INTERNATIONAL ECONOMIC REVIEW

May 2018
Vol. 59, No. 2

DOI: 10.1111/iere.12273

ECONOMICS: BETWEEN PREDICTION AND CRITICISM*

BY ITZHAK GILBOA, ANDREW POSTLEWAITE, LARRY SAMUELSON, AND DAVID SCHMEIDLER[1]

HEC, Paris-Saclay, France, and Tel-Aviv University, Israel; University of Pennsylvania, U.S.A.; Yale University, U.S.A.; Tel-Aviv University, Israel

We suggest that one way in which economic analysis is useful is by offering a critique of reasoning. According to this view, economic theory may be useful not only by providing predictions, but also by pointing out weaknesses of arguments. It is argued that when a theory requires a nontrivial act of interpretation, its roles in producing predictions and offering critiques vary in a substantial way. We offer a formal model in which these different roles can be captured.

If you put two economists in a room, you get two opinions, unless one of them is Lord Keynes, in which case you get three opinions. (attributed to Winston Churchill)

1. INTRODUCTION

One view of economics is that it is a predictive science, with the goal of generating predictions that can be tested by observations. This view has many variants. It can refer to the standard conceptualization of models as approximations of reality, as is often the case in the natural sciences. It can focus on qualitative instead of quantitative predictions. It can refer to predictions that are generated by analogies instead of by general rules. It can focus on notions of explanation and understanding, which may lead to predictions in a yet unspecified way. Indeed, the recent literature on the methodology of economics has offered a variety of ways in which economic analysis in general, and economic models in particular, can be understood and used. (See Gibbard and Varian, 1978; Aumann, 1985; McCloskey, 1985; Hausman, 1992; Maki, 1994, 2005; Cartwright, 1998, 2010; Sugden, 2000; Rubinstein, 2006; Grune-Yanoff and Schweinzer, 2008; Grune-Yanoff, 2009; and Gilboa et al., 2014, among others.)

It is easy to find arguments that economics has achieved great success (e.g., Litan, 2014) as well as great failure (e.g., Desai, 2015) in prediction. But critics claim that much of economics fails to generate *any* predictions. In this article, we argue that there is another view of economics as a useful academic discipline, which has merit even in cases where it fails to commit to

*Manuscript received February 2016; revised November 2016.

[1] An earlier version of this article circulated under the title "A Model of Modeling." We are thankful to many colleagues and friends with whom we have discussed the issues addressed here over the years and to two referees for their comments and questions. We also thank Eva Gilboa-Schechtman for conversations that partly motivated and greatly contributed to this work. We gratefully acknowledge ISF Grant 204/13 (Gilboa and Schmeidler), ISF Grant 704/15 and ERC Grant 269754 (Gilboa), and NSF Grants SES-0961540 (Postlewaite) and SES-1459158 (Samuelson). Please address correspondence to: Andrew Postlewaite, Economics Department, University of Pennsylvania, 3718 Locust Walk 424 McNeil, Philadelphia, PA 19104-6297. E-mail: *apostlew@econ.upenn.edu*.

specific predictions. This view suggests that one role of economics is to critique reasoning about economic questions. If, for example, the government intends to increase the tax rate on a certain good and expects a certain revenue based on the tax rate and the *current* volume of trade in the respective market, one would do well to mention that the quantity of the good demanded is likely to change as a result of the tax. Such a comment would fall short of calculating the actual tax revenue expected. Indeed, an economist may find it difficult to calculate the elasticity of demand, let alone the general equilibrium effects of such a tax. Nonetheless, it might be extremely valuable to identify the fallacy in the naive reasoning based on the current quantity demanded.

Economic analysis may thus be useful simply as a form of criticism. It may critique reasoning at the very basic level of testing logical deductions; at a conceptual level, say, by identifying equilibrium effects (as in the example above); and in other ways, such as confronting intuition with specific models or with empirical findings.[2]

The distinction between critique and qualitative predictions is not always clear. In the tax example, by pointing out that the quantity demanded is likely to respond to a price hike, the economist may make the implicit prediction that tax revenue will be lower than calculated based on current quantity demanded. But the economist might also be aware of various anomalies for which demand curves might be upward-sloping, or for which the general equilibrium effect of taxation might be qualitatively different from its partial equilibrium effect. Thus, she may restrict her claim to the critique of a proposed line of reasoning without venturing even a qualitative prediction.

What distinguishes the use of economic analysis for prediction and for critique? One obvious distinction has to do with the context in which they appear: A critique comes in response to an existing statement or prediction, whereas a prediction need not have such a predecessor. Or, put differently, a prediction can be viewed as a reply to an open-ended question, "What is likely to occur?" as opposed to a critique, which can be thought of as a reply to a yes/no question, "Does the following argument apply?"

There is also a structural distinction between prediction and critique, which is not context-dependent. Economic analysis allows a particular situation to be modeled in different ways. It is often the case that more than one formal model can be viewed as a plausible abstraction of a given reality. One can imagine situations where, according to all such abstractions, a certain outcome follows, as well as situations where there are some such abstractions, and finally situations where there are no such abstractions. We will say that a theory predicts an outcome if, according to all plausible abstractions, that outcome follows. A theory critiques an argument if it shows that it only holds under certain abstractions, or, in the extreme case, that no plausible abstraction supports it.

For example, imagine an experimental ultimatum game. In one abstraction, it can be modeled as a game with monetary payoffs only, where game theory predicts low offers to be made and accepted. In another abstraction, it can be modeled as a game in which utility functions are also affected by, say, pride or fairness considerations. Game theory per se will therefore not make a unique prediction in this case. However, it may be used to critique a prediction that is based on monetary payoffs only.

In Section 2, we offer a simple model in which this distinction can be made precise. Our model makes reference to models and to theories. Our analysis is most obviously applicable to economic analysis that is based on explicit models, including most prominently work in economic theory. However, we believe that virtually all of economic analysis is based on models, even if the latter are often left unspecified and must be inferred from the context. We thus view our analysis as being applicable to economics in general, with the first step on the path to a critical assessment often being to make clear the implicit underlying model. Indeed, models appear throughout the sciences, and we view our analysis as being applicable to scientific research more generally. We explain as we proceed why we think our analysis is often especially relevant for economics.

[2] As such, the role of economists can be viewed as helping society reach rational policies in Habermas's (1984) sense of "communicative rationality."

We use the model in Section 3 to prove two results about the complexity of testing whether a theory predicts an outcome or, conversely, critiques it. Section 4 then discusses several applications of our model to questions in the methodology of economics and decision sciences. Section 5 concludes.

2. A MODEL OF ECONOMIC MODELING

Our objective is to develop a model of economic modeling within which the discussion above can be made more precise. As is often the case with modeling in economic theory, our goal is not to provide the most detailed and accurate account of the reality in question. Rather, we wish to abstract away from many details that seem secondary to the discussion in Section 1, and focus on what appears to be the crux of the issue. One benefit of the simplified and highly idealized model is that it will be sufficiently abstract to point out some analogies between seemingly unrelated distinctions. Indeed, Section 4 argues that the formal model can capture not only the distinction between economics as a predictive science and as a criticism, but also three other distinctions related to economic and decision theory.

Importantly, this is a theoretical article that does not aim to perform the empirical analysis of the way economists think about their work. We make no claims about the prevalence of the different interpretations of economic analysis or about what their relative importance is or should be. The article provides the conceptual framework for such analysis, but it makes neither descriptive not normative claims about the state of the art.

We present the components of this model in Subsections 2.1–2.5. We complete the model in Subsection 2.6.

2.1. *Descriptions.* The first component of our model of economic modeling is a description. A description can be thought of as a list of statements in a given language, describing a certain state of affairs. Importantly, descriptions are used both for the formal entities in an economist's model and for their informal interpretation in terms of real entities. For example, if an economist describes a bank run phenomenon as a coordination game, she will have a model with "players," "strategies," and "payoffs." This model might consist of a game that includes agents $P1$ and $P2$ who can either withdraw their money, W, or leave it, L, with the outcomes as in the following table:[3]

	P2	
P1	L	W
L	10, 10	0, 8
W	8, 0	2, 2

The economist's interpretation of this model will typically refer to various features of reality, such as investors who may or may not leave their money in banks. In our model of economic modeling, this interpretation will be captured by a mapping between two descriptions—one of the entities in the economist's model (such as "players") and one of the various real entities to which she refers (such as "investors").

A description begins with a finite set of entities E. The set E can be thought of as a set of letters in an alphabet. For the bank run model, the entities include $P1$, $P2$, L, and W.[4] Formally, we define the set of possible lists of entities of a model, E^*, as

$$E^* = \cup_{k\geq 1} E^k,$$

and let a typical element of E^* be denoted by e.

[3] The relationship to a bank run is that an agent contemplates withdrawing her money only because of uncertainty as to whether the other agent might.

[4] This is not the complete set of entities for the bank run example but illustrates the basic idea. We describe in more detail several examples below.

The next component of a description is a finite set of predicates F, with a typical element denoted by f. Like E, the set F is a formal set of letters (disjoint from E). Predicates are used to attribute properties to entities or to establish relationships between entities. In the economist's bank run model, $P1$ and $P2$ are *players in a game* and L and W are *actions* the players may take. The description of the bank run model would include predicates *players in a game* and *actions*. Predicates will come in different varieties, referred to as k-place predicates for various values of k. A 1-place predicate will indicate whether an entity has a particular property; for example, predicate f might be used to designate $P1$ a *player in a game.* A 2-place predicate will describe relationships between pairs of entities. A 3-place predicate will describe relationships between triples of entities and so on.

A description will link entities with predicates: $P1$ is a player in the game, W is an action a player can take, etc. Formally, a *description* is a triple $\mathbf{d} \equiv (E, F, d)$, where E is a set of entities, F is a set of predicates, and d is a function[5]

$$d : E^* \times F \to \{0, 1, \circ, *\}.$$

For a k-tuple $e \in E^k$ and a predicate f, $d(e, f)$ is intended to say whether the predicate f applies to the list of entities e. The values 1 and 0 are naturally intended to capture *true* and *false*. If $P1, P2, W, L \in E$, and *player in a game* and *action* are predicates, $d(P1, \textit{player in a game}) = 1$ and $d(P1, \textit{action}) = 0$. The value $*$ is used for unknown values, and it will allow us to start with a description that is partial and augment it by additional facts. Finally, the special value $\circ$ is interpreted as saying that the question of whether f applies to e is not meaningful. The value $\circ$ allows us to describe all (finite-place) predicates with a single function d instead of having to consider sets of predicates with a varying number of places. In particular, the value $\circ$ will be used when the number of places in the predicate f differs from the number of entities k. For example, we have $f(e) = \circ$ for any $e \in E^2$ whenever f is a 1-place predicate. More generally, for every $f \in F$, there exists a unique $m \geq 1$ such that, for every $k \neq m$, every $e \in E^k$, and every d, we have $d(e, f) = \circ$. This $m = m(f)$ will be called the degree of f.[6]

2.1.1. *Compatibility.* We will be interested in "extending" a description by adding additional structure to it. Toward that end, we define the notion of compatibility of descriptions.

