A Model of Modeling *

Itzhak Gilboa[†], Andrew Postlewaite[‡], Larry Samuelson,[§] and David Schmeidler[¶]

August 13, 2014

Abstract

We propose a formal model of scientific modeling, geared to applications of decision theory and game theory. The model highlights the freedom that modelers have in conceptualizing social phenomena using general paradigms in these fields. It may shed some light on the distinctions between (i) refutation of a theory and a paradigm, (ii) notions of rationality, (iii) modes of application of decision models, and (iv) roles of economics as an academic discipline. Moreover, the model suggests that all four distinctions have some common features that are captured by the model.

*We are thankful to many colleagues and friends with whom we have discussed the issues addressed here over the years. We also thank Eva Giboa-Schechtman for conversations that partly motivated and greatly contributed to this work. We gratefully acknowledge ISF Grant 204/13 (Gilboa and Schmeidler), ERC Grant 269754 (Gilboa), and NSF Grants SES-0961540 (Postlewaite) and SES-1153893 (Samuelson).

[†]HEC, Paris, and Tel-Aviv University, tzachigilboa@gmail.com.

[‡]University of Pennsylvania. apostlew@econ.sas.upenn.edu.

[§]Department of Economics, Yale University, Larry.Samuelson@yale.edu.

[¶]The InterDisciplinary Center in Herzliya, and TAU, davidschmeidler@gmail.com.

Contents

A Model of Modeling

1 Introduction

1.1 Motivation

The use of formal models to describe, explain, and predict phenomena is as old as is science, and, indeed, many would consider formal modeling to be an essential ingredient of scientific activity. In contrast, thinking formally about models is a more recent phenomenon, dating back to the late 19th century. Moreover, discussions of formal modeling in economics have rarely been related to logic or to model theory in a formal way. Indeed, a formal discussion of modeling requires cumbersome machinery, and much can be said and understood without it. Yet, our own experience with methodological discussions led us to believe that a formal model of modeling might clarify some arguments relating to the way economics and related academic disciplines interact with reality. In particular, such a model can sharpen the debate about economics as a positive science, as well as highlight some distinctions about the way it can be used for normative purposes. The purpose of this paper is, therefore, to offer a simple model of modeling and use it to capture some of the on-going debates about the role and success of economics and its foundations.

We begin in Subsection 1.2 by raising four seemingly unrelated questions relating to: (i) the role of decision and game theory; (ii) the notion of rationality; (iii) the role of models as decision aids; and (iv) the role of economics as an academic discipline. In each case, we present a distinction between two rather different extreme positions.

Sections 2-3 introduce a model of modeling that is geared to capture applications of decision theory and game theory in economics and related disciplines. The model offers a general framework that can be used to discuss both descriptive and normative applications. Section 4 revisits the four distinctions introduced in Section 1.2 and argues that they are analogous, and that their common aspects are captured by the formal model of Sections 2-3. Section 5 concludes.

1.2 Four Distinctions

This section discusses four seemingly unrelated issues, touching on the way models and theories can be applied. We will later use the formal model to suggest that the four distinctions bear some similarity to each other.

1.2.1 Theories and Paradigms

The dominant methodology of economics since the 1920s-30s has championed observability and falsifiability, following from the logical positivists' and from Popper's (1934) philosophy of science. Graduate as well as undergraduate textbooks highlight the importance of relating theoretical concepts, such as utility maximization, to observations, such as choice behavior. The "revealed preference paradigm" insists that only choice observations should be used for defining theoretical concepts, and students are inculcated with the need to test theories by their observable implications, to shun disputes between observationally equivalent theories, and so forth. Moreover, the need to have axiomatic foundations is tightly related to the logical positivistic dictum of relating theoretical to observable terms, and to the desire to delineate a theory's scope of applicability.

However, economic theory has also been criticized for making too few concrete, falsifiable predictions, and for being able to explain any phenomena post-hoc. In particular, the rise of decision and game theory as the basic modeling tools in economics over the past decades has been accompanied by questions about whether these "theories" are indeed falsifiable theories in the Popperian sense, or, perhaps, some entities that are closer to "paradigms". The following example illustrates.

Example: Framing effects We say that framing effects arise if two seemingly equivalent descriptions of alternatives lead people to make different choices between those alternatives. For example, it is typically said to be a result of a framing effect that more people opt to be organ donors if the default is to be a donor and one must actively choose to opt out, than if the default is not to be a donor and one must actively choose to join in. In standard models, in which choosing is effortless and costless, these two problems appear to be equivalent, and should we observe differences in the responses they induce, we conclude that framing effects are at work.

However, when modeling the choice problem, one can add a predicate that designates "default choice on form", and embed the problem in a more elaborate one, where another agent has designed the form and decided which option should appear as default. In such a problem the symmetry between the two descriptions breaks down: indeed, if the decision maker believes that the choice of a default on the form indicates some beliefs about the norm in the society she belongs to, the two choice situations are no longer the same, and need not induce the same action.

Similarly, a physician who describes a procedure in terms of a probability p of fatality instead of probability $1 - p$ of survival may be revealing information about her subjective beliefs, the trust she puts in these statistical estimates, and so forth.[1] More generally, the choice of the representation of the problem may be informative above and beyond the information contained in the representation itself.

Importantly, in many examples such as Monty Hall 3-door problem (see Raiffa and Metcalfe, 2002), the textbook solution to the puzzle is to include all information in one's model: not only the information one acquired, but also the way it was acquired; not only the signal sent, but who sent it, what other signals they could have sent, and so forth. Once one has done so, two descriptions of alternatives can always be viewed as not being equivalent.

[1]This point was made by Ehud Kalai.

Similarly, any apparent violation of consequentialism—the principle that only outcomes matter, and not the path by which they are reached—can be rescued by redefining outcomes to included additional elements of their history, perhaps through emotions generated by that history. Violations of expected utility theory based on small probabilities, including much of the motivation for Prospect Theory (Kahneman and Tversky, 1979), can be incorporated via a model in which people view small probabilities as being overwhelmed by measurement error.

It follows that the status of the most fundamental building blocks of economic theory is somewhat ambiguous. They are called "theories", as in "decision theory" and "game theory", and when taught in classes they are coupled with well-defined terms of refutation. Careful and non-trivial mathematical analysis describes what is assumed to be observable and which observations will and will not be compatible with the theory. Yet, when applied to real-life problems, it appears that the basic terms used in these "theories" are open to a wide range of interpretations. The freedom in selecting an interpretation often makes one wonder, are these theories truly refutable? Or, are they refutable in any but the simplest laboratory experiments that are specifically designed to test them? And if this is not the case, what are these constructs? Are they better conceived of as "paradigms", and, if so, what is the distinction between a theory and a paradigm?

1.2.2 Objective and Subjective Rationality

The advances in axiomatic decision theory in recent decades raise the question of rationality: what does the concept of rationality attempt to capture? Are all deviations from classical axioms irrational? Are they to be distinguished according to degree or kind? How should we use the term "rational", if at all?

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.

These two notions of rationality seem to raise more questions than they resolve. For example, who is a reasonable person? Who attempts to convince her and how? How do we tell if a person has indeed been convinced? While these questions and many others need to be addressed in order to render the concepts meaningful, the two definitions are nonetheless useful in discussing decision making. In particular, unlike a definition based on departures from the classic axioms, they do not dub "irrational" a mode of behavior that people cannot or do not wish to abandon, such as using a system in chess that is suboptimal. Rather, the notions reserve the term for behaviors that might change as a result of analysis that is shared with the decision makers, that is, as a result of the interaction between theory and practice.

It is natural to assume that objective rationality entails subjective rationality: if one can be convinced that a decision is correct, it should be the case that one won't be convinced that the same decision is wrong. However, it should be expected that there are subjectively rational decisions that are not objectively rational: a person may make a decision that seems right to her without being able to convince others that it is indeed the correct decision (for her, taking into account her goals and preferences). Thus, objective rationality is typically expected to violate the completeness axiom of decision theory: based on evidence, logic, statistics, and scientific analysis, one expects to be able to resolve many issues, but not all. By contrast, the classical defense of the completeness axiom, namely, that decisions need to be made, suggests that it is reasonable to demand that subjective rationality be complete: should one venture beyond the safe realm of objective scientific knowledge, one might hope to be able to deal with the trickier questions on which science remains silent.

