Theoretical Economics 12 (2017), 493–511 1555-7561/20170493

# General revealed preference theory

Christopher P. Chambers
Department of Economics, University of California, San Diego

Federico Echenique
Division of the Humanities and Social Sciences, California Institute of Technology

Eran Shmaya
Kellogg School of Management, Northwestern University

We generalize the standard revealed preference exercise in economics, and prove a sufficient condition under which the revealed preference formulation of an economic theory has universal implications and when these implications can be recursively enumerated. We apply our theorem to two theories of group behavior: the theory of group preference and the theory of Nash equilibrium.

Keywords. Revealed-preference theory.

JEL classification. D0.

## 1. Introduction

Economic theories have observable and unobservable components, and one can say that an observable data set is consistent with the theory if *there exists* some specification of the unobservables that is consistent with the theory. The statement that observables are consistent if there exists unobservables that are consistent with the theory is the "as if" or "revealed preference" formulation of the theory. However, the revealed preference formulation of a theory may not be useful as an empirical test of the theory.

For example, consider the theory of utility maximization. One can observe the choices made by an agent, and ask if there exists some utility function (unobservable) that is consistent with the choices and with the theory of utility maximization. To actually use this formulation as a test, one would have to check all possible utility functions and see if they can explain the data. This is a problem because there are infinitely many utility functions. After one has checked any finite set of utilities and verified that none of them can explain the data, one cannot conclude that the data are inconsistent with the

Christopher P. Chambers: cpchambers@ucsd.edu
Federico Echenique: fede@hss.caltech.edu
Eran Shmaya: e-shmaya@kellogg.northwestern.edu
Chambers and Echenique acknowledge support from the NSF through Grant SES-0751980. We are grateful to seminar audiences at Arizona State University and Brown University. An anonymous referee provided extremely helpful feedback that greatly improved the paper. We are especially grateful to Kim Border, Faruk Gul, and Philippe Mongin for many detailed comments and suggestions.


DOI: 10.3982/TE1924

theory. Another example is profit maximization in a model of industrial organization: one may observe production and pricing decisions, and ask if there is some specification of firm technologies such that the observations are consistent with Nash equilibrium. Again, this revealed preference formulation does not give a practical (or effective) test of the theory because there are infinitely many possible technologies.

The contribution of our paper is to give a sufficient condition on a theory, under which the revealed preference formulation of the theory enables a practical test. The existence of a test will translate into a "universal" formulation of the theory. "Practical" will translate into "effective." We shall discuss these two terms.

The first is universality. We said that observable data are consistent with the theory if *there exists* some specification of the unobservables that makes the data consistent with the theory. This formulation of the revealed preference question is existential because it starts with "there exists." Because it is existential, no finite number of utility functions that fail to explain the data constitute evidence that the data are inconsistent with the theory.

A testable formulation will instead start with "for all," and therefore be a universal formulation. The following examples from Popper (1959) illustrate the basic ideas. Suppose that theory E claims, "There is a black swan," while theory U says, "All swans are white." Theory E, an existential theory, is not falsifiable because no matter how many finite data sets of non-black swans we find, it is still possible that there is a black swan somewhere. Theory U, a universal theory, is falsifiable because the observation of a single non-white swan contradicts the theory.

The second term is effectiveness. For universal theories, effectiveness means that there has to be an algorithm that detects, after a finite number of steps, whether a data set is inconsistent with the theory. An effective test comes with an algorithm that one can run on a data set and that will stop after a finite number of steps when the data are incompatible with the theory. Later (see Example 12) we provide an example of a universal theory that is not effective in this sense. For such a theory, even if a data set is inconsistent with the theory, there may not be a way to demonstrate the inconsistency.

Our paper provides a very general result on revealed preference theory. Our result says that whenever an economic model has a certain kind of axiomatization, its revealed preference formulation can be translated into a universal and effective test for the theory. Essentially our result says the following. Consider the theory, with its observable and unobservable components. An axiomatization of the theory can talk about observable and unobservable components. Whenever the theory has a universal and effective axiomatization, then there is a "projection" of the axiomatization onto the theory's observable components. The projection gives a universal and effective test for the theory. Moreover, as we illustrate in this paper, this condition on a theory is widely applicable in economics.

In fact, many papers in revealed preference theory set out to accomplish universal and effective tests for particular economic models. One of the best known examples is Brown and Matzkin (1996), who show that there exists a test for general equilibrium theory. There are many other papers with a similar agenda, but no general result that encompasses all of them. Our main theorem gives a general result that is applicable to all economic environments studied in the revealed preference literature.

We proceed to illustrate our framework and our result by considering the example of utility maximization in more detail. The idea is that there is a theoretical object (a preference or a utility) that is not observable, but that places restrictions on observable data. The theory was originally developed for consumer choice (Samuelson 1938, Houthakker 1950): the restrictions placed on data by the theory are captured by the *strong axiom of revealed preference* (SARP). SARP is the universal and effective test we are talking about. In principle, one needs to check all utility functions before one decides that a data set is inconsistent with the theory of utility maximization, but the result of Samuelson and Houthakker means that one can check SARP instead of checking all possible utility functions. The purpose of our paper is to give a general result, in the same spirit as Samuelson and Houthakker's, establishing the existence of universal and effective tests for the revealed preference formulation of an economic theory.

We use a simple and abstract formulation of the problem of consumer choice (Richter 1966). Assume that one can observe the binary comparisons (weak and strict) between objects made by an agent. Let us refer to the observed comparisons as $R$ and $P$: $R$ is a binary relation, often called a revealed weak preference, and $P$ is a revealed strict preference.

The theory of preference maximization says that the agent has a weak order (complete and transitive relation) governing these comparisons (what economists call a rational preference). Each weak order $\leq$ is associated with its strict part, $<$. Thus, we posit that *there exists* a pair of binary relations $\leq$ and $<$, which are theoretical (and hence not directly observable) for which the following axioms are satisfied:

Axiom 1 (Completeness). $\forall x \forall y (x \leq y \vee y \leq x)$.

Axiom 2 (Transitivity). $\forall x \forall y \forall z (x \leq y \wedge y \leq z \rightarrow x \leq z)$.

Axiom 3 ($<$ strict part of $\leq$). $\forall x \forall y (x < y) \leftrightarrow (x \leq y) \wedge \neg(y \leq x)$.

Axiom 4 (The $R$-rationalization). $\forall x \forall y (x \; R \; y \rightarrow x \leq y)$.

Axiom 5 (The $P$-rationalization). $\forall x \forall y (x \; P \; y \rightarrow x < y)$.

We wish to emphasize that the theory hypothesizes the existence of unobservable $\leq$ and $<$ for which the axioms in the list are satisfied, and that all of Axioms 1–5 are universal. The first point means that the theory does not directly provide a test of when observed revealed preference relations $R$ and $P$ are inconsistent with preference maximization. There are infinitely many weak orders $\leq$. So for any $\leq$ that cannot explain $R$ and $P$, there may exist a different weak order that can explain it. In other words, this formulation of the theory does not provide a proof that an inconsistent pair $R$ and $P$ is indeed inconsistent with preference maximization.

The second point to emphasize is that all of Axioms 1–5 are universal: This is clear as they begin with the universal quantification $\forall$. Following Popper, then, *if one could* *observe* $\leq$ and $<$, so that all of $R$, $P$, $\leq$, and $<$ were observable entities, then the theory described by Axioms 1–5 (the axioms of rationalization with weak order) would be universal and, therefore, falsifiable.