Two descriptions $\mathbf{d} = (E, F, d)$ and $\mathbf{d}' = (E, F, d')$ are *compatible* if

(i) for every $e \in E^*$ and every $f \in F$,

$$d(e, f) = \circ \qquad \Longleftrightarrow \qquad d'(e, f) = \circ$$

and

(ii) for every $e \in E^*$, every $f \in F$, and every $x, y \in \{0, 1\}$, if

$$d(e, f) = x \qquad \text{and} \qquad d'(e, f) = y,$$

then $x = y$.[7]

[5] It may often be convenient to expand the range of the function d to a set $X = X_0 \cup \{\circ, *\}$, where X_0 contains more elements than simply $\{0, 1\}$. This may, in particular, allow quantitative theories to be discussed more elegantly. However, it will suffice for our discussion to take the range of X to be $\{0, 1, \circ, *\}$.

[6] We implicitly assume therefore that there is a degree of commonality in the use of language so that the same predicate f will not be used by different descriptions to have a different number of places.Observe also that we do not insist that all m-tuples of entities be meaningful for a predicate of degree m. For example, it makes sense to ask whether a player prefers one outcome to another, but not whether she prefers an outcome to a strategy, or a player to a player.

[7] Here and in the sequel, we refer to strict equalities and universal quantifiers when defining our concepts. Naturally, this is an idealization. In a more realistic model, strict equalities should be replaced by sufficiently good approximations and universal quantifiers by some statistical measures, as in the concept of "Probably Approximately Correct."

Thus, compatible descriptions agree on which statements are meaningful (condition (i)), and among these, they cannot assign to the same statement two incompatible values in $\{0, 1\}$ (condition (ii)). However, compatibility allows the descriptions to differ if one of them assumes the value $*$ and the other takes some value $x \in \{0, 1\}$. Equivalently, the value $*$, designating an unknown value in $\{0, 1\}$, is considered to be compatible with any value in $\{0, 1\}$. In this sense, one can "extend" a description that is silent on some aspect to specify that aspect.

2.1.2. *Extensions.* For two descriptions $\mathbf{d} = (E, F, d)$ and $\mathbf{d}' = (E, F, d')$, we say that $\mathbf{d}'$ is *an extension* of $\mathbf{d}$, denoted by $\mathbf{d}' \rhd \mathbf{d}$, if

(i) for every $e \in E^*$ and every $f \in F$,

$$d(e, f) = \circ \qquad \Longleftrightarrow \qquad d'(e, f) = \circ$$

and

(ii) for every $e \in E^*$, every $f \in F$, and every $x \in \{0, 1\}$,

$$d(e, f) = x \qquad \Longrightarrow \qquad d'(e, f) = x.$$

Thus, a description $\mathbf{d}'$ extends a description $\mathbf{d}$ if the two agree on what is and what is not a meaningful statement (condition (i)), and, for all meaningful statements, any statement that appears in $\mathbf{d}$ appears in $\mathbf{d}'$ as well (condition (ii)). As an illustration, consider an initial model of bank runs whose description includes two outcomes o_1 and o_2 and a predicate g that indicates preference. The description $\mathbf{d}$ of this model might have $d(P1, o_1, o_2, g) = *$ (meaning that it is unknown whether $P1$ prefers o_1 to o_2). An extension $\mathbf{d}'$ might have $d'(P1, o_1, o_2, g) = 1$ (meaning that at d' it is known that $P1$ prefers o_1 to o_2).

Clearly, if $\mathbf{d}' \rhd \mathbf{d}$, then $\mathbf{d}$ and $\mathbf{d}'$ are compatible. However, compatibility allows d' to assume a specific value $d'(e, f) = x$ for $x \in \{0, 1\}$, whereas d is silent about it, that is, $d(e, f) = *$, as well as vice versa. By contrast, for $\mathbf{d}'$ to be an extension of $\mathbf{d}$, only the former is allowed: d' specifies more values in $\{0, 1\}$ than does d.

In other words, the extension relation is a subset of the compatibility relation, defined by having a smaller set of $*$ values (relative to set inclusion). Thus, if $\mathbf{d}' \rhd \mathbf{d}$, we will also refer to $\mathbf{d}'$ as "larger" than $\mathbf{d}$, in the sense that

$$d^{-1}(0) \cup d^{-1}(1) \quad \subset \quad d'^{-1}(0) \cup d'^{-1}(1).$$

(And, if $\mathbf{d}$ and $\mathbf{d}'$ are compatible, the above inclusion is equivalent to $\mathbf{d}' \rhd \mathbf{d}$). We can also define a "minimal extension" of a description $\mathbf{d}$ that satisfies a certain property, where, again, minimality will be with respect to set inclusion applied to $d^{-1}(0) \cup d^{-1}(1)$.

2.2. *Models.* The definitions above hold for any two disjoint sets E, F. However, we will henceforth save the notations E, F for the sets of entities and of predicates (respectively) in the economist's model. We similarly will want descriptions of real-world problems we confront and hope to understand better with the use of economists' models. For convenience, when we discuss descriptions of reality, we will introduce new sets of entities and of predicates. We will thus refer to a description $\mathbf{d} \equiv (E, F, d)$, defined for the formal entities and predicates, E and F, as a model.

When dealing with a model $\mathbf{d} = (E, F, d)$, the sets E or F denote the abstract notation in the economist's model. They can therefore consist of any mathematical symbols, though, naturally, economists tend to use mnemonic notation that suggests certain interpretations. For example, the set of alternatives that a decision maker can choose might be referred to as a "strategy set" and a typical element thereof as "s."

EXAMPLE 1 (THE DICTATOR GAME). We use the dictator game as a running example. Our description of this game would have the set of entities

$$E = \{P1, P2, 100, \ldots, 0, (100;0), \ldots, (0;100)\},$$

and the set of predicates

$$F = \{\mathit{Player}, \mathit{Strategy}, \mathit{Outcome}, \mathit{Result}, \succsim, \mathit{May}\}.$$

The single-place predicate *Player* is designed to convey the information that an entity is (or is not) a player. A description $\mathbf{d}^1 = (E, F, d^1)$ will satisfy

$$d^1(P1, \mathit{Player}) = 1,$$
$$d^1(P2, \mathit{Player}) = 1,$$

and, for any other $e \in E$, $d^1(e, \mathit{Player}) = 0$, indicating that it is meaningful to ask whether such entities are players, but that they are not. For any $e \in E^k$ with $k > 1$, we have $d^1(e, \mathit{Player}) = \circ$, indicating that it is not meaningful to ask whether a pair or triple (or so on) of entities is a player.

The possible choices available to $P1$ are given by the two-place predicate *Strategy*:

$$d^1(P1, 100, \mathit{Strategy}) = 1$$
$$\vdots$$
$$d^1(P1, 0, \mathit{Strategy}) = 1,$$

whereas $d^1(P2, e, \mathit{Strategy}) = 0$ for all $e \in E$ indicates that $P2$ is a dummy player in the game. Next, we define preferences, as in

$$d^1((P1, (100;0), (99;1)), \succsim) = 1,$$
$$d^1((P1, (99;1), (100;0)), \succsim) = 0,$$

indicating that $P1$ strictly prefers the allocation $(100;0)$ to $(99;1)$. Similar statements for any other pair of outcomes indicate that $P1$ always prefers the split that gives him more. Explicitly,

$$d^1(P1, (n, 100-n), (m, 100-m)), \succsim) = 1 \text{ if } n \geq m,$$
$$d^1(P1, (n, 100-n), (m, 100-m)), \succsim) = 0 \text{ if } n < m.$$

We have $d^1(P1, \succsim) = \circ = d^1(P1, (99, 1), \succsim)$, indicating that it is meaningless to ask how the preference relation $\succsim$ ranks $P1$, and that it is meaningless to ask whether a player prefers an outcome without asking to what it is to be compared.

Similarly, *Outcome* is a 1-place predicate indicating which entities are outcomes, and *Result* is a 2-place predicate indicating, for every player-1 strategy and outcome, whether the outcome is the result of the strategy. For the moment we let $d^1(e, \mathit{May}) = *$ for all e, indicating that d^1 says nothing about the predicate *May*, which is meant to capture the predictions of the model. We refer to this description as $\mathbf{d^1}$.

EXAMPLE 2 (THE DICTATOR GAME, CONTINUED). The description $\mathbf{d}^1$ makes no statement as to the outcome of the game. This would be appropriate if the goal of the description is simply to present the game. We consider here a description $\mathbf{d}^2 = (E, F, d^2)$ that comments on outcomes.

Outcomes are described by the predicate *May*. For example, an assertion that the dictator will split the surplus evenly would have

$$d^2((50, 50), May) = 1,$$

with $d^2(e, May) = 0$ for all other outcomes e (i.e., all other e with $d^1(e, Outcome) = 1$). Similarly, a description can let *May* have the value 1 for more than one outcome, indicating that more than one outcome is possible. This will be useful for solution concepts that do not choose a unique outcome, including those that are simply agnostic about some outcomes. Observe that the description $\mathbf{d}^2$ is compatible with $\mathbf{d}^1$ and it is an extension of $\mathbf{d}^1$.

EXAMPLE 3 (THE DICTATOR GAME, CONTINUED). There are many other extensions of $\mathbf{d}^1$. For example, an assertion that the dictator will keep all the surplus would have

$$d^3((100, 0), May) = 1,$$

with $d^3(e, May) = 0$ for other outcomes e. This gives rise to a description $\mathbf{d}^3$ that is compatible with and extends $\mathbf{d}^1$, but is neither compatible with nor extends $\mathbf{d}^2$.

2.3. *Descriptions of Reality.* Typically, a scientific paper will not formally describe reality as separate from its model—indeed, the model is often taken to be the formal description of reality. However, since our goal here is to model the act of modeling, we need to treat the reality that the economist considers as a formal object separate from the model that she constructs. To this end, we introduce new notation for the sets for the entities and predicates used in descriptions of real-world problems.

Assume, then, a set of entities E_R that are supposed to capture the objects in the "real world" being modeled. For example, analyzing an international crisis, the United States and Russia might be such entities. Although the economist will refer to these countries informally, say, in the introduction of her paper, we will use formal elements, *United States* and *Russia*, as members of the set E_R. More generally, entities in E_R could be thought of as objects that are referred to in a daily newspaper. To avoid confusion, we will assume that real objects are distinct from formal ones, that is, that $E \cap E_R = \varnothing$.

