined for economic prediction, as in case-based theories or in similar techniques in statistics (kernel estimation and classification, *k*-nearest neighbor, and so on).

According to this view, a formal model is not to be interpreted as a rule-based statement, preceded by universal quantifiers, but as a 'theoretical case' whose similarity to any prediction problem should be judged alongside the similarity of other theoretical cases, as well as empirical and experimental ones. This model of the way economists interpret formal models was designed to explain certain phenomena, such as the relative equanimity with which economists accept apparent violations of their models, their preferences for highly simplified models, etc.

For an example of the main argument, we may consider the ultimatum-game model. The backward induction solution applied to the game with purely material payoffs suggests that the proposer captures (nearly) all the surplus. As is well known, experimental evidence (starting with Guth et al., 1982) shows that this is often not the case. If we consider the universal statement that backward induction (computed for purely monetary payoffs) will be played, the experiments provide counterexamples to the theory. Under this view, this evidence should force one to abandon the universal statement. However, if the theoretical model is not viewed as a universal statement but as a 'theoretical case,' it is not meant to produce predictions via logical deduction, but rather via analogies. Alongside the theoretical case there are many others, including experimental and empirical evidence.[18] When faced with a concrete prediction problem or with the task of explaining observed behavior, the economist is expected to ask herself which cases it is similar to, and to make predictions according to the outcomes of these cases. In particular, the economist might consider a real 'ultimatum game' played for millions of dollars, and pose the question, 'Is this case more similar to the theoretical analysis I had on my whiteboard or to the experiments I read about?' Her final judgment may be either one. But at no point does she have to deal with a theory that was refuted: the theoretical analysis made no claim to be a general rule to begin with, and thus it could not have been refuted. Cases can refute rules, but they cannot refute other cases.

The distinction between rule-based, deductive inference, and case-based, analogical one is, again, a matter of interpretation. Formal models in economics are typically not accompanied by a user's manual. This holds for the way in which a rule-based model should be used, for the way a case-based one should be used, as well as to the question, which type of model it is. Specifically, if one takes a deductive inference interpretation, one might wish to see a characterization of the scope of phenomena to which the model can be applied, though this is rarely specified. If, on the other hand, one takes an analogical interpretation, one would need some guidelines as to how one is to assess the similarity of the case to the prediction or explanatory problem at hand, and how similar the case is to other analogies. Again, economic papers typically leave this to the discretion of the reader. Finally, even the question, 'How should I interpret this model?' is rarely addressed explicitly. As a result, a model that was originally suggested as a general rule might well be reinterpreted as a potential analogy.

For example, the original or 'classical' view of game theory was that the game theoretic model contains all the information relevant to the interaction, and hence can be viewed as a literal description of the interaction.[19]In this view, a model of the ultimatum game is a rule-based statement inviting deductive inferences. Observations to the contrary indicate that the model's domain of applicability must be contracted, with new models arising to handle the excluded situations. As evidence mounts, the applicability of the model may ultimately be considered relevant only when Doctor Spock (the Vulcan, not the pediatrician) plays with his clone. In contrast, game theoretic models are now typically viewed as models that capture some elements of a strategic interaction while excluding others. The model now becomes a theoretical case. The model is then always relevant, serving to point out as starkly as possible how backward induction reasoning may influence behavior. This observation is combined with those arising from other cases to reach a prediction tailored to the setting at hand.

5. Economic models vs. models of economics

Gilboa et al. (2021) points out that many results in economic theory do not deal with economic phenomena such as production, trade, and consumption, but with scientific phenomena of modeling, fitting data, and so forth. For example, Arrow's impossibility result (Arrow, 1950) shows that a preference-aggregation function that satisfies certain conditions has to be dictatorial. Thus, no democracy has managed to implement a function that satisfies these conditions. But the theorem is hardly about the workings of current democracies. Rather, it is about the ability of social scientists to design systems of aggregation of preferences. Similarly, the no-trade theorem of Milgrom and Stokey (1982) states that under certain assumptions, including common knowledge of rationality, the arrival of new information cannot explain trade. The theorem should not be read as a statement about financial markets, and the fact that trade exists is hardly proof that Milgrom and Stokey are wrong. Rather, their result was about economic research, suggesting that the standard assumptions are probably too strong, as they lead to implausible results.

Thus, Gilboa et al. (2021) distinguishes between (i) 'economics,' which deals with economic phenomena, (ii) economic methods, which develop technical tools for economists to use, and (iii) 'economic methodology,' which deals with the scientific phenomena of economists studying economics. The development of economic methods makes no claims about reality, and could just as well (and often does) occur in mathematics, statistics, philosophy, and other disciplines. Much of the work in economic methods occurs within economics because it is motivated by a specific vision of how the resulting methods can be used to enhance the study of economics. Similarly, work in economic methodology could be conducted in the philosophy of science, but is often undertaken by economists in a quest to guide economic research.

We are concerned here with economics and economic methodology. As Gilboa et al. (2021) notes, the distinction between economics and economic methodology can be a matter of interpretation. For example, a result about the existence of an equilibrium concept may be viewed as a statement about the long-run behavior of an economic system, as well as a statement supporting the use of the concept as a tool of analysis. An impossibility theory may be interpreted as the methodological statement that certain models are incapable of modeling certain phenomena, or as the positive statement that certain economic institutions or policies are doomed to fail.

