THE AXIOMATIC STRUCTURE OF EMPIRICAL CONTENT

CHRISTOPHER P. CHAMBERS, FEDERICO ECHENIQUE, AND ERAN SHMAYA

Abstract. We provide a framework for studying the empirical content of an economic theory. We define the **empirical content** of a theory to be the least weakening of the theory that makes only falsifiable claims. We prove that the empirical content of a theory is exactly captured by a certain kind of axiomatization, one that uses axioms which are universal negations of conjunctions of atomic formulas.

Date: May 10, 2012.

Chambers is affiliated with the Department of Economics, University of California, San Diego. Echenique is affiliated with the Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena CA 91125. Shmaya is affiliated with the Kellogg School of Management at Northwestern University, 2001 Sheridan Road, Evanston, IL 60208-2001. We thank Nageeb Ali, Yaron Azrieli, Bettina Klaus, Matias Iaryczower, Tomasz Strzalecki, and Leeat Yariv for comments. Chambers and Echenique acknowledge support from the NSF through grant SES-0751980. Emails: `cpchambers@ucsd.edu` (Chambers), `fede@caltech.edu` (Echenique), and `e-shmaya@kellogg.northwestern.edu` (Shmaya).

Contents

1. Introduction

The purpose of this study is to understand the notions of falsifiability and empirical content, independently of their specific meaning in particular economic theories. A researcher often understands the testable implications of his particular theory. We study the common formal structure of all particular instances of empirical content.

We introduce a framework for studying the empirical content of an economic theory, and describe the kinds of axioms that characterize empirical content. Much of modern decision theory is motivated by the idea that axioms provide the testable implications of a theory. We take this motivation very seriously, and seek to understand the properties of axioms that allow them to achieve this goal. Our framework presents general notions of data, theory and axioms.

We start from data, and we describe theories as collections of possible data generating processes. For example if one can observe revealed preference relations, then a theory is a collection of processes that may generate different revealed preference relations. An example of a theory is the set of data-generating processes associated with the maximization of some utility function.

Next, we describe the empirical content of a theory. One can some times weaken a theory in ways that have no observable consequences; this happens when the theory makes some non-testable claims. The theory of utility maximization is a case in point. We define the empirical content of a theory T as the *most* we can non-observably weaken T.

Our main result ties empirical content to a certain kind of axioms, called "UNCAF." These axioms rule out certain basic claims, and they rule them out for all elements of any given universe. We shall explain the definition of UNCAF below; suffice it to say here that the classical weak axiom of revealed preference is an example of an UNCAF axiom. The completeness axiom (saying that any two alternatives must be comparable) is not an UNCAF axiom.

We prove that the empirical content of a theory T is axiomatized by the UNCAF axioms that are true in T. This means that UNCAF axioms capture all the empirical content of a theory, and that a theory that has an UNCAF axiomatization is one that only makes testable claims: we call these theories "totally falsifiable." As an example, we describe below the UNCAF axioms that are true in the theory of utility maximization, and show how they are a generalization of the strong axiom of revealed preference.

Decision theorists often regard some axioms as "technical," and others as substantial. For example, an axiom that says that preference is continuous may be technical. We provide *relative* versions of our results, where the testable implications of a theory may be discussed relative to such non-testable, technical, properties.

The body of the paper has two parts, Section 2 and the rest. In Section 2 we present a verbal description of our results. We have written Section 2 as a short essay, one that we hope is readable and interesting to a broad audience of economists. It does not require more than familiarity with some of the themes (not the results) of first-year graduate-level microeconomic theory (as found in, say, chapter 2 in Mas-Colell, Whinston, and Green (1995)). When we talk about axioms, these will be familiar from introductory microeconomics, as are the motivating examples: the theory of rational choice and utility maximization. Section 2 is a self-contained, and fairly complete, exposition of our research. It describes our main results and explains why they are true.

The remainder of the paper contains a mathematical exposition of our results, together with some applications that are more involved than the basic applications in Section 2. The mathematics used is simple, but may be unfamiliar to many economists. We borrow basic ideas from mathematical logic and model theory; these are areas of mathematics studying the relation between the formal structure of mathematical statements and the mathematical objects they are meant to apply to. We borrow the basic mathematical framework, which we lay out in Appendix B, but no knowledge of any results from logic of model theory are needed to understand our paper. We hope that Section 2 makes clear that the ideas are simple and intuitive.

2. A verbal exposition of our results

2.1. **The concepts.** There are three important concepts we need to explain before we describe our main results. The first is the primitive of our model: the things we can observe are the primitive. The second is what we mean by a data set: data sets are finite and consist of partial observations. The third is our notion of a theory: a theory is a formal way of hypothesizing that certain relationships hold between objects of interest. A theory is comprised of a collection of possible data generating processes, representing possible relationships between the observables. The economist claims that one of these processes generates data. A theory can be consistent with some data sets but not others. A theory is not, however, identified with the data sets with which it is compatible.

2.2. **The concepts: data.** We shall introduce our main ideas by means of a familiar economic example: the theory of rational choice. Vaguely speaking, the theory says that "choice" is made in accordance with a complete and transitive preference relation, but if we want to test the theory we shall have to be more precise. In particular, we need to clearly state in what form our potential observations come when we test the theory. The empirical content of a theory depends on what we may be able to observe.

Let us assume then that we observe comparisons between pairs of alternatives: we observe *revealed weak preference* and *revealed strict preference.* For example, x may be revealed weakly preferred to y if it is chosen in a direct comparison, but x is revealed strictly preferred to y if an individual is willing to choose x over y and some non-negligible amount of money. We use the symbols $\succeq$ to denote revealed weak preference and $\succ$ to denote revealed strict preference.

A data set is a snapshot of the universe which happens to be available to a researcher. In this example, a data set is given by a finite set of alternatives, and two binary relations, $\succeq$ and $\succ$, defined on the set of alternatives. As an example, consider the data set with set of alternatives $D^1 = \{a, b, c\}$ where we observe that a is weakly revealed preferred to b, and b to c, but we do not observe any strict comparisons. In symbols, $\succeq^{\mathcal{D}^1}$ is the relation given by $a \succeq^{\mathcal{D}^1} b$ and $b \succeq^{\mathcal{D}^1} c$, while $\succ^{\mathcal{D}^1}$ is empty.

The example demonstrates an important feature of data. We might theorize that $a \succeq^{\mathcal{D}^1} b$ implies either $b \succeq^{\mathcal{D}^1} a$ or $a \succ^{\mathcal{D}^1} b$, but often data will not contain this kind of information. In fact, consumption data sets (the most common real-world application of revealed preferences) will typically not contain this information. A data set affirms the existence of the relationships we observe, but it does not mean that the relationships we do not observe do not exist; a phenomenon we call *partial observability.*[1]

To sum up, data sets are always finite and observations are partial. They consist of a finite set of observed objects, and the observed relations among the objects: what exactly can be observed is a primitive notion, and it depends on the theories under consideration. Partial observability is a simple, but crucial, idea. It means that not observing a theoretical relation does not mean that it does not hold; it is an obvious property of economic data sets, and we shall see that it affects the kinds of axioms that capture testability.

[1]In the words of Carl Sagan: "absence of evidence is not evidence of absence."

2.3. **The concepts: theory.** A theory is an abstraction that can generate some sets of data but not others. We define a theory as a *collection of data generating processes* that is internally consistent (closed under isomorphisms). What we mean is the following: a data generating process is a grand set of alternatives together with relations among these alternatives. A data set is generated by such a process if its observed set of alternatives is a subset of the grand set, and its observed relations are subsets of the "theoretical" relations.

To continue with our revealed-preference example, a theory is a collection of sets X, each one with relations $\succeq^X$ and $\succ^X$ which describe potential observations of weak and strict revealed preference. A data set is generated by the theory if there is some particular X, $\succeq^X$ and $\succ^X$, in the theory, such that the observations in the data set are a subset of X, and where x is weakly revealed preferred to y only if $x \succeq^X y$; similarly for strict revealed preference.

The theory of rational choice is the class of triples $(X, \succeq^X, \succ^X)$ such that $\succeq^X$ is a complete and transitive preference relation, and $\succ^X$ is the strict preference derived from $\succeq^X$. We denote the theory of rational choice by T_R. Note that T_R can generate the data set $\mathcal{D}^1$ described above: $\mathcal{D}^1$ has observed objects $D^1 = \{a, b, c\}$, where a is revealed weakly preferred to b, and b to c. The data $\mathcal{D}^1$ can be generated by T_R because it can be generated, for example, by the set X of all letters in the English alphabet, with $\succeq^X$ being the lexicographic order on X. Of course, there are many other data generating processes in T_R that could generate $\mathcal{D}^1$.

We emphasized that $\mathcal{D}^1$ is silent on some aspects of the relationship between objects a and b; these aspects are not observed, but we do not view this partial observability as a conflict with T_R. Some "theoretically true" relations are simply not observed in the data set.

Another example is the theory of utility maximization, denoted T_U, which is the set of triples $(X, \succeq^X, \succ^X)$ in T_R such that there is $u : X \to \mathbf{R}$ with $x \succeq^X y$ iff $u(x) \geq u(y)$. The theory T_U could also have generated D^1.