Objective and subjective rationality are related to (at least one possible interpretation of) Classical and Bayesian statistics. The former aspires to derive objective conclusions from data, while the latter allows and even welcomes subjective inputs. Correspondingly, Classical statistics remains silent on many issues, whereas Bayesian approaches provide quantified beliefs for any proposition of interest.

1.2.3 On the Use of Decision Models

Decision theory sells its products in two main markets. The first is the academic world of economics, finance, political science, and other disciplines interested in modeling human behavior in order to better understand social phenomena. The second is the world of practitioners who need to make decisions and who seek to make better decisions. These might include investors in the stock market as well as patients who have to choose medical treatments. Naturally, participants in the first market have a greater demand for descriptively accurate theories, whereas the latter have a greater demand for normatively compelling theories.

Considering the domain of practical decision making, 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.

However, it would be premature, and probably a mistake, to discard decision theory altogether and relegate all decisions to intuition alone. Psychological studies provide us with plenty of examples in which intuition can lead us astray, in systematic and predictable ways, resulting in decisions that would be considered mistakes by the decision makers themselves. Some of these mistakes can be avoided by adopting decision theoretical tools. For example, the very use of formal models renders one immune to framing effects, and the use of subjective probabilities ensures that one doesn't fall prey to the "conjunction fallacy" (Tversky and Kahneman, 1981).

This discussion suggests that, when the theory is incapable of providing clear predictions and recommendations on its own, it can still be used as a check on the reasoning of the decision maker. In fact, any application of decision models that uses subjective inputs such as utilities or subjective probabilities can be viewed as posing a battery of questions in order to elicit parameters that will be used in making the decision. The resulting analysis may also involve meta-preferences, where the decision maker might express preferences over the type of model she wishes to use, the decision theoretic axioms she would not like to violate, and so on.

An extreme version of such a dialog would relegate decision theory models to the endorsement of independently-obtained decisions. Imagine a decision theorist who is asked to advise an investor on an R&D venture, saying to her potential client, "Look, let me be frank: my models can't find the right answer. In fact, I'm sure that you know much more about this problem than

I do. Your intuition is probably very good, or else you wouldn't have any money to invest in the first place. So I'm in no position to offer a model that would replace your intuition. But my models can detect some common errors and biases. So let's do the following: you write here your decision, and now we'll try to work together to justify it using standard models that I teach in classes. If we have a reasonable model, I'll sign this form, stating that I endorse your decision. If not, and we just can't find any model that supports your intended decision, or if, in order to do so, we have to use some crazy scenarios, weird preferences, or outlandish parameters, then we admit failure and you go back to the drawing board."

Thus, at one extreme we can think of a model used to find the shortest path on a map, where the decision theorist can rest assured that, provided some parameters, her model provides the best decision for the decision maker. At the other extreme, the decision theorist only tests whether the decision maker is about to make a decision that would be a grave mistake in his own eyes. In between, one can imagine a whole gamut of dialogs, varying in the relative roles of intuition and formal analysis.

1.2.4 The Role of Economics

The role played by economic models, especially in modern microeconomics, has been discussed already by Friedman (1953), but has drawn growing attention in recent decades. (See Gibbard and Varian, 1978, McCloskey, 1985, Hausman, 1992, Maki, 1994, 2005, Cartwright, 1998, in press, Sugden, 2000, Rubinstein, 2006, Grune-Yanoff and Schweinzer, 2008, and Grune-Yanoff, 2009, among others.) One view is that the essence of economics is to make predictions. Another view holds that the goal of economic analysis is explanation rather than prediction (see, for example, Aumann, 1985). Understanding and explaining are closely related to prediction: first, explanations are powerful tools to have in one's kit to be used in future prediction problems. Our taste for understanding may well have evolved thanks to the role of explanation in prediction. Second, predictions are useful in testing and choosing among competing explanations. Nonetheless, explanation may be viewed as a separate role of scientific disciplines, economics included.

Without attempting to judge the success of economics in providing predictions or explanations, we point out that there is yet another view of economics as a useful academic discipline. This view suggests that one possible 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, thereby invalidating the revenue calculation. 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. Yet, such an economist would earn her keep simply by pointing out the fallacy in a given calculation.

Thus, economic analysis might be useful simply as a form of criticism. It may critique reasoning at the very basic level of testing logical deductions, at the level of equilibrium analysis (as in the example above), as well as in other ways such as confronting intuition with empirical findings. Often, the distinction between critique and qualitative predictions may be fuzzy. In the tax example, by pointing out that the quantity demanded is likely to respond to a price hike, the economist makes the implicit prediction that tax revenue will be lower than calculated based on current demand. But the economist might 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. In such cases the economist might be careful not to make any predictions, quantitative or qualitative, yet make a contribution to public debate by critique of reasoning.

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. Our main point 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, while 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.

The social and political nature of economic questions encourages lay people, including journalists and politicians, to more readily offer new economic ideas than they are to offer mathematical or physical innovations. Some of these ideas turn out to be very good and to influence economic thought. But others are not as successful. Thus, without minimizing the success of economics as a predictive science, it appears that it is more important as a critical field than are the natural sciences.

2 A Model of Modeling

To model the act of modeling, we need a formal structure that can encompass a variety of theories from which a social scientist may choose, a model of the reality she attempts to study, and a connection between the two.

We begin with a few examples in which various problems can be formally described in a language that makes an explicit reference to the predicates used. This allows us to describe a theory's prediction or recommendation as a conditional statement, taking a set of predicate values as given and concluding additional values from it, suggesting what might or should happen. We then move on to the general model, in which predicates are indexed by elements of an abstract set.

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).

2.1 Examples

2.1.1 Decisions under Certainty

Consider first a single-decision-maker problem under certainty. The formal model would typically describe a set of possible alternatives, A, and the decision maker's preferences. For example, these preferences can be represented by a binary relation, $\succsim$, which is a subset of $A \times A$.

We fit this model within a unified framework by viewing the model as being composed of two predicates: a one-place predicate that describes the alternatives, and a two-place predicate that describes the preferences. For example, $Act\,(a)$ might mean that $a \in A$, that is, that a is a possible alternative, and $\succsim (a, b)$ might mean that a is at least as desirable as b, or that $(a, b) \in \succsim$.

A theory in this example could be that preferences are transitive. This can be viewed as a rule that says that, for every a, b, c, $\succsim (a, b)$ and $\succsim (b, c)$ imply $\succsim (a, c)$. Thus, the theory takes a collection of statements, given by the predicates Act and $\succsim$, and adds some statements, in this case given by the predicate $\succsim$. We would expect that the new statements do not contradict the given ones. Hence, if we start with the statements $\succsim (a, b)$ and $\succsim (b, c)$ then we may add $\succsim (a, c)$; but if we were to start with $\succsim (a, b)$, $\succsim (b, c)$, and the negation of $\succsim (a, c)$ (assuming such a negation is a valid statement), the theory cannot add the statement $\succsim (a, c)$.

2.1.2 Decisions under Uncertainty

Next consider a decision problem under uncertainty. The structure now will have to distinguish between acts that can be chosen and outcomes that can be experienced. Moreover, one would have to have a third type of element to describe possible eventualities, such as states of the world. Hence, the model might have three one-place predicates, *State*, *Outcome*, and *Act*. The predicate $\succsim (a, b)$ would be meaningful if a and b are acts, that is, if $Act\,(a)$ and $Act\,(b)$ hold. One may also define $\succsim$ over outcomes, but it would be a mistake to define it over, say, an act and an outcome, or over two states.