Reality is described by a set of predicates, F_R (where, as in the case of the formal language, we assume $F_R \cap E_R = \varnothing$). To avoid confusion, we also use a different set of predicates for reality and for the model: $F \cap F_R = \varnothing$.[8] There might be cases in which one would be tempted to use the same predicate in describing both the real world and its model. For example, if we wish to state the fact that the unemployment rate has increased, the term "increase" is a natural choice both for the description of reality and in the model. However, there are cases where the mapping between real and formal predicates is far from clear. For example, when modeling a political science problem as a game, it is not always obvious who the players are: countries or their leaders. Similarly, in many decision problems, one has a choice regarding the definition of an outcome: A (final) outcome in one model may be an act with uncertain results in another. We therefore assume that the sets of predicates F and F_R are disjoint.

Given a set of entities E_R and predicates F_R, facts that are known about reality will be modeled by a description $\mathbf{d}_R = (E_R, F_R, d_R)$. We refer to $\mathbf{d}_R$ as a description of reality.

EXAMPLE 4 (THE DICTATOR GAME, CONTINUED). A description of a real dictator game would refer to specific people who interact in the game, perhaps in a laboratory experiment. The set of entities could be

$$E_R = \{Mary, John, \$100, \ldots, \$0, (\$100; \$0), \ldots, (\$0, \$100)\},$$

[8] To be precise, we assume that $(E \cup E_R) \cap (F \cup F_R) = \varnothing$, so that the four sets are pairwise disjoint.

with the set of predicates

$$F_R = \{Participant, Keeps, Allocation, Result_R, Prefers, Possible\_Outcome\}.$$

The sets of entities and of predicates are clearly in 1 to 1 correspondence with the respective sets in the formal model. The reality in the laboratory experiment is that Mary and John are the participants, that Mary can choose how many dollars to keep, and so forth. Note that, as the word "results" seems to be natural both in reality and in the formal model, we use a predicate $Result_R \in F_R$ that is formally distinct from $Result \in F$.

2.4. *Abstractions and Interpretations.* We now turn to the relationship between reality and the model. The economist's conceptualization of the problem is viewed as including

(i) a description of reality, $\mathbf{d}_R = (E_R, F_R, d_R)$;
(ii) a model $\mathbf{d} = (E, F, d)$;
(iii) a pair of functions (ϕ_E, ϕ_F) such that:
- $\phi_E : E_R \to E$ is a bijection;
- $\phi_F : F_R \to F$ is a bijection;
- for every $f_R \in F_R$, the degree of $f = \phi(f_R)$ (according to $\mathbf{d}$) is the same as the degree of f_R (according to $\mathbf{d}_R$);
- for every $e \in E^*$ and $f \in F$, we have $d(e, f) = d_R(\phi_E^{-1}(e), \phi_F^{-1}(f))$, or equivalently, for every $e_R \in E_R^*$ and $f_R \in F_R$, we have $d(\phi_E(e_R), \phi_F(f_R)) = d_R(e_R, f_R)$.

The pair of bijections ϕ_F, ϕ_E will be jointly referred to as ϕ, defined as the union of the ordered pairs in ϕ_F and in ϕ_E. The function ϕ can be thought of as an abstraction of reality, where it takes real-life objects (such as "*USA*") and thinks of them as formal objects in a model ("*Player*1").

Formally, we define an abstraction to include both the domain and the range of these functions. Thus, an *abstraction* is a triple

$$A = (\mathbf{d}_R = (E_R, F_R, d_R), \mathbf{d} = (E, F, d), \phi = \phi_E \cup \phi_F).$$

The insistence that ϕ be a bijection between reality and a model stands in apparent contrast to the idea that a model should be an approximation of reality. However, the bijection is between the model and a *description* or reality. The abstraction occurs in the course of the description, where many presumably unimportant aspects of reality are excluded. Those aspects that remain are retained because they are thought to be important. The bijection ensures that these are mapped into corresponding elements in the model and that the model has no unnecessary features.

A common view of science suggests that there is a phenomenon of interest in reality that is modeled by the scientist. Thus, one starts with a description of reality $\mathbf{d}_R$ and looks for an appropriate abstraction $A = (\mathbf{d}_R, \mathbf{d}, \phi)$ to describe it formally. The practice in economic theory is sometimes reversed: It is not uncommon for an economist to come up with a model and to have his peers suggest to him that the formal model has better "real-life examples" than those he has started out with. In this case, the model $\mathbf{d}$ is the starting point, and one looks for a description of reality $\mathbf{d}_R$ that can serve as an example of the model. Hence, we can refer to ϕ^{-1} as the interpretation of the (formal) model.

EXAMPLE 5 (THE DICTATOR GAME, CONTINUED). We begin with the model $\mathbf{d}^1$ constructed in Example 1 and the sets of entities and predicates (E_R, F_R) from Example 4. We extend these building blocks to an abstraction by specifying d_R and ϕ.

The function $\phi = \phi_E \cup \phi_F$ connects predicates in the obvious way:

$$\phi_F(Participant) = Player$$

$$\phi_F(Keeps) = Strategy$$
$$\phi_F(Allocation) = Outcome$$
$$\phi_F(Prefers) = \succsim$$
$$\phi_F(Possible\_Outcome) = May.$$

The entities might be usefully connected in more than one way. If we assume that Mary is the dictator, then we have

$$\phi_E(Mary) = P1$$
$$\phi_E(John) = P2.$$

Similarly,

$$\phi_E(\$100, \$0) = (100, 0)$$
$$\vdots$$
$$\phi_E(\$0, \$100) = (0, 100).$$

We would then ensure that $d_R(e_R, f_R) = d(\phi_E(e_R), \phi_F(f_R))$, and thence that we have an abstraction, by specifying

$$d_R(Mary, Participant) = 1$$
$$d_R(John, Participant) = 1$$
$$d_R(Mary, (\$100; \$0), (\$99, \$1), prefers) = 1,$$

and so on.

As is always the case, there are many ways to construct a model designed to examine a particular situation or answer a particular question. For example, the analyst might exclude the receiver from the analysis. *John* would then be deleted from the set of entities E_R in the description of reality and $P2$ would be deleted from the set of entities E in the formal model. It is also not obvious that people care only about their own monetary payoffs, and hence Mary's preferences (in the description of reality) and $P1$'s preferences (in the model) might be some other ordering over the 101 outcomes. Hence, the analyst must choose from many possible descriptions of reality and many matching models.

Importantly, not all bijections ϕ yield reasonable models. There are mappings that would be contrary to common sense. For example, a real-life entity such as "a state" or "a leader" may be mapped to a theoretical entity named "*Player*1." But an inanimate object such as "money" might not, in most reasonable models.[9]

We will assume that a set of acceptable abstractions is exogenously given. For a given description of reality, $\mathbf{d}_R = (E_R, F_R, d_R)$, denoting a phenomenon of interest, a bijection $\phi = \phi_E \cup \phi_F$ maps (E_R, F_R) onto sets (E, F), generating the abstraction $A = (\mathbf{d}_R, \mathbf{d}, \phi)$. We will denote the set of *acceptable abstractions* for $\mathbf{d}_R$ by $\mathcal{A}(\mathbf{d}_R)$.

As we note in Subsection 5.1, we believe that one characteristic distinguishing economics from the natural sciences is that the latter appear to more readily settle on a relatively narrow set of acceptable abstractions. Within economics, advances often come in the form of identifying new

[9] There are exceptions to this rule. For example, electric current may be modeled as a congestion game, where an electron is mapped to a player.

abstractions to be added to the "acceptable" category. The determination of which abstractions are acceptable is to a large extent a social process. We have nothing to say about this process but consider it eminently worthy of study.

2.5. *Theories.*

2.5.1. *Extending descriptions.* We define a theory as a mapping between models that guarantees extension. To make this precise, let $D(E, F)$ be the set of descriptions $\mathbf{d} = (E, F, d)$ with sets of entities and predicates (E, F). Then, a theory is a function $T : D(E, F) \to D(E, F)$ such that, for all $\mathbf{d} \in D(E, F)$, we have $T(\mathbf{d}) \rhd \mathbf{d}$. Thus, a theory takes an existing model and adds to its information about various predicates. As a descriptive theory, T should be viewed as saying "If $\mathbf{d}$ is the case, then $T(\mathbf{d})$ will also be true." One can think of $\mathbf{d}$ as the question, or the prediction problem, and of $T(\mathbf{d})$ as the answer, or the solution to the problem.

Observe that the notion of an extension allows for the possibility that $T(\mathbf{d}) = \mathbf{d}$, suggesting that the theory adds nothing to $\mathbf{d}$ or is silent about it. This might be the case if the theory is simply irrelevant for the description $\mathbf{d}$, if $\mathbf{d}$ is already fully specified (has no $*$ values) or if the description $\mathbf{d}$ contains information that refutes the theory, thereby rendering its predictions dubious.

Note that the domain of the theory consists only of models (E, F, d) in $D(E, F)$ and not descriptions of reality (E_R, F_R, d_R). Thus, we do not allow theories to make direct reference to proper names in the world. A theory can reflect a statement "If Player 1 . . . , then . . . " but not "If Mary . . . , then . . . "

Although our main interest is in the question of prediction, theories in our model of economic modeling can also be normative, that is, to provide recommendations. When a theory T is interpreted normatively, it can be viewed as saying, "In case one (a person, a society, etc.) is faced with $\mathbf{d}$, then one should do what is specified in $T(\mathbf{d})$." This alternative interpretation is allowed at no additional cost and will be used in Section 4.[10]

EXAMPLE 6 (THE DICTATOR GAME, CONTINUED). Suppose that we are interested in the theory T specifying that players in the dictator game choose the actions corresponding to a subgame perfect equilibrium or, equivalently, that the dictator chooses a utility-maximizing action. We will then have

$$T(\mathbf{d}^1) = \mathbf{d}^3.$$

The model $\mathbf{d}^1$ specifies preferences but says nothing about behavior, and the theory T extends this description by specifying that the dictator will choose her most preferred allocation. By contrast, the model $\mathbf{d}^2$, specified in Example 2, indicates that the dictator splits the surplus evenly, while maintaining the description of preferences given by $\mathbf{d}^1$ (Example 1), and hence is inconsistent with the theory T. We would then have

$$T(\mathbf{d}^2) = \mathbf{d}^2.$$

We will also have

$$T(\mathbf{d}^3) = \mathbf{d}^3,$$

because the description $\mathbf{d}^3$ (Example 3) already describes the utility-maximizing strategy of $P1$, and theory T cannot add any information to it.