Note that theories are collections of data-generating processes, so we can compare them using set inclusion. We have $T_U \subseteq T_R$, meaning that T_R is weaker than T_U: the theory of utility maximization "claims more" than rational choice theory.

2.4. **Empirical content.** With the notions of data and theory in place, we can define empirical content. We shall see first what it means for a data set to falsify a theory, and then observe that one may sometimes weaken a theory in ways that have no observable consequences. This happens when, in some sense, the theory makes non-testable claims. The empirical content of a theory is the result of the most one can non-observably weaken a theory. The empirical content of a theory T is the weakest theory which contains T and is observationally equivalent to T.

A theory can generate some data sets and not others. We say that a data set *falsifies* a theory if it could not have been generated by any of the data-generating processes in the theory. For example, the theory T_R of rational preference is consistent with the data set $\mathcal{D}_1$ described above, and so D^1 does not falsify rational choice. On the other hand, consider the data set $\mathcal{D}^2$, where $D^2 = \{a, b, c\}$ as with D^1, but where we observe no weak comparisons, and instead observe that

$$a \succ^{\mathcal{D}^2} b \succ^{\mathcal{D}^2} c \succ^{\mathcal{D}^2} a.$$

No process in T_R could have generated D^2 because the "theoretical" strict preference $\succ^X$ in such a process would need to exhibit the cyclic comparisons $a \succ^X b \succ^X c \succ^X$; this is impossible for a rational preference relation.

The idea of falsification allows us to compare the observable restrictions of two theories. The theory of utility maximization, T_U, is more restrictive than T_R, in the sense that all the triples $(X, \succeq^X, \succ^X)$ in T_U are also in T_R. We can also say that T_R is weaker than T_U. However, *every data set that falsifies* T_U *also falsifies* T_R. We can see why by thinking of lexicographic preferences in a Euclidean space, the canonical example of a rational preference with no utility representation. On any finite set, the lexicographic relation is representable by a utility function, but it has no such representation on the whole space. More generally, it follows from standard utility representation theorems that a finite data set falsifies T_U if and only if it also falsifies T_R. We conclude that the empirical contents of T_U and T_R are the same: the two theories are falsified by the same data sets, so we must conclude that they are observationally equivalent. In other words, the additional claim made by T_U, that $\succeq^X$ is representable by a utility, is not testable.

Does the theory T_R make any non-testable claims? The answer turns out to be yes. Given our primitive assumption that we can only observe revealed weak and revealed strict preference, the theory T_R makes non-testable claims. Consider the data-generating process $(\mathbf{R}, \geq, >')$: the set of alternatives is the set of real numbers, $\mathbf{R}$; $\geq$ is the usual order on $\mathbf{R}$, and $>'$ is defined as $x >' y$ if $x > y + 5$. The process $(\mathbf{R}, \geq, >')$ is not in T_R, as $>'$ is not the strict preference relation associated to $\geq$. However, no data set generated by the process could falsify T_R. Any data set that could be observed under $(\mathbf{R}, \geq, >')$ would have to be consistent with T_R, meaning that it could also be generated by one of the processes in T_R. For example, the data set with alternatives $\{3, 5\}$, $5 \geq 3$ and $>' = \emptyset$ is compatible with T_R, because the data are silent on whether 5 is strictly preferred to 3 or 3 is weakly preferred to 5. This silence, on the other hand, is the essence of partial observability.[2]

We have explained how rational choice, T_R, is weaker than T_U, yet observationally equivalent. The theory of rational choice, in turn, imposes some non-testable restrictions. As a result, we could further weaken T_R and obtain a third theory, even weaker than T_R but observationally equivalent to T_R and T_U. How far can we go? We define the empirical content of a theory as the weakest theory that contains it and is observationally equivalent.

A theory T' that contains T, $T' \supseteq T$, is a weakening of T; it is less restrictive, or "claims" less. The empirical content of T, denoted $\mathrm{ec}(T)$, is the weakest $T' \supseteq T$ with the property that any data set that falsifies T also falsifies T'. In other words, $\mathrm{ec}(T)$ is the weakest theory with the property of being observationally equivalent to T.

Our notion of empirical content captures the sense in which T_R and T_U are observationally equivalent, and applies to any arbitrary theory. The empirical content of any T is the weakening that exactly retains all empirical implications of T. There are however theories that, unlike T_R and T_U, cannot be further weakened without observable consequences.

A theory is *totally falsifiable* if it is identical to its empirical content ($T = \mathrm{ec}(T)$). In other words, T is totally falsifiable if all its assertions are falsifiable; thus a totally falsifiable theory cannot be weakened without observable consequences. The theories of rational choice and utility maximization make some non-testable restrictions, therefore T_R and T_U are proper subsets of $\mathrm{ec}(T_R)$ and $\mathrm{ec}(T_U)$, respectively. Neither T_R nor T_U are totally falsifiable. There is a special role in our analysis for totally falsifiable theories; as we shall see, they are always axiomatizable, and posses an axiomatic characterization by UNCAF axioms.

Our remarks on the empirical content of T_R and T_U, and their relation to finiteness of data sets and partial observability are already present implicitly or explicitly in the previous literature. Our goal is to provide an analytic framework that supports these intuitions and explore their role in the axiomatizations of the theories. Such a framework should allow us to formalize the assertions that the theories T_R and T_U have the same empirical content, and to distinguish between the situations in which absence of preference is observed and not observed.

[2]The general problem behind the example relates to empirically distinguishing semiorders from rational preferences. Such a distinction is not possible using revealed preference data.

2.5. **Axioms.** The main result of our paper is that the empirical content of a theory is described by certain kinds of axioms. We describe a property of axioms we call UNCAF (universal negation of conjunction of atomic formulas), which means that they rule out some basic claim about all objects in a data-generating process. Our main result is that the empirical content of a theory is the theory of the UNCAF axioms that hold true in the theory.

An axiom is a mathematical statement in a language. In our case, the language is dictated by what we can observe. Consider our examples of T_R and T_U. The following axiom is "true" in these theories because it is a simple consequence of transitivity and the definition of strict preference:

$$\text{for all } x, y \text{ and } z, \text{ it is not true that } (x \succeq y \text{ and } y \succeq z \text{ and } z \succ x);$$

or, more succinctly,

$$(\text{for all } x, y, z) \text{ not } (x \succeq y \text{ and } y \succeq z \text{ and } z \succ x).$$

Using mathematical symbols, we can rewrite the axiom. Use $\forall$ to mean "for all," $\neg$ to mean "not," and $\wedge$ for "and." Then we obtain

$$\forall x \forall y \forall z \neg \left((x \succeq y) \wedge (y \succeq z) \wedge (z \succ x) \right) \tag{1}$$

Axiom (1) is true in T_R and T_U in the sense that it holds in any data-generating process in T_R or T_U.

Here is another example of an axiom, which expresses the property of nonsatiation:

$$(\text{for all } x)(\text{there is } y) \text{ such that } (y \succ x).$$

A preference is satiated when there is an alternative that is better than any other alternative. The non-satiation axiom says that for any alternative x there is some alternative y which is strictly better than x. Using the symbol $\exists$ for "there is," we can write it as:

$$\forall x \exists y (y \succ x). \tag{2}$$

We shall see that Axiom (1) is testable, while Axiom (2) is not, and trace the testability of an axiom to its formal structure (its syntax). Let us start with Axiom (1). We can easily find a data set that is incompatible with Axiom (1): the axiom is basically telling us what the dataset should be. Choose a set $D = \{a, b, c\}$, and suppose that

$$a \succeq^{\mathcal{D}} b \succeq^{\mathcal{D}} c \succ^{\mathcal{D}} a.$$

Evidently Axiom (1) is incompatible with data set D. Importantly, the axiom cannot be a true statement of any data generating process that could have generated D.

The structure, or form (or syntax), of Axiom (1) is why it is testable. It starts with "for all,"the *universal quantifier*, meaning that the statement should apply to *all* the elements in a data generating process; then a negation; and then a conjunction of statements of the form '$y \succeq z$' or '$z \succ x$'. The latter statements are "basic" in this language, in mathematics they are called atomic formulas. The statements '$y \succeq z$' or '$z \succ x$' stand for indivisible claims about the observable relations among elements in a data set or a data-generating process: they are indivisible (or atomic) claims in the sense that they are not made up of simpler statements about observables. Axioms like (1) are universal statements (they only use the $\forall$ quantifier) which negate a conjunction of basic statement about observables: "universal negation of conjunctions of atomic formulas," or UNCAF axioms for short.

Axiom (2), on the other hand, is not UNCAF. We can see that, as a result, the axiom makes a non-testable claim. The quantifier "for all x" is followed by "there is y," and one cannot transform the statement into something that applies to all elements of a universe of objects. There is a choice in what y can be, and a choice is bad for testing. Consider how a data set might fail to be compatible with axiom (2). We would need an alternative a for which there is no alternative that is strictly better. Of course, there may be a data set in which a is the best possible alternative; but a finite data set can *never rule out* the existence of a better alternative than a *somewhere* in the data-generating process that gave rise to the data. We may have simply failed to observe the better element. There may or may not be a better alternative to a, but a data set can never guarantee that there is not a better alternative.