In selecting a preferred interpretation, context plays an important role. For example, because Arrow's impossibility theorem was stated in a modern democracy, the proof that certain assumptions imply dictatorship can only be viewed as an impossibility result. But had the same result been proved three millennia earlier, it could well have been viewed as an argument against any attempt to establish democratic systems. Along similar lines, should someone wish to establish a stock market in which only open-source software is allowed to trade, the result of Milgrom and

Stokey (1982) might appear relevant to such a market as a piece of economics, rather than methodology.

There are examples of results that can be reasonably interpreted only as methodological under any plausible context. For example, Afriat's theorem (1967) explicitly deals with databases of consumer choices and with utility functions that can rationalize them. As such, it is hard to imagine an interpretation of the theorem that would *not* be about the working of economists. However, such examples are rare. Indeed, one may argue that any result that is currently interpreted as 'economics' can be reinterpreted as 'methodology' in a slightly different context. To make this argument it suffices to note that the authors of a paper can always add a comment, 'In our opinion, these conclusions are highly unreasonable, and they only go to show how flawed is the system of assumptions our profession currently espouses.' Piccione and Rubinstein (2007) offer two different interpretations of their model, where the latter tends to view the exercise as an illustration of a methodological point.

Finally, we observe that the third, 'analytical' interpretation of models suggested in Section 2 above relates to the distinction between economics and methodology: when an economist reads a model as 'merely' testing the consistency of standard assumptions with an observed phenomenon, she invariably makes a methodological claim. Specifically, if standard assumptions are proved consistent with the phenomenon at hand, this weakens the implicit or explicit argument for changing the paradigm within which theory is conducted.

6. Prediction vs. critique

As mentioned above, several authors, including Hausman (1992) and Sugden (2000), have pointed out that the analysis of economic models may not be constrained by observations as one might have expected. Along similar lines, Gilboa et al. (2018) suggests that some economic models and results are designed to critically examine reasoning, without necessarily offering any predictions. Specifically, Gilboa et al. (2018) presents a model of economic modeling, where a given phenomenon needs to be reconciled with at least one permissible interpretation of a 'conceptual framework,' or where a given prediction/recommendation is shown to follow from all such permissible interpretations.[20] Economics is used as critique when a given statement – whether a prediction or a recommendation – is tested for consistency with an acceptable (conceptual) framework. If there is no permissible interpretation of the framework that is compatible with the statement, the latter is criticized as unreasonable. However, if such an interpretation exists, the analysis may still be inconclusive. Only if *all* permissible interpretations lead to the same conclusion is the statement supported. Thus, according to this conceptualization, the distinction between prediction and critique is akin to the distinction between universal and existential quantifiers.

Using economic models for critique and interpreting them as 'analytical' models (as in Section 2 above) are closely related. In fact, in both cases the formal exercise is the same: one asks whether a given statement about the world is compatible with (at least) one permissible interpretation of a framework. The difference between the two notions is the balance of power, as it were, between the statement (about the world) and the (conceptual) framework. In the case of an analytical model, the statement about the world is a fact, presumably a robust observation. Scientists are not supposed to argue with facts, and this means that the statement has considerable power.[21] The framework is then under attack, and scientists wish to know whether it can accommodate the fact, with the alternative being that the framework is modified or discarded. An analytical model might show that the framework can still accommodate the fact. In doing so it does not claim to make prediction about the world, only to defend the framework.[22] But if no such analytical model is found the framework becomes more dubious.

By contrast, there are scenarios in which the statement is not considered to be a fact, but rather a prediction or a recommendation. In these cases the framework might be more powerful. For example, assume that the framework involves equilibrium analysis, and the statement is that 'printing money' will solve the government's budgetary problems. In this case, the equilibrium framework is rather powerful, being accepted by most, and the prediction that is incompatible with any permissible interpretation of the framework is under attack. This would be an example of interpreting the model as a critique.[23]

The focus on qualitative prediction versus critique is again a matter of interpretation. It is hard to identify clear-cut examples of critiques of reasoning that do not make some qualitative predictions as well. And, rather naturally, this interpretation can change according to context, as well as from one person to another. For example, an economist[24] might analyze a model and conclude that, ceteris paribus, an increase in the money supply would result in the same increase in prices. This prediction might fail for a variety of reasons. Asked to defend her analysis, the economist might say, 'Well, we don't have ceteris paribus in real life. But my point was that increasing the money supply wouldn't simply increase the government's budget in real terms.' That is, a model that was originally stated as a predictive one can be re-interpreted as a critique of an implicit argument.

A central example of multiple interpretations is given by models of competitive equilibria within which existence results and the two welfare theorems are proved. The first welfare theorem is sometimes regarded as a positive model of the economy, predicting that markets will clear and outcomes will be efficient. One description of the work of Gerard Debreu is that 'Debreu's mathematical models provided proof of how prices affect the supplies of goods bought and sold.' (Gallagher, 2005). Building on this interpretation, the first welfare theorem's conclusion that competitive equilibria are efficient provides the foundation for a view typically associated with conservative policy makers that markets on the whole produce efficient outcomes, and should be left to operate free of regulation or interference. Of course, those advancing such views recognize that a host of market imperfections prevent a literal application of the welfare theorem, but they argue that these imperfections are typically not of first-order importance.