[10] Note that the way we refer to a "theory" here might be closer to an everyday usage of the term "paradigm" in economics (though not precisely identical to Kuhn's, 1962, original usage) or to "conceptual framework" (Gilboa and Schmeidler, 2001).

2.5.2. *Rules.* Our model of modeling describes theories in a concrete, extensional manner. For example, our model would describe Nash equilibrium theory by spelling out its predictions in a specific game. If we want this theory to apply to all games, then we would need to list every possible game, along with its Nash equilibrium.

This is clearly an inefficient way of characterizing theories. In practice, a theory is typically described not by listing its implication in every circumstance to which it could be applied but by specifying a rule that allows one to deduce that implication. Moreover, extensional descriptions of theories do not allow us to distinguish between equivalent descriptions of the same theory. Finally, extensional descriptions do not allow us to compare theories in terms of being more or less complex. As a result, it will often be useful to describe theories by a set of rules. This section explains how rules are used to specify theories, and we make important use of rule-based descriptions in the course of proving Propositions 1 and 2.

A rule r specifies an antecedent criterion that a description may (or may not) satisfy and identifies an extension of those descriptions that satisfy the antecedent. Hence, a rule r gives rise to a function (also denoted by r) that associates with any description (E, F, d) an extension $r(E, F, d) = (E, F, d') \rhd (E, F, D)$. If a description (E, F, d) fails the antecedent of the rule r, then $r(E, F, d) = (E, F, e)$.

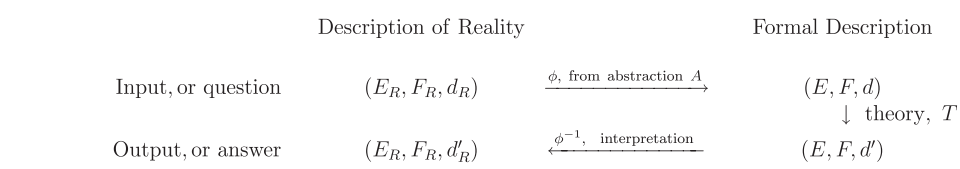
EXAMPLE 7 (BACKWARD INDUCTION). Consider descriptions $\mathbf{d} = (E, F, d)$ that capture finite extensive form games with perfect information and no ties. The Backward Induction solution can be viewed as a theory T that provides a unique prediction for each such game. Clearly, its domain is infinite. Yet, it can be succinctly described by a simple rule. Consider a 2-place predicate, *Predicted*, applying to a node n in the extensive form of the game and an outcome, z, and interpreted as stating that all players predict that, should the node n be reached, the game will evolve to outcome z. With this predicate, we can state a rule that says that if (a description $\mathbf{d}$ implies that) at node n, belonging to player i, each successor of n is predicted (according to *Predicted*) to have a particular outcome, then it is predicted (according to *Predicted*) that i's play at n will select the successor whose predicted outcome maximizes Player i's utility. Applying this rule once generates a description $r(\hat{\mathbf{d}})$ that assigns Predicted values to the nodes before the leaves. Applying this rule inductively generates a description $\mathbf{d}'$ that augments the description of the game $\mathbf{d}$ to have *Predicted* values for all nodes, including the root of the tree, and this prediction will be the Backward Induction solution.

As the Backward Induction example illustrates, when we define a theory by rules, we use them as soon as their antecedents hold, irrespective of the compatibility of the theory with these antecedents. For example, the rule of the backward induction prediction is used in every node of the game tree, including those nodes that will eventually be ruled out by the backward induction theory.

2.6. *Using Models and Theories to Examine Reality.* Suppose that we have a description of reality (E_R, F_R, d_R), from which we would like to draw some conclusions. For example, the description may identify a person, a set of feasible alternatives, and preferences, and our task may be to characterize the person's behavior. Alternatively, the description may specify a game, and our task may be to apply an appropriate notion of equilibrium behavior.

We draw such conclusions as the joint product of an abstraction and a theory. More concretely, we use a description of reality $\mathbf{d}_R = (E_R, F_R, d_R)$, an abstraction $A \in \mathcal{A}(\mathbf{d}_R)$, and a theory T to proceed as follows:

1. Recall that the abstraction $A = ((E_R, F_R, d_R), (E, F, d), \phi)$ defines a model (E, F, d), satisfying $d(e, f) = d_R(\phi^{-1}(e), \phi^{-1}(f))$.
2. The theory T then gives a model (E, F, d') that extends (E, F, d).
3. The inverse of the abstraction, namely, the interpretation of the model, ϕ^{-1}, then gives a description of reality (E_R, F_R, d'_R) that satisfies $d'_R(e_R, f_R) = d'(\phi(e_R), \phi(f_R))$.



NOTES: The point of departure is a description of reality $\mathbf{d}_R = (E_R, F_R, d_R)$. The function ϕ associated with the abstraction A associates a formal description (E, F, d) with this description of reality, with d_R and d related via $d(e, f) = d_R(\phi^{-1}(e), \phi^{-1}(f))$. The theory T then extends the formal description (E, F, d) to a description (E, F, d'). We can then again use the interpretation ϕ^{-1} to find an associated description of reality (E_R, F_R, d'_R), satisfying $d'_R(e_R, f_R) = d'(\phi(e_R), \phi(f_R))$. We refer to $\mathbf{d}'_R = (E_R, F_R, d'_R)$ as the $A-T$ extension of (E_R, F_R, d).

FIGURE 1

HOW A MODEL AND A THEORY ARE USED TO DRAW CONCLUSIONS ABOUT REALITY

Figure 1 illustrates this process. It is a straightforward calculation to verify that $\mathbf{d}'_R$ is indeed an extension of $\mathbf{d}_R$. We refer to (E_R, F_R, d'_R) as the $A-T$ extension of (E_R, F_R, d_R).

3. TESTING AND APPLYING THEORIES

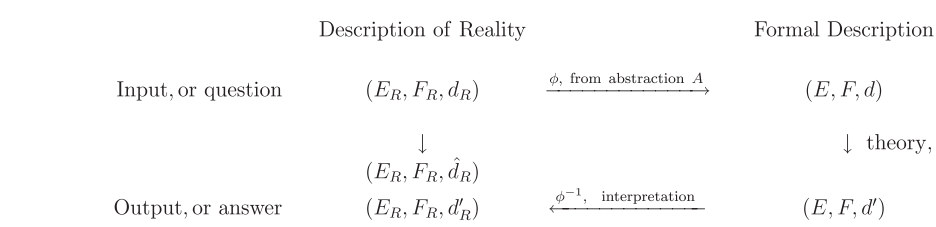
3.1. *The Question.* Suppose that we have a description of reality $\mathbf{d}_R = (E_R, F_R, d_R)$. We have seen that it can be extended by analysis, involving an abstraction A and a theory T. We wish to compare this extension with extensions that are not the result of analysis.

What other sources of extensions are there? The simplest one is the accumulation of data. New observations are added to the description of reality, extending $\mathbf{d}_R$ to some $\hat{\mathbf{d}}_R$. In this case, the question we would be interested in would be whether the new observations are in line with the theory's predictions, or do they refute the theory? Does the theory predict precisely these observations, or can it at least be reconciled with them?

Alternatively, the extension $\hat{\mathbf{d}}_R$ may reflect normative considerations, intuition or introspection, ethical principles, or ideology. In this case, the question would be whether these nontheoretical inputs can be reconciled with the theory. Or is it perhaps the case that the theory can even inform intuition by pointing out the "correct" extension $\hat{\mathbf{d}}_R$? In the next subsections, we define these notions more formally.

EXAMPLE 8 (THE DICTATOR GAME, CONTINUED). Suppose that someone proposed running an experiment to "test" theory T described in Example 6, specifying that players in the dictator game choose the actions corresponding to a subgame perfect equilibrium or, equivalently, that the dictator chooses a utility-maximizing action. Let us consider an abstraction A with ϕ mapping the experiment into description $\mathbf{d}^1$ and then apply theory T. As pointed out in that example, $T(\mathbf{d}^1) = \mathbf{d}^3$. Suppose that in the experiment, the dictators systematically choose splits that do not give them all the money. We can deduce from this that there is some problem in our construction, but we cannot necessarily deduce that theory T is "wrong." We might argue instead that the abstraction A is the problem: The subjects' preferences were not such that $P1$ prefers (100,0) to any split that gives a positive amount to $P2$.

EXAMPLE 9 (THE ULTIMATUM GAME). Consider the Ultimatum Game with two players, $P1$ and $P2$, in which $P1$ proposes a split of 100 to $P2$. If $P2$ agrees, the split will be implemented; if not, both players get 0. Suppose that someone proposed running an experiment to "test" theory T. Consider an abstraction A with ϕ mapping the experiment into a description similar to d^1, but with the strategies *acc* (accept) and *rej* (reject) added as strategies for $P2$, and preferences for $P2$ analogous to those for $P1$ (outcomes in which she gets more money are always preferred to outcomes in which she gets less). Applying theory T to this description, we get the analog of the result in the Dictator Game: $P1$ proposes the split (100,0) and $P2$ accepts the split. Suppose that in the experiments, we often see subjects in the role of $P2$ rejecting proposed splits of



NOTES: The point of departure is a description of reality (E_R, F_R, d_R). A process of data collection, experimentation, introspection, intuition, or so on extends this description to the description $(E_R, F_R, \hat{d}_R)$. As before, the function ϕ, which is part of the abstraction A, associates a formal description (E, F, d) with this description of reality. The theory T then extends the formal description (E, F, d) to a description (E, F, d'), and the interpretation ϕ^{-1} is then used to obtain the $A-T$ extension of (E_R, F_R, d). We then say that the theory T is A-compatible with $\hat{\mathbf{d}}_R$ if the $A-T$ extensions $\mathbf{d}'_R$ and $\hat{\mathbf{d}}_R$ are compatible descriptions, and the theory T A-necessitates $\hat{\mathbf{d}}_R$ if the $A-T$ extension $\mathbf{d}'_R$ is an extension of $\hat{\mathbf{d}}_R$.

FIGURE 2

HOW DATA, NORMATIVE CONSIDERATIONS, OR OTHER CONSIDERATIONS ARE USED TO EVALUATE A THEORY

$(99, 1)$, and the description of reality $d_R = (E_R, F_R, d_R)$ is extended to incorporate this data. As in the previous example, we can conclude that there is some problem in our construction, but we should not conclude that the theory (that players choose actions corresponding to a subgame perfect equilibrium) is wrong. As in that example, the problem probably lies in the players' preferences posited in the abstraction.