Note that predicates are defined over some entities, or n-tuples of entities, but not on all. One role of one-place predicates is to identify entities that can be meaningfully used in other predicates. For example, if $Act\,(a)$ and $Act\,(b)$, then $\succsim (a, b)$ is a meaningful predicate (whether it is true, false, or unknown). But, in standard decision theory, if $State\,(s)$ and $State\,(t)$, $\succsim (s, t)$ is meaningless, as would be $\succsim (a, t)$ or $\succsim (s, b)$.

In this set-up, a theory could satisfy, apart from transitivity, also Savage's axioms P2, P3, and P4, which are naturally defined as adding statements to the theory based on a set of given statements.

2.1.3 Game Theory

When a game-theoretic model is considered, at least one additional predicate is needed – say, *Player* – and some predicates would need to have more places. For example, it does not suffice to say $Act\,(a)$, as we need to specify which player can choose a. Thus we may use a predicate $Act\,(i, a)$, where $Player\,(i)$ holds, to indicate that player i can choose strategy a. Similarly, preferences need to be assigned to players, and so do beliefs, if they are to be included in the model.

For example, consider the pure Nash equilibrium theory in the "Battle of the Sexes". We might start with the description of the game by the following set of statements:

$Player\,(P1)$
$Player\,(P2)$
$Act\,(P1, A1)\,, Act\,(P1, B1)$
$Act\,(P2, A2)\,, Act\,(P2, B2)$

$Outcome\left(o_1\right), Outcome\left(o_2\right), Outcome\left(o_3\right), Outcome\left(o_4\right)$

$Result\left(A1, B1, o_1\right), Result\left(A1, B2, o_2\right)$

$Result\left(A2, B1, o_3\right), Result\left(A2, B2, o_4\right)$

$\succsim\left(P1, o_1, o_4\right), \succsim\left(P1, o_4, o_2\right), \succsim\left(P1, o_2, o_3\right), \succsim\left(P1, o_3, o_2\right), \ldots$(and its transitive closure)

$\succsim\left(P2, o_4, o_1\right), \succsim\left(P2, o_1, o_2\right), \succsim\left(P2, o_2, o_3\right), \succsim\left(P2, o_3, o_2\right), \ldots$(and its transitive closure)

Suppose that a single-place predicate May defined on outcomes is used to describe the theory's prediction or recommendation. Then, the pure-strategy Nash equilibrium theory would add to the list of statements above the following two:

$May\left(o_1\right), May\left(o_4\right)$

indicating that the only two outcomes that may obtain, according to the pure Nash equilibrium theory, are outcomes o_1 and o_4.

2.1.4 General Theories

The standard description of a strategic game is a triple $\left(N, \left(S_i\right)_{i\in N}, \left(h_i\right)_{i\in N}\right)$ with $h_i : \Pi_{i\in N} S_i \to \mathbb{R}$. The language described above is too primitive to describe, say, Nash equilibrium predictions for any such game $\left(N, \left(S_i\right)_{i\in N}, \left(h_i\right)_{i\in N}\right)$. One would need a more elaborate language to describe quantifiers, conditional statements, and so forth. However, for the purposes of this paper this additional structure is not necessary. Instead, we will imagine that each game of the form $\left(N, \left(S_i\right)_{i\in N}, \left(h_i\right)_{i\in N}\right)$ is spelled out by a distinct set of entities. Similarly, for a given game form one may consider many possible preferences. Again, the way these would fit into our model would be by listing different entities for any specification of the preferences.

To illustrate, we consider a dictator game in which Player 1 can determine how to split \$100 between herself and Player 2, who is a "dummy" player in the sense that he has no choices to make. Assume first that Player 1's preferences are determined solely by her own monetary payoffs in a monotone way. The game can be described as follows.

$Player\,(P1)\,, Player\,(P2)$

$Act\,(P1,0)\,, Act\,(P1,1)\,, ..., Act(P1,100)$

$Outcome\,((0;0))\,, Outcome\,((100;0)), Outcome\,((99;1))\,, ..., Outcome\,((0;100))$

$Result\,(0,(100;0))\,, Result\,(1,(99;1)), ..., Result\,(100,(0;100))$

$\succsim(P1,(100;0)\,,(99;1))\,, \succsim(P1,(99;1)\,,(98;2)), ..., \succsim(P1,(1;99)\,,(0;100))$

(and its transitive closure)

And its prediction would be

$May\,((100;0))$.

However, if the preferences of the player are not necessarily based on monetary payoffs alone, the same game form may give rise to a different game. In fact, for any preference relations over the outcomes $(100;0)\,,(99;1)\,, ..., (1;99)\,,(0;100)$ one would have a different set of statements as above (describing $\succsim(P1,\cdot,\cdot)$) and the equilibrium outcome will be described by May.

Note that the formulation we use here makes no claim to be succinct. In fact, if we only consider strict preferences of player $P1$ the simple dictator game is here modeled by 101! different sets of entities, each describing a different game. We will later on have to deal with this inefficient representation of the theory.[2]

2.2 Descriptions

We now turn to a more formal development of the model. We begin with the concept of a description. One might be tempted to refer to a description as a model, and the two are often equated in casual discourse. However, a description is a purely formal object. For example, a description alone gives no indication as to whether it is intended to be used to describe the motion

[2]It will turn out that in this very simple example one need not have different games: as the formal terms denoting outcomes have no meaning in and of themselves, one may refer to the player's most preferred outcome as $(100;0)$, to the second as $(99;1)$ and so forth. However, in a slightly richer model these outcomes will also be involved in other predicates and then the multiplicity of entities will be crucial.

of heavenly bodies, transactions between corrupt politicians, or the mating behavior of spiders. We promote a description to the status of a model only when it is accompanied by an indication of how the description is related to reality (cf. Section 2.4).

2.2.1 The Formal Elements

A description begins with a set of *entities* E, with a typical element denoted by e. It can be thought of as a set of letters in an alphabet. We assume that E is a subset of a universal set of entities $\hat{E}$. We assume that $\hat{E}$ is infinite, so that the analyst will never be in want of letters to designate new entities. In a typical description, however, E will be a strict subset of $\hat{E}$, reflecting the modeler's desire to focus on objects that are particularly important and ignore less relevant considerations, and E will often be finite.

The next component of a description is a 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), and F is a subset of an infinite universal set of predicates $\hat{F}$. The examples of Section 2.1 consider a variety of predicates. While a handful of predicates may suffice to describe all decision problems under uncertainty, or all games, in general there may be more predicates involved. Once again, however, the set of predicates in a description will typically be a strict subset of $\hat{F}$, and will often be finite.

We do not assume that the sets E or F limit their interpretations in any way. In reality, these sets contain formal symbols that are "nicknames", suggesting certain interpretations, such as "player" or "strategy". However, in the formal model all elements of E and of F are devoid of any interpretation.

A k-place predicate can be thought of as a subset of E^k, or a function $f : E^k \to \{0, 1\}$. It will be convenient to extend this definition in a few ways, to allow for more flexibility:

1. First, the set of possible values may be a general set X_0 rather than $\{0, 1\}$. This would allow us to discuss quantitative theories more elegantly.

2. More importantly, we wish to be able to distinguish between a predicate being known not to hold and being undefined for a given k-tuple of entities. For example, in the example of decision under uncertainty above, we may know that $\succsim (a, b)$ is false (if, say, b is strictly preferred to a), and that $\succsim (a, t)$ is meaningless. Hence, it makes sense to add to the range of a function a value that denotes "meaninglessness", say $\circ$.

3. Next, it will be useful to state that it is meaningful to ask what is the value of a function f for a k-tuple of entities, but that this value is not known. Thus, we add a value, denoted $*$, to the range of theories.

4. Finally, it will save on notation if we don't have to worry about the number of arguments each function has (or the number of places in each predicate). To this end, we extend each function f with m arguments to be defined on all of $\cup_{k\geq 1} E^k$, with the convention that $f\left(e^k\right) = \circ$ for all $e^k \in E^k$ $k \neq m$.