An alternative view is that, rather than predicting that market outcomes will be efficient, the first welfare theorem sets the terms of the policy debate, placing the burden of proof on interveners to identify a market failure and to argue that the proposed intervention will lead to an improvement. In this view, the first welfare theorem becomes an exercise in critique, arguing that we should not be persuaded by a claim that 'this regulation will improve welfare' that cannot identify a market failure and that does not examine whether the regulation introduces further distortions. When viewed as critique, the first welfare theorem cannot be invoked to conclude that the labor market is working well and requires no intervention. Moreover, an understanding of the welfare theorems is consistent with a proposal to increase the minimum wage. However, applying the first welfare theorem as critique, one would reasonably expect the proposer to point to some glitch in the market (perhaps monopsony power on the part of employers, or an inability for workers to borrow against their human capital in order to acquire skills) and to argue that the increase will not have prohibitively deleterious employment effects (perhaps, again, because of monopsony power, or because workers will become more valuable by acquiring more skill). The key point is that the use of the model can shift between prediction and critique, depending on the application.

The second welfare theorem also admits varying interpretations. In some contexts, this can be viewed as a positive result, making the observation that policies designed to adjust endowments need not sacrifice Pareto efficiency, or as a normative result, offering the advice that incomes policies should be pursued by reallocating endowments. In other circumstances, the second welfare theorem acts as a critique. When evaluating suggestions that low-income people be given food stamps, or young people be given educational loans, or senior citizens have access to subsidized mass transit, the critique offered by the second welfare theorem is that one should simply give the recipients money, and then let utility maximization and markets work. One might reasonably respond that it is naive to expect markets to work so magically. Perhaps young people are not the best judge of how to spend a windfall, or it is too difficult to target those senior citizens who use mass transit, and too large a distortion to subsidize all senior citizens. However, the value of the second welfare theorem as critique is again that it forces the proponent of the intervention to be precise about the policy goal, and to explain how the costs and benefits of the proposed policy might combine to produce a more effective package than the market.

7. Dueling interpretations

We have briefly surveyed five distinctions that can help us understand how economists think about models, and have pointed out that, according to each distinction, a model may be interpreted in more than one way. Moreover, interpretations may vary across contexts, time and people. How do interpretations vary, and why? We offer several examples.

First, a model originally presented or viewed as positive can be reinterpreted in reaction to dissatisfaction with its plausibility or predictive success. Expected utility theory had its birth in the work of Jacob Bernoulli as a positive model, designed to explain preferences over lotteries. By the time this theory reached its culmination in the work of Savage, doubts about its predictive success were emerging. As we noted in Section 2, Savage's response was to interpret the model normatively. Similarly, models of reputations in repeated games grew out of a positive motivation, prompted by counterintuitive outcomes in models such as the chain store game (Kreps et al., 1982; Selten, 1978). As the literature developed, however, Wilson (1985) cautioned that it was 'too easy' to explain behavior as reputation building, and that the growing suite of reputation models could better be viewed as tests of consistency rather than positive explanations. In the other direction, backward induction has its origins in dynamic programing as a normative model of decision making, while early applications of backward induction in industrial organization were positive in nature. More recently, models of backward induction and the common knowledge of rationality are more appropriately viewed as analytical models (Aumann, 1995) or critique (Binmore, 1997). In addition, as noted in Section 4, empirical challenges to the predictions of simple models of the ultimatum game prompted these models to be viewed as theoretical cases rather than positive models. More generally, the notion that economic models are only supposed to investigate the implications of axioms is often mentioned in response to a criticism of the unrealistic nature of economic models (see Gilboa et al., 2019). Perhaps more importantly, much of the discussion of the role of economics models, including Sugden (2009) and Gilboa et al. (2014, 2018), can be viewed as a somewhat apologetic effort to explain why economists appear to be content with models that do not seem to be up to the standards of other sciences.

Next, models that were initially presented as 'straightforward' predictive science can find themselves re-interpreted as methodological contributions. In many cases, this occurs because a paper offers a paradigmatic example of a new type of analysis. For example, Angrist and Krueger (1991) discussed compulsory school attendance, but the importance of the paper lies, in hindsight, mostly in its methodological contribution. Akerlof's (1978) lemons model is couched in terms of the market for used cars, but its lasting contribution is the demonstration that familiar techniques can be combined to extract surprising conclusions about markets with asymmetric information. Alternatively, Akerlof's paper can be viewed as a critique of the view that market equilibria with many agents should be efficient. In the process, it makes a fundamental methodological contribution: asymmetric information is important. Friedman's (1971) study of collusion in oligopoly was one of several early illustrations of trigger strategies in repeated games. While originally viewed as a positive analysis of firm behavior, the ultimate impact of the work was methodological, leading to a rich set of methods for dealing with repeated games.

In each of these cases, the methodological contribution received a boost from the fact that the paper appeared to capture something essential about its originally intended application. In other cases, a paper makes an important methodological contribution while its original application has faded from view. Though the relative importance of signaling and productivity enhancement in education is still a subject of contention and research, the Spence (1973) model did not ultimately revolutionize the way economists think about education. Nonetheless, it remains an influential methodological contribution.

A more curious phenomenon is the opposite shift: models that were viewed by many as methodological contributions when first introduced have subsequently been viewed as economic models. Many thought that the importance of Muth's (1961) paper on rational expectations was that it 'put discipline' on what beliefs should be allowed in economic models, motivated by the plethora of equilibrium outcomes consistent with *some* beliefs. Subsequently, some economists viewed rational expectations from a positive point of view – economic agents can foresee the consequences of a policy and evaluate choices taking that into account (Lucas's island model 1972 is an example.)