Now, what does revealed preference theory say about Axioms 1–5? It says that $R$ and $P$ are consistent with Axioms 1–5 for some weak order $\leq$, if and only if $R$ and $P$ jointly satisfy the following countably infinite list of axioms:

The Strong Axiom of Revealed Preference (SARP). *We have that*

$$\forall x_1 \dots \forall x_k \quad \neg \bigwedge_{i=1}^{k} (x_i \, S_i \, x_{(i+1)_{\operatorname{mod} k}})$$

*for every $k$ and every $S_1, \dots, S_k$, where $S_1 = P$ and $S_i \in \{R, P\}$ for all $i \in \{2, \dots, k\}$.*

Here we wish to again emphasize several points. First of all, the strong axiom (formally, a countably infinite collection of axioms) is also universal, but unlike Axioms 1–5, it does not refer to the unobservable objects $\leq$ or $<$. It is a statement only about the observable $R$ and $P$. Second, there is an algorithm that decides whether an observable data set satisfies the strong axiom of revealed preference. So the strong axiom constitutes a universal and effective axiomatization of the theory of utility maximization.[1]

The purpose of our paper is to generalize these results beyond utility maximization. We prove that if a theory hypothesizes the existence of a collection of unobservable relations, but it does so in such a way that the theory would have a universal and effective axiomatization were these relations observable, then the theory has an equivalent universal and effective axiomatization purely in terms of observables. Put differently, if the theory has a universal and effective axiomatization when unobservable relations are assumed to be observable, then there is a universal and effective axiomatization that only refers to observables.

One such equivalent universal axiomatization consists of all the logical consequences of the original theory that are universal and refer only to observables. As in the case of preference maximization and SARP, it is straightforward to establish that the universal consequences referring only to observables must be satisfied by the theory. The converse, that if all universal consequences are satisfied, then there exists unobservable relations such that the original universal axiomatization is satisfied, relies on the axiom of choice.

This result is not trivial; it is possible to write down theories involving unobservables whose projection onto observables has no axiomatization whatsoever. For example, consider a theory claiming that every man in the universe can be matched to exactly one woman. We can describe the revealed preference formulation of the theory as involving statements about who is a man, who is a woman, and who is matched to whom.

[1]We could repeat the exercise assuming only that revealed weak preference $R$ were observable, and seek rationalization by a linear order (complete, transitive, and antisymmetric). This would involve introducing only one new symbol, $\leq$, which would be required to satisfy antisymmetry. Further, the requirement of $P$-rationalization would be dropped. The resulting version of the strong axiom would be similar, except that *all* instances of $S_i$ would be $P$.

Suppose now that the matching itself is not observable. Given that the universe of men and women is infinite, there is no axiom we could write down that would preclude the set of men and women from each being infinite, but of different cardinalities. And this could never be observed with finite data. The reason our result would fail in this case is because if the matching function were observable, then the theory would hypothesize existential statements; that for each man, *there exists* a unique woman to whom he is matched. We discuss this in more detail in Remark 5.

After laying out the formal structure of the model, and presenting the main result (Theorem 1), we demonstrate applications for two economic approaches to collective decision making. We first discuss an application to preference aggregation in which group preferences are assumed to be some function of individual preferences. Then we turn to a framework in which group choice is modeled as the outcome of strategic interaction between the agents, and discuss the testable implications of Nash equilibrium.

Section 6 describes the meaning of our results for testing and falsifying data sets.

## 2. Main results

### 2.1 *Preliminary definitions*

Our results are about axiomatizations of possible data. We use model theory to study these ideas. The framework used here is developed in more detail in Chambers et al. (2014), where it is used for a different purpose. At the end of this paper, we discuss the relation to Chambers et al. (2014) in more detail.

The following presentation of results is terse, but (we hope) fairly self-contained. In Section 6, we take stock and interpret our findings.

We collect the symbols that we need into a language. The language is a primitive, and specifies the properties and relations (both observable and unobservable) that one can make statements about. A *relational first-order language* $\mathcal{L}$ is given by a set $\mathcal{R}$ of *relation symbols* and a positive integer $n_R$, the *arity of* $R$, for every $R \in \mathcal{R}$. For example, if we wish to talk about preference, we may use a language with a single binary symbol $R$. We can then write axioms and make sense of when a set $X$ and a specific binary relation on $X$ satisfy these axioms. The example in the Introduction used the language $\langle R, \leq \rangle$, in which the arity of $R$ and $\leq$ is 2 (both are binary relation symbols). The statements Axioms 1–5 in the Introduction are axioms in language $\langle R, \leq \rangle$.

A structure is a universe of possible objects (called a domain) and an interpretation of the relation symbols of the language in that universe. An $\mathcal{L}$-*structure* $\mathcal{M}$ is given by a nonempty set $M$ called the *domain* of $\mathcal{M}$, and for every $n$-ary relation symbol $R \in \mathcal{R}$, an $n$-ary relation,[2] the *interpretation* $R^{\mathcal{M}}$ over $M$ of $R$. When the language $\mathcal{L}$ is understood, we refer to an $\mathcal{L}$-structure simply as a *structure*. Structures provide the appropriate framework for interpreting the symbols in the language. For example, when $\mathcal{L} = \langle \succeq \rangle$ has a single binary relation, then one possible structure is $(\mathbf{R}, \geq)$; the structure of the real numbers with the usual greater-than binary relation. Another example is $(2^X, \supseteq)$, the power set of $X$ with set containment.

[2]An $n$-ary relation $R^{\mathcal{M}}$ over $M$ is a subset $R^{\mathcal{M}} \subseteq M^n$.

Suppose that $\mathcal{M}$ and $\mathcal{N}$ are $\mathcal{L}$-structures with universes $M$ and $N$, respectively. Then $\mathcal{M}$ and $\mathcal{N}$ are *isomorphic* if there exists a bijective map $\eta : M \to N$ that preserves the interpretations of all relation symbols, i.e., such that

$$(m_1, \dots, m_n) \in R^{\mathcal{M}} \leftrightarrow \big(\eta(m_1), \dots, \eta(m_n)\big) \in R^{\mathcal{N}}$$

for every $n$-ary relation symbol $R$ of $\mathcal{L}$ and $m_1, \dots, m_n \in M$.

Given a language $\mathcal{L}$, we can write sentences using the relation symbols in $\mathcal{L}$. In addition to the relation symbols specified by $\mathcal{L}$, we shall use certain logical symbols. These symbols are fixed, and we are allowed to use them regardless of the language under consideration. The logical symbols are the *quantifiers* "exists" ($\exists$) and "for all" ($\forall$), "not" ($\neg$), the *logical connectives* "and" ($\wedge$) and "or" ($\vee$), a countable set of *variable symbols* $x, y, z, u, v, w, \dots,$ parentheses ( and ), and the equality symbol $=$. Certain strings of symbols can be put together to form *sentences*, or axioms. Rules for forming sentences are given in, for example, Marker (2002). Such rules are intuitive and immediately recognizable: The string $\forall x \exists y\ x\ R\ y$ is a legitimate sentence; the string $\forall y \exists R x$ is not. We refer to rules of forming legitimate sentences as rules of *syntax*.