Thus, we consider a set X_0 (containing at least two values), and define

$$X = X_0 \cup \{\circ, *\}$$

We also set

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

while conserving on notation by denoting a typical element of $\mathcal{E}$ by e, and define a *description* to be

$$d : F \times \mathcal{E} \to X$$

the property that for every $f \in F$ there exists $m \geq 1$ such that, for every $k \neq m$, and every $e \in E^k$, we have $d(f, e) = \circ$. Unless $d \equiv \circ$, there exists exactly one such m for every $f \in F$, and it will be called the *degree* of f according to d.

Let $D = D(F, E)$ be the set of all descriptions for the set of entities E and predicates F.

The dictator game example In this game we would have

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

$F = \{Player, Act, Outcome, Result, \succsim, May\}$

Thus, the single-place predicate $Player$ is described by

$d(Player, (P1)) = 1, d(Act, (P1)) = 0,$

$d(Player, (0)) = 0, d(Act, (0)) = 1, ...,$

$d(Player, (P1, 0)) = \circ, ...$

saying that $P1$ is a player but not an act, while 0 is an act but not a player, and so forth. Also, the statement $d(Player, (P1, 0)) = \circ$ suggests that, as $Player$ is a single-place predicate, it is a categorical mistake to apply it to a pair of entities, and such an application results in a meaningless statement.

Similarly, a three-place predicate $\succsim$ would be given by values such as

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

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

$d(\succsim, (P1, (99; 1))) = \circ, ...$

reflecting the fact that player $P1$ strictly prefers the outcome $(100; 0)$ to $(99; 1)$, that it is meaningless to ask whether she prefers an outcome (without asking to what), etc.

2.2.2 Compatibility

Two descriptions $d, d' \in D$ are *compatible* if:

(i) for every $f \in F$ and every $e \in \mathcal{E}$,

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

(ii) for every $f \in F$, every $e \in \mathcal{E}$, and every $x, y \in X_0$, if

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

then $x = y$.[3]

[3]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 have to agree on the set of meaningful statements (condition (i)), and among these they cannot assign to the same statement two incompatible values in X_0. However, compatibility allows the descriptions to differ if one of them assumes the value $*$ and the other takes some concrete value $x \in X_0$. Put differently, the value $*$, designating unknown value in X_0, is considered to be compatible with any value in X_0.

The dictator game example The description above does not specify what the outcome of the game is. It thus has

$d(May, (0; 100)) = d(May, (1; 99)) = ...d(May, (100; 0)) = *$

suggesting that which act will be chosen by $P1$ is not yet known. Any description that agrees with d on the values $0, 1, \circ$, but that further specifies $0, 1$ values for $d(May, (i; 100 - i))$ will be compatible with d.

2.2.3 Extensions

For two descriptions $d, d' \in D$ we say that d' is an *extension* of d, denoted $d' \rhd d$, if:

(i) for every $f \in F$ and every $e \in \mathcal{E}$,

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

(ii) for every $f \in F$, every $e \in \mathcal{E}$, and every $x \in X_0$,

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

Thus, a description d' extends a description d if the two agree on what is and what isn't a meaningful statement (condition (i)), and, for all meaningful statements, any statement that appears in d has to appear in d' as well (condition (ii)).

Clearly, if $d' \rhd d$ then d and d' are compatible. However, compatibility implies that d' may assume a specific value $d'(f, e) = x$ for $x \in X_0$ whereas d is silent about it, that is, $d(f, e) = *$, *as well as vice versa.* For d' to be an

	Representation
Language	(F, E)
Input, or question	d
Output, or answer	d'

Figure 1: An illustration of a description. The language of the description is a set of entities E and a set of predicates F describing these entities. The description d identifies which pairs (f, e) are true, false, meaningless, or meaningful but unknown, and can be viewed as providing the input or posing the question that serves as the point of departure for the analysis. Possible outputs or answers consist of extensions d' of d.

extension of d, however, only the former is allowed: d' specifies more values in X_0 than does d.

Note that this formulation allows us to include, in a description d, claims about certain statements being true as well as being false, and demanding that any extension of the description respect these statements. For example, if we wish to say that a is strictly preferred to b we may have $d_1(\succsim, (a, b)) = 1$ and $d_1(\succsim, (b, a)) = 0$. This would be distinguished from the pair $d_2(\succsim, (a, b)) = 1$ and $d_2(\succsim, (b, a)) = *$, where a is weakly preferred to b but nothing is said about the converse preference. A theory that would imply that b is weakly preferred to a ($d'(\succsim, (b, a)) = 1$) would be a contradiction to (and not an extension of) d_1 but compatible with (and an extension of) d_2.

When considering the set of extensions of a description d, we will partially order them by inclusion between the sets of arguments (f, e) for which they assign specific values $x \in X_0$. That is, if the inverse image of $*$ is smaller (in the sense of set inclusion), the description will be defined to be larger. With respect to this relation, it will be meaningful to refer to a minimal extension of d that satisfies certain conditions.

Figure 1 illustrates the concept of a description.

The dictator game example Given

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

$F = \{Player, Act, Outcome, Result, \succsim, May\}$

suppose that a description d provides all information about the game form and the player's preferences, and assume that these are monotone in her own monetary payoff. Then the game prediction will be given by an extension d' of d such that

$d\,(May, (100; 0)) = 1.$

2.3 Reality

Typically, a scientific paper will not formally describe reality as separate from its model—indeed, the model is typically taken to be the formal description of reality. However, since our goal here is to (formally) model the act of (formal) modeling, we need to treat the informal reality the scientist considers as a separate formal object from the model that the scientists constructs. To this end, we will use a similar language to that used above. We start with a set of entities E_R that are supposed to capture concepts in the "real world" being modeled. For example, analyzing an international crisis, the US and Russia might be such entities. Thinking about an economic problem, entities might include the demand for goods, the rate of inflation, a nation's debt, and so forth. Entities in E_R could be thought of as objects that are referred to in a daily newspaper. These are also models of sorts: for example, the "United States of America" is not a palpable object that one can touch or smell; it is a social construct that we have come to refer to as a "real" object despite the fact that it is part of a mental description of our sense data.[4] Because real objects are distinct from formal ones, $E \cap E_R = \varnothing$.

[4]In this sense, even physical objects such as the computer we're typing on are, in the final analysis, mental descriptions, or parts of models that we construct in order to make sense of data and other mental phenomena.

	Reality	Representation
Language	(F_R, E_R)	(F, E)
Input, or question	d_R	d
Output, or answer	d'_R	d'

Figure 2: We model reality as a set of entities E_R and predicates F_R, with d_R characterizing what is known and an extension d'_R characterizing a possible output of answer.

Reality is described by a set of predicates that are also defined in natural language, F_R (where, again, we assume $F_R \cap E_R = \varnothing$). In principle, this set is also disjoint from the set of formal predicates, that is, $F \cap F_R = \varnothing$.[5] 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" may be clearly used both in natural language and in the formal model. However, there are cases where the mapping between real and formal predicates is far from clear. For example, in a political science problem, it is not always obvious who are the agents: are states to be considered players who make decisions, or should their leaders assume these roles? Or, to consider another example, one scientist may view a state of affairs as a final outcome, whereas another conceptualizes it as an act in a model that involves uncertainty. For simplicity, we will therefore assume that the sets of predicates are disjoint.

Facts that are known about reality will be modeled by descriptions $d_R \in D(F_R, E_R)$, defined as above. Similarly, given two descriptions $d_R, d'_R \in D(F_R, E_R)$, we use the definitions of compatibility and extension as above. Figure 2 adds our model of reality to our illustration.

[5] We also assume that $(E \cup E_R) \cap (F \cup F_R) = \varnothing$.

The dictator game example Reality would refer to specific people who interact in the game, say, in a lab experiment. The payoff and acts are the same as in the formal game, but we would use new entities to retain the standard that all of the sets E, F, E_R, F_R are pairwise disjoint. Thus,

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

$F_R = \{Player_R, Act_R, Outcome_R, Result_R, \succsim_R, May_R\}$.

2.4 Models

2.4.1 Connecting Reality and Description

A model consists of (i) a set of entities in the real world and a description of their properties with a set of predicates, (ii) a set of formal entities and a description of their properties with a set of predicates, and (iii) a map allowing one to conceptualize the real-world entities and properties in terms of their formal counterparts.