Kydland and Prescott (1982) is a similar example. At one time, it was thought to be 'obvious' that business cycles that caused large fluctuations in consumption were inefficient, and a central role of government should be to smooth consumption. Kydland and Prescott showed that in a highly simplified 'representative agent' model, optimal consumption over time did *not* feature smooth consumption. Even if one did not take seriously the model because of its simplicity, it was clear that one had to revisit government's role in the economy. That is a methodological contribution. Beyond this, however, a large subsequent literature treated the model, and numerous variants, as positive economics.

Similarly, the Ricardian equivalence papers of Barro (1974), and especially Bernheim and Bagwell (1988), can be interpreted as highlighting the difficulties of examining economic policy by working with models based on intertemporal optimization on the part of agents with perfect foresight and perfect capital markets. Economic policy turns out to be utterly ineffective in such a setting, despite the observation that policy obviously has an impact on our economy. However, at least some economists embraced such models as positive contributions, arguing for a passive government.

Results in economic theory are sometimes referred to as 'positive' or 'negative,' roughly corresponding to a statement that something will or will not be the case. (The use of 'positive' in this context is different from 'positive' in the sense of 'descriptive,' as opposed to 'normative.') However, a result that appears to be positive can be interpreted as negative, and vice versa. For example, if it is shown that under assumptions a $A_1, \ldots, A_k$ conclusion C follows, the result would be positive if there are few assumptions that are considered to be weak. By contrast, if the assumptions are rather strong and there are many of them, the result may take a negative flavor. Clearly, if the assumptions imply the conclusion, then one cannot conclude that if one of the assumptions fails the conclusion doesn't hold. But readers of the result may tend to think that the converse implication is more likely than they used to think. This is another manifestation of Grice's (1975) principle, in this case applied to mathematical inquiry: because it is commonly accepted that a mathematical result is more elegant if it uses fewer assumptions, a reader of the theorem might rationally assume that, if one of the assumptions $A_1, \ldots, A_k$ could be dropped, the author of the paper would have done it already.[25] This is by no means a trivial assumption. Yet, it makes sense that the reader would consider it more likely that all assumptions are necessary for the conclusion or at least for the proof. Hence, the reader might feel that the result is 'negative,' perhaps being surprised that so many strong assumptions are needed to establish the conclusion.

Whether a result of the type 'A implies C' is read as positive or negative depends on the reader's reference point, as it were: if the conclusion C seemed somewhat special, and the assumptions A are weak, readers might be surprised and believe that C is actually the case more than they used to. If, by contrast, C was assumed natural, and it turns out that rather stringent assumptions A are needed to derive it, readers might take the result as a 'negative' one.

The reader's reference point, and hence the classification of a result as positive or negative depends in turn upon the prevailing wisdom. A result is more likely to be regarded as positive if people expect it to hold or if prevailing research has been intent on establishing the result. A result is more likely to be viewed as negative if the prevailing wisdom is that it would not hold. Myerson and Satterthwaite (1983) showed the impossibility of efficient trade when there is asymmetric information at a time when many researchers were searching for just such a mechanism, giving us an important negative result.[26] The initial folk theorems were positive contributions, confirming results that (as the name suggests) people expected to hold. As folk theorems have expanded to ever-more-esoteric domains, such as community enforcement with private monitoring, they are more likely to be regarded as negative results. The ability to accomplish seemingly impossible goals suggests to many that something is missing from the model.

There is a natural tendency to view positive (in the sense that something will be the case) results as either positive (in the sense of descriptive) or normative and negative results as either analytical, critique, or methodological. For example, some hailed Abreu & Sen's (1991) finding that any social choice function can be virtually implemented as a normative result. Under this view, the virtual implementation result solved the implementation problem, providing a recipe for constructing mechanisms. Others view it as a negative result and as a critique of implementation theory, demonstrating that something must be missing from a theory that produces such prescriptions. Similarly, some view belief-free equilibria in repeated games of private monitoring as a positive result, showing that any feasible, individually-rational payoff profile is a plausible outcome in such games. Others again view this as a negative result, indicating that something is missing from the model.

In some cases, even the author's intended interpretation of a theory or model is unclear. Was McAfee's (1983) analysis of how the American economy would have developed had Columbus not discovered the New World intended merely as humor, or is it a critique of the counterfactual style of analysis common in economic history? Was Blinder's (1974) analysis of the economics of brushing teeth meant to be a satire of the tendency of economists to overreach, or was it meant to be an illustration of the power and scope of economic reasoning? Given the large literature, albeit much of it outside economics, on the determinants and implications of brushing teeth, the latter interpretation may be relevant. Similarly, was Waldfogel's (1993) analysis of the deadweight loss of Christmas meant to be a spoof of economists' tendency to focus on consequentialist models of utility functions incorporating only material rewards, or was it again meant to illustrate the power and scope of economic reasoning? Whatever the original attention, the number of subsequent comments and responses, along with Waldfogel (2009), suggests that the latter interpretation has taken hold.

8. Conclusion

Why do different interpretations of economic models, and the possibility of shifting interpretations, matter? One goal in exploring these distinctions is to help consumers or observers of economic theory understand the seemingly crazy things done by economic theorists. Understanding the importance of models as critiques can pave the way for the use of models in policy analysis, without demanding predictions. Understanding the importance of explanation can help asses which models to use when assessing new problems.