In a given structure, sentences can be either true or false. Again we skip the formal definition of what it means for a sentence to be true in a structure since it is intuitively clear: The sentence $\forall x \exists y\ x\ R\ y$ is true in the structure $(\mathbf{R}, \geq)$: For every $x \in \mathbf{R}$ there exists $y \in \mathbf{R}$ such that $x \geq y$. The same sentence is false in the structure with domain $X = \{1, 2, 3\}$ when the relation symbol $R$ is interpreted as $>$: it is not true that for every $x \in X$ there exists $y \in X$ such that $x > y$.

### 2.2 *Main result*

The objects we study in this paper are classes of structures (over some language) that are closed under isomorphism. Such a class of structures captures our idea of a theory. For example, the theory of preference maximization encompasses all structures $\mathcal{M} = (M, R^{\mathcal{M}})$ for which the observed $R^{\mathcal{M}}$ can be extended to a linear order over all elements of $M$. We caution that the term "theory" is somewhat misleading because it means something else in model theory. For this reason, we do not use the term "theory" in our formal definitions.

We say that a class of structures $T$ over some language is *(formally) axiomatized* by a collection of sentences $\Sigma$ if $T$ consists exactly of the structures for which each sentence in $\Sigma$ is valid. Given two classes of structures, $T$ and $T'$, where $T \subseteq T'$, we say that $T$ is (formally) axiomatized by a collection of sentences $\Sigma$ *with respect to* $T'$ if $T$ consists of exactly those structures in $T'$ for which each sentence in $\Sigma$ is valid.

A *universal sentence* is a sentence that includes only universal quantifiers.[3] The axioms of reflexive, complete, transitive, and antisymmetric relations in the Introduction are all universal sentences in the language with two binary relation symbols $\leq, R$. A universal axiomatization is an axiomatization that consists entirely of universal sentences.

[3]These quantifiers can only come at the very beginning of the sentence.

Finally, a set $\Sigma$ of sentences is called *recursively enumerable* (r.e.) if there exists a Turing machine that enumerates over the elements of $\Sigma$. If a recursively enumerable $\Sigma$ axiomatizes a class of structures $T$, we say that $T$ admits an *effective axiomatization*. A Turing machine is a formalization of the intuitive idea of algorithm or effective procedure, without any requirements about computational resources. So when we say that there is a Turing machine that enumerates over the elements of $\Sigma$, we mean that there is a procedure that outputs an exhaustive list $\varphi_1, \varphi_2, \dots$ of all the elements of $\Sigma$. This also means that if $\varphi \in \Sigma$, then there is a way to demonstrate this membership. In Section 6, we argue that existence of effective axiomatization captures our idea of falsifiability of a theory. For example, the collection of axioms in SARP is a r.e. set of sentences. See Sipser (2012) for formal definitions.

Let $\mathcal{F} = \langle R_1, \dots, R_N \rangle$ and $\mathcal{L} = \langle R_1, \dots, R_N, Q_1, \dots, Q_K \rangle$ be languages, where all the $R_n$ and $Q_k$ are relation symbols. Note that $\mathcal{F} \subseteq \mathcal{L}$. The languages $\mathcal{F}$ and $\mathcal{L}$ capture the difference between observable and unobservable objects. The relations $R_n$ are assumed to be observable in the data, while the relations $Q_k$ are unobservable. In our applications below, we choose $\mathcal{F}$ and $\mathcal{L}$ with this interpretation in mind.

Let $T$ be a class of $\mathcal{L}$-structures, closed under isomorphism. Define $F(T)$ to be the class of $\mathcal{F}$-structures $(X^*, R_1^*, \dots, R_N^*)$ for which there exist relations $Q_1^*, \dots, Q_K^*$ such that $(X^*, R_1^*, \dots, R_N^*, Q_1^*, \dots, Q_K^*) \in T$. That is, $F(T)$ is the *projection* of $T$ onto the language $\mathcal{F}$.[4]

We are now in a position to state our theorem.

Theorem 1. *Let $T$ be a class of structures that is closed under isomorphism. If $T$ admits a formal universal axiomatization, then $F(T)$ admits a formal universal axiomatization. Moreover, if $T$ admits a universal effective axiomatization, then $F(T)$ admits a universal effective axiomatization.*

As an example of Theorem 1, recall the revealed preference example in the Introduction. In that example, $\mathcal{F} = \langle R, P \rangle$, which consists of the observed relation and $\mathcal{L} = \langle R, P, \leq, < \rangle$, and also includes the unobserved preference relation of the agent as well as its strict part. The class of structure $T$, which represents the theory of preference maximization, is axiomatized by Axioms 1–5 in the Introduction. The class $F(T)$ is axiomatized by the strong axiom of revealed preference, which is in fact a r.e. sequence of axioms. The following corollary, which follows immediately from Theorem 1, extends the theorem to the case of axiomatization of a class of structures with respect to a larger class.

Corollary 2. *Let $T$ be a class of $\mathcal{L}$-structures and let $T'$ be a class of $\mathcal{F}$-structures that are closed under isomorphism. If $T$ admits a formal universal axiomatization, then $F(T) \cap T'$*

[4] If $T$ has a finite first-order axiomatization, then $F(T)$ is an example of an *existential second-order* theory for language $\mathcal{F}$ in that it allows existential quantification over relations. That is, if $\sigma$ is a first-order $\mathcal{L}$-axiom axiomatizing a class of structures $T$, then $F(T)$ is axiomatized by

$$\exists Q_1 \dots \exists Q_K \sigma.$$

*admits a formal universal axiomatization with respect to $T'$. Moreover, if $T$ admits a universal effective axiomatization, then $F(T) \cap T'$ admits a universal effective axiomatization with respect to $T'$.*

## 3. Proof of Theorem 1

### 3.1 *Preliminaries*

We recall some terminology from model theory used in the proof. An *atomic formula* over a language $\mathcal{L}$ is a string of the form $P(x_1, \dots, x_n)$, where $P$ is an $n$-ary relation symbol in $\mathcal{L}$ and $x_1, \dots, x_n$ are variable symbols. As usual, when $n = 2$, we sometimes write $x\ P\ y$ for $P(x, y)$. A *quantifier-free formula* is a string of symbols that is composed of atomic formulas and the connective symbols $\neg, \vee, \wedge, \rightarrow$ under the rule of syntax. For example the string $\neg(x \succeq y) \rightarrow (y \succeq z)$ is a quantifier-free formula with variables $x, y, z$ in the language with a binary predicate $\succeq$. Every universal sentence can be written in the form $\forall x_1 \dots \forall x_n\ \varphi(x_1, \dots, x_n)$, where $\varphi(x_1, \dots, x_n)$ is a quantifier-free formula with variables $x_1, \dots, x_n$.

If $\varphi(x_1, \dots, x_n)$ is a quantifier-free formula in a language $\mathcal{L}$, and $\mathcal{M}$ is an $\mathcal{L}$-structure with domain $M$, then for every $m_1, \dots, m_n \in M$, there is a well defined sense in which the expression $\varphi(m_1, \dots, m_n)$, obtained by substituting the elements $m_i$ for the variables $x_i$, *is true* in $\mathcal{M}$. Again we provide an example instead of formal definition: if $\varphi(x, y, z) = (x \succeq y) \wedge (y \succeq z)$, then $\varphi(4, 3, 1)$ is true in the structure $(\mathbf{R}, \geq)$ (since $4 \geq 3$ and $3 \geq 1$) but $\varphi(1, 7, 5)$ is false. In particular, for atomic formulas, $P(m_1, \dots, m_n)$ is true in $\mathcal{M}$ if and only if $(m_1, \dots, m_n) \in P^{\mathcal{M}}$. With this notation, a universal sentence $\forall x_1 \dots \forall x_n$ $\varphi(x_1, \dots, x_n)$, where $\varphi(x_1, \dots, x_n)$ is a quantifier-free formula with variables $x_1, \dots, x_n$, is true in $\mathcal{M}$ if and only if $\varphi(m_1, \dots, m_n)$ is true in $\mathcal{M}$ for every $m_1, \dots, m_n \in M$.