Formally, a *model* is a quintuple $M = (F_R, E_R, F, E, \phi = \phi_F \cup \phi_E)$ such that:

- $F_R \subset \hat{F}_R$; $E_R \subset \hat{E}_R$; $F \subset \hat{F}$; $E \subset \hat{E}$
- $\phi_F : F_R \to F$ is a bijection.
- $\phi_E : E_R \to E$ is a bijection;

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.

Modeling may take various other forms. For example, the modeler may ignore some values of the description of reality (replacing values $x \in X_0$ by the value $*$). Conversely, it is possible that the formal modeling would add assumptions that have no factual counterparts in reality. For example, if a scientist makes assumptions about players' utility functions, which may not be observable in reality, we would need to include more propositions in the model than are given to us in reality.

2.4.2 Acceptable Models

Importantly, not all bijections ϕ yield reasonable models. First, predicates typically have a number of places defined for them. For example, the symbol "$\succsim$" cannot be applied to a single entity, and thus it would be inappropriate to map a single-place predicate from F_R to $\succsim$. Second, 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 "$Player1$". But an inanimate object such as "money" might not, in most reasonable models.[6]

For a given (F_R, E_R), denoting a phenomenon of interest, a bijection $\phi = \phi_F \cup \phi_E$ defines the sets F, E (as its image). We can therefore uniquely identify a model for (F_R, E_R) by its bijection ϕ. We will denote the set of *acceptable models* for (F_R, E_R) by $\Phi = \Phi(F_R, E_R)$.

For simplicity, we will consider sets of bijections that are determined by predicate bijections only. This corresponds also 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 US as $Player1$ but not as $Player2$.

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 models will be defined by mappings between predicates, allowing for all the bijections between entities. Intuitively, entities are devoid of any content, and anything

[6]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.

we know about them is reflected in the predicates they satisfy.

Formally, let there be given a description of reality, $d_R \in D(F_R, E_R)$, and a set of bijections

$$\Phi_F \subset \left\{ \phi_F \,\middle|\, \begin{array}{l} \phi_F : F_R \to F, \\ \phi_F \text{ is a bijection} \end{array} \right\}.$$

The set of acceptable models $\Phi(F_R, E_R)$ consists of the bijections obtained by the union of a $\phi_F : F_R \to F$ in Φ_F and any bijection $\phi_E : E_R \to E$.

The dictator game example In this example there is little doubt about the identity of the players and their moves. Hence, for any $\phi_F \in \Phi_F$ we should have

$\phi_F(Player_R) = Player$

$\phi_F(Act_R) = Act$

and so forth.

Further, if in the experiment $Mary$ was the dictator, we should have

$\phi_F(Mary) = P1$.

However, the modeler has several decisions to make now. First, she has to decide whether to include the recipient in her model. Formally, she may choose a smaller set of entities, $E' \subset E$, dropping $P2$ (as well as dropping $John$ out of E_R) and analyzing the game as a one-person decision problem. Alternatively, she may leave $P2$ in E and $John$ in E_R, and have

$\phi_F(John) = P2$.

This, however, does not exhaust her choices: one can make the case that it is not obvious that people care only about their own monetary payoffs, and that, for that reason, player $P1$'s preferences might be some other ordering over the 101 outcomes. Hence, the modeler may decide to which such game form the real-life game should be mapped.

2.5 Theories

We can finally define a theory as a mapping between (formal) descriptions that guarantees extension: given a set of descriptions $D = D(F, E)$, a *theory* is a function $T : D \to D$ such that, for all $d \in D$, $T(d) \rhd d$. Hence, a theory takes an existing description that represents reality and adds additional information about various predicates. Note that the domain of the theory are only formal descriptions $D(F, E)$ and not descriptions of reality, $D_R = D(E_R, F_R)$. 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 the President of the US... then...".[7]

2.5.1 Extensionality

Observe that in the above, theories are described in a very concrete, extensional manner. For example, the Nash equilibrium theory was described by spelling out its predictions in a specific game, and, should it be applied to all games, one would need to consider its specific predictions in each and every game. Clearly, this is not the way we usually think of the theory. We cannot conceive of all possible games, and we often find it much simpler to use concise mathematical descriptions of the theory, using variables, parameters, and so forth. However, for simplicity we will not formally describe these descriptions of theories until Section 3.4.1.

Extensional descriptions of theories do not allow us to distinguish between equivalent descriptions of the same theory. Often, the equivalence between two such descriptions may be far from obvious. Indeed, a major role of axiomatizations is to explore such equivalences. In the present paper, axiomatizations, as well as other mathematical results, cannot be modeled. We implicitly assume that the reasoners who think about the models are

[7] As mentioned above, this formal constraint can be circumvented by defining a restricted set of models Φ, whereby a real-life predicate denoting the President maps to a formal predicate corresponding to it. However, it seems more natural to define theories on formal elements alone, thereby suggesting a certain claim to generality.

unboundedly rational and can translate every theory to all its implications in all possible instances. Such reasoners would not have understood why, for example, Savage had to publish his axiomatic derivation of expected utility theory. Clearly, our model can be extended to be more realistic by allowing different representations of a theory, and capturing the framing effects for which axiomatizations are interesting.

2.5.2 Using Theories to Examine Reality

Given a model $M = (F_R, E_R, F, E, \phi = \phi_F \cup \phi_E)$, a description of reality, $d_R \in D(F_R, E_R)$, and a theory T, one can extend d_R by the model M and the theory T, to define the M-T-extension of d_R, $d'_R(d_R, M, T)$ as follows:

(i) define a description $d \in D(F, E)$ by $d(f, e) = d_R\left(\phi^{-1}(f), \phi^{-1}(e)\right)$ for all $(f, e) \in F \times E$; denote this description by $d_R \circ \phi^{-1}$;

(ii) apply T to obtain an extension of d, $d' = T(d)$

(iii) define a description $d'_R = d'_R(d_R, M, T) \in D(F_R, E_R)$ by $d'_R(f, e) = d'(\phi(f), \phi(e))$ for all $(f, e) \in F_R \times E_R$; denote this description $d' \circ \phi$.

Figure 3 illustrates a trivial but important point given by the following.

Proposition 1 *For every model $M = (F_R, E_R, F, E, \phi)$, description $d_R \in D(F_R, E_R)$, and theory T, the M-T-extension of d_R,*

$$d'_R(d_R, M, T) = T\left(d_R \circ \phi^{-1}\right) \circ \phi \in D(E_R, E_R)$$

is an extension of d_R.

The dictator game example In this example the theory T that players play a subgame perfect equilibrium will boil down to utility maximization by player $P1$. As there is no uncertainty involved, the theory T will only take a description of player $P1$'s preferences in d and will add the statement that she will choose the best outcome based on these preferences. That is, should

$$\begin{array}{lccccl} & \text{Reality} & & \text{Representation} & \\ \text{Language} & (F_R, E_R) & & (F, E) & \\ \text{Input, or question} & d_R & \xrightarrow{\phi} & d & = d_R \circ \phi^{-1} \\ & & & \downarrow & T \\ \text{Output, or answer} & d'_R & \xleftarrow{\phi^{-1}} & d' & = T(d) = T(d_R \circ \phi^{-1}) \end{array}$$

Figure 3: An illustration of how a theory is used to draw conclusions about reality. Beginning with a description of reality d_R, we use the model M to construct the formal description d satisfying $d(f, e) = d_R(\phi^{-1}(f), \phi^{-1}(e))$. The theory T then gives the extension $T(d)$, at which point we again use the model to construct the description of reality given by $d'_R = d'_R (d_R, M, T) = T \left(d_R \circ \phi^{-1}\right) \circ \phi$. d'_R is the M-T-extension of d_R.

$i \in \{0, ..., 100\}$ satisfy

$$d\left(\succsim, \left(P1, (i; 100 - i), (j; 100 - j)\right)\right) = 1$$

for all $j \in \{0, ..., 100\}$, the theory will add to d the value

$$d\left(May, (i; 100 - i)\right) = 1.$$

Note that the previous lines describe an informal algorithm, by which we convey the value of T for every d and every replica of the game form (one for each possible preference relation of $P1$). As promised, we will later introduce more concise representations of such theories.