These distinctions can also help economists understand one another. In particular, disagreement over models often reflects different views as to what the models are trying to accomplish (or *should* be trying to accomplish), and adverse reactions to papers sometimes reflects a mistaken view of what the paper is trying to accomplish. A paper meant to be normative or analytical can be inadvertently criticized for failing to be positive. On the other hand, a motivation of a paper as analytical is more convincing if one can argue that the paper makes a methodological contribution.

We also think that understanding these distinctions can make research in economics more effective. Understanding that theories can serve as either rules or cases can serve as a reminder to provide some guidance as to how the theory is to be interpreted and used. Understanding that theories can make methodological contributions can make it less likely that such contributions are neglected. Understanding that theories can have multiple interpretations can encourage researches to explore the applications of their theories, while being clear about the boundaries.

Notes

1. 'The combined assumptions of maximizing behavior, market equilibrium, and stable preferences, used relentlessly and unflinchingly, form the heart of the economic approach as I see it.' (Becker, 1976).
2. Among a legion of critics, Daniel Kahneman and Amos Tversky built Nobel-prize worthy careers on criticism of standard economic theories.
3. There is also less agreement on the operational definitions of these concepts than, say, of concepts such as utility or subjective probability.
4. The most celebrated contender to replace a standard theory is, arguably, Kahneman and Tversky's Prospect Theory. However, this theory has also been modified and refined, and there is no consensus on the functional form that is supposed to be 'the theory.' For an ongoing debate see Bernheim and Sprenger (2020) and Abdellaoui et al. (2020), who criticize the former paper but also admit that Cumulative Prospect Theory has many descriptive failures. Similarly, the semi-hyperbolic model for consumption over time has been criticized for its empirical failures (Benhabib et al., 2010). Finally, many experimental results seem to be too fragile to serve as foundations for robust theories (see, for example, Grimm & Mengel, 2011 on the Ultimatum Game, and Hertwig et al., 2004 on the role of small probabilities).
5. While 'neuroeconomics' is a respectable subfield that is recognized by the *JEL* classification system, it is hardly the method of choice for most papers that appear in top journals, and it receives very little attention in graduate school curricula.
6. We prefer to avoid the term 'paradigm' altogether, as it is used somewhat differently in philosophy of science, in psychology, and economists' everyday parlance. Instead, we use the more cumbersome 'conceptual framework,' a formal structure that doesn't specify the real-life objects to which its entities relate.
7. A conceptual framework provides a language and tools for analysis, but needs an interpretation of formal concepts to be refutable. For example, expected utility maximization is a conceptual framework, but it does not become refutable until one specifies states, outcomes, acts, and permissible arguments of the utility function.Even the weak axiom of revealed preference requires an assumption that various choices, each made in its own set of circumstances, are made with the same preferences. In contrast to our use, the term 'theory' is often used to mean a conceptual framework (as in 'expected utility theory'), or to denote the entire body of work in the field (along the lines of 'economic theory assumes…').
8. Similarly, a team designing an ocean liner might examine blueprints of existing ships – positive models. The next step might be drawing blueprints for the proposed ship – a normative model. The Harland and Wolff shipyard that built the Titanic would then draw chalk cross sections of the proposed vessel on the floor of a cavernous attic, at a scale of 1:1 for widths and 1:4 for lengths, to ensure feasibility – an analytical model.
9. We discuss methodological contributions in Section 5, noting that analytical models are often motivated as making methodological contributions.
10. Formally speaking, a test of consistency allows for rather fanciful assumptions as long as they are within the paradigm. For example, as long as agents are assumed to be Bayesian, one can assign to them any beliefs one wishes. But the practice of economic theory imposes additional restrictions of plausibility, which make the 'mere test of consistency' closer to the notion of 'credibility.'
11. 'A model which took account of all the variation of reality would be of no more use than a map at the scale of one to one.' (Robinson, 1962). This sentiment reappears rather colorfully in Carroll's (1893) *Sylvie and Bruno Concluded* and Borges' (1998) 'On Exactitude in Science.'
12. Fisher (1989) made a similar point with regard to oligopoly theory.

    > …it should be plain that (with or without game theory) the status of the theory of oligopoly is that of exemplifying theory. We know that a lot of different things can happen. We do not have a full, coherent, formal theory of what must happen or a theory that tells us how what happens depends on well-defined, measurable variables.

    More recently, Rodrik (2015) and Aydinonat (2018) emphasize the tendency of economics to rely on a collection of models to make a point rather than insisting on a single 'correct' model.
13. The legend, as told, seems to suggest that the debate was descriptive in nature. Evidently, Savage's own main contribution (Savage, 1954) had a mostly normative flavor. Yet, there is evidence that Savage viewed expected utility maximization as a valid descriptive theory. Indeed, in 1948–1951 he was promoting expected utility maximizaiton under risk as a tool to be used for economic models (see Moscati, 2019); and, if we are to trust the legend, it took him a day to come back with the normative reponse.
14. We do not distinguish here between these two concepts.
15. Fudenberg et al. (2022) explore one way of quantifying this sacrifice.
16. Other disciplines also investigate equivalent models (for example, see Feynman, 2012), but economics stands out for its emphasis on such contributions.
17. By 'making decisions' we refer to the completeness axiom. With a finite set of alternatives, a binary order is complete and transitive if and only if it can be represented by maximization of a utility function.