UNCAF axioms have two important features, and they are related to the work of Popper (1959) in the philosophy of science. The first feature is that the axiom has only "for all" quantifiers, meaning that it is a universal statement. It is a statement that is required to hold for all instances of its variables. Popper famously regarded universality to be a basic property of scientific theories. Popper presents as examples a theory (really an axiom in our framework, but he called it a theory) that claims "all swans are white" and a theory saying "there is a black swan." Presuming we can observe and describe non-white swans, the first theory is universal and falsifiable: a data set of a single non-white swan falsifies it. The second theory is existential and not falsifiable, as a data set of finitely many white swans does not preclude the existence of a black swan. Our axiom (2) is an economic instance of Popper's black swan example.

The second feature of UNCAF axioms is the negation of the conjunction of certain possible observations. This requirement is not present in Popper's analysis. It results from our insistence on partial observability. For example, consider the completeness axiom in decision theory:

$$(\text{for all } x \text{ and } y)(x \succeq y \text{ or } y \succeq x),$$

saying that any two alternatives must be comparable using the preference $\succeq$. The completeness axioms satisfies Popper's requirement of being a universal statement. One could only, however, falsify completeness by observing that some alternative is *not* preferred to another. We do not normally observe this in economic choice data. Under the assumptions of our running example, only weak and strict revealed preference is observable, and under these primitive assumptions one could never contradict the completeness axiom. The axiom is not UNCAF and not testable.

Another example is, using mathematical notation,

$$\forall x \forall y \left((x \succ y) \leftrightarrow (x \succeq y) \wedge (\neg (y \succeq x)) \right), \tag{3}$$

which expresses that $\succ$ is the strict preference associated to $\succeq$. Axiom (3) is not an UNCAF axiom and one can see how, similarly to the completeness axiom, it will not be testable using data on revealed preferences.

We have focused here on the best known axioms in microeconomics, axioms that are familiar to any graduate student. But our discussion is obviously also relevant for modern research. To give some examples from influential papers, the main axioms in Kreps (1979) and Gul and Pesendorfer (2001) are UNCAF. Some of the main axioms in Dekel, Lipman, and Rustichini (2009) are not UNCAF.[3] In addition, the working paper version of our paper develops detailed applications to models in behavioral economics.

2.6. **Main result.**

> *The empirical content of a theory T is axiomatized by the UNCAF axioms that are true in T. A totally falsifiable theory is one that has an UNCAF axiomatization.*

We should stress that we do not assume that all theories can be axiomatized; indeed, some theories cannot. Our result implies that axiomatizability is a consequence of total falsifiability.

[3]Axioms 3, 7 and 11 in Dekel, Lipman, and Rustichini (2009) are not UNCAF.

The idea behind our result is simple. One can describe the empirical content of a theory by the data sets that would falsify the theory. Every data set that would falsify T_U would also falsify T_R, that is why they have the same empirical content. For each falsifying data set, one can write an axiom ruling out this data set. Such an axiom would need to abstract from the particulars of the data set: it cannot depend on the particular observed objects in the data set. The axiom must capture the properties of the data set which make abstract sense, and which apply broadly. This exercise is similar to a decision theorist abstracting from a thought experiment: maybe the experiment involves urns, and balls of different colors. The theorist would not write an axiom that talks about urns and balls, but would instead abstract the essence of the choices made in the thought experiment, and use it to write an axiom. Our main result performs a similar abstraction from the observations in any data set.

So the abstract essence of a data set motivates an axiom. The axiom takes the UNCAF form because (1) it applies to any element of a data set or a data-generating process, and (2) it rules out (hence the negation) a particular configuration of observed relations among the variables; these observed relations are the basic or atomic formulas in the UNCAF axiom.

The idea of writing an axiom to rule out (essentially) each possible data set may seem naive. However, as we proceed to explain, that is precisely what the strong axiom of revealed preference achieves. It may also suggest that we need too many axioms (see Corollary 23).

2.7. **Illustration of empirical content.** We illustrate our main result by investigating the empirical content of the theory of rational choice, T_R. When specialized to T_R, most economists are familiar with our result, at least under some form. The UNCAF axiomatization of $\mathrm{ec}(T_R)$ is a generalization of the strong axiom of revealed preference, SARP. Despite the singular noun in its name, SARP is actually a collection of axioms, not a single axiom.

We have an axiomatization of T_R, but it involves non-UNCAF axioms. That is why $\mathrm{ec}(T_R)$ is a weaker theory than T_R. The theory T_R is axiomatized by completeness and transitivity of weak preference, and by an axiom that defines strict preference. Completeness, and the definition of strict preference, are not UNCAF, as we have shown in Section 2.5.

What then are the UNCAF axioms that hold true in T_R. Well, each falsifying data set must exhibit some kind of cyclic choice behavior: a cycle involving at least one strict revealed preference choice. For example, D^2 in the preceding discussion exhibited

$$a \succ^{\mathcal{D}^2} b \succ^{\mathcal{D}^2} c \succ^{\mathcal{D}^2} a.$$

The data set D^2 suggests an axiom

$$\forall x, y, z \neg (x \succ y \wedge y \succ z \wedge z \succ x)$$

This reasoning applies to cycles of any length, as long as they include at least one strict revealed preference comparison. We obtain that the collection of all UNCAF axioms that hold true in T_R is: for all $n \geq 1$,

$$\forall x_1 ... \forall x_n \neg \left(\bigwedge_{i=1}^{n} (x_i R_i x_{i+1}) \wedge (x_n \succ x_i) \right), \tag{4}$$

where for all i, R_i is either $\succ$ or $\succeq$. The empirical content of T_R, $\mathrm{ec}(T_R)$, is exactly the theory of the data-generating processes in which the above UNCAF axioms are true.

2.8. **Relative theories.** A researcher often wants to take some theory as given (or as baseline) and impose additional hypotheses. For example, modern decision theory papers usually take completeness and continuity axioms as given, and investigate the consequences of some new axiom. The motivation is often to investigate the empirical consequence of some new "substantive" axiom, taking some convenient technical properties as given.

Our theory is applicable to the additional empirical content that results from the *relative* restrictions that a theory imposes on a baseline theory. If our analysis would rule out any axiomatization that uses a continuity axiom, it would not be very useful.

We illustrate relative theories by using our example; we subsume T_R in a theory where the absence of preference can be observed. The reason T_R is not totally falsifiable is that we cannot observe when $x \succ y$ is not true. The situation changes if we instead assume that the absence of preference can be revealed in the data. We introduce two symbols: $\tilde{\succeq}$ to represent absence of weak revealed preference, and $\tilde{\succ}$ to represent absence of strict revealed preference. Consider a theory specified by the axioms

$$x \tilde{\succeq} y \leftrightarrow \neg x \succeq y \quad \text{and} \quad x \tilde{\succ} y \leftrightarrow \neg x \succ y.$$

Then, relative to this theory, the theory of rational preference maximization becomes totally falsifiable. To see why, note that completeness, transitivity, and $\succ$ being the strict part of $\succeq$ can now be formulated with the following four UNCAF axioms:

(1) $\forall x \forall y \neg (x \tilde{\succeq} y \wedge y \tilde{\succeq} x)$
(2) $\forall x \forall y \forall z \neg (x \succeq y \wedge y \succeq z \wedge x \tilde{\succeq} z)$
(3) $\forall x \forall y \neg (x \succeq y \wedge y \tilde{\succeq} x \wedge x \tilde{\succ} y)$
(4) $\forall x \forall y \neg (x \succ y \wedge x \tilde{\succeq} y \wedge y \succeq x)$

The first axiom expresses completeness, the second expresses transitivity; and the last two express, in UNCAF form,

$$\forall x \forall y \left((x \succ y) \leftrightarrow (x \succeq y) \wedge (\neg (y \succeq x)) \right).$$

2.9. **Joint hypotheses.** Totally falsifiable theories have interesting properties. For example, the intersection of two theories may have strictly more empirical content than the two theories viewed in isolation (in symbols, we may have $\mathrm{ec}(T_0 \cap T_1)$ be a proper subset of $\mathrm{ec}(T_0) \cap \mathrm{ec}(T_1)$), *but only if T_0 and T_1 are not totally falsifiable.* Hence, for theories that impose some non-testable restrictions, the "joint hypothesis" $T_0 \cap T_1$ may be strictly stronger than what results from the empirical contents of both theories.

2.10. **Discussion.** At this point we want to pause, take stock, and go over what we have and have not done here. We have introduced a workable formal notion of empirical content, and characterized the form (the syntax) of the axioms that capture empirical content.