3 Testing and Applying Theories

3.1 Model-Dependent Definitions

Let there be given a description of reality $d_R \in D(F_R, E_R)$ and a theory T. We are interested in comparing the predictions or recommendations of T, that is, the theoretical extension of d_R, with other extensions thereof. For example, given the data d_R, theory T may suggest predictions that are an extension of d_R. These may be contrasted with new observations, which are an extension of the data d_R obtained empirically. Or, to consider a normative interpretation, given a description of a problem d_R, theory T may provide a recommendation, which assigns values to decision variables, thereby defining an extension of d_R. The latter can be compared with the decision maker's intuition, which is a non-theoretical extension of d_R.

To make this idea precise, we begin with a specific model

$$M = (F_R, E_R, F, E, \phi = \phi_F \cup \phi_E) .$$

Recall that such a model defines the M-T-extension of d_R, $d'_R(d_R, M, T)$ to be

$$d'_R = T\left(d_R \circ \phi^{-1}\right) \circ \phi \in D(F_R, E_R) .$$

Let us now assume that there exists another extension d''_R of d_R. We view this extension as arising out of new data, or normative considerations, intuition or introspection, ethical principles, ideology, or some other process. For simplicity, assume that there is a function $N : D(F_R, E_R) \to D(F_R, E_R)$ such that, for all $d_R \in D(F_R, E_R)$, $N(d_R) \rhd d_R$. The extended description $d''_R = N(d_R)$ might include additional observations, people's intuitions, decision maker's hunches, and so forth. Importantly, the function N is not supposed to be theory-driven, but rather to be contrasted with the theoretical extension.

How do we use the process N and the resulting description d''_R to evaluate a theory? The following two definitions will be useful. 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).

A theory T is *M-compatible with d''_R* if the the M-T-extensions d'_R and d''_R are compatible descriptions.

$$
\begin{array}{rccccl}
 & \text{Reality} & & \text{Representation} & & \\
\text{Language} & (F_R, E_R) & & (F, E) & & \\
\text{Input, or question} & d_R & \xrightarrow{\phi} & d & = d_R \circ \phi^{-1} & \\
N & \downarrow & & \downarrow & T & \\
 & d''_R & & & & \\
\text{Output, or answer} & d'_R & \xleftarrow{\phi^{-1}} & d' & = T(d) = T(d_R \circ \phi^{-1}) & .
\end{array}
$$

Figure 4: Illustration of how data, normative considerations, or other considerations are used to evaluate a theory. Given a description d_R, the model M and theory T are used to construct its M-T-extension d'_R, as described in Figure 3. The process N, perhaps representing the collection of additional data, generates the extension d''_R. We say that the theory T is weakly compatible with d''_R if there is at least one acceptable model M for which the M-T-extension d'_R is compatible with d''_R, and the theory T is strongly compatible with d''_R if it is the case that for every acceptable model M, the M-T-extension d'_R is compatible with d''_R.

For a given $(f,e) \in F_R \times E_R$ such that $d_R(f,e) = *$, and a value $x \in X_0$, define $d''_R(f,e,x)$ by the minimal extension of d_R such that $d''_R(f,e) = x$. A theory T M*-necessitates* (f,e,x) if d'_R is an extension of $d''_R(f,e,x)$.

3.2 Model-Independent Definitions

The important point in Subsection 3.1 is that in the absence of a specific model M, it is not clear whether a theory is or is not compatible with observed data, nor whether it necessitates certain conclusions. It might do so for some models but not for others. We therefore need to consider the way one deals with the freedom in selecting M.

Three possibilities arise:

(i) Using a strong notion in which a concept applies for all acceptable models;

(ii) Using a weak notion in which a concept applies for at least one acceptable model;

(iii) Suggesting some aggregation over models, weighing the set of models for which the concept applies vis-a-vis the set for which it does not.

We begin with the first two options. First, consider the notion of compatibility. Given a description of reality $d_R \in D(F_R, E_R)$, an extension thereof d''_R, and a theory T, T is *strongly compatible* with d''_R if it is M-compatible with d''_R for every model $M = (F_R, E_R, F, E, \phi)$ derived from an acceptable $\phi \in \Phi$. T is *weakly compatible* with d''_R if it is M-compatible for at least one such model M.

Given a description of reality $d_R \in D(F_R, E_R)$, a pair $(f,e) \in E_R \times F_R$ such that $d_R(f,e) = *$, a value $x \in X_0$, a conclusion (f,e,x) and a theory T, we say that T *strongly necessitates* (f,e,x) if, for every $\phi \in \Phi$, T M-necessitates (f,e,x) for $M = (F_R, E_R, F, E, \phi)$. T *weakly necessitates* (f,e,x) if, there exists $\phi \in \Phi$, such that T M-necessitates (f,e,x) (for M as above).

Figure 4 illustrates these concepts. Note that our notions of compatibility – both strong and weak – demand that the extension obtained from the theory, d'_R, be compatible with the extension obtained otherwise, say, by the function N. Thus, if the diagram in Figure 4 commutes for all (some) model $\phi \in \Phi$, we would surely have that T is strongly (weakly) compatible with $d''_R = N(d_R)$. However, these notions of compatibility do not require equality of d''_R and d'_R, and are thus a weaker condition than that Figure 4 commute.

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.

3.3 Analogies: Aggregation of Models

Rather than 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, Postlewaite, Samuelson and Schmeidler [9]. We suggest in that paper 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 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.

This view suggests that economic models are not quite scientific in the Popperian sense. For example, there is no sense in which a theory can be wrong in this view of economics. Upon observing data that contradicts the prediction of a theory, the economist simply adds the data to her stock of cases, alongside the theory. Both are used as inputs when the economist is next called upon to make a prediction, with relative weights determined by their relatively similarity to the next prediction problem. Moreover, economists typically do not offer a well-defined algorithm for how conclusions are to be drawn from their cases, let alone for how conflicting conclusions of such cases are to be aggregated. If such algorithms were provided, case-based (analogical) reasoning could be refutable just as standard, rule-based reasoning is perceived to be. Without such algorithms, economic modeling is "pre-scientific".

We could extend the current model to capture such case-based reasoning. Toward this end, an economic model, or a "theoretical case" might be stated as a theory that generates predictions in a formal context, and the analogical mapping would be defined by the function ϕ of a model M. The aggregation of many cases would correspond to aggregation of models.

3.4 Complexity Results

How easy is it to identify whether a theory is strongly or weakly compatible with data, or whether a theory strongly or weakly necessitates a conclusion?

3.4.1 Identifying Theories

To answer this question in an intuitive way, we first require a more parsimonious description of a theory. A theory can be identified by a set of rules, where each rule has antecedents and consequents and each rule is understood to apply to all descriptions that satisfy the antecedents. Consider, for example, transitivity of a preference relation $\succsim$. Transitivity is naturally defined by the conditional statement that, (for all a, b, c) *if* $\succsim (a, b)$ and $\succsim (b, c)$ *then* $\succsim (a, c)$. In the formulation we use, this takes the form that, for every d such that

$$\begin{aligned} d\left(\succsim, (a,b)\right) &= 1 \\ d\left(\succsim, (b,c)\right) &= 1 \end{aligned}$$

(the extension) $T(d)$ satisfies

$$\begin{aligned} d\left(\succsim, (a,b)\right) &= 1 \\ d\left(\succsim, (b,c)\right) &= 1 \\ d\left(\succsim, (a,c)\right) &= 1. \end{aligned}$$

This specifies some of the values of $T(d)$ for many descriptions d, which can vary in terms of preferences ($\succsim$) between other alternatives (apart from a, b, c), as well as in terms of many other predicates. It will be convenient to specify transitivity by the rule above, with the understanding that it defines $T(d)$ as the minimal extension of d that satisfies the conditional statement.[8]

Formally, a *rule* is a pair of descriptions $r = (d, d')$ such that $d' \rhd d$. For a set of rules $\mathcal{R}$, the theory $T = T(\mathcal{R})$ defined by $\mathcal{R}$ is the function $T : D \to D$ defines as follows: For every d_1

(i) If $d_1 \not\rhd d$, $T(d_1) = d_1$;

(ii) If $d_1 \rhd d$, $T(d_1)$ is the minimal joint extension of d_1 and of d' for each $r = (d, d') \in \mathcal{R}$.