3.2. *Abstraction-Dependent Definitions.* We next formalize some of the ideas implicit in the previous two examples. We begin with an abstraction A, consisting of the description of reality $\mathbf{d}_R = (E_R, F_R, d_R)$, a model $\mathbf{d} = (E, F, d)$, and a function ϕ. We combine this abstraction with a theory T. Together, these define the $A-T$ extension of (E_R, F_R, d_R), which is a description

$$\mathbf{d}'_R = (E_R, F_R, d'_R)$$

satisfying $d'_R(e_R, f_R) = d'(\phi(e_R), \phi(f_R))$.

Our task is now to compare $\mathbf{d}'_R$ with another extension of $\mathbf{d}_R, \hat{\mathbf{d}}_R$. We consider two possibilities for this comparison. The first will be used to determine when a (descriptive) theory T is refuted by the data or when a (normative) theory T cannot justify a decision. The second will be useful to indicate when a theory implies a certain conclusion (prediction or recommendation).

DEFINITION 1. Given an abstraction $A = (\mathbf{d}_R = (E_R, F_R, d_R), \mathbf{d} = (E, F, d), \phi)$ and an extension $\hat{\mathbf{d}}_R$ of $\mathbf{d}_R$, we say that

(1.1) A theory T is A-compatible with $\hat{\mathbf{d}}_R$ if the $A-T$ extension $\mathbf{d}'_R$ and $\hat{\mathbf{d}}_R$ are compatible descriptions.

(1.2) A theory T A-necessitates $\hat{\mathbf{d}}_R$ if the $A-T$ extension $\mathbf{d}'_R$ is an extension of $\hat{\mathbf{d}}_R$.

Figure 2 illustrates these definitions. When the extension $\hat{\mathbf{d}}_R$ is obtained from $\mathbf{d}_R$ by adding a single specification of the form $\hat{d}_R(e, f) = x \in \{0, 1\}$, we say that theory T A-necessitates (e, f, x) (as well as that theory T A -necessitates $\hat{\mathbf{d}}_R$).

Example 8 described a case in which the abstraction A associated $\mathbf{d}_R$, the dictator game experiment, with $\mathbf{d}^3$, the model in which player 1's preferences are specified as preferring outcomes that gave him more money. The theory T that players actions' are consistent with subgame perfect equilibrium yielded the extension in which player 2 received 0 in the outcome. The $A-T$ extension $\mathbf{d}'_R$ is *not* compatible with $\hat{\mathbf{d}}_R$; in the extension of the experiment to include the experimental results, subjects in the role of player 1 systematically gave positive amounts to their partners, whereas $\mathbf{d}'_R$ prescribed that this not happen.

If one employed the abstraction A' that associated $\mathbf{d}_R$ with $\mathbf{d}^1$, that left preferences for player 1 over outcomes unspecified, the $A'-T$ extension would have made no "prediction" about outcomes, and hence the theory T would have been A' compatible. The theory T would not, however A'-necessitate $\hat{\mathbf{d}}_R$.

EXAMPLE 10 (THE SUNK COST FALLACY). The classical sunk cost fallacy is illustrated by someone who has paid \$50 for a ticket to a concert but forgets to take the ticket to the concert. Only when arriving at the concert does he realize his mistake. He can buy a replacement ticket for \$50 but chooses not to "pay twice" to hear the concert. An economist's conventional response to this decision is to label it as misguided, since at the time the concert goer chooses whether to buy a replacement ticket, the cost of the initial ticket has been sunk. The economist's argument can be thought of as beginning with a description of reality, $\mathbf{d}_R$, and then extending the description to $\hat{\mathbf{d}}_R$, in which the agent chooses not to buy the replacement ticket (the extension based on the agent's observed behavior). The economist then analyzes the following abstraction A of the problem. The agent prefers $(-50, \textit{Yes})$ to $(0, \textit{No})$, where the first entry represents the cost of the ticket and the second represents whether he has a ticket or not. The economist then considers an extension $\mathbf{d}$ of the abstraction that reflects the theory T that agents play in accordance with subgame perfection, that is, the agent chooses the most preferred outcome. This extension, of course, has the agent purchasing the replacement ticket. Thus, the $A-T$ extension $\mathbf{d}'_R$ is incompatible with $\hat{\mathbf{d}}_R$.

This example bears some similarity with Examples 8 and 9 above in that one can understand the $A-T$ extension of the description of reality, $\mathbf{d}_R$, to be the result of an "incorrect" choice of the abstraction associating $\mathbf{d}_R$ to a model $\mathbf{d}$. A plausible interpretation of the incompatibility of the $A-T$ extension and the extension $\hat{\mathbf{d}}_R$ that encompasses the empirical observations is that the players' preferences in the abstraction were incorrect.

However, one might argue that this example differs in a substantive way from Examples 8 and 9. An economist observing the incompatibilities in Examples 8 and 9 might conclude that the incompatibility most likely arose because of the preferences associated with the abstractions. In the sunk cost fallacy problem, however, the economist might think that the problem was not necessarily in the preferences embodied in the abstraction, but rather in the theory T. The economist might argue that the abstraction captures the agent's preference accurately, but the agent is making a mistake in his choice. Indeed, the problem is called the sunk cost *fallacy* to emphasize that the agent is making *a* mistake. In Examples 8 and 9, the economist might rethink his analysis of the dictator and ultimatum games but decide in the sunk cost fallacy game that he should spend his time educating the decision maker.

We are not arguing that the sunk cost fallacy *necessarily* implies that the problem lies in the theory T. The agent who forgot the ticket and chose not to buy a replacement might resist the economist's characterization of this decision as a fallacy as follows: "I didn't buy the replacement ticket because I wouldn't have enjoyed the concert. I would have kept thinking that forgetting to bring the ticket was stupid, so I decided to spend the evening at a bookstore rather than go to the concert." If one takes this argument at face value, the incompatibility does not lie in the theory T, but as in Examples 8 and 9, it lies in the abstraction to a model that with a description of preferences less nuanced than appropriate. The economist may or may not be able to convince the decision maker that such preferences are "confused" and should be modified.

We emphasize that whether most economists would cling to the backward induction solution in Examples 8 and 9 but not in 10 is an empirical question. In this article, we do not purport to predict how economists would react to these challenges, just as we do not attempt to predict what people would do in these games. Our only goal is to provide a model in which these discussions can be conducted.

3.3. *Abstraction-Independent Definitions.* It is clear from Subsection 3.2 that in the absence of a specific abstraction A, one cannot ask whether a theory is or is not compatible with observed

data, nor whether it necessitates certain conclusions. It might do so for some abstractions but not for others. Can we formulate counterparts of these ideas that are not dependent on specific abstractions?

Three possibilities arise:

(i) using a strong notion in which a concept applies for all acceptable abstractions;
(ii) using a weak notion in which a concept applies for at least one acceptable abstraction;
(iii) suggesting some aggregation over abstractions, weighing the set of abstractions for which the concept applies vis-à-vis the set for which it does not.

We consider the first two possibilities in this section.

DEFINITION 2. Given a description of reality $\mathbf{d}_R = (E_R, F_R, d_R)$ and an extension $\hat{\mathbf{d}}_R$ of $\mathbf{d}_R$, we say that

(2.1) A theory T is strongly compatible with $\hat{\mathbf{d}}_R$ if for every acceptable abstraction $A = ((E_R, F_R, d_R), (E, F, d), \phi) \in \mathcal{A}(\mathbf{d}_R)$, the $A-T$ extension $\mathbf{d}'_R$ and $\hat{\mathbf{d}}_R$ are compatible descriptions.
(2.2) A theory T is weakly compatible with $\hat{\mathbf{d}}_R$ if there exists an acceptable abstraction $A = ((E_R, F_R, d_R), (E, F, d), \phi) \in \mathcal{A}(\mathbf{d}_R)$, such that the $A-T$ extension $\mathbf{d}'_R$ and $\hat{\mathbf{d}}_R$ are compatible descriptions.
(2.3) A theory T strongly necessitates $\hat{\mathbf{d}}_R$ if for every acceptable abstraction $A = ((E_R, F_R, d_R), (E, F, d), \phi) \in \mathcal{A}(\mathbf{d}_R)$, the $A-T$ extension $\mathbf{d}'_R$ necessitates $\hat{\mathbf{d}}_R$.
(2.4) A theory T weakly necessitates $\hat{\mathbf{d}}_R$ if there exists an acceptable abstraction $A = ((E_R, F_R, d_R), (E, F, d), \phi) \in \mathcal{A}(\mathbf{d}_R)$, such that the $A-T$ extension $\mathbf{d}'_R$ necessitates $\hat{\mathbf{d}}_R$.

Theories that say little can be compatible with many descriptions. Clearly, a more general theory, that is, one that extends more descriptions and/or extends them further, is more easily refutable. In the following, we focus on the questions of necessitation, asking whether a theory can weakly or strongly necessitate a given extension. Similar questions can be posed for compatibility.

3.4. *Analogies: Aggregation of Models.* Instead of seeking results that hold for at least one acceptable model or results that hold for all models, we might look for ways to aggregate models. We have pursued this approach in Gilboa et al. (2014). In that paper, we argue that economic reasoning is often case based. Moreover, economic models may be viewed as theoretical cases. According to this view, each model is only a source of analogy that suggests possible predictions. A practitioner called upon to answer an economic question or to make a prediction may use various theoretical models as well as empirical and experimental evidence, intuition and thought experiments, historical studies, and other sources of inspiration. The practitioner aggregates the "cases" with the help of a similarity function, effectively taking a weighted average of their predictions to generate a prediction in the current problem.

We could extend the current model to capture such case-based reasoning. Toward this end, a "theoretical case" is a model, analyzed by a theory that generates an extension thereof. Analogies are given by an abstraction ϕ. The aggregation of many cases would correspond to aggregation of abstractions.

3.5. *Complexity Results.*

3.5.1. *Interpretations of a theory.* Our model of economic modeling begins with a description of reality $\mathbf{d}_R$ and exploits an abstraction to map this $\mathbf{d}_R$ into a description $\mathbf{d}$. As mentioned above, however, economic modeling often operates in reverse: The economist may construct a model aimed not at a specific real-world problem but rather at a phenomenon that might be found in a number of real-world problems. Spence's (1973, 1974) signaling model is an exemplar. In

its simplest form, the model considers an agent who has either high or low ability. Firms value an agent's ability, but ability is unobservable. The agent can, however, make an observable investment in an attribute, and the (unobservable) investment cost to the agent is inversely correlated with her ability. The attribute in which the agent invests is in itself of no value to the firm, but the agent's investment level is observable. Now, if the agent is of high ability, she can signal her ability by making a sufficiently large investment; had she been of low ability her (assumed) higher cost of investment would have deterred her from making that investment.