Let $\mathcal{F}$ be a relational language and let $\mathcal{M}$ be an $\mathcal{L}$-structure with domain $M$. A substructure $\mathcal{M}'$ of $\mathcal{M}$ is a structure $M'$ whose domain satisfies $M' \subseteq M$ and such that

$$R^{\mathcal{M}'} = \{(m_1, \dots, m_n) \in M'^n | (m_1, \dots, m_n) \in R^{\mathcal{M}}\}$$

for every $n$-ary relation symbol $R$.

We use the following theorem of Tarski (1954).

Theorem 3. *Let $\mathcal{F}$ be a relational language and let $T$ be a class of structures of $\mathcal{F}$. Then $T$ admits a universal axiomatization if and only if the following conditions are satisfied.*

*(i) The class $T$ is closed under isomorphism.*

*(ii) The class $T$ is closed under substructure.*

*(iii) For every $\mathcal{F}$-structure $\mathcal{M}$, if $\mathcal{M}' \in T$ for all finite substructures $\mathcal{M}'$ of $\mathcal{M}$, then $\mathcal{M} \in T$.*

The proof uses basic ideas from sentential logic.[5] A sentential logic is given by a set $\mathcal{S}$ of *sentence symbols*. A *(well founded) formula* of $\mathcal{S}$ is a string built from sentence

[5]For more on this, see, e.g., Chapter 1 of Enderton (2001).

symbols and the connective symbols $\neg, \vee, \wedge, \rightarrow$, using the rules of syntax. For example, if $p$, $q$, and $r$ are sentence symbols, then $(p \rightarrow q) \wedge r$ is a well founded formula. A *truth assignment* $v$ for $\mathcal{S}$ is a function $v : \mathcal{S} \rightarrow \{T, F\}$ (where $T$ stands for TRUE and $F$ stands for FALSE). Every such assignment can be extended uniquely to a truth assignment $\bar{v}$ defined over all formulas. For example, if $v(p) = v(q) = F$ and $v(r) = T$, then $\bar{v}(p \vee q \vee r) = T$. A set $\Gamma$ of formulas is *satisfiable* if there exists some truth assignment such that $\bar{v}(\gamma) = T$ for every $\gamma \in \Gamma$. The compactness theorem for sentential logic, which is equivalent to the axiom of choice, asserts that a set $\Gamma$ of formulas is satisfiable if every finite subset of $\Gamma$ is satisfiable.

Proof of Theorem 1. We prove that $F(T)$ satisfies the conditions of Tarski's theorem. Closure of $F(T)$ under isomorphism and substructure follows from the corresponding properties of $T$. We show that condition (iii) holds.

Let $\mathcal{M}$ be an $\mathcal{F}$-structure and assume that $\mathcal{M}' \in F(T)$ for all finite substructures $\mathcal{M}'$ of $\mathcal{M}$. To show that $\mathcal{M} \in F(T)$, we have to provide interpretations $Q^{\mathcal{M}}$ in $\mathcal{M}$ of the unobservable relation symbols $Q \in \mathcal{L}$ such that the axioms in $\Sigma$ are satisfied.

Let $\mathcal{S}$ be the sentential logic whose set of sentence symbols are all formal expressions of the form $P(m_1, \dots, m_n)$, where $P \in \mathcal{L}$ and $m_1, \dots, m_n \in M$. Let $\Gamma$ be the set of all formulas of $\mathcal{S}$ of the form $\varphi(m_1, \dots, m_k)$ for some axiom $\forall x_1 \dots \forall x_k\ \varphi(x_1, \dots, x_k) \in \Sigma$ with a quantifier-free $\varphi$ and some $m_1, \dots, m_k \in M$. We claim that there exists a truth assignment $v : \mathcal{S} \rightarrow \{F, T\}$ such that the following statements hold:

(i) For every observable relation $P \in \mathcal{F}$, it holds that

$$v\big(P(m_1, \dots, m_n)\big) = T \quad \text{if and only if} \quad P(m_1, \dots, m_n) \text{ is true in } \mathcal{M}.$$

(ii) For every $\gamma \in \Gamma$, $\bar{v}(\gamma) = T$, where $\bar{v}$ is the extension of $v$ to formulas of $\mathcal{S}$.

By the compactness theorem of sentential logic, it is sufficient to show that this can be done for every finite subset of $\Gamma$. Let $\Gamma'$ be a finite subset of $\Gamma$ and let $\mathcal{M}'$ be the finite substructure of $\mathcal{M}$ whose domain consists of all elements of $M$ that appear in some formula of $\Gamma'$. Since $\mathcal{M}' \in F(T)$, there exist interpretations $Q^{\mathcal{M}'}$ of the unobservable relation symbols $Q$ in $\mathcal{M}'$ such that all axioms of $\Sigma$ are satisfied in the augmented $\mathcal{L}$-structure $\mathcal{M}'$. In particular $\varphi(m_1, \dots, m_k)$ is true in $\mathcal{M}'$ for every axiom $\forall x_1 \dots \forall x_k$ $\varphi(x_1, \dots, x_k) \in \Sigma$ and every $m_1, \dots, m_k \in M'$. Therefore, if we set the truth value of the sentential symbol $Q(m_1, \dots, m_n)$ of $\mathcal{S}$ to be the truth value of $Q(m_1, \dots, m_n)$ in $\mathcal{M}'$ (i.e., $v(Q(m_1, \dots, m_n)) = T$ if and only if $Q(m_1, \dots, m_n)$ is true in $\mathcal{M}'$), then all formulas in $\Gamma'$ are satisfied, as desired.

We showed that $F(T)$ admits a universal axiomatization. In particular, the set of all universal $\mathcal{F}$-sentences that are true in all structures of $F(T)$ is such an axiomatization. Since the projection from $T$ to $F(T)$ preserves truthfulness of $\mathcal{F}$-sentences, it follows that these are the universal $\mathcal{F}$-sentences that are true in all structures of $T$. Since $\Sigma$ axiomatizes $T$, these are the universal $\mathcal{F}$-sentences that are the *semantic implications* of $\Sigma$, i.e., the set of all sentences that are true in every structure in which all the axioms of $\Sigma$ are true. Since the set of semantic implications of a recursively enumerable set is recursively enumerable, the second statement in the theorem follows. $\square$

Remark 4. The existence of formal axiomatization in Theorem 1 also follows from an unpublished result of van Benthem (1975); see, e.g., Swijtink (1976).

Remark 5. For readers who are familiar with model theory, we note that Theorem 1 remains valid when the language $\mathcal{F}$ also includes constant and function symbols. However, the unobservable symbols $Q_1, \dots, Q_K$ must be relation symbols. As an example of what might go wrong with unobservable functions, assume that $\mathcal{L} = \langle R, f\rangle$ and $\mathcal{F} = \langle R\rangle$, where $R$ is an unary relation symbol and $f$ is a function symbol, and let $T$ be the class of $\mathcal{L}$-structures $\mathcal{M}$ with the effective universal axiomatization

$$\forall x \quad f\big(f(x)\big) = x,$$

$$\forall x \quad \big(R(x) \to \neg R\big(f(x)\big)\big) \wedge \big(\neg R(x) \to R\big(f(x)\big)\big).$$

Then $F(T)$ is the class of structures $(X, R^*)$ for which the sets $R^*$ and $(R^*)^c$ have the same cardinality. This class of structures is not even formally axiomatizable. In particular, this example is related to the example from the Introduction of existence of a bijective matching.