The dictator game example The informal representation of preference-maximization used above can now be made formal: for every $i \in \{0, ..., 100\}$ *if*

$$d\left(\succsim, \left(P1, (i; 100 - i), (j; 100 - j)\right)\right) = 1$$

for all $j \in \{0, ..., 100\}$, *then*

$$d\left(May, (i; 100 - i)\right) = 1.$$

[8]The availability of concise descriptions of this nature will be crucial for our complexity results.

3.4.2 Results

Given that theories can be concisely represented by rules, we find that determining whether a theory weakly or strongly necessitates an observation is a computationally difficult problem.

Proposition 2 *Given a description of reality $d_R \in D(F_R, E_R)$, a pair $(f, e) \in F_R \times E_R$ such that $d_R(f, e) = *$, a value $x \in X_0$, a conclusion (f, e, x), a theory T, and a set Φ, it is NP-Hard to determine whether T weakly necessitates (f, e, x).*

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

Proposition 3 *Given a description of reality $d_R \in D(F_R, E_R)$, a pair $(f, e) \in E_R \times F_R$ such that $d_R(f, e) = *$, a value $x \in X_0$, a conclusion (f, e, x), a theory T, and a set Φ, it is NP-Hard to determine whether T strongly necessitates (f, e, x).*

4 Applications

In this section we attempt to clarify the distinctions we introduced in Subsection 1.2 with the help of our formal model. In all four cases, there is a "classical" view of the role of science: one starts with a description, d_R, uses a model M and a theory T to obtain its extension $d'_R(d_R, M, T)$, and refers to this extension as the prediction or recommendation of the theory. Ideally, there is little doubt about the appropriate model M, and thus it is quite clear what descriptions d'_R are necessitated by T. A reasoner (predictor) or a decision maker may have her own, non-theoretical extension of d_R, d''_R, which may be based on intuition, habit, etc. It is then easy to verify whether d''_R is compatible with T or even necessitated by it.

However, in all four applications it may well be the case that T does not strongly necessitate d''_R, or that it does not strongly necessitate any non-trivial extension of d_R at all. Indeed, T may not even be strongly compatible with any such extension. In this case one may use theory T only to verify, post-hoc, if there is a way to justify d''_R by *some* model. That is, to test whether intuition or habit that are independent of the theory are weakly necessitated by it.

4.1 Theories and Paradigms

As indicated above, an empirical test of a theory T can be conceptualized as posing a problem in a description d_R (which might be empirical or experimental), applying a model M to obtain a set of the theory's predictions in an extension $d'_R(d_R, M, T)$, and contrasting these predictions with observed facts d''_R. Having done this, how do we draw conclusions about the theory? We consider two possibilities.

4.1.1 Strong Compatibility

One view is that the theory will be considered well-established if no acceptable model $M \in \mathcal{M}$ results in refutation, that is, if the observed phenomena d''_R are strongly compatible with T. This could arguably be the case for Newtonian physics within some domain restrictions. Within certain ranges of measurable parameters (such as energy, mass, etc.), bodies obey Newton's laws, whether these bodies are planets or billiard balls. Refining the theory's domain, such as limiting it to certain levels of energy, could be viewed as making it (strongly) compatible with predictions it does not (weakly) necessitate.

In this Popperian description of the refinement of the theory, the choice of the model $M \in \mathcal{M}$ plays a minor role. An established theory is compatible with each of the models in $\mathcal{M}$, and no model is particularly important in interpreting the theory.

4.1.2 Weak Compatibility

Many theories in economics are not subjected to such a strong compatibility test. To go back to the dictator game example, the theory that agents maximize their respective utility functions might be viewed as violated by this example, if utility is only defined over monetary outcomes. But the observations may be explained if the "dictator" preferences can be non-monotonic in her own payoff, possibly taking into account the other player's payoff via considerations of equity, altruism, and so forth. In fact, the way we have described the dictator game above, it will be weakly compatible with any possible observation, since any observed choice may well be a maximizer of player $P1$'s preferences.

Observe that the freedom one has in modeling a laboratory experiment of the dictator game does not end with choosing player $P1$'s preferences. As mentioned above, one may or may not include player $P2$ as a formal entity in the game. Moreover, slightly more involved games, such as the ultimatum game, will also offer more freedom in the definition of an outcome. Specifically, an outcome might describe the path leading to it, where, for example, player $P2$ can view a certain monetary split differently depending on who determined the split and what other options she had had, allowing for emotional payoffs such as humiliation, vengeance, regret, and so forth.

4.1.3 Is Economic Theory Vacuous?

Allowing ourselves a casual observation, 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 a greater freedom in selecting a model ϕ (in the sense of our formal model). 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 models $\mathcal{M}$. 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 the refutability of economics and its status as a science, as well as some of debates revolving around behavioral economics have to do with the distinction between weak and strong compatibility (or necessitation). Detractors of the field point to phenomena where some acceptable models are not compatible with the data. Some responses have been along the Popperian lines (Popper, 1934), 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 are 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 re-definition 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.[9]

This discussion suggests that decision and game theory should be viewed as conceptual frameworks or as paradigms rather than as specific theories. As a rough approximation, one can view theories as refuted as soon as *one* of the appropriate models in Φ is at odds with observations. By contrast, a paradigm is rejected only when *all* such models are contradicted by evidence. Stated differently, we expect theories to be strongly compatible with the data, whereas paradigms need only be weakly compatible with the data.

[9] At the same time, re-definition 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 model $\phi \in \Phi$ but not given others.

However, this approach, and the examples in 1.2.1, 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 models Φ, which is taken to be exogenous in this paper. It is common sense that has to determine the scope of Φ. We believe that in this sense our model can capture 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 isn't whether it has been refuted, but rather, whether the set of mappings Φ hasn't become embarrassingly large.

4.2 Objective and Subjective Rationality

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 model $M \in \mathcal{M}$, 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 model $M \in \mathcal{M}$ 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.

4.3 The Role of Decision-Aid Models

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 models $\mathcal{M}$, 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 counter-intuitive 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 d''_R of 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.

4.4 The Role of Economics

The classical model of science would leave no room for intuition. A description of reality, d_R, is given, it is modeled by a model M about which there is little room for debate, economic theory T is applied to generate predictions and recommendations, and these are mapped back into reality. The corresponding predictions and recommendations are strongly necessitated by T, either because there is but one model in $\mathcal{M}$, or because the various models in $\mathcal{M}$ aren't so different from each other and end up converging on their predictions and recommendations.

By contrast, it is possible that the theory can be applied in a variety of ways, and that no non-trivial extension of d_R is strongly necessitated by T. In this case, politicians and journalists might still come up with predictions and policy proposals. The economist can then check, post-hoc, whether these are consistent with economic lore, that 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, and there is no reasonable model M that supports it, one may well wonder whether these are reasonable guidelines to follow.

5 Conclusion

5.1 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 model M. 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 a model, but cannot derive any new conclusions from the union of the theories.

5.2 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" rather than "is". But what is this "ought"? Who has the right to determine what ought to be done? Are social scientists such as economists to be viewed as religious preachers? If so, by what authority do they obtain this role?

It is our view that normative economics, as well as normative statements in decision or game theory, should ultimately derive their validity from the people or societies they discuss. Theorists may serve an important role in helping these individuals or societies figure out their own preferences, but they should not determine these preferences. The present paper suggests that there may be more than one way in which the dialog between a theorist and her subject may be conducted. On the one end, there is the objective model, in which the theorist can convince any reasonable person, compute the optimal decision, and generate refutable (but hopefully unrefuted) predictions that may guide decision making. At the other extreme the theorist only tests the coherence of decision makers' decisions, of economic predictions and so forth. We suggest that students and practitioners may benefit from awareness of these two extreme models, and from finding value in analysis even if it does not lead to refutable, quantitative predictions.