Spence motivated his model with a very simple story of workers investing in education, with the assumption that higher ability workers had lower costs of acquiring any given level of education. The model was not meant to be a serious model of human capital investment, as it ignored the kind of school the agent went to, her choice of topics, reasons other than subsequent wages she might have chosen to go to school, and so on. Rather, the model was meant to demonstrate how nonpayoff-relevant choices—the investments—might serve in equilibrium as signals about unobservable payoff-relevant characteristics.

Beginning with the signaling model, economists interpret the model with a description of reality $\mathbf{d}_R$. For example, the model might have an interpretation that suggests how a company chooses whether to pay dividends today. The agent in the model could be mapped into the company, which has private information about whether profits tomorrow will be high or low. The firm in the model could be mapped into an investor who might purchase stock in the company. With this mapping, the company in the real world might signal that profits tomorrow will be high by paying a dividend today, where paying a dividend today is more costly to the company if its profits tomorrow are going to be low.

Spence's signaling model has proven useful in a vast array of economic problems because it admits many different interpretations. Firms can offer warranties to signal the quality of goods sold, entrepreneurs can signal the quality of a proposed enterprise by taking a large equity interest in the enterprise, a suitor can signal his long-term interest by signing a prenuptial agreement that triggers a generous payment upon divorce, the suitee can signal her long-term interest by tearing up a proposed prenuptial agreement from the suitor, and so on. The diversity of the interpretations of the signaling model is possible because of the simplicity of the basic model. Had one been interested in a model of educational investment choice, it would have been useful to extend the model to include some of the neglected considerations mentioned above. Adding these considerations to the model would undoubtedly have made the model more useful for understanding education choices. However, including more details in the model makes it more complex and necessarily decreases the interpretations of the model. For example, if the model had included the various reasons an agent might invest in education, there might not be an interpretation that describes the dividend policy question.

3.5.2. *Determining the compatibility of a theory.* Making a theory more accurate by adding more details not only decreases the number of interesting interpretations, it makes determining the compatibility of the theory with a description of reality $\mathbf{d}_R$ more difficult. As the number of elements of a theory increases, the number of possible abstractions increases exponentially. Consequently, identifying whether the theory is strongly or weakly compatible with data and whether the theory strongly or weakly necessitates a conclusion can be expected to get ever more difficult as the theory becomes more complex. We turn next to a formal analysis of how much more difficult is that task.

Recall that we only deal with acceptable abstractions $A \in \mathcal{A}(\mathbf{d}_R)$, where the set $\mathcal{A}(\mathbf{d}_R)$ is assumed to be exogenously given. We further assume that acceptability places restrictions only on the mapping between predicates. This corresponds to an intuitive notion of universality: If, for example, a theory makes claims about players in a game, and we have established which real-life entities might be modeled as players, it seems natural that each such real-life entity can be mapped to each player in the game. In particular, we rule out cases in which it is acceptable to model, say, the United States as *Player 1* but not as *Player 2*.

This condition might appear restrictive in the following sense: Suppose that a theorist wishes to model one state as a single player, but another state as a collection of players. This might appear reasonable, if, for example, the former is a dictatorship and the latter is a democratic country with various institutions. However, in this case, there would be predicates that distinguish the two countries. Thus, we find it natural that acceptable abstractions will be defined by mappings between predicates, allowing for all the bijections between entities. Intuitively, entities are devoid of any content, and anything we know about them is reflected in the predicates they satisfy.

For any description of reality $\mathbf{d}_R = (E_R, F_R, d_R)$, there will thus be many acceptable abstractions. Even if there is only one obvious way to map predicates into one another, as in our example of the dictator game, there will be many ways to map entities into one another, and the number of such ways grows rapidly as does the number of entities. This is the source of our complexity results, the proofs of which are in the Appendix.

PROPOSITION 1. *Consider a description of reality* $\mathbf{d}_R = (E_R, F_R, d_R)$*, a pair* $(e, f) \in E_R \times F_R$ *such that* $d_R(e, f) = *$*, a value* $x \in \{0, 1\}$*, and a set of acceptable abstractions* $\mathcal{A}(\mathbf{d}_R)$*. Then, it is NP-Hard to determine whether a theory* T *weakly necessitates* (e, f, x)*.*

Next, we show that a similar conclusion applies to strong necessitation.

PROPOSITION 2. *Consider a description of reality* $\mathbf{d}_R = (E_R, F_R, d_R)$*, a pair* $(e, f) \in E_R \times F_R$ *such that* $d_R(e, f) = *$*, a value* $x \in \{0, 1\}$*, and a set of acceptable abstractions* $\mathcal{A}(\mathbf{d}_R)$*. Then, it is NP-Hard to determine whether a theory* T *strongly necessitates* (e, f, x)*.*

Making use of theories in a computationally simple way will thus require either that we restrict attention to models with relatively small sets of entities or that we find some additional structure that can limit the set of acceptable abstractions.

4. APPLICATIONS

This section examines applications of our model of modeling. We begin with the distinction between prediction and critique, suggested in Section 1, which motivated our work on this topic. We then examine three further distinctions that, with the help of our model, we argue are analogous to that between prediction and critique.

4.1. *Prediction and Critique in Economics.* The classical view of science leaves no room for intuition. A description of reality, $\mathbf{d}_R$, is given, it is modeled by an abstraction A about which there is little room for debate, a theory T is applied to generate predictions and recommendations, and these are mapped back into reality. The corresponding implications are strongly necessitated by T, either because there is but one acceptable abstraction $A \in \mathcal{A}(\mathbf{d}_R)$ or because the various acceptable abstractions are not so different from each other and yield the same predictions and recommendations.[11] In this case, economic reasoning allows us to make predictions.

By contrast, it is possible that the theory can be applied in a variety of ways, that is, there are many and varied acceptable abstractions. It may then be that no nontrivial extension of $\mathbf{d}_R$ is strongly necessitated by T. In this case, politicians and journalists might nonetheless come up with predictions and policy proposals and might call upon economists to evaluate them. Despite her inability to point to strongly necessitated conclusions, an economist can still be useful. In particular, the economist can check whether these are consistent with economic reasoning, that

[11] In particular, we can have a very large number of acceptable abstractions that yield the same results for reasons of symmetry. For example, if a large number of identical objects in reality are mapped onto an equally large number of identical entities in the model, the number of possible mappings is exponentially large, but, due to symmetry, one of them is enough for the analysis.

is, whether they are weakly necessitated by T. A positive answer does not amount to a support of the proposed policy or prediction, as it merely verifies its consistency with the economic principles embodied in T. However, a negative answer is a cause for concern: If the prediction or recommendation is not weakly necessitated by T, then there is no acceptable abstraction $A \in \mathcal{A}(\mathbf{d}_R)$ that supports it, and one may well wonder whether these are reasonable guidelines to follow. In this case, economic reasoning is valuable as a source of critique. Moreover, economic analysis can serve the purpose of critique even when a conclusion is weakly, but not strongly, necessitated by a theory. By pointing to abstractions under which a conclusion does not follow, one may expose hidden assumptions that are needed to support it.

4.2. *Is Economic Theory Vacuous?* It appears that over the past several decades, there has been a shift in microeconomic theory from general equilibrium models to game-theoretic models. This change has been accompanied by greater freedom in selecting an acceptable abstraction ϕ. For example, a "good" in general equilibrium theory may be a concrete product, as well as an Arrow security, but the concept cannot easily accommodate social and psychological phenomena. By contrast, an "outcome" in a game is a more flexible notion. Similarly, concepts such as "player," "strategy," and "state of the world" suggest a rich set of acceptable abstractions $\mathcal{A}(\mathbf{d}_R)$. As we have seen above, this freedom may render the theory "vacuous." An example such as the dictator game may refute a particular assumption about the determinants of players' utility, but it cannot shake the foundations of decision or game theory.

It seems that some of the discussions about whether economics is refutable or vacuous, and its status as a science, as well as some of debates revolving around behavioral economics, are discussions of the distinction between weak and strong compatibility (or necessitation). Detractors of the field point to phenomena where some acceptable abstractions are not compatible with the data. Some responses have been along the Popperian lines (Popper, 1959), attempting to redefine the scope of the theory. For example, it has often been argued that people's behavior in ultimatum or dictator games would conform to standard theory if the stakes were high enough. This is reminiscent of the restriction of Newtonian physics to certain levels of energy. However, another response to the experimental challenge has been the redefinition of terms as indicated above. This is a switch from strong compatibility to weak compatibility, which is more frequent in economics than in, say, physics.[12]

This discussion suggests that decision and game theory should be viewed as conceptual frameworks instead of as specific theories. As a rough approximation, one can view theories as refuted as soon as one of the appropriate abstractions in $\mathcal{A}(\mathbf{d}_R)$ is at odds with observations. By contrast, a conceptual framework is rejected only when all such abstractions are contradicted by evidence. Stated differently, we expect theories to be strongly compatible with the data, whereas conceptual frameworks need only be weakly compatible with the data.

This approach raises the question of refutability: Are the foundations of modern economic theory tautologically true? Can one imagine any set of observations that would not be compatible with decision and game theory?

The answer depends, of course, on what is considered "acceptable," that is, on the choice of the appropriate set of abstractions $\mathcal{A}(\mathbf{d}_R)$, which is taken to be exogenous in this article. It is common sense that has to determine the scope of $\mathcal{A}(\mathbf{d}_R)$. We believe that our model partly captures a phrase by Amos Tversky: "Theories are not refuted; they are embarrassed." Indeed, decision theory can typically be shown to be weakly compatible with the data, and the question is not whether it has been refuted but, rather, whether the set of mappings $\mathcal{A}(\mathbf{d}_R)$ has not become embarrassingly large or counterintuitive.

[12] At the same time, redefinition of terms is by no means restricted to economics or to its foundations. For example, one may view part of Freud's contribution as changing the unit of analysis, having goals and beliefs, from a unified self to ego, id, and super-ego. Similarly, in defending evolutionary reasoning, one often needs to explain that the unit of analysis is not the organism but the gene. In both cases, as in the rational choice paradigm, a "theory" may appear to be refuted given one abstraction ϕ but not given others.

4.3. *Objective and Subjective Rationality.* One view of rationality relates the concept of a rational decision to the robustness of the decision, reflected in the ability to convince others that the decision is reasonable. Specifically, it is suggested that decisions can be rational in two separate but related ways: A decision is objectively rational if any "reasonable person" can be convinced that it is the correct decision. A decision is subjectively rational for a "reasonable" person if she cannot be convinced that this is a wrong decision for her. In both cases, by "convincing," we refer to reasoning that does not resort to new information; that is, to the type of reasoning that the decision maker could have come up with on her own.