To some extent we formalize existing ideas on the desirability of universal axioms, but in so doing we discover that they need to be adapted to the reality of partial observability. As a consequence, UNCAF axioms are needed to understand empirical content. Universality is not enough. To continue with the example of rational choice, and according to Popper's viewpoint, completeness and transitivity would already be the empirical content of rational preference maximization. Yet acyclicity (SARP), rather than completeness and transitivity, constitutes the "correct" notion of empirical content of rational preference maximization. The defect in the criterion of universality lies in its implicit presumption that all *potentially* observable data are *actually* observable.

The implications and importance of our result are numerous. Most decision theoretic and indeed most choice theoretic studies operate under the implicit presumption that all data are actually observable. This renders such models at the least impractical and at the worst useless for testing in applied environments. The value of having an axiom in the UNCAF form cannot be underestimated. Take the case of the sure thing principle, or the independence axiom, both of decision theory. Each of these axioms are in the UNCAF form. They preclude the existence of certain collections of observations simultaneously holding, and can hence be falsified. In fact, each of these has been falsified in experimental settings. One could argue that the falsification of these two axioms has been the driving force behind modern decision theory, leading to models of ambiguity and other behavioral theories. Thus, we suggest that whenever possible, the UNCAF implications of a theory should be described. This holds especially for some of the more sophisticated modern theories of choice.

We should point out some limitations of our study. Our results follow the revealed preference tradition in economics: testing has a non-parametric and non-stochastic meaning. Much of empirical economics, in contrast, uses statistical techniques to estimate parameters and attach a probabilistic value to conclusions on whether a data set is consistent with the theory. A data set for us either falsifies a theory or it does not, and we have not worried about estimation or inference.

The paper focuses on the most basic theories and axioms of individual decision making in microeconomics. We chose to focus on these theories for pedagogical reasons: all economists will be familiar with our examples. We should emphasize that our results are much more widely applicable. They should be of interest to producers and consumers of axioms in economics; a community that probably includes everyone who works on individual decision making, and who is already friendly to the revealed preference approach we have used as motivation.[4]

2.11. **Previous literature.** We are not the first to formally discuss notions of falsifiability and empirical content in an abstract sense. Results exist in the mathematical psychology literature, as well as among philosophers. Adams, Fagot, and Robinson (1970) seems to be the first work discussing empirical content in a formal sense (see also Pfanzagl, Baumann, and Huber (1971) and Adams (1992)). This work defines two theories to be empirically equivalent if the set of all formulas (of a certain type) consistent with one theory is equivalent to the set of all formulas (of a certain type) consistent with the other. Just as in our work, the notion of empirical equivalence necessarily depends on what is allowed as data. The distinction is that these works do not provide a general characterization of the axiomatic structure of empirical content, but rather focus on characterizing the empirical content of specific theories. Pfanzagl, Baumann, and Huber (1971) (p. 106-119) for example, simply define testable formulas to be exactly the universal formulas.

[4]The working paper version develops a detailed application to behavioral economics. There are also applications to game theory, but we have chosen not to emphasize those since they revealed-preference theory has made less headway with game-theoretic models.

Simon and Groen (1973) present a formal study of the testable implications of scientific theories (see also Simon (1979, 1983, 1985); Rynasiewicz (1983); Shen and Simon (1993)) The focus in their work is when a theory that involves theoretical terms can be reduced to statements about observables by a process known as a Ramsey elimination. Apart from the questions that they investigate, the main difference from our work lies in their definition of data. They consider substructures (in the sense of mathematical logic) to be data. Our notion of data, on the other hand, is broader. The notion of substructure does not allow for partial observability, which is a crucial component of our theory (and a feature of economic data sets).

Finally, some of our formal arguments are close to results by Tarski (1954). Tarski's main results deal with languages involving no constant or function symbols. In such a framework, he characterizes those theories that have a universal axiomatization. As we demonstrate below, the issue of universal axiomatization is related to falsification, but Tarski never explored this aspect of the results. We elaborate more on the relation of our theorem with Tarski's in Section 8. In all, our results are hardly novel contributions to Mathematical Logic or Model Theory. Rather, we have formalized some questions that economists in particular care about, and obtained a characterization of the empirical content of a theory.

The discussion in Brown and Kubler (2008b) also provides a general framework for falsification in economic theories. The focus is on mathematical environments that admit quantifier elimination, and on economic theories that can be expressed using these environments. The main example is to theories that can be expressed using polynomial inequalities, and an existential quantifier that ranges over theoretical terms. One can then eliminate the theoretical terms by quantifier elimination and obtain a formulation (a test) purely on observables. The book Brown and Kubler (2008a) contains a collection of papers where this idea is carried out for particular economic theories.

The problems we discuss are very general, but it seems that mostly economists and psychologists have worked on formalizing them. The formalization is an exercise in the axiomatic method, hence it comes naturally to economic theorists and mathematical psychologists.

2.12. **Outline of the rest of the paper.** The rest of the paper is a mathematical presentation of our results. It requires a basic familiarity with standard concepts in mathematical logic, provided in Appendix B. Any reader who understands the definitions in the appendix will be able to follow the mathematical exposition of our results.

Section 3 discusses our general notion of theory, building from concepts in model theory. Section 4 discusses our semantic notions of data, falsifiability, total falsifiability, and falsifiable closure. Section 5 contains our main results: syntactic characterizations of the notions presented in Section 4. The culmination of this section is Section 7.3, where we present our general results relating to relative notions of falsifiability. In Section 8, we present some related works involving Tarski. Section 9 is devoted to an application to Afriat's theorem in our context. Section 10 discusses the relation of our work to the work of Herbert Simon, mainly as expressed in Simon and Groen (1973). Lastly, Section 11 concludes. Appendix A shows how our results on falsifiability can be presented by the dual notion of verifiability, and Appendix B discusses the basic notions from mathematical logic and model theory which are required to understand our paper.

3. Theories and structures

We use standard notions from mathematical logic and model theory. To make our paper self-contained, we have included an appendix with the relevant definitions: see Appendix B. The definitions are taken quite literally from Marker (2002). At the very least, the reader should be familiar with the notions of language, structure, truth, and isomorphism of structures.[5]

The primitive notion in our framework is that of a language. The language we choose should correspond to what we believe to be observable. For example, for studying the basic theory of rational choice, we want a language that–at a minimum–allows us to express the observation "x is preferred to y." Thus we need a language which includes a binary relation symbol intended to represent (revealed) preference. Now, if we can observe the *absence* of preference, "x is not preferred to y," we need to include a separate relation symbol corresponding to the absence of preference. This is an important point because the absence of preference does not need to follow from the absence of an observed preference. To incorporate the observation of absence of preference, we need to incorporate this extra relation symbol. Our notion of data set (below) allows us to distinguish between the absence of observation and the observation of absence; the distinction turns out to be important.

[5]A classical reference is Chang and Keisler (1990).

Remark 1. We use the term 'class' for a collection that can be described by some formula in the language of set theory, but which may be 'too large' to be a set. Thus we can talk about the 'class of all sets' and 'the class of all structures of a language $\mathcal{L}$', even though these classes are not themselves sets. For a formal treatment, see Levy (2002).

Definition 2. Let $\mathcal{L}$ be a language. A *theory* T over $\mathcal{L}$ is a class of structures that is closed under isomorphism. Elements of T are called *models* of T.

Remark 3. Marker and other model theory textbooks only study first-order theories (See Definition 12 below). In our definition of theory we follow Tarski (1954).

We present a first example as an illustration of our basic definitions. It is essentially the example we discussed in the introduction.

Example 4. Consider the language $\mathcal{L} = \langle \succeq, \succ \rangle$ with two binary relations:

- $\succeq$, which is intended to express revealed weak preference,
- and $\succ$, which is intended to express revealed strict preference.

A structure of L is a triple $\mathcal{M} = (M, \succeq^{\mathcal{M}}, \succ^{\mathcal{M}})$, where M is a set, and $\succeq^{\mathcal{M}}$ and $\succ^{\mathcal{M}}$ are binary relations on M.

The *theory of rationality* is the theory of weak-order maximization, denoted by T_{wo}. This is specified as the class of all structures $(M, \succeq^{\mathcal{M}}, \succ^{\mathcal{M}})$ for which $\succeq^{\mathcal{M}}$ is complete and transitive, and for all $x, y \in M$, $x \succ^{\mathcal{M}} y$ if and only if $x \succeq^{\mathcal{M}} y$ and it is not the case that $y \succeq^{\mathcal{M}} x$. That is, $\succeq^{\mathcal{M}}$ expresses weak preference, and $\succ^{\mathcal{M}}$ is the strict part of $\succeq^{\mathcal{M}}$.

We distinguish T_{wo} from the theory of utility maximization, which is the class of $\mathcal{L}$-structures T_u for which there exists a real-valued function $u : M \to \mathbf{R}$ such that $x \succeq^{\mathcal{M}} y \leftrightarrow u(x) \geq u(y)$ and $x \succ^{\mathcal{M}} y \leftrightarrow u(x) > u(y)$. We can also define the "vacuous theory," T_v: The class of all the structures of $\mathcal{L}$. Finally, for future reference, we use the notation T_w for the theory of all the structures of $\mathcal{L}$ that satisfy the axioms (4) in the Introduction. Note that

$$T_u \subseteq T_{wo} \subseteq T_w \subseteq T_v.$$

So we can express that one theory is more restrictive than another by set containment.