18. See also Maki (2005) for an early contribution to the duality between models and experiments.
19. This view is clearly stated by Kohlberg and Mertens (1982, footnote 3, p. 1005):

    > We adhere to the classical point of view that the game under consideration fully describes the real situation-that any (pre)commitment possibilities, any repetitive aspect, any probabilities of error, or any possibility of jointly observing some random event, have already been modeled in the game tree…. In principle, in situations where those restrictions are not met, the game tree is just used as a shorthand notation for the rules of a much bigger "extended game" (cf. loc. cit.), and it is the stability of the equilibria of the extended game that has to be analyzed.

20. In that paper we use the term 'theory' for what we dub here a 'conceptual framework,' that is, a claim that becomes refutable (when interpreted positively) only in conjunction with an interpretation, which maps formal mathematical concepts into real-world ones. It is assumed that the set of such mappings that are permissible, or considered reasonable, is exogenously given.
21. This statement should be qualified. When a presumed 'fact' is in too stark a conflict with deeply entrenched beliefs, it is the fact that might be questioned.
22. In the language of the preceding section we will refer to this interpretation as 'methodological,' as it is a statement about the work of economists, not about the economy.
23. Notice that it would be interpreted as 'economics' rather than 'methodology' in terms of the preceding Section. That is, as a critique it makes a statement about what will or will not happen in the economy, rather than about the method that economists should use.
24. Or a philosopher such as David Hume.
25. Grice's principle (Grice, 1975) for natural language states that people opt for the simplest utterance that conveys their message. Thus, a more complex utterance implicitly suggests that the simpler ones do not hold.
26. In a similar vein, Gibbard (1973) and Satterthwaite (1975) showed the impossibility of non-manipulable voting schemes at a time many researchers were looking for such schemes.

Acknowledgments

The final version of this paper was completed after David Schmeidler, colleague, mentor, and dear friend, passed away (March 17, 2022). We are grateful to two anonymous referees and to the editors for comments and references.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

We gratefully acknowledge support from the Investissements d'Avenir ANR -11- IDEX-0003/Labex ECODEC No. ANR -11-LABX-0047 as well as ISF Grant 1443/20, the AXA Chair for Decision Sciences at HEC, the Foerder Institute at Tel-Aviv University, the Sapir Center for Economic Developmemt (Gilboa) and NSF grant SES-1851449 (Postlewaite).

Notes on contributors

Itzhak Gilboa's research focuses on decision under uncertainty, and has worked extensively with David Schmeidler on axiomatic foundations of non-Bayesian decision theory. He has also contributed to complexity in game theory, evolutionary game theory, and social choice. He has published numerous articles and six books, of which three are textbooks in decision theory: "Theory of Decision under Uncertainty", two textbooks "Rational Choice", and "Making Better Decisions". Gilboa holds the AXA Chair of Decision Sciences at HEC, Paris, the Chair of Economic Theory and Decision Theory at Tel-Aviv University. He served in various editorial positions, including a co-editorship of Econometrica, and was elected an International Honorary Member of the American Academy of Arts and Sciences.

Andrew Postlewaite is the Harry P. Kamen Professor of Economics and Professor of Finance at the University of Pennsylvania. Postlewaite Earned his B.A. from Illinois Wesleyan University in 1965 and his Ph.D. in Applied Mathematics from Northwestern University in 1974. He previously was Assistant and Associate Professor at the University of Illinois, and has held visiting positions at Cal Tech, the Federal Reserve Banks of Minneapolis and Philadelphia, Harvard, Paris School of Economics, Princeton, Stanford, University of Toulouse and Yale. He works primarily in economic theory and game theory. He has served as Editor of the International Economic Review, co-editor of Econometrica and founding Editor of American Economic Journal: Microeconomics. He is a Fellow of the Econometric Society and the American Academy for the Arts and Sciences.

Larry Samuelson is the A. Douglas Melamed Professor of Economics at Yale University. Samuelson earned his B.A. from the University of Illinois in 1974 and his Ph. D. in Economics from the University of Illinois in 1978. He has previously held faculty positions at the University of Florida, Syracuse University, Penn State University and the University of Wisconsin. He works in economic theory, with a particular interest in game theory. His areas of specialization include the evolutionary foundations of economic behavior, the theory of repeated games, and nonBayesian models of behavior. He has served as a co-editor of Econometrica. the American Economic Review, and the American Economic Review: Insights, and the president of the Game Theory Society, as well as in various roles in the Econometric Society. He is a Fellow of the Econometric Society and the American Academy for the Arts and Sciences.

David Schmeidler received his Ph.D. in Mathematics from the Hebrew University of Jerusalem under the supervision of Robert Aumann. He has held positions of Professor of Economics, Statistics, and Management at Tel-Aviv University, The Ohio State University, and Reichman University. He has made contributions to the theory of coalitional ("cooperative") game theory, including the concept of the nucleolus, to general equilibrium theory, theory of economic justice, and notably to decision theory in general and to decision under uncertainty in particular. He pioneered non-Bayesian axiomatic decision theory, which has influenced research in many fields of economics. David Schmeidler passed away on March 17, 2022.

References

Abdellaoui, M., Li, C., Wakker, P. P., & Wu, G. (2020). *A defense of prospect theory in Bernheim & Sprenger's experiment* (Working Paper).