5.3 The Discipline of Economics

Research is a social phenomenon, involving people deciding 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 paper 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 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 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.

5.4 Economics as Criticism and as Case-Based Reasoning

Relative to the view 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: while 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.

5.5 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. While we do believe this to be the case, it is important to point out that the choice of a model and re-definition 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...). Clearly, this theory has counter-examples in observed data. However, Browning, Chiappori and Weiss (2014, Chapter 3) argue 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 much more reasonable a hypothesis than if the household is viewed as a single unit. That is, while the standard approach is to map a specific household to a single "consumer", this paper suggests 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, while 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.

6 Appendix: Proofs

Proof or Proposition 2.

We reduce the Clique problem to Weak Necessitation. Let there be given an undirected graph (V, A) with $|V| = n$ and a number k $(1 < k \leq n)$. The set $A \subset V \times V$ denotes the set of arcs.[10] We assume that $(v, v) \notin A$ for all v and that $(v, w) \in A$ implies $(w, v) \in A$. Construct the following problem. Set $X_0 = \{0, 1\}$. 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 $F_R = \{B, A\}$ where B, K are single-place predicates and L, A are two-place predicates. We abuse notation and use A for a predicate in our model because it will be identical to the arcs in the graph (V, A). We set $\Phi_F = \{\phi_F\}$ where $\phi_F(B) = K$ and $\phi_F(A) = L$. 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 $d_R(A, (v, w)) = 1$ iff $(v, w) \in A$. For $v \in V$ set also $d_R(A, (v, y)) = d_R(A, (y, v)) = 0$. Likewise, $d_R(A, (y, y)) = 0$.

[10] We avoid the term "edge" as E stands for entities in our model.

Finally, set $d_R(B,e) = *$ for all $e \in E_R$. Let $E = \{1,...,n+1\}$.

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

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

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

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

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 M-necessitates $(B,y,1)$ for $M = (E_R,F_R,E,F,\phi)$. 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 $T(d)(K,n+1) = 1$ and the mapping back implies that the extension of d_R, d'_R, satisfies $d'_R(B,y) = 1$.

Conversely, if T weakly necessitates $(B,y,1)$, 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 $\{1,...,k\}$ 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.□

Proof of Proposition 3.

We will reduce the (closed) Hamiltonian path problem and prove that Strong Necessitation is co-NPC. That is, given an undirected graph (V,A) we will construct $d_R \in D(E_R,F_R)$, a pair (f,e) with $d_R(f,e) = *$, a value $x \in X_0$, a conclusion (f,e,x), and a theory T such that T strongly necessitates (f,e,x) if and only if the original graph does *not* have a closed Hamiltonian path. Again, set $X_0 = \{0,1\}$. As in the proof of the previous result, $F =$

$\{K, L\}$ and $F_R = \{B, A\}$ where B, K are a single-place predicates and L, A are two-place predicates. Again, we abuse notation and use A for a predicate in our model because it will be identical to the arcs in the graph (V, A). We set $\Phi_F = \{\phi_F\}$ where $\phi_F(B) = K$ and $\phi_F(A) = L$.

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

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

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

To see this, consider a permutation of the nodes $\phi_E : E_R(\equiv V) \rightarrow E(\equiv \{1, ..., n\})$. If this permutation defines a closed Hamiltonian path, the theory cannot be applied (because $d(L, (i, i+1)) = 1$ for all i) and it doesn't provide any non-trivial extension of d. Consequently, nothing can be added to d_R and, in particular, we remain with $d_R(B, v) = * \ \forall v \in V$. Thus, if the graph (V, A) contains a Hamiltonian path, at least one possible model (permutation of the nodes) will not result in $d_R(B, v) = 1$ and therefore $(B, v_1, 0)$ 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 $(i, i+1) \notin A$, or $d(L, (i, i+1)) = 0$, and thus the theory would yield $d(K, j) = 0$ for all j. Mapping this conclusion back, we obtain $d_R(B, v) = 0$ for all v, and, in particular, for v_1. As this holds for every mapping ϕ, the conclusion $(B, v_1, 0)$ is strongly necessitated by T.

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

References

[1] Enriqueta Aragones, Itzhak Gilboa, Andrew Postlewaite, and David Schmeidler. Rhetoric and Analogies. , mimeo., 2001.

[2] Robert J. Aumann. What Is Game Theory Trying to Accomplish? in K. Arrow and S. Honkapohja (Eds.), Frontiers of Economics (pp. 5–46). 1985.

[3] Martin Browning, Pierre-Andre Chiappori and Yoram Weiss. *Economics of the Family.* Cambridge University Press, Cambridge, 2014.

[4] S. Brue. *The Evolution of Economic Thought, 5th edition.* Houghton Mifflin Harcourt, 1993.

[5] Nancy Cartwright. Capacities. *in J. Davis, W. Hands, and U. Maki (Eds.), The handbook of economic methodology* (pp. 45–48). Cheltenham: Edward Elgar, 1998.

[6] Nancy Cartwright. Models: Parables vs. Fables. *in Beyond Mimesis and Convention: Representation in Art and Science, Roman Frigg, Matthew Hunter eds.* In Press.

[7] Milton Friedman. The Methodology of Positive Economics. In *Essays in Positive Economics*, pp. 3–43, 1953.

[8] Allan Gibbard and Hal R. Varian. Economic Models. *The Journal of Philosophy*, 75(11): 664–677, 1978.

[9] Itzhak Gilboa, Andrew Postlewaite, Larry Samuelson and David Schmeidler. Economic Models as Analogies. *Economic Journal*, Forthcoming.

[10] Itzhak Gilboa and David Schmeidler. *A Theory of Case-Based Decisions.* Cambridge University Press, Cambridge, 2001.

[11] Donald Gillies. The Duhem Thesis and the Quine Thesis. *in Martin Curd and J.A. Cover ed. Philosophy of Science: The Central Issues*, New York: Norton 302–319, 1998.

[12] Werner Guth, Rolf Schmittberger, and Bernd Schwarze An Experimental Analysis of Ultimatum Bargaining. *Journal of Economic Behavior and Organization* , 3(4): 367–388, 1982.

[13] Till Grune-Yanoff. Learning from Minimal Economic Models. *Erkenntnis*, 70(1): 81–99, 2009.

[14] Till Grune-Yanoff and Paul Schweinzer The Roles of Stories in Applying Game Theory. *Journal of Economic Methodology* , 15: 131–146, 2008.

[15] Daniel Hausman. *The Inexact and Separate Science of Economics*. Cambridge University Press, Cambridge, 1992.

[16] Daniel Kahneman and Amos Tversky. Prospect Theory: An Analysis of Decision Under Risk. *Econometrica*, 47: 263–291, 1979.

[17] Thomas S Kuhn. *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago, 1962.

[18] Uskali Maki. Isolation, Idealization and Truth in Economics. *Poznan Studies in the Philosophy of the Sciences and the Humanities* 38: 147–168, 1994.

[19] Uskali Maki. Models are Experiments, Experiments are Models. *Journal of Economic Methodology* 12: 303–315, 2005.

[20] Deirdre McCloskey. *The Rhetoric of Economics*. University of Wisconsin Press, Madison, 1985.

[21] Karl R. Popper. *Logik der Forschunghe: The Logic of Scientific Discovery*. London: Hutchinson and Co. 1934 (English Edition 1958).

[22] Howard Raiffa and David Metcalfe. *Negotiation Analysis: The Science and Art of Collaborative Decision Making.* . Cambridge: Harvard University Press, 2002.

[23] Ariel Rubinstein. Dilemmas of an Economic Theorist. *Econometrica* 74: 865–883, 2006 (Presidential Address to the Econometric Society, Madrid 2004).

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

[25] Daniel Kahneman and Amos Tversky. The Framing of Decisions and the Psychology of Choice. *Science*, 211: 453–458, 1981.