Things change if we assume that we can observe the absence of preference. Consider the language $\mathcal{L}' = \langle \succeq, \tilde{\succeq} \rangle$ with two binary relations:

- $\succeq$, which is intended to express revealed weak preference,
- and $\tilde{\succeq}$, which is intended to express the absence of weak preference.

Now the theory of weak orders is axiomatized by:

(1) $\forall x \forall y, (x \succeq^{\mathcal{M}} y) \vee (x \tilde{\succeq}^{\mathcal{M}} y)$
(2) $\forall x \forall y, \neg[(x \succeq^{\mathcal{M}} y) \wedge (x \tilde{\succeq}^{\mathcal{M}} y)]$
(3) $\forall x \forall y \forall z, \neg[(x \succeq y) \wedge (y \succeq z) \wedge (x \tilde{\succeq} z)]$
(4) $\forall x \forall y, \neg[(x \tilde{\succeq} y) \wedge (y \tilde{\succeq} x)]$.

The first axiom expresses that there must be either preference or absence of preference between all pairs. The second axiom expresses consistency between preference and absence of preference: if there is a preference between x and y, there cannot be absence of a preference. The third formalizes transitivity, and the last formalizes completeness.

That is, the theory of weak orders is the class of all structures $(M, \succeq^{\mathcal{M}}, \succ^{\mathcal{M}})$ for which axioms (1)-(4) are true.

4. Empirical Content: Semantics

Definition 5. Let $\mathcal{L}$ be a language. A *data set* $\mathcal{D}$ over $\mathcal{L}$ is given by:

(1) A non-empty set D (the domain of $\mathcal{D}$)
(2) An n-ary relation $P^{\mathcal{D}}$ over D for every n-ary relation symbol P of $\mathcal{L}$
(3) A function $f^{\mathcal{D}} : \text{Dom}(f^{\mathcal{D}}) \subseteq D^n \to D$ for every n-ary function symbol f of $\mathcal{L}$.
(4) A set $C(\mathcal{D})$ of constant symbols of $\mathcal{L}$ and an element $c^{\mathcal{D}} \in D$ for every $c \in C(D)$.

A data set $\mathcal{D}$ is *finite* if the domain D and the sets $\{P | P^{\mathcal{D}} \neq \emptyset\}$, $\{f | \text{Dom}(f^{\mathcal{D}}) \neq \emptyset\}$, and $C(\mathcal{D})$ of, respectively, relation symbols, function symbols and constant symbols that *appear in* $\mathcal{D}$ are finite.

Our definition of data set reflects the crucial property of partial observability. As we explained in the introduction, and discuss in detail in Section 8, a data set does not impose that one observe all the theoretically possible relations among objects in the data set. This imposition would result in a rather unrealistic notion of data set, and our definition avoids it.

Definition 6. Let $\mathcal{L}$ be a language. A structure $\mathcal{M}$ of $\mathcal{L}$ *contains* a data set $\mathcal{D}$, denoted $\mathcal{D} \subseteq \mathcal{M}$ if the following conditions are satisfied:

(1) $D \subseteq M$, where D and M are the domains of $\mathcal{D}$ and $\mathcal{M}$.
(2) $P^{\mathcal{D}} \subseteq P^{\mathcal{M}}$ for every relation symbol P
(3) $f^{\mathcal{D}}$ is the restriction of $f^{\mathcal{M}}$ to $\text{Dom}(f^{\mathcal{D}})$ for every function symbol f.
(4) $c^{\mathcal{D}} = c^{\mathcal{M}}$ for every constant symbol $c \in C(\mathcal{D})$.

Observe that we do not require $P^{\mathcal{D}}$ to be the restriction of $P^{\mathcal{M}}$ to D (and similarly for functions). Consider the language in Example 4, and the structure $\mathcal{M} = (\mathbf{R}, \geq, >)$ of T_{wo}, where $\geq$ is the usual order on $\mathbf{R}$. Then the data set $\mathcal{D}$ with domain $\{1, 2, 3\}$ and the binary relations $\geq^{\mathcal{D}} = \{(2, 1)\}$, $>^{\mathcal{D}} = \varnothing$ is contained in $\mathcal{M}$.

Definition 7. Let $\mathcal{L}$ be a language.

(1) A data set $\mathcal{D}$ *falsifies* a theory T if no model of T contains $\mathcal{D}$.
(2) Let $\mathcal{M}$ be a structure. A theory T is *falsifiable at* $\mathcal{M}$ if $\mathcal{M}$ contains a data set that falsifies T.

A theory T is falsified at a structure $\mathcal{M}$ if some claim that T makes is incompatible with data that could be observed if $\mathcal{M}$ represented the universe.

Definition 8. A theory T is *falsifiable* if there exists some data set that falsifies T.

A theory T is falsifiable if T makes at least one claim that can be demonstrated to be false. Consider Example 4. The theory T_u of utility maximization is falsifiable: the data set $\mathcal{D} = (D, \succeq^{\mathcal{D}}, \succ^{\mathcal{D}})$ with domain $D = \{a, b\}$ and where $\succeq^{\mathcal{D}} = \emptyset$ and $\succ^{\mathcal{D}} = \{(a, b), (b, a)\}$ falsifies T_u.

On the other hand, while T_u is falsifiable, not *all* its claims are falsifiable. For an example, consider the structure $\mathcal{M}_{lex} = (\mathbf{R}^2_+, \geq_{lex}, >_{lex})$, where $\geq_{lex}$ is the lexicographic order on $\mathbf{R}^2_+$. It is well-known that $\mathcal{M}_{lex} \notin T_u$, but no finite data set in $\mathcal{M}_{lex}$ falsifies T_u.

Thus, we may be interested in theories all of whose claims are falsifiable, and more importantly, in the empirical content of a theory such as T_u. These observations motivate the following definitions.

Definition 9. A theory T is *totally falsifiable* if T is falsifiable at every structure which is not a model of T.

Definition 10. Let T be a theory. The *empirical content* of T, denoted $\text{ec}(T)$ is the class of all structures $\mathcal{M}$ such that T is not falsifiable at $\mathcal{M}$.

From Lemma 21 it follows that $\text{ec}(T)$ is a theory (*i.e.* closed under isomorphism). The theory $\text{ec}(T)$ captures our idea of empirical content. In particular, T is totally falsifiable if and only if $\text{ec}(T) = T$.

Example 11. Consider again Example 4. Then $\text{ec}(T_u) = \text{ec}(T_{wo}) = T_w$. Thus, the theory of utility maximization and the theory of preference maximization are empirically indistinguishable. In addition, the empirical content of T_u and T_{wo} is, in a sense, contained in axioms (4) in the introduction.

5. Empirical Content: Syntax

We now formalize the assertions that can be expressed using the language $\mathcal{L}$ to describe properties of $\mathcal{L}$-structures. This follows the details in Appendix B. The only departure we make from classical model theory is the inclusion of a symbol '$\neq$' in our meta-language, which is always interpreted in the "correct" way. Hence, expressions in our language are strings of symbols built from the symbols of $\mathcal{L}$, variable symbols $v_1, v_2, \ldots,$ the equality and inequality symbols $=, \neq$, Boolean connectives $\neg, \vee, \wedge$, quantifiers $\exists, \forall$ and parentheses $($, $)$. As we allow the symbol $\neq$ to appear in our sentences, we need to make small changes in our definitions of formula, sentence, and truth. The changes necessary should be obvious to those familiar with mathematical logic; again, details are presented in Appendix B.

Definition 12. For a set Γ of sentences of $\mathcal{L}$, let $\mathcal{T}(\Gamma)$ be the theory of all structures $\mathcal{M}$ of $\mathcal{L}$ such that all the formulas in Γ are true in $\mathcal{M}$. Theories of the form $\mathcal{T}(\Gamma)$ for some set Γ of formulas are called *first-order theories*. We also say that Γ *axiomatizes* $\mathcal{T}(\Gamma)$.

Example 13. In Example 4, the theory T_{wo} is a first order theory. The theory T_u is not a first order theory. That T_u has no first order axiomatization may not be immediately obvious, but follows from classical results in model theory.

Definition 14. Let $\mathcal{L}$ be a language. A *universal negation of a conjunction of atomic formulas (UNCAF)* sentence of $\mathcal{L}$ is a sentence of the form

$$\forall v_1 \forall v_2 \ldots \forall v_n \neg \left(\phi_1 \wedge \phi_2 \cdots \wedge \phi_m\right)$$

where $\phi_1, \phi_2, \ldots, \phi_m$ are atomic formulas with variables $v_1, \ldots, v_n$.

For a theory T denote by uncaf(T) the set of UNCAF formulas that are true in all models of T. The following theorem is our main result. It established the syntactic characterization of the semantic concept of empirical content. An immediate corollary is that totally falsifiable theories are exactly those which have an UNCAF axiomatization. This is our main result.