Abreu, D., & Sen, A. (1991). Virtual implementation in Nash equilibrium. *Econometrica*, *59*(4), 997–1021. https://doi.org/10.2307/2938171

Afriat, S. N. (1967). The construction of utility functions from expenditure data. *International Economic Review*, *8*(1), 67–77. https://doi.org/10.2307/2525382

Akerlof, G. (1978). The market for "Lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, *84*(3), 488–500. https://doi.org/10.2307/1879431

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'ecole americaine. *Econometrica*, *21*(4), 503–546. https://doi.org/10.2307/1907921

Angrist, J. D., & Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, *106*(4), 979–1014. https://doi.org/10.2307/2937954

Arrow, K. J. (1950). A difficulty in the concept of social welfare. *Journal of Political Economy*, *58*(4), 328–346. https://doi.org/10.1086/256963

Aumann, R. J. (1985). What is game theory trying to accomplish? In K. Arrow, & S. Honkapohja (Eds.), *Frontiers of economics* (pp. 5–46). Basil Blackwell.

Aumann, R. J. (1987). Correlated equilibrium as an expression of Bayesian rationality. *Econometrica*, *55*(1), 1–18. https://doi.org/10.2307/1911154

Aumann, R. J. (1995). Backward induction and common knowledge of rationality. *Games and Economic Behavior*, *8*(1), 6–19. https://doi.org/10.1016/S0899-8256(05)80015-6

Aydinonat, N. E. (2018). The diversity of models as a means to better explanations in economics. *Journal of Economic Methodology*, *25*(3), 237–251. https://doi.org/10.1080/1350178X.2018.1488478

Barro, R. (1974). Are government bonds net wealth? *Journal of Political Economy*, *82*(6), 1095–1117. https://doi.org/10.1086/260266

Becker, G. (1976). *The economic approach to human behavior*. University of Chicago Press.

Benhabib, J., Bisin, A., & Schotter, A. (2010). Present-bias, quasi-hyperbolic discounting, and fixed costs. *Games and Economic Behavior*, *69*(2), 205–223. https://doi.org/10.1016/j.geb.2009.11.003

Bernheim, B. D., & Bagwell, K. (1988). Is everything neutral? *Journal of Political Economy*, *96*(2), 308–338. https://doi.org/10.1086/261538

Bernheim, B. D., & Sprenger, C. (2020). On the empirical validity of cumulative prospect theory: Experimental evidence of rank-Independent probability weighting. *Econometrica*, *88*(4), 1363–1409. https://doi.org/10.3982/ECTA16646

Binmore, K. G. (1997). Rationality and backward induction. *Journal of Economic Methodology*, *4*(1), 23–41. https://doi.org/10.1080/13501789700000002

Blinder, A. (1974). The economics of brushing teeth. *Journal of Political Economy*, *82*(4), 887–891. https://doi.org/10.1086/260243

Borges, J. L. (1998). *On exactitude of science*. Penguin Books.

Box, G. E. P. (1979). Robustness in the strategy of scientific model building. In R. L. Launer, & G. N. Wilkinson (Eds.), *Robustness in statistics* (pp. 201–236). Academic Press.

Carroll, L. (1893). *Sylvie and bruno concluded*. Macmillan and Co.

Cartwright, N. (2010). Models: Parables vs. fables. In R. Frigg, & M. Hunter (Eds.), *Beyond mimesis and convention: Representation in art and science*. Springer.

Dawkins, R. (1976). *The selfish gene*. Oxford University Press.

Dowe, D. L., Gardner, S., & Oppy, G. R. (2007). Bayes not bust! Why simplicity is no problem for Bayesians. *The British Journal for the Philosophy of Science*, *58*(4), 709–754. https://doi.org/10.1093/bjps/axm033

Feynman, R. (2012). *Feynman: Knowing versus understanding* [Video]. YouTube. https://youtu.be/NM-zWTU7X-k

Fisher, F. M. (1989). Games economists play: A noncooperative view. *The RAND Journal of Economics*, *20*(1), 113–124. https://doi.org/10.2307/2555655

Friedman, J. (1971). A noncooperative view of oligopoly. *International Economic Review*, *12*(1), 106–122. https://doi.org/10.2307/2525500

Fudenberg, D., Kleinberg, J., Liang, A., & Mullainathan, S. (2022). Measuring the completeness of economic models. *Journal of Political Economy*, *130*(4), 956–990. https://doi.org/10.1086/718371

Gallagher, N. (2005). *Gerard Debreu, 1983 nobel prize winner and UC Berkeley professor emeritus, dies in Paris at age 83*. https://www.berkeley.edu/news/media/releases/2005/01/05_debreu.shtml

Gibbard, A. (1973). Manipulation of voting schemes: A general result. *Econometrica*, *41*(4), 587–601. https://doi.org/10.2307/1914083

Gilboa, I., Postlewaite, A., Samuelson, L., & Schmeidler, D. (2014). Economic models as analogies. *Economic Journal*, *124*(578), F513–F533. https://doi.org/10.1111/ecoj.12128

Gilboa, I., Postlewaite, A., Samuelson, L., & Schmeidler, D. (2018). Economics: Between prediction and criticism. *International Economic Review*, *59*(2), 367–390. https://doi.org/10.1111/iere.2018.59.issue-2

Gilboa, I., Postlewaite, A., Samuelson, L., & Schmeidler, D. (2019). What are axiomatizations good for? *Theory and Decision*, *86*(3–4), 339–359. https://doi.org/10.1007/s11238-018-09685-1

Gilboa, I., Postlewaite, A., Samuelson, L., & Schmeidler, D. (2021). Economic theory: Economics, methods and methodology. *Revue Économique*, *73*(6), 891–913.