Remark 6. By a theorem of Craig (1953), $T$ admits a universal r.e. axiomatization if and only if it admits a universal recursive axiomatization.[6] Therefore, in Theorem 1 "universal r.e. axiomatization" can be replaced by "universal recursive axiomatization." We use r.e. axiomatization because this is the concept that captures our idea of falsifiability.

Remark 7. Regarding the assumption of universality, Sneed (1971, p. 54) offers an example, attributed to Dana Scott, in which $F(T)$ has no formal axiomatization when $T$ is not universally axiomatizable (Tuomela 1973, pp. 60–61, establishes that there is no axiomatization of $F(T)$, infinite or otherwise).

### 3.2 *Example: Individual rational choice*

As an illustration of Theorem 1, consider the revealed preference formulation of the theory of individual rational choice.

When choice is primitive, the application of our theorem is a bit involved. This is because we need to be able to describe budget sets, and we require symbols for standard set-theoretic operations. To this end, we define the language of choice as $\mathcal{F} = \langle A, B, \in, C\rangle$. The relations $A$ and $B$ are unary relations: $A(x)$ stands for "$x$ is an alternative"; $B(x)$ stands for "$x$ is a budget set." The relations $\in$ and $C$ are binary, $\in$ is the typical set-theoretic relation, and $C$ is a binary relation, where $C(x, y)$ means "$x$ is chosen from $y$."

The theory of choice, $T_C$, consists of the class of all structures for which there is some global set of alternatives $X$, some family of sets $\mathcal{B} \subseteq 2^X \setminus \{\varnothing\}$ (the budgets), and a *nonempty* choice function $c : \mathcal{B} \to 2^X \setminus \{\varnothing\}$ (satisfying the usual properties) for which all relations are interpreted properly and all structures isomorphic to these.

[6]In fact, Craig's theorem is not stated for universal axiomatization, but his argument can be adapted to universal sentences.

Define the theory of rationalizable choice, $T_R$, to be the subtheory of $T_C$ where, for each structure $\mathcal{M}$, the associated choice function is rationalizable by a weak order.[7] That is, there exists a weak order $R$ on the global set of alternatives $X$ for which $c(B) = \{x \in B : \forall y \in B, x\ R\ y\}$.

The following theorem is well known (for example, see Richter (1966)), but we establish it here using our framework. In the proof of the theorem, $T_R = F(T) \cap T_C$, where $T$ is a theory in an augmented language to which we apply Theorem 1 (and Corollary 2).

Proposition 8. *The theory $T_R$ is universally and r.e. axiomatizable with respect to $T_C$.*

Proof. Introduce the language $\mathcal{L} = \langle A, B, \in, C, R\rangle$, where all relations $A$, $B$, $\in$, and $C$ are as in $\mathcal{F}$, and $R$ is a binary relation. Consider the $\mathcal{L}$-theory $T$ axiomatized by the sentences

(i) $\forall x \forall y \forall z(\in(x, z) \wedge \in(y, z) \wedge C(x, z)) \rightarrow R(x, y)$,

(ii) $\forall x \forall y \forall z(\in(x, z) \wedge \in(y, z) \wedge R(x, y) \wedge C(y, z)) \rightarrow C(x, z)$,

(iii) $\forall x \forall y(A(x) \wedge A(y)) \rightarrow (R(x, y) \vee R(y, x))$,

(iv) $\forall x \forall y \forall z(A(x) \wedge A(y) \wedge A(z)) \rightarrow ((R(x, y) \wedge R(y, z) \rightarrow R(x, z)))$.

Note that any structure $\mathcal{M} \in T_C$ is a member of $T_R$ if and only if there exists a binary relation $R$ on the global set of alternatives for which, for all budgets $B \in \mathcal{B}$, $x \in c(B) \rightarrow \forall y \in B\ x\ R\ y$, and $\forall x \in B$, if $y \in c(B)$ and $x\ R\ y$, then $x \in c(B)$. To see this, note that if $R$ rationalizes $c$, then clearly the preceding two conditions are satisfied for $R$. Alternatively, suppose these conditions are satisfied for some $R$. We claim that $R$ rationalizes $c$. To see this, note that if $B \in \mathcal{B}$, $x, y \in B$, and $x \in c(B)$, then clearly $x\ R\ y$. Alternatively, suppose that $x\ R\ y$ for all $y \in B$. Then because $c$ is nonempty, there exists $y^* \in c(B)$. Then in particular $x\ R\ y^*$. Conclude $x \in c(B)$.

Now note that $T_R = F(T) \cap T_C$. Since $T$ is universally axiomatizable, it follows from Corollary 2 that $T_R$ is universally and r.e. axiomatizable with respect to $T_C$. □

## 4. Rationalizing group preferences

In this section, we assume a fixed set of agents $N$. We observe a social ranking, and want to test whether that social ranking is consistent with the hypothesis that the individuals have preferences that are linear orders and aggregate using some specified social choice rule. For the example of Pareto rule with two agents, axiomatizations are known (see, in particular, Section 3.2 of Trotter 2001 and the references therein; also Baker et al. 1972).[8] Other rules, such as majority rule, are less well understood; most results assume a variable set of agents (McGarvey 1953, Deb 1976, Shelah 2009).

We work with neutral preference aggregation rules that satisfy independence of irrelevant alternatives. By working with such preference aggregation rules, we need not

[7] A weak order is complete and transitive.

[8] Sprumont (2001) considers a related question in an economic environment.

specify what the global set of alternatives is in advance. A set of agents $N = \{1, \dots, n\}$ is fixed and finite. A *preference aggregation rule* is, therefore, defined to be a mapping $f$ carrying any set of alternatives $X$ and any $n$ vector of linear orders[9] (termed a *preference profile*) over those alternatives $(\geq_1, \dots, \geq_n)$ to a binary relation over $X$. We write $f(\geq_1, \dots, \geq_n)$ for the binary relation that results (suppressing notation for dependence on $X$). We assume the following property.

Definition 9 (Neutrality and independence of irrelevant alternatives). For all sets $X$ and $Y$, for all $A \subseteq X$ and $B \subseteq Y$ for which $|A| = |B|$, for all bijections $\sigma : A \to B$, and all preference profiles $(\geq_1, \dots, \geq_n)$ over $X$ and $(\geq_1', \dots, \geq_n')$ over $Y$, if for all $i \in N$ and $x, y \in A$, $x \geq_i y \leftrightarrow \sigma(x) \geq_i' \sigma(y)$, then $x\ f(\geq_1, \dots, \geq_n)\ y \leftrightarrow \sigma(x)\ f(\geq_1', \dots, \geq_n')\ \sigma(y)$ for all $x, y \in A$.

Our assumption embeds the standard hypotheses of neutrality and independence of irrelevant alternatives. Neutrality means that social rankings should be independent of the names of alternatives, while independence of irrelevant alternatives means that the social preference between a pair of alternatives should depend only on the individual preferences between that pair.