Theorem 15. *For every theory T one has $ec(T) = \mathcal{T}\left(uncaf(T)\right)$.*

Corollary 16. *A theory T is totally falsifiable if and only if it admits an UNCAF axiomatization.*

6. Proof of Theorem 15

Let $\mathcal{L}$ be a language and $\mathcal{D}$ a finite data set. For every $d \in D \setminus C(\mathcal{D})$ let v_d be a variable, and let z_d for every $d \in D$ be the term given by $z_d = c$ if $d = c^{\mathcal{D}}$ for some $c \in C(\mathcal{D})$ and $z_d = v_d$ if $d \in D \setminus C(\mathcal{D})$. Let $\phi_{\mathcal{D}}$ be the following UNCAF formula of $\mathcal{L}$:

(5)

$$\phi_{\mathcal{D}} = \forall \bar{v} \neg \bar{\phi}_{\mathcal{D}}(\bar{v}), \text{ where}$$
$$\bar{\phi}_{\mathcal{D}}(\bar{v}) = \left(\bigwedge (z_d \neq z_{d'}) \bigwedge P\left(z_{d_1}, \ldots, z_{d_n}\right) \wedge \bigwedge f\left(z_{d_1}, \ldots, z_{d_n}\right) = z_{f^{\mathcal{D}}(d_1, \ldots, d_n)} \right),$$

The first conjunction ranges over all pairs $d \neq d' \in D$; the second conjunction ranges over all relation symbols P that appear in $\mathcal{D}$ and every $(d_1, \ldots, d_n) \in P^{\mathcal{D}}$; and the third conjunction ranges over all function symbols f that appear in $\mathcal{D}$ and every $(d_1, \ldots, d_n) \in \mathrm{Dom}(f^{\mathcal{D}})$.

Lemma 17. *Let $\mathcal{D}$ be a finite data set. Then $\phi_{\mathcal{D}}$ is not true in $\mathcal{M}$ if and only if $\mathcal{D}$ is contained in some isomorphic copy of $\mathcal{M}$.*

Proof of Proposition 15. We divide the proof into two steps:

Step 1: If $\mathcal{M} \in \mathcal{T}(\mathrm{uncaf}(T))$ then $\mathcal{M} \in \mathrm{ec}(T)$.

Let $\mathcal{D}$ be a data set that falsifies $\mathrm{ec}(T)$. Then from Lemma 17, and the fact that T is closed under isomorphism it follows that $\phi_{\mathcal{D}} \in \mathrm{uncaf}(T)$. Therefore $\mathcal{M} \models \phi_{\mathcal{D}}$, as by hypothesis $\mathcal{M} \in \mathcal{T}(\mathrm{uncaf}(T))$. By Lemma 17 again it follows that $\mathcal{M}$ does not contain $\mathcal{D}$. Therefore $\mathcal{M}$ does not contain any data set that falsifies $\mathcal{D}$, so that T is not falsifiable at $\mathcal{M}$, *i.e.* $\mathcal{M} \in \mathrm{ec}(T)$ as desired.

Step 2: If $\mathcal{M} \notin \mathcal{T}(\mathrm{uncaf}(T))$ then $\mathcal{M} \notin \mathrm{ec}(T)$.

Let $\phi \in \mathcal{T}(\mathrm{uncaf}(T))$ be not true in $\mathcal{M}$. Let $\bar{v} = (v_1, \ldots, v_n)$ be the variables of ϕ so that $\phi = \forall \bar{v} \neg \bar{\phi}(\bar{v}) \in \mathcal{T}(\mathrm{uncaf}(T))$ where $\bar{\phi}(\bar{v})$ is a conjunction of atomic formulas.

Since ϕ is not true in $\mathcal{M}$, it follows that then $\bar{\phi}[\bar{d}]$ is true in $\mathcal{M}$ for some $\bar{d} = (d_1, \ldots, d_n)$. Let $\mathcal{D}$ be the finite data set defined as follows: The domain $D \subseteq \mathcal{M}$ of $\mathcal{D}$ is the set of all elements of the form $t[d_1, \ldots, d_k]$ where t is some term that appears in $\bar{\phi}$. For every relation symbol P,

$$P^{\mathcal{D}} = \{(t_1[d_1, \ldots, d_k], \ldots, t_n[d_1, \ldots, d_k]) \,|\, P(t_1, \ldots, t_n) \text{ appears in } \bar{\phi}\}.$$

For every function symbol f,

$$\mathrm{Dom}(f^{\mathcal{D}}) = \{(t_1[d_1,\ldots,d_k],\ldots,t_n[d_1,\ldots,d_k]) \,|\, f[t_1,\ldots,t_n] \text{ appears in } \bar{\phi}\},$$

and for every $(t_1,\ldots,t_n)$ such that the atomic formula $t = f(t_1,\ldots,t_n)$ appears in $\bar{\phi}$

$$f^{\mathcal{D}}\,(t_1[d_1,\ldots,d_k],\ldots,t_n[d_1,\ldots,d_k]) = t[d_1,\ldots,d_k].$$

If there are two different atomic formulas that appear in $\bar{\phi}$ with the same arguments of f then we choose one of them arbitrarily to define the corresponding value of $f^{\mathcal{D}}$.

Then $\mathcal{D}$ is a data set that is contained in $\mathcal{M}$ and $\bar{\phi}[d_1,\ldots,d_k]$ is true in every structure that contains $\mathcal{D}$, and, in particular, ϕ is not true in any structure that contains $\mathcal{D}$. But ϕ is is true in every model of T, and therefore $\mathcal{D}$ falsifies T. Thus, we proved that $\mathcal{M}$ contains the data set $\mathcal{D}$ that falsifies T and therefore $\mathcal{M} \notin \mathrm{ec}(T)$. □

Proof of Lemma 17. If a structure $\mathcal{M}$ contains $\mathcal{D}$ then substituting d for v_d we get that $\bar{\phi}_{\mathcal{D}}[\bar{d}]$ is false in $\mathcal{M}$ and therefore $\phi_{\mathcal{D}}$ is not true in $\mathcal{M}$. Since truth is preserved under isomorphism, it follows that if an isomorphic copy of $\mathcal{M}$ contains $\mathcal{D}$ then $\phi_{\mathcal{D}}$ is not true in $\mathcal{M}$.

Assume now that $\mathcal{M}$ is a structure of $\mathcal{L}$ such that $\phi_{\mathcal{D}}$ is not true in $\mathcal{M}$, and assume without loss of generality that the domains M and D of $\mathcal{M}$ and $\mathcal{D}$ are disjoint (otherwise replace $\mathcal{M}$ with an isomorphic structure). Let $\bar{m} = (m_d)_{d\in D}$ be elements of $\mathcal{M}$ such that $\bar{\phi}_{\mathcal{D}}[\bar{m}]$ is false in $\mathcal{M}$. Consider the isomorphic structure of $\mathcal{M}'$ which is obtained by replacing every element m_d with d. Then $\bar{\phi}_{\mathcal{D}}[\bar{d}]$ is false in $\mathcal{M}'$. It follows that all the corresponding substitutions of $\bar{d}$ in the atomic formulas in the conjunctions that makes up $\phi_{\mathcal{D}}$ in (5) are true. In particular, $(d_1,\ldots,d_n) \in P_{\mathcal{M}'}$ for every relation symbol P that appears in $\mathcal{D}$ and every $(d_1,\ldots,d_n) \in P^{\mathcal{D}}$. Thus, $P^{\mathcal{D}} \subseteq P_{\mathcal{M}'}$ for every relation symbol P that appears in $\mathcal{D}$, and so property (2) in Definition 6 is satisfied. The other properties are proved by similar argument. Therefore $\mathcal{M}'$ is an isomorphic copy of $\mathcal{M}$ that contains $\mathcal{D}$. □

7. Properties of empirical content

7.1. **Closure properties.** The following lemmas establish some simple properties which are useful later.

Lemma 18. *If a theory T is falsifiable at a structure $\mathcal{M}$ then ec(T) is also falsifiable at $\mathcal{M}$.*

Proof. Let $\mathcal{D}$ be a finite data set that is contained in $\mathcal{M}$ and falsifies T. By Definition 10 no model of $\mathrm{ec}(T)$ contains $\mathcal{D}$ (since $\mathcal{D}$ falsifies T). By Definition 7 this means that $\mathcal{D}$ falsifies $\mathrm{ec}(T)$. Since $\mathcal{M}$ contains $\mathcal{D}$ it follows that $\mathrm{ec}(T)$ is falsifiable at $\mathcal{M}$. □

The following lemmas establish some simple properties which are useful later.