The Sunk Cost Fallacy examined in Example 10 illustrates subjective and objective rationality. The agent who chooses not to buy a replacement ticket to the concert might alter his choice after taking an economics course that discusses sunk costs. With the help of an economics course and sufficient discussion, the agent's preferences might conform to those in the abstraction the economist has associated with $\mathbf{d}_R$. If we find the agent's counterargument in the example plausible, we would deem the agent as subjectively rational. Only if we think that all reasonable people should be convinced by the economist's argument, we would say that an agent who forgoes the concert is not even subjectively rational.

The concept of objective rationality relates to strong necessitation: Independently of the decision maker's hunches and intuition, based on hard evidence, the theory makes specific recommendations. As these recommendations are valid for every abstraction $A \in \mathcal{A}(\mathbf{d}_R)$, they should be able to convince every reasonable person.

By contrast, in subjective rationality, the decision maker makes various choices that need not be supported by a model. Rather, she makes decisions based on her intuition, and the question becomes: Can she be convinced that she is wrong? If the decision maker can point to at least one abstraction $A \in \mathcal{A}(\mathbf{d}_R)$ that justifies (necessitates) her choices, she can defend them and cannot be convinced that she was in the wrong. Thus, subjective rationality applies to all choices that are weakly necessitated by the theory.

4.4. *The Role of Decision-Aid Models.* There is a tendency, especially among lay people, to expect decision theoretical models to come up with the "correct answer." Presumably, decision theory is supposed to provide mathematical models that, taking into account the relevant data and perhaps some subjective parameters, will compute correct predictions given possible outcomes of all available actions and eventually find the best decision. Indeed, this high standard is often attained, in particular, in domains such as statistics or operations research. For example, theory can help one identify which of two drugs has higher efficacy or how to find a shortest path between two points on a map.

Unfortunately, not all problems can be neatly resolved by theoretical models. Unknown fundamental mechanisms, high degrees of complexity, and unavailable data can each hamper a model's performance. Decisions involving human and social factors might encounter all of these difficulties, rendering theory almost useless in predicting phenomena such as wars and stock market crashes.

The classical model is often encountered in operations research problems that have a relatively small component of individual input. For example, if Mary wishes to find the shortest driving distance between two points, she may ignore intuition and let theory guide her. Reality would consist of the map and related information; there will be a relatively tight set of reasonable abstractions $\mathcal{A}(\mathbf{d}_R)$, and a simple algorithm would be the recommendation of theory T under each of them. That is, the claim that a certain path is optimal will be strongly necessitated by T. Mary would do well to follow that recommendation even if some turns along the path might seem counterintuitive to her.

By contrast, if Mary wishes to invest her savings, she may find that there are too many ways to model the world's financial markets. Mary may adhere to a theory of optimal portfolio management, T, but in the absence of a choice (an extension $\mathbf{d}''_R$ of $\mathbf{d}_R$) that is strongly necessitated by the theory, she may be at a loss. As a result, she may choose simply to follow her intuition. However, it would be useful for her to test, post hoc, whether there exists a model that justifies

this choice, that is, whether her choice is weakly necessitated by T. If it is not, Mary might wonder why and whether she can still do better after all.

5. CONCLUSION

5.1. *Other Sciences.* The standard view of "science" brings to mind an academic discipline engaging in the construction of formal models that provide predictions. However, there are respectable academic disciplines that are considered useful without being scientific in this sense, ranging from mathematics to history and philosophy.

A general point this article attempts to make is that domains that are considered scientific can also be useful as criticism. Physics provides laws of preservation that check fanciful ideas such as perpetual motion machines, whereas biology makes us doubt physical immortality. In both cases, it is useful to find the flaw in an argument even if it is not replaced by any quantitative prediction. Similarly, economics often proves useful without necessarily making predictions.

Beyond this claim, it is worthwhile to ask whether economics is viewed as a critique more often than are the natural or life sciences and, if so, why. We believe that this is indeed the case for two reasons. The first relates to the nature of economic questions: Dealing with social and political issues, lay people such as journalists and politicians often offer new economic ideas and predictions more readily than, say, mathematical or physical innovations. The reader has probably never had a cab driver suggest his or her personal theory of planetary motion or cold fusion but might have heard any number of such theories about the 2007 financial meltdown. Without minimizing the success of economics as a predictive science, it appears that the context of the debate makes it more important as a critical field than are the natural sciences.

The second reason has to do with the nature of economic questions and answers: The reliance of modern economic analysis on general tools such as decision and game theory generates a richness of possible interpretations—there are many acceptable abstractions. Thus, there are more cases in which there is a distinction between predicting (an outcome) and critiquing (an argument). For fields in which the theoretical terms are more clearly mapped onto real ones the distinction between the two is less important, and, indeed, most critiques would also generate predictions.

5.2. *Probabilistic Abstractions.* It would be interesting to extend the deterministic model that we have set out to allow probabilistic abstractions. In our examples based on the Dictator Game, one part of the abstractions dealt with player 1's preferences. The examples considered two abstractions that entailed different specifications of these preferences: One abstraction left the preferences completely open, whereas another specified that for any two different outcomes, the player preferred the outcome that gave him the larger amount of money.

The extension of the real-world description $\mathbf{d}_R$ to $\hat{\mathbf{d}}_R$ that incorporated the results of a Dictator Game experiment was surely compatible with the abstraction of the model that did not specify preferences, since in that case, no prediction was made. (That is, the extension of the model given the theory that the agent always chose the best outcome is the trivial extension.) Abstractions that specify nontrivial preferences (i.e., preferences such that there exist two outcomes o_1 and o_2 with o_1 strictly preferred to o_2) make possible incompatibility of the $A-T$ extension with $\hat{\mathbf{d}}_R$.

The problem, though, is that any such specification is likely to result in incompatibility if the number of subjects in the experiment is large. If one's belief is that few subjects who find themselves in the role of player 1 will give nothing to player 2, one would not want to specify deterministic preferences that make this impossible. Rather, one would like an abstraction in which relatively few, say less than 10%, of the subjects in the player 1 role choose to take the entire amount of money.

Asking whether more than 10% of the subjects chose (100, 0) is a yes-or-no question. Replications of the experiment might give rise to different extensions of $\mathbf{d}_R$, and the $A-T$ extension in which the abstraction has fewer than 10% of the agents caring only about their monetary

outcome might be compatible with some extensions of d_R and not others. The interpretation of a statement like "The $A-T$ extension of the reality description $\mathbf{d}_R$ is incompatible with $\hat{\mathbf{d}}_R$" thus becomes more complicated than in our model in this article. One would naturally ask how likely it was that the incompatibility would arise in a replication of the experiment.

5.3. *Multiple Theories.* In the discussion above, we refer to a single theory that can be compared to data or to intuition. One may consider several theories that compete in their attempt to generate predictions or recommendations. However, our basic model involved no loss of generality: Given several distinct theories, one may consider their union as a "grand theory," and relegate the choice of a theory to the choice of the abstraction A. To this end, it suffices that the sets of entities to which theories apply be disjoint. Figuratively, it is as if we guarantee that each scientist has access to her own set of variables, and we consider the entirety of their research papers as a single theory. This single theory generates only the extensions that are unions of extensions suggested by the single (original) theories, so that a practitioner can choose which theory to use by choosing an abstraction but cannot derive any new conclusions from the union of the theories.

5.4. *Normative Economics.* The social sciences differ from the natural sciences, inter alia, in that the former deal with subjects who can be exposed to and understand the theories developed about them. When focusing on descriptive theories, this distinction explains economists' focus on equilibria: Nonequilibrium predictions would be self-refuting prophecies, offering the economist a more or less sure way to be wrong. Moreover, this distinction also gives rise to normative considerations in the social sciences, considerations that are meaningless in the natural ones. It seems, however, that there is more than one way to understand what normative science is. The textbook approach says that "normative" refers to "ought" instead of "is." But what is this "ought"?

One possibility is illustrated by the Sunk Cost Fallacy of Example 10. The agent who chooses not to buy a replacement ticket to the concert might alter his choice after taking an economics course that discusses sunk costs. "Ought" in this case means that with sufficient discussion, the agent's preferences will conform to those in the abstraction the economist has associated with $\mathbf{d}_R$. As we suggested there, the agent may not be convinced by the economist's arguments, but the economist's arguments are typically of the form "Rational agents ought to ignore sunk costs."

One does not have to accept the economist's normative judgments. As discussed in Subsection 4.3, an economist's—or anyone's—normative prescriptions are valid only to the extent that others find the arguments behind the prescription compelling. We suggested there that the dialog between a theorist and her subject may not result in the subject accepting the economist's argument to ignore sunk costs.

5.5. *The Discipline of Economics.* Research is a social phenomenon, in which people decide which topics to study, what to publish, and so forth. In studying methodology, one observes this phenomenon and tries to understand it, thereby engaging in social science. One's interest may have a normative flavor—typically referred to as "philosophy of science"—or a stronger descriptive bent—closer to the "sociology of science." Both tendencies can be viewed as belonging to the realm of social science, broadly construed.

As a descriptive social science, the sociology of economics cannot expect to have formal, mathematical models that provide perfect descriptions of reality. As in other social sciences, such as economics itself, one expects models that are rather imperfect to help in understanding reality. We view our task in this article as theoretical: Our goal is to offer new models that may enhance understanding of social phenomena, in the case at hand, of formal modeling in some realms of the social sciences. Empirical work is needed to test which model best fits observations. Hence, we do not purport to argue here that our view of economics as critique

better explains economic research than the more classical view of economics as a "Popperian" science. We offer another way to conceive of observations, but we do not claim to have made an empirical investigation of the relative success of this conceptualization.

The above notwithstanding, we offer here a tentative conjecture that our model might fit some observations better than the classical one. Economic theory offers many qualitative predictions that are supposed to hold under ceteris paribus assumptions. These ceteris paribus conditions tend to be very hard to observe in real life, rendering such predictions dubious from an empirical viewpoint. By contrast, ceteris paribus arguments are valid as criticism: To challenge a way of reasoning, they are very powerful, even if nothing is held fixed in reality. That is, they can serve for gedankenexperiments when natural experiments are hard to identify.

This way of looking at economics can be applied to our model as well. Indeed, our model can be viewed as a form of critique: It criticizes the demands on economics to make predictions, by pointing out that economists can be useful without making predictions.