Given a preference aggregation rule $f$, we will say that a binary relation $R$ on a set $X$ is *$f$-rationalizable* if there exists a profile of linear orders $(\geq_1, \dots, \geq_n)$ for which $R = f(\geq_1, \dots, \geq_n)$.

Let $\mathcal{F} = \langle R \rangle$ be a language involving one binary relation symbol. Given $f$, a structure $(X, R^X)$ is $f$-rationalizable if $R^X$ is $f$-rationalizable. Call the class of such $\mathcal{F}$-structures the theory of $f$-rationalizable preference, or $T_f$.

Proposition 10. *For any preference aggregation rule $f$ satisfying neutrality and independence of irrelevant alternatives, $T_f$ is universally and r.e. axiomatizable.*

Proof. Since $f$ is neutral and satisfies independence of irrelevant alternatives, we can conclude that there is a collection of sets $\mathcal{N}_f \subset 2^N$ for which for any set $X$, any profile of linear orders $(\geq_1, \dots, \geq_n)$ over $X$, and any pair $x, y \in X$ for which $x \neq y$, $x\ f(\geq_1, \dots, \geq_n)\ y$ if and only if

$$\{i \in N | x \geq_i y\} \in \mathcal{N}_f.$$ [10]

By neutrality and independence of irrelevant alternatives, it follows that either for every $(\geq_1, \dots, \geq_n)$, $f(\geq_1, \dots, \geq_n)$ is reflexive or for every $\geq_1, \dots, \geq_n$, $f(\geq_1, \dots, \geq_n)$ is irreflexive.[11]

Consider the language $\mathcal{L} = \langle R, \geq_1, \dots, \geq_n \rangle$, the $\mathcal{L}$-theory $T$ axiomatized by universal sentences that assert the completeness, transitivity, and antisymmetry of each $\geq_i$,

[9] A linear order is complete, transitive, and antisymmetric.

[10] To see why there is such a collection $\mathcal{N}_f$, let $X$ be any binary set $\{x, y\}$. Define $\geq_E$ to be the preference profile for which for all $i \in E$, $x \geq_i y$, and for all $i \notin E$, $\neg(x \geq_i y)$. Define $\mathcal{N}_f = \{E \subset 2^N : x\ f(\geq_E)\ y\}$. Finally, by neutrality and independence of irrelevant alternatives, the characterization holds across preference profiles and sets $X$.

[11] A binary relation $R$ is reflexive if for every $x \in X$, $x\ R\ x$, and is irreflexive if for every $x \in X$, $\neg(x\ R\ x)$.

and the following sentence that asserts $f$-rationalizability. We consider the two possible cases: the case where $f$ always outputs a reflexive relation, and the case where it always outputs an irreflexive relation,

$$\forall x \forall y \quad R(x,y) \leftrightarrow \bigvee_{E \in \mathcal{N}_f} \left( \bigwedge_{i \in E} (x \geq_i y) \wedge \bigwedge_{i \notin E} (\neg x \geq_i y) \right) \vee (x = y)$$

in the reflexive case and

$$\forall x \forall y \quad R(x,y) \leftrightarrow \neg(x = y) \wedge \bigvee_{E \in \mathcal{N}_f} \left( \bigwedge_{i \in E} (x \geq_i y) \wedge \bigwedge_{i \notin E} (\neg x \geq_i y) \right)$$

in the irreflexive case.

Finally, note that $T_f = F(T)$, so the result follows by Theorem 1. □

## 5. Rationalizing strategic group behavior

In this section, we look at Nash equilibrium behavior. We assume that we observe a collection of game forms and a choice made from each game form. We ask whether there could exist *strict* preferences for a collection of agents over those game forms that generate the observed choices as Nash equilibrium behavior. We show, using Theorem 1, that this theory has a universal axiomatization.

We first have to set up our framework. Instead of focusing on Nash equilibrium specifically, we work with a general collection of theories of group choice. Nash equilibrium, strong Nash equilibrium, and Pareto optimal choice are special cases. We fix a finite set of agents $N = \{1, \dots, n\}$ and a collection $\Gamma \subseteq 2^N \setminus \{\varnothing\}$. The elements of $\Gamma$ are the sets of agents that can deviate from a profile of strategies.

A *game form* is a tuple $(S_1, \dots, S_n)$ of nonempty sets, where we think of $S_i$ as the set of strategies available to agent $i$. For each profile of preferences $(\geq_1, \dots, \geq_n)$ over $\prod_{i \in N} S_i$, a game form $(S_1, \dots, S_n)$ defines a normal-form game

$$(S_1, \dots, S_n, \geq_1, \dots, \geq_n).$$

We define a $\Gamma$-Nash equilibrium of a game $(S_1, \dots, S_n, \geq_1, \dots, \geq_n)$ to be $s \in \prod_{i \in N} S_i$ for which for all $\gamma \in \Gamma$ and all $s' \in \prod_{i \in N} S_i$, if there exists $j \in \gamma$ for which $(s'_\gamma, s_{-\gamma}) \succ_j s$, then there exists $k \in \gamma$ for which $s \succ_k (s'_\gamma, s_{-\gamma})$. If we think of $\Gamma$ as a collection of "blocking" coalitions, a $\Gamma$-Nash equilibrium $s$ is a strategy profile whereby no group $\gamma \in \Gamma$ is willing to jointly deviate, where at least one agent $i \in \gamma$ strictly wants to deviate.

The following cases are special:

- Nash equilibrium results when $\Gamma = \{\{i\} : i \in N\}$.
- Pareto optimality results when $\Gamma = \{N\}$.
- Strong Nash equilibrium results when $\Gamma = 2^N \setminus \{\varnothing\}$.

Other kinds of theories are permissible. For example, by setting $\Gamma = \{G : |G| > |N|/2\}$, we get a kind of majority rule core.

We imagine that we observe a collection of game forms, and *some* strategy profiles that are chosen from each. We do not necessarily observe the entire collection of strategy profiles that could potentially be chosen.

We ask when the strategy profiles are rationalizable by a list of preference relations; obviously, if we make no restriction on preferences, then *every* strategy profile is rationalizable by complete indifference. To this end, we require that preferences be strict over strategy profiles.

Define the *language of group choice* $\mathcal{F}$ to include the following relations:

- For each $i \in N$, one unary relation $S_i$, where $S_i(y)$ is intended to mean that $y$ is a set of strategies for $i$.
- For each $i \in N$, one unary relation $s_i$, where $s_i(x)$ means that $x$ is a strategy for $i$.
- The typical set theoretic binary relation $\in$, meant to signify membership in a set.
- A $2n$-ary relation $R$, where $R(y_1, \dots, y_n, x_1, \dots, x_n)$ means that $(x_1, \dots, x_n)$ is observed as being chosen from game form $(y_1, \dots, y_n)$.

The *theory of group choice* $T_G$ is the class of all structures for the preceding language constructed in the following way. For each agent $i \in N$, there is a *global strategy space* $\mathcal{S}_i \neq \varnothing$ for which the following objects are the elements of the universe:

- Each nonempty $S_i \subseteq \mathcal{S}_i$.
- Each $s_i \in \mathcal{S}_i$.