Lemma 19. *If $T_1 \subseteq T_2$ are theories and T_2 is falsifiable at a structure $\mathcal{M}$ then T_1 is also falsifiable at $\mathcal{M}$.*

Lemma 20. *If T_1, T_2 are theories that are falsifiable at a structure $\mathcal{M}$ then $T_1 \cup T_2$ is falsifiable at $\mathcal{M}$.*

Lemma 21. *If a theory T is falsifiable at a structure $\mathcal{M}$ then T is falsifiable at every isomorphic copy $\mathcal{M}'$ of $\mathcal{M}$.*

Proof of Lemma 19. Let $\mathcal{D} \subseteq \mathcal{M}$ be a finite data set that falsifies T_2. Then $\mathcal{D}$ falsifies T_1. □

Proof of Lemma 20. Let $\mathcal{D}_1$ and $\mathcal{D}_2$ be finite data sets that are contained in $\mathcal{M}$ and falsify T_1 and T_2 respectively. Let $\mathcal{D}_1 \cup \mathcal{D}_2$ be the data set with domain $D_1 \cup D_2$ and such that $p^{\mathcal{D}_1 \cup \mathcal{D}_2} = p^{\mathcal{D}_1} \cup p^{\mathcal{D}_2}$ for every relation symbol p, $f^{\mathcal{D}_1 \cup \mathcal{D}_2} = f^{\mathcal{D}_1} \cup f^{\mathcal{D}_2}$ for every function symbol f and $C(\mathcal{D}_1 \cup \mathcal{D}_2) = C(\mathcal{D}_1) \cup C(\mathcal{D}_2)$. Note that $f^{\mathcal{D}_1} \cup f^{\mathcal{D}_2}$ defines a function because $\mathcal{D}_1$ and $\mathcal{D}_2$ are contained in $\mathcal{M}$. Then $\mathcal{D}_1 \cup \mathcal{D}_2$ falsifies $T_1 \cup T_2$. □

Proof of Lemma 21. Let $\eta : \mathcal{M}' \to \mathcal{M}$ be an isomorphism, and let $\mathcal{D} \subseteq \mathcal{M}$ be a finite data set with domain D that falsifies T. Let $\mathcal{D}' \subseteq \mathcal{M}'$ be the data set with domain $D' = \eta^{-1}(D)$, and such that the relations and functions of $\mathcal{D}'$ are the pullbacks by η of the corresponding relations and functions of $\mathcal{D}$, $C(\mathcal{D}') = C(\mathcal{D})$ and $c^{\mathcal{D}'} = \eta^{-1}\left(c^{\mathcal{D}}\right)$ for every $c \in C(\mathcal{D})$. Then it follows from the fact that T is closed under isomorphisms that $\mathcal{D}'$ falsifies T. □

The following proposition says that the operator $T \mapsto \mathrm{ec}(T)$ over theories T has the properties of a topological closure. The theory $\varnothing$ is the theory which contains no structures.

Proposition 22. *The empirical content has the following properties.*

- **Extensiveness:** *$T \subseteq ec(T)$ for every theory T.*
- **Idempotence:** *$ec(ec(T)) = ec(T)$ for every theory T.*
- **Preservation of Nullary Union:** *$ec(\varnothing) = \varnothing$.*

Preservation of Binary Union: *$ec(T_1 \cup T_2) = ec(T_1) \cup ec(T_2)$ for all theories T_1, T_2.*

Proof. Extensiveness follows from the fact that T is not falsifiable at its own models. Idempotence follows from Lemma 18: If $\mathcal{M} \notin \mathrm{ec}(T)$ then T is falsifiable at $\mathcal{M}$ and therefore $\mathrm{ec}(T)$ is falsifiable at $\mathcal{M}$, *i.e.* $\mathcal{M} \notin \mathrm{ec}(\mathrm{ec}(T))$. Preservation of nullary union follows as every model contains a data set falsifying $\varnothing$. Preservation of binary union follows from Lemma 20. □

The following corollary deals with finite axiomatizations. One should not necessarily expect a theory to have a finite axiomatization, as it is equivalent to a uniform bound on the size of a falsifying data set. For example in classical demand theory (Section 9.1), the theory axiomatized by the weak axiom of revealed preference can always be falsified by two observations; the strong axiom, on the other hand, is an infinite collection of axioms, and there is no bound on a falsifying data set. We took the main idea in Corollary 23 from Vaught (1954); the proof follows from the proof of Theorem 16 and it is omitted.

Corollary 23. *Let $\mathcal{L}$ be a language with finitely many symbols, and T a theory over $\mathcal{L}$. Then T admits an axiomatization by finitely many UNCAF sentences if and only if the following condition is satisfied: There is an n such that, for every structure $\mathcal{M}$, if every finite sub data-set of $\mathcal{M}$, whose domain has at most n elements, is contained in some model of T then $\mathcal{M}$ is a model of T.*

7.2. **Joint hypotheses.** We present a trivial example establishing that the falsifiable closure operator does not commute with respect to intersection. While the empirical content of the intersection of two totally falsifiable theories is the intersection of the closures, this is not true of theories that are not totally falsifiable.

Example 24. Let the language $L = \langle R, S \rangle$ involve two unary relations. T' is the vacuous theory of all structures with two unary relations. T_1 is the theory axiomatized by $\forall x, R(x)$. T_2 is the theory axiomatized by $\forall x, R(x) \rightarrow \neg S(x)$. Note that the empirical content of T_1 is T', while the empirical content of T_2 is T_2 itself. Consequently, the intersection of the empirical contents is T_2.

However, the UNCAF axiom $\forall x, \neg S(x)$ is true in $T_1 \cap T_2$, while it is not true in either T_1 or T_2. Consequently the empirical content of $T_1 \cap T_2$ is a proper subtheory of the intersection of the individual empirical contents.

The example is trivial, but captures the essence of a familiar problem. It is possible that two theories imposed jointly imply stronger hypotheses than just those which follow logically from each of the two theories. Our results imply that this only happens for theories which are not totally falsifiable.

7.3. **Relative notions.** It is often useful to have a relative notion of falsifiability. In some cases, there is a theory which we postulate to be a "base" theory, and we want to test some additional hypothesis (a stronger theory). For example, consider the theories in Example 4. We may ask about additional empirical content in the T_u, relative to T_{wo}; and conclude that the hypotheses that T_u adds to T_{wo} have no additional empirical content. We may also be interested in controlled economic experiments, in which some hypotheses are necessarily satisfied by control.

The theories we have been describing up until now must necessarily be completely specified, and everything that these theories postulate must be open to testing–including the primitives. Our results do not require such a detailed description.

To take a trivial example, we may know that there are at least three alternatives over which an agent forms a preference. We could formalize this by ensuring that all structures in our theory have domains with at most three elements. It turns out that so long as our theory is not vacuous, this theory could never be totally falsifiable. The reason is that, if we are given any model $\mathcal{M}$ of our theory, and consider a substructure $\mathcal{M}^* \subseteq \mathcal{M}$ of this theory with a domain containing only two elements, then $\mathcal{M}^*$ is clearly not a model of our theory. But our theory is also not falsifiable at $\mathcal{M}^*$, as $\mathcal{M}^* \subseteq \mathcal{M}$. This is only an example, of course, but it illustrates the need to allow for some hypotheses to be taken as "given."[6]

To discuss relative notions of falsifiability, in this section we fix two theories $T \subseteq T'$. We assume that T' is a "base", or known, theory. We say that T is *falsifiable with respect to* T' if T is falsifiable at some model of T'. Thus a theory T is falsifiable with respect to a weaker theory T' if some claim that T makes in addition to T' is incompatible with data that could be observed if T' were true. T is *totally falsifiable with respect to* T' if T is falsifiable at every model of T' which is not a model of T. The *empirical content of* T *in* T', denoted $\mathrm{ec}_{T'}(T)$, is given by $\mathrm{ec}_{T'}(T) = T' \cap \mathrm{ec}(T)$, the class of all models $\mathcal{M}$ of T' such that T is not falsifiable at $\mathcal{M}$. Note that T is totally falsifiable with respect to T' if and only if $\mathrm{ec}_{T'}(T) = T$. We have the following theorem:

[6]We may also decide to take some mathematical objects as given, so that our axiomatization only needs to characterize economically meaningful hypotheses. Modern decision theory papers often proceed in this way, assuming in one axiom relatively "technical" assumptions (e.g continuity), and then studying a second "substantive" axiom. Our frameworks allows the study of the additional empirical content resulting from the substantive axiom.

Theorem 25. *Suppose $T \subseteq T'$. Then T is totally falsifiable with respect to T' if and only if there exists a set Σ of UNCAF sentences of $\mathcal{L}$ such that $T = T' \cap \mathcal{T}(\Sigma)$.*

Proof of Theorem 25. If T is totally falsifiable with respect to T', then by Proposition 15

$$T = \mathrm{ec}_{T'}(T) = T' \cap \mathrm{ec}(T) = T' \cap \mathcal{T}(\Sigma),$$

where $\Sigma = \mathrm{uncaf}(T)$.