Gould, S. J., & Lewontin, R. C. (1979). The spandrels of san marco and the panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London Series B, Biological Sciences*, *205*(1161), 581–598. https://doi.org/10.1098/rspb.1979.0086

Grice, P. (1975). Logic and conversation. In P. Cole, & J. Morgan (Eds.), *Syntax and semantics. 3: Speech acts* (pp. 41–58). Academic Press.

Grimm, V., & Mengel, F. (2011). Let me sleep on it: Delay reduces rejection rates in ultimatum games. *Economics Letters*, *111*(2), 113–115. https://doi.org/10.1016/j.econlet.2011.01.025

Grune-Yanoff, T., & Schweinzer, P. (2008). The roles of stories in applying game theory. *Journal of Economic Methodology*, *15*(2), 131–146. https://doi.org/10.1080/13501780802115075

Guth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, *3*(4), 367–388. https://doi.org/10.1016/0167-2681(82)90011-7

Hausman, D. M. (1992). *The inexact and separate science of economics*. Cambridge University Press.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, *15*(8), 534–539. https://doi.org/10.1111/j.0956-7976.2004.00715.x

Keynes, J. M. (1936). *The general theory of employment, interest and money* (14th ed., 1973). Macmillan.

Kohlberg, E., & Mertens, J.-F. (1982). On the strategic stability of equilibria. *Econometrica*, *27*(2), 245–252. https://doi.org/10.2307/1912320

Kreps, D. M., Milgrom, P., Roberts, J., & Wilson, R. (1982). Rational cooperation in the finitely repeated prisoner's dilemma. *Journal of Economic Theory*, *27*(2), 245–252. https://doi.org/10.1016/0022-0531(82)90029-1

Kydland, F., & Prescott, E. (1982). Time to build and aggregate fluctuations. *Econometrica*, *50*(6), 1345–1370. https://doi.org/10.2307/1913386

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In I. Lakatos, & A. Musgrave (Eds.), *Criticism and the growth of knowledge* (pp. 91–196). Cambridge University Press.

Lucas, R. (1972). Expectations and the neutrality of money. *Journal of Economic Theory*, *4*(2), 103–124. https://doi.org/10.1016/0022-0531(72)90142-1

Maki, U. (2005). Models are experiments, experiments are models. *Journal of Economic Methodology*, *12*(2), 303–315. https://doi.org/10.1080/13501780500086255

McAfee, R. P. (1983). American economic growth and the voyage of Columbus. *The American Economic Review*, *73*(4), 735–740. https://www.jstor.org/stable/1816570

Milgrom, P. R., & Stokey, N. (1982). Information, trade, and common knowledge. *Journal of Economic Theory*, *26*(1), 17–27. https://doi.org/10.1016/0022-0531(82)90046-1

Moscati, I. (2019). *Measuring utility: From the marginal revolution to behavioral economics*. Oxford University Press.

Muth, J. (1961). Rational expectations and the theory of price movements. *Econometrica*, *29*(3), 315–335. https://doi.org/10.2307/1909635

Myerson, R. B., & Satterthwaite, M. A. (1983). Efficient mechanisms for bilateral trading. *Journal of Economic Theory*, *29*(2), 265–281. https://doi.org/10.1016/0022-0531(83)90048-0

Nash, J. F. (1951). Non-cooperative games. *Annals of Mathematics*, *54*(2), 286–295. https://doi.org/10.2307/1969529

Piccione, M., & Rubinstein, A. (2007). Equilibrium in the jungle. *The Economic Journal*, *117*(522), 883–896. https://doi.org/10.1111/j.1468-0297.2007.02072.x

Robinson, J. (1962). *Essays in the theory of economic growth*. Palgrave Macmillan.

Rodrik, D. (2015). *Economics rules: Why economics works, when it fails, and how to tell the difference*. Oxford University Press.

Satterthwaite, M. (1975). Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, *10*(2), 187–217. https://doi.org/10.1016/0022-0531(75)90050-2

Savage, L. J. (1954). *The foundations of statistics*. Dover.

Selten, R. (1978). The chain store paradox. *Theory and Decision*, *9*(2), 127–159. https://doi.org/10.1007/BF00131770

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, *25*(3), 289–310. https://doi.org/10.1214/10-STS330

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*(1), 99–118, https://doi.org/10.2307/188485232–72.

Simon, H. A. (2001). Science seeks parsimony, not simplicity: Searching for pattern in phenomena. In *Simplicity, inference and modelling: Keeping it sophisticatedly simple* (pp. 32–72). Cambridge University Press.

Spence, A. M. (1973). Job market signaling. *Quarterly Journal of Economics*, *87*(3), 355–375. https://doi.org/10.2307/1882010

Sugden, R. (2000). Credible worlds: The status of theoretical models in economics. *Journal of Economic Methodology*, *7*(1), 1–31. https://doi.org/10.1080/135017800362220

Sugden, R. (2009). Credible worlds, capacities and mechanisms. *Erkenntnis*, *70*(1), 3–27. https://doi.org/10.1007/s10670-008-9134-x

Waldfogel, J. (1993). The deadweight loss of Christmas. *American Economic Review*, *83*(5), 1328–1336.

Waldfogel, J. (2009). *Scroogenomics: Why you shouldn't buy presents for the holidays*. Princeton University Press.

Wilson, R. (1985). Reputations in games and markets. In A. Roth (Ed.), *Game theoretic models of bargaining with incomplete information* (pp. 27–62). Cambridge University Press.