The relations $S_i$, $s_i$, and $\in$ are all interpreted properly. Finally, for each game form $\prod_i S_i^*$ and strategy profile $(s_1^*, \dots, s_n^*)$, $R(S_1^*, \dots, S_n^*, s_1^*, \dots, s_n^*)$ implies that $S_i(S_i^*)$, $s_i(s_i^*)$ and, last, that $s_i^* \in S_i^*$. This latter requirement means that only strategy sets can go in the first $n$ places in $R$, and that only strategies can go in the last $n$ places. The phrase $R(S_1^*, \dots, S_n^*, s_1^*, \dots, s_n^*)$ means that strategy profile $(s_1^*, \dots, s_n^*)$ is chosen from game form $(S_1^*, \dots, S_n^*)$—this explains the requirement that $s_i^* \in S_i^*$.

The *theory of* $\Gamma$*-rationalizable choice* $T_\Gamma \subseteq T_G$ is the theory of group choice for which for each $i \in N$, there exists a linear order $\geq_i$ over $\prod_{i \in N} \mathcal{S}_i$ for which, for all game forms $(S_1^*, \dots, S_n^*)$, $R(S_1^*, \dots, S_n^*, s_1^*, \dots, s_n^*)$ implies that $(s_1^*, \dots, s_n^*)$ is a $\Gamma$-Nash equilibrium of the normal-form game $(S_1^*, \dots, S_n^*, \geq_1, \dots, \geq_n)$.

Proposition 11. *The theory of* $\Gamma$*-rationalizable choice is universally and r.e. axiomatizable with respect to the theory of group choice.*

Proof. Consider the language $\mathcal{L}$ that includes all relations in $\mathcal{F}$, but also includes, for each agent $i$, a $2n$-ary relation $\geq_i$.

Consider the class of structures $T$ axiomatized by the following sentences: For each $\gamma \in \Gamma$ and $k \in \gamma$,

$$\forall x_1 \ldots \forall x_n \forall y_1 \ldots \forall y_n \forall z_1 \ldots \forall z_n$$
$$\bigwedge_{i\in\gamma} \in(z_i, y_i) \wedge \bigwedge_{i\in N} \in(x_i, y_i) \wedge \neg \bigwedge_{i\in\gamma}(x_i = z_i) \wedge R(y_1, \ldots, y_n, x_1, \ldots, x_n) \wedge \geq_k\big((z_\gamma, x_{-\gamma}), x\big)$$
$$\rightarrow \bigvee_{i\in\gamma\setminus\{k\}} \geq_i\big(x, (z_\gamma, x_{-\gamma})\big)$$

and the universal axioms that express that $\geq_k$ is a linear order (complete, transitive, reflexive, and antisymmetric). As $T$ has a universal axiomatization, so does $F(T)$. Since the axiomatization of $T$ is finite, $F(T)$ has a recursively enumerable universal axiom by Theorem 1, and $T_\Gamma = F(T) \cap T_G$. So by Corollary 2, $T_\Gamma$ has a r.e. universal axiomatization within $T_G$. □

## 6. Discussion

The notions of universal and r.e. universal axiomatization capture the idea of falsifiability of a theory in the following way: Suppose that a scientist postulates a theory $T$ by providing a universal axiomatization $\Sigma$ for $T$. Suppose that we observe the elements $a_1, \ldots, a_n$ of some structure $\mathcal{M}$, and the relationships between them. We call these observations a *data set*. If there exists some universal axiom $\forall x_1, \ldots, \forall x_n\ \varphi(x_1, \ldots, x_n) \in \Sigma$ such that $\varphi(a_1, \ldots, a_n)$ is not true, then $T$ has been falsified. Thus, a violation of an axiom of the theory can be demonstrated by presenting a data set. The fact that the theory is given by a universal axiomatization means that any violation of the theory can be demonstrated. In the terminology introduced in Chambers et al. (2014), such a theory is identical to its empirical content.[12]

If, in addition, $\Sigma$ is recursively enumerable, then the scientist can describe $\Sigma$ by providing the algorithm (or Turing machine) that generates $\Sigma$. In this case, if a data set falsifies the theory, then this falsification can be demonstrated by pointing to the index of the axiom that is violated in the recursive enumeration of the axioms. The following example illustrates this issue.

Example 12. Consider a language with a single binary relation symbol $L$, where $x\ L\ y$ is supposed to represent the relationship $x$ loves $y$. A *love cycle of size* $k$ is a sequence of people $x_1, \ldots, x_k$ such that, for every $1 \leq i, j \leq k$, it holds that $x_i$ loves $x_j$ if and only if $j = (i + 1) \mod k$. Let $C \subseteq \mathbb{N}$. Suppose that a scientist postulates the theory that there are no love cycles of size $k$ for any $k \in C$. Such a theory has a universal axiomatization. If $C$ is recursively enumerable, then the scientist can describe the theory by providing a computer program that enumerates over $C$. If the theory is incorrect, i.e., if there exists a love cycle of size $k$ for some $k \in C$, then an antagonist can demonstrate this violation and falsify the theory by pointing to a data set that violates the theory (i.e., to a love

[12]Under a caveat that in the current paper we assume that absence of relationship can be observed. See Section 4.3 in Chambers et al. (2014).

cycle of size $k$ for some $k \in C$) and, by pointing to the index of $k$ in the enumeration of $C$ to show that $k \in C$. If, however, $C$ is not recursively enumerable,[13] then, while the scientist's theory has a universal axiomatization, he has no way to formally describe the theory. An antagonist who has an access to a love cycle of size $k$ for some $k \in C$ may not be able to demonstrate that $k \in C$. $\diamondsuit$

Example 12 shows that a theory may have a universal axiomatization, but the ability of a theory to be falsified in practice depends on the existence of an effective or recursively enumerable axiomatization. However, if a data set is consistent with the theory, then this consistency cannot necessarily be effectively demonstrated, since such a demonstration requires checking that the countably infinite list of axioms is satisfied. Consider Example 12 again. If $C$ is recursively enumerable but not recursive, then the fact that a given data set is consistent with the theory cannot necessarily be demonstrated since if $C$ is not recursive, the researcher has no way to effectively demonstrate for a given love cycle of size $k$ that $k \notin C$. Note that, by Remark 6, the theory also admits a recursive axiomatization, but this additional feature does not seem to be related to falsifiability. In particular, as the example shows, it does not mean that the researcher can demonstrate that a given data set is consistent with the theory.

Thus, recursively enumerable universal theories have the property that any violation of the theory can be demonstrated. In logical terminology, if a sentence $\varphi$ is an element of $\Sigma$, then there exists a formal deduction (a proof) for this fact. Gradwohl and Shmaya (2015) go one step further and require in addition that this proof be short.

Consider now the case in which a theory $T$ admits a *finite* universal axiomatization. This is the case of the consumer choice example discussed in the Introduction, as well as the applications to group choice and Nash equilibrium developed above, and to all the other natural economic applications of which we are aware.

In this case, there is a way to check whether a given data set is consistent with the theory (and, a fortiori, to demonstrate consistency of the data set with the theory): go over all the axioms and check that they are all satisfied. Moreover, in the framework of Section 2.2, if $T$ is an $\mathcal{L}$ theory that admits a finite and universal axiomatization, then there is a way to check whether a given finite data set is consistent with $F(T)$: one has to go over all possibilities for the unobserved relations $Q_1, \dots, Q_K$ and check whether it is possible to define them in a way that is consistent with all of the axioms. In logical terminology, if $T$ has a finite universal axiomatization, then the set of semantic implications of $F(T)$ is recursive.

Summing up, for the case in which $T$ has a finite universal axiomatization (which it does in all our examples), there is a way to check whether a finite data set is consistent with the theory $F(T)$, and Theorem 1 implies that if the theory is incorrect, then one can point to a finite data set that is not consistent with it.