5.6. *Economics as Criticism and as Case-Based Reasoning.* Relative to the view posited by Gilboa et al. (2014) that economic models are theoretical cases, the view of economics as critique is even more modest: In the latter, the goal of economic modeling is only to test whether certain reasoning is valid, without making any predictions (case-based or rule-based). However, our focus on a single theory T does seem to attribute greater importance to the theory than the analogical (case-based) model. This seems to be compatible with the notion of critique: Although it does not need to proactively generate predictions, it aims to be a more objective standard for testing predictions. To consider an extreme example, assume that one's theory consists of no more than logical deductions. In and of themselves, such deductions make no predictions; specific assumptions about the reality modeled would be needed to reach any conclusions. However, logic enjoys a very high degree of objectivity when it comes to testing the validity of arguments.

5.7. *Freedom of Modeling in Economics.* Most of the examples above suggest that decision and game theory are closer to being "paradigms" or "conceptual frameworks" than specific theories and that this is much less true of more classical microeconomic theory. Although we do believe this to be the case, it is important to point out that the choice of a model and redefinition of terms is not restricted to game or decision theory applications. Consider, for example, basic consumer theory, according to which consumers choose a bundle of goods so as to maximize a utility function given a budget constraint, and that they therefore satisfy the axioms of revealed preferences (WARP, SARP, etc.). Clearly, this theory has counterexamples in observed data. However, Chiappori (1988, 1992; see also Browning et al., 2014, Chapter 3) argues that if a household's expenses are split between members of the household—specifically, between a wife and a husband—then utility maximization may be a much more reasonable hypothesis than if the household is viewed as a single unit. That is, although the standard approach is to map a specific household to a single "consumer," they suggest that individuals within households are to be mapped to different "consumers." One can easily imagine how similar redefinitions of terms might be important in assessing theories in other fields in economics. For example, should growth be measured for a country or a region thereof? Or perhaps a set of countries? What counts as "money"? Thus, although decision and game theory are probably the most prominent examples in which weak and strong compatibility with the data vary, they are not the only ones.

APPENDIX: PROOFS

A.1. *Proof of Proposition 1.* We reduce the Clique problem to Weak Necessitation. Let there be given an undirected graph (V, Q) with $|V| = n$ and a number k $(1 < k \leq n)$. The set $Q \subset V \times V$ denotes the set of edges. We assume that $(v, v) \notin Q$ for all v and that $(v, w) \in Q$ implies $(w, v) \in Q$. Construct the following problem. There is one predicate of degree 1 and

one predicate of degree 2 both for the real and the theoretical model. Formally, $F = \{K, L\}$ and $F_R = \{B, Q\}$ where B, K are single-place predicates and L, Q are two-place predicates. We abuse notation and use Q for a predicate in our model because it will be identical to the edges in the graph (V, Q). We allow $\mathcal{A}(\mathbf{d}_R)$ to include only the abstractions generated by ϕ_F where $\phi_F(B) = K$ and $\phi_F(Q) = L$. (Note that this $\mathcal{A}(\mathbf{d}_R)$ can be succinctly described.) Thus, there exists only one acceptable mapping of predicates, and different mappings will differ in their permutation of entities.

Define $E_R = V \cup \{y\}$ where $y \notin V$. Let d_R be a description (applying to reality) of the edges in the graph, which says nothing about the predicate B. That is, for $v, w \in V \subset E_R$, we have $d_R((v, w), Q) = 1$ iff $(v, w) \in Q$. For $v \in V$, set also $d_R((v, y), Q) = d_R((y, v), Q) = 0$. Likewise, $d_R((y, y), Q) = 0$. Finally, set $d_R(e, B) = *$ for all $e \in E_R$. Let $E = \{1, \ldots, n+1\}$.

The theory T is given by a single rule: If $\{1, \ldots, k\}$ are all pairwise linked according to L, then predicate K applies to element $(n+1)$. Formally, if

$$d((i, j), L) = 1$$

for all $i, j \leq k$, then $T(d)$ satisfies

$$T(d)(n+1, K) = 1.$$

We argue that the original graph has a clique of size k if and only if there exists a bijection ϕ_E such that $T - A$-necessitates $(B, y, 1)$ for $A = ((E_R, F_R, d_R), (E, F, d), \phi)$. Indeed, if such a clique exists, ϕ_E can be defined by any permutation of the nodes that places the clique nodes in the first k places, any permutation that places the rest in the next $(n-k)$ places, and that maps y to $(n+1)$. Theory T can then be used to derive the extension according to which $T(d)(n+1, K) = 1$, and the mapping back implies that the extension of d_R, d'_R, satisfies $d'_R(y, B) = 1$.

Conversely, if T weakly necessitates $(y, B, 1)$, it must be the case that $\phi_E(y) = n+1$ (as $n+1$ is the only entity in E for which T may yield such a conclusion). This means that the entities $\{1, \ldots, k\}$ are images of nodes in the original graph V (and none of them is an image of y) and thus ϕ_E^{-1} identifies a clique in V.

Finally, observe that the construction of the Weak Necessitation problem can be done in polynomial time.

A.2. *Proof of Proposition 2.* We will reduce the (closed) Hamiltonian path problem and prove that Strong Necessitation is co-NPC. That is, given an undirected graph (V, Q), we will construct $d_R \in D(E_R, F_R)$, a pair (e, f) with $d_R(e, f) = *$, a value $x \in \{0, 1\}$, a conclusion (e, f, x), and a theory T such that T strongly necessitates (e, f, x) if and only if the original graph does not have a closed Hamiltonian path. As in the proof of the previous result, $F = \{K, L\}$ and $F_R = \{B, Q\}$ where B, K are a single-place predicates and L, Q are two-place predicates. Again, we abuse notation and use Q for a predicate in our model because it will be identical to the arcs in the graph (V, Q). We allow $\mathcal{A}(\mathbf{d}_R)$ to include only the abstractions generated by ϕ_F where $\phi_F(B) = K$ and $\phi_F(Q) = L$. (Note that this $\mathcal{A}(\mathbf{d}_R)$ can be succinctly described.)

Define $E_R = V$ and $E = \{1, \ldots, n\}$ for $n = |V|$. Set

$$\begin{aligned} d_R(v, B) &= * \qquad \forall v \in V, \\ d_R((v, w), Q) &= 1 \qquad \forall (v, w) \in Q, \\ d_R((v, w), Q) &= 0 \qquad \forall (v, w) \notin Q. \end{aligned}$$

The theory T will be defined by n^2 rules, each of which might indicate that a Hamiltonian path has not been found. Specifically, for $i, j \in E$ rule r_{ij} says that if $d((i, i+1), L) = 0$ (with

$n+1=1$), then $d(j,K)=0$. Select $v_1 \in V$. We claim that T strongly necessitates $(v_1, B, 0)$ if and only if the graph (V, Q) does not have a Hamiltonian path.

To see this, consider a permutation of the nodes $\phi_E : E_R(\equiv V) \to E(\equiv \{1,\ldots,n\})$. If this permutation defines a closed Hamiltonian path, the theory cannot be applied (because $d((i, i+1), L) = 1$ for all i) and it does not provide any nontrivial extension of d. Consequently, nothing can be added to d_R, and, in particular, we remain with $d_R(v, B) = * \; \forall v \in V$. Thus, if the graph (V, Q) contains a Hamiltonian path, at least one possible model (permutation of the nodes) will not result in $d_R(v, B) = 1$ and therefore $(v_1, B, 0)$ is not strongly necessitated by T.

Conversely, if a Hamiltonian path does not exist, then, for any permutation of the nodes, there exists at least one i for which $(i, i+1) \notin Q$, or $d(L, (i, i+1)) = 0$, and thus the theory would yield $d(j, K) = 0$ for all j. Mapping this conclusion back, we obtain $d_R(v, B) = 0$ for all v, and, in particular, for v_1. As this holds for every mapping ϕ, the conclusion $(v_1, B, 0)$ is strongly necessitated by T.

Finally, observe that the construction is carried in polynomial time.

REFERENCES

Aumann, R. J., "What Is Game Theory Trying to Accomplish?" in K. Arrow and S. Honkapohja, eds., *Frontiers of Economics* (Oxford: Basil Blackwell, 1985), 5–46.

Browning, M., P. Chiappori, and Y. Weiss, *Economics of the Family* (Cambridge: Cambridge University Press, 2014).

Cartwright, N., "Capacities," in J. Davis, W. Hands, and U. Maki, eds., *The Handbook of Economic Methodology* (Cheltenham: Edward Elgar, 1998), 45–8.

———, "Models: Parables vs. Fables," in R. Frigg, and M. Hunter eds., *Beyond Mimesis and Convention: Representation in Art and Science* (Berlin/New York: Springer Science & Business Media, 2010), 19–32.

Chiappori, P., "Rational Household Labor Supply," *Econometrica* 56 (1988), 63–90.

———, "Collective Labor Supply and Welfare," *Journal of Political Economy* 100 (1992), 437–67.

Desai, M., *Hubris: Why Economists Failed to Predict the Crisis and How to Avoid the Next One* (New Haven: Yale University Press, 2015).

Gibbard, A., and H. R. Varian, "Economic Models," *The Journal of Philosophy* 75 (1978), 664–77.

Gilboa, I., and D. Schmeidler, *A Theory of Case-Based Decisions* (Cambridge: Cambridge University Press, 2001).

———, A. Postlewaite, L. Samuelson, and D. Schmeidler, "Economic Models as Analogies," *Economic Journal* 128 (2014), F513–33.

Grune-Yanoff, T., "Learning from Minimal Economic Models," *Erkenntnis* 70 (2009), 81–99.

———, and P. Schweinzer, "The Roles of Stories in Applying Game Theory," *Journal of Economic Methodology* 15 (2008), 131–46.

Habermas, J., *Theory of Communicative Action, Vol. I: Reason and the Rationalization of Society*, English edition (Boston: Beacon, 1984).

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

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

Litan, R. E., *Trillion Dollar Economists* (New York: John Wiley and Sons, Inc., 2014).

Maki, U., "Isolation, Idealization and Truth in Economics," *Poznan Studies in the Philosophy of the Sciences and the Humanities* 38 (1994), 147–68.

———, "Models Are Experiments, Experiments Are Models," *Journal of Economic Methodology* 12 (2005), 303–15.

McCloskey, D., *The Rhetoric of Economics* (Madison: University of Wisconsin Press, 1985).

Popper, K. R., *The Logic of Scientific Discovery*, English edition 1958 (1934, English edition, London: Hutchinson and Co., 1959).

Rubinstein, A., "Dilemmas of an Economic Theorist," *Econometrica* 74 (2006), 865–83.

Spence, A. M., "Job Market Signaling," *Quarterly Journal of Economics* 87 (1973), 225–43.

———, *Market Signaling* (Cambridge: Harvard University Press, 1974).

Sugden, R., "Credible Worlds: The Status of Theoretical Models in Economics," *Journal of Economic Methodology* 7 (2000), 1–31.