## 7. Relation to previous literature

One relevant antecedent to our paper is Brown and Matzkin (1996). These authors exploit a famous model-theoretic result, the Tarski–Seidenberg theorem (Tarski 1951), to

[13] For example, if $C$ is the set of codes of computer programs that do not halt.

study the empirical restrictions placed by the competitive equilibrium hypothesis on observable data. Roughly, they use Afriat's theorem to show that testing the consistency of data with the theory of Walrasian equilibrium boils down to verifying whether there exists a solution to a finite collection of polynomial inequalities. The Tarski–Seidenberg theorem establishes that the existence of a solution to this set of inequalities is equivalent to the satisfaction of another set of polynomial inequalities (which can be algorithmically determined) in which no theoretical variables appear. The second system of inequalities only depends on data. What we accomplish is similar in spirit to what Brown and Matzkin do in their paper, but our techniques are different (our result follows from Tarski's theorem on universal axiomatization, not on the theory of quantifier elimination in systems of real polynomials).

Many papers in revealed preference theory are interested in the specific form of the axiomatization of $F(T)$, but there are also many studies that are primarily interested in the existence of a test. We have already mentioned Brown and Matzkin (1996), but there are many other papers based on developing so-called Afriat inequalities. Two recent developments are Quah (2012) and Polisson and Quah (2013), which seek to show that certain economic theories are falsifiable by establishing what in our papers would be universal axiomatizations of $F(T)$.

Our paper is also related to a model-theoretic literature that studies when a theory can be given an axiomatization using additional relations. See, for example, Craig and Vaught (1958), who provide conditions under which a class of structures is of the form $F(T)$ for some finitely axiomatizable $T$. The work also contains a finite model-theoretic result related to Theorem 1.

The type of issues we discuss here have previously been studied by philosophers of science. Without going into full detail, Ramsey (1931) was one of the first to discuss the elimination of "theoretical" terms from scientific theories. Various authors give different interpretation to the notion of "Ramsey elimination."

Herbert Simon wrote a sequence of papers on falsifiability and empirical content. For example, Simon (1985) discusses some of the issues we discuss here: Simon argues that the theory of rational choice is falsifiable, even though its usual formulation existentially quantifies over unobservables (what he calls theoretical). As Simon (1985) states, "although existential quantification of an observable is fatal to the falsifiability of a theory, the same is not true when the existentially quantified term is a theoretical one."

While this may seem obvious, it has led to a large degree of confusion among economists. For example, Boland (1981) argued that the theory of rational choice is not falsifiable precisely because of its existential formulation over unobservables. Mongin (1986) counters this argument. He observed that if $T$ has universal $\mathcal{F}$-implications, then $F(T)$ has universal $\mathcal{F}$-implications and, hence, satisfies the logical condition for falsifiability. Our result establishes that all of its implications are falsifiable in the full sense of falsifiability, not only the logical one.

In Chambers et al. (2014), we have used a similar framework as here to investigate the notion of empirical content. That paper looks at the syntactic counterpart to the semantic notion of empirical content. Here we have instead focused on conditions under which a theory has a universal axiomatization.

## References


Baker, Kirby A., Peter C. Fishburn, and Fred S. Roberts (1972), "Partial orders of dimension 2." *Networks*, 2, 11–28. [503]

Boland, Lawrence A. (1981), "On the futility of criticizing the neoclassical maximization hypothesis." *American Economic Review*, 71, 1031–1036. [509]

Brown, Donald J. and Rosa L. Matzkin (1996), "Testable restrictions on the equilibrium manifold." *Econometrica*, 64, 1249–1262. [494, 508, 509]

Chambers, Christopher P., Federico Echenique, and Eran Shmaya (2014), "The axiomatic structure of empirical content." *American Economic Review*, 104, 2303–2319. [497, 507, 509]

Craig, William (1953), "On axiomatizability within a system." *Journal of Symbolic Logic*, 18, 30–32. [502]

Craig, William and Robert L. Vaught (1958), "Finite axiomatizability using additional predicates." *Journal of Symbolic Logic*, 23, 289–308. [509]

Deb, Rajat (1976), "On constructing generalized voting paradoxes." *Review of Economic Studies*, 43, 347–351. [503]

Enderton, Herbert B. (2001), *A Mathematical Introduction to Logic*. Academic Press, San Diego. [500]

Gradwohl, Ronen and Eran Shmaya (2015), "Tractable falsifiability." *Economics and Philosophy*, 31, 259–274. [508]

Houthakker, Hendrik S. (1950), "Revealed preference and the utility function." *Economica*, 17, 159–174. [495]

Marker, David (2002), *Model Theory: An Introduction*, volume 217 of Graduate Texts in Mathematics. Springer-Verlag, New York. [498]

McGarvey, David C. (1953), "A theorem on the construction of voting paradoxes." *Econometrica*, 21, 608–610. [503]

Mongin, Philippe (1986), "Are 'all-and-some' statements falsifiable after all?: The example of utility theory." *Economics and Philosophy*, 2, 185–195. [509]

Polisson, Matthew and John K.-H. Quah (2013), "Revealed preference tests under risk and uncertainty." Discussion Papers in Economics 13/24, Department of Economics, University of Leicester. [509]

Popper, Karl R. (1959), *The Logic of Scientific Discovery*. Hutchinson & Co., London. [494]

Quah, John K.-H. (2012), "A revealed preference test for weakly separable preferences." Department of Economics Discussion Paper Series 601, University of Oxford. [509]

Ramsey, Frank P. (1931), "Theories." In *The Foundations of Mathematics and Other Logical Essays*, 212–236, Routledge & Kegan Paul, London. [509]

Richter, Marcel K. (1966), "Revealed preference theory." *Econometrica*, 34, 635–645. [495, 503]

Samuelson, Paul A. (1938), "A note on the pure theory of consumer's behaviour." *Economica*, 5, 61–71. [495]

Shelah, Saharon (2009), "What majority decisions are possible." *Discrete Mathematics*, 309, 2349–2364. [503]

Simon, Herbert A. (1985), "Quantification of theoretical terms and the falsifiability of theories." *British Journal for the Philosophy of Science*, 36, 291–298. [509]

Sipser, Michael (2012), *Introduction to the Theory of Computation*. Cengage Learning, Boston. [499]

Sneed, Joseph D. (1971), *The Logical Structure of Mathematical Physics*. Springer, Dordrecht. [502]

Sprumont, Yves (2001), "Paretian quasi-orders: The regular two-agent case." *Journal of Economic Theory*, 101, 437–456. [503]

Swijtink, Zeno G. (1976), "Eliminability in a cardinal." *Studia Logica*, 35, 71–89. [502]

Tarski, Alfred (1951), "A decision method for elementary algebra and geometry." RAND Corporation Report Series. [508]

Tarski, Alfred (1954), "Contributions to the theory of models I." *Indagationes Mathematicae*, 57, 572–581. [500]

Trotter, William T. (2001), *Combinatorics and Partially Ordered Sets: Dimension Theory*, volume 6 of Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press. [503]

Tuomela, Raimo (1973), *Theoretical Concepts*, volume 10 of LEP Library of Exact Philosophy. Springer-Verlag, Wien. [502]

van Benthem, Johan (1975), "A problem concerning expansions and its connections with modal logic." Working paper. [502]

---

Co-editor Faruk Gul handled this manuscript.

Manuscript received 2 August, 2014; final version accepted 7 June, 2016; available online 21 June, 2016.