Assume now that $T = T' \cap \mathcal{T}(\Sigma)$ for some set Σ of UNCAF sentences. In particular, every sentence in Σ is true in every model of T and therefore $\Sigma \subseteq \mathrm{uncaf}(T)$. It follows that

$$T' \cap \mathrm{ec}(T) = T' \cap \mathcal{T}\left(\mathrm{uncaf}(T)\right) \subseteq T' \cap \mathcal{T}(\Sigma) = T,$$

where the first equality follows from Proposition 15 and the inclusion from the fact that $\Sigma \subseteq \mathrm{uncaf}(T)$. Since in addition $T \subseteq T' \cap \mathrm{ec}(T)$, we get that $T = T' \cap \mathrm{ec}(T)$, so that T is totally falsifiable with respect to T'. □

Proposition 26. *Let $T \subseteq T'$ be theories. Then $\mathrm{ec}_{T'}(T)$ is the smallest theory that contains T and is totally falsifiable with respect to T'.*

Proof. From the fact that ec is idempotent and monotone (Proposition 22), we conclude that

$$\mathrm{ec}_{T'}(\mathrm{ec}_{T'}(T)) = \mathrm{ec}(\mathrm{ec}(T) \cap T') \cap T' \subseteq \mathrm{ec}(\mathrm{ec}(T)) \cap T' = \mathrm{ec}(T) \cap T' = \mathrm{ec}_{T'}(T).$$

Moreover, $\mathrm{ec}_{T'}(T) \subseteq \mathrm{ec}(\mathrm{ec}_{T'}(T))$ by extensiveness of ec, so that $\mathrm{ec}_{T'}(T) \subseteq T'$ implies $\mathrm{ec}_{T'}(T) \subseteq \mathrm{ec}_{T'}(\mathrm{ec}_{T'}(T))$. Therefore $\mathrm{ec}_{T'}(T)$ is totally falsifiable with respect to T'. Assume now that $T \subseteq \tilde{T} \subseteq T'$ and $\tilde{T}$ is totally falsifiable with respect to T'. Then

$$\mathrm{ec}_{T'}(T) \subseteq \mathrm{ec}_{T'}(\tilde{T}) = \tilde{T},$$

where the first inclusion follows from monotonicity of the closure and the fact that $T \subseteq \tilde{T}$ and the equality from the fact that $\tilde{T}$ is totally falsifiable with respect to T'. □

Example 27. Consider again the language $\mathcal{L}' = \langle \succeq, \tilde{\succeq} \rangle$ of Example 4. We define the theory of orders, T_o, as the class of all structures satisfying

$$\forall x \forall y, [(x \succeq^{\mathcal{M}} y) \leftrightarrow \neg(x \tilde{\succeq}^{\mathcal{M}} y)].$$

In a slight abuse of notation, we use the same notation for language $\mathcal{L}'$ as for $\mathcal{L}$; so we denote the theory of weak order maximization by T_{wo}, and utility representation by T_u.

Then $\mathrm{ec}_{T_o}(T_u) = \mathrm{ec}_{T_o}(T_{wo}) = T_{wo}$. That is, if we assume that every pair is either ranked or unranked, then the theory of weak order is totally falsifiable. The theory of weak order is the empirical content of the theory of utility maximization. The idea that numerical representation of preference is without empirical content is well-known, but it is comforting that our formal notion coincides with our intuition in this case.

The example formalizes the discussion of relative theories in the introduction. If we take as "given" that every pair is either ranked or unranked, then all additional restrictions imposed by the theory of weak order are falsifiable.

8. Relation to Tarski

The following corollary reformulates Corollary 16 in a way that makes the relation to Tarski's Theorem (Tarski (1954)) more transparent.

Corollary 28. *Let $\mathcal{L}$ be a language and T a theory over $\mathcal{L}$. Then T admits an axiomatization by UNCAF sentences if and only if the following condition is satisfied: For every structure $\mathcal{M}$, if every finite sub data-set of $\mathcal{M}$ is contained in some model of T then $\mathcal{M}$ is a model of T.*

An UNCAF sentence is a special case of a *universal sentence*, i.e. a sentence of the form

$$\forall v_1 \dots v_n \phi(v_1, \dots v_n),$$

where ϕ is quantifier-free formula. A theory T admits a universal axiomatization if $T = \mathcal{T}(\Sigma)$ for some set Σ of universal sentences. Tarski (1954) proved the following theorem:

Theorem 29. *Let $\mathcal{L}$ be a language without constants and function symbols and let T be a theory over $\mathcal{L}$. Then T admits axiomatization by universal sentences if and only if the following conditions are satisfied:*

(1) *T is closed under substructures.*
(2) *For every structure $\mathcal{M}$, if every finite substructure of $\mathcal{M}$ is a model of T, then $\mathcal{M}$ is a model of T.*

The similarity of our condition in Corollary 28 and Tarski's second condition is clear: In our framework data sets replace substructures. Our notion of data sets have an important feature. One may only be able to observe some relations among the data, not all of them. For example, for data on revealed preferences, if one observes that x is revealed preferred to y, and that y is revealed preferred to

z, one may not know (not observe) the direction of revealed preference between x and z. Our notion of data sets accommodates this feature of real-world data sets.

Remark 30. The reason we are able to prove a theorem axiomatizing theories with function symbols whereas Tarski could not is that the notion of data set allows a function to be defined on a subdomain of the domain under consideration. In general; however, if we consider a function restricted to an arbitrary subset of a domain, the function may not take values in that subset, and hence the resulting object will not be a substructure. In a sense, the distinction between functions and relations in mathematical logic is made because of the way these objects relate across structures: in our context, they can be considered the same type of object (any function is a relation).

We now turn to formalize the relationship between the syntactic notions of UNCAF and universal axiomatization.

Let us say that a language $\mathcal{L}$ *supports negation of relations* if its relation symbols are divided into pairs $(P, \tilde{P})$. The idea is that $\tilde{P}$ should represent the relation 'P does not hold'. If $\mathcal{L}$ supports negation of relations, we denote by $\Lambda_{\mathcal{L}}$ the set of sentences of the form

$$\forall v_1 \dots \forall v_n \neg P(v_1, \dots, v_n) \leftrightarrow \tilde{P}(v_1, \dots, v_n)$$

for all n-ary relation symbols p in the language. We say that a theory T *respects negation of relations* if $T \subseteq \mathcal{T}(\Lambda_{\mathcal{L}})$, so that $\tilde{P}$ is interpreted as 'P does not hold' in all models of T.

Lemma 31. *Let $\mathcal{L}$ be a language that supports negation of relations. Then for every universal sentence ϕ in $\mathcal{L}$ there exist UNCAF sentences $\phi_1, \dots, \phi_n$ such that $\Lambda_{\mathcal{L}} \vdash \phi \leftrightarrow \phi_1 \wedge \dots \wedge \phi_n$.*

Corollary 32. *Let $\mathcal{L}$ be a language that supports negation of relations, and let $T \subseteq \mathcal{T}(\Lambda_{\mathcal{L}})$. Then there exists a set of universal sentences Σ such that $T = \mathcal{T}(\Lambda_{\mathcal{L}}) \cap \mathcal{T}(\Sigma)$ if and only if there exists a set of UNCAF sentences Σ' such that $T = \mathcal{T}(\Lambda_{\mathcal{L}}) \cap \mathcal{T}(\Sigma)$. T admits a universal axiomatization relative to $\mathcal{T}(\Lambda_{\mathcal{L}})$ if and only if T admits an UNCAF axiomatization relative to $\mathcal{T}(\Lambda_{\mathcal{L}})$.*

Thus, for theories that respect negation of relations our theorem and Tarski's provide the same type of axiomatization.

Proof of Lemma 31. We give a purely syntactic proof: Consider the universal sentence $\forall \bar{v} \bar{\phi}(\bar{v})$, where ϕ is quantifier free and $\bar{v}$ are the variables that appear in ϕ. Writing $\bar{\phi}$ in its conjunctive normal form, we get that ϕ is equivalent to a formula of the form

$$\forall \bar{v} \bigwedge_{i=1}^{m} \bigvee_{j=1}^{n} \phi_{i,j}$$

where each $\phi_{i,j}$ is a *literal*, i.e. an atomic formula or a negation of an atomic formula. Changing the order of the conjunction and the universal quantifier we obtain a formula of the form

$$\bigwedge_{i=1}^{m} \forall \bar{v} \bigvee_{j=1}^{n} \phi_{i,j}.$$

Using De Morgan's law and replacing each $\phi_{i,j}$ with its negation we get a formula of the form

$$\bigwedge_{i=1}^{m} \forall \bar{v} \neg \bigwedge_{j=1}^{n} \phi_{i,j}. \tag{6}$$

Finally, under $\Lambda_{\mathcal{L}}$ every literal is equivalent to an atomic formula: for every term $t_0, t_1, \ldots, t_k$, $\neg f(t_1, \ldots, t_k) = t_0$ is equivalent to $f(t_1, \ldots, t_k) \neq t_0$, and $\neg P(t_1, \ldots, t_k)$ is equivalent to $\tilde{P}(t_1, \ldots, t_k)$. Therefore we can change the formulas $\phi_{i,j}$ in (6) to atomic formulas and so we arrive at a conjunction of UNCAFs, as desired. □

In fact, for the theory of falsifiability, it is often important that our theory support negation of relations. Recall Popper's theory "all swans are white." Clearly, such a theory could never be falsified if it were impossible to observe a swan which was *not* white. The following example is our example of weak order maximization, recast in a language involving only one relation.

Example 33. Let $\mathcal{L} = \langle R \rangle$ be a language involving only one binary relation, interpreted as weak preference. Consider the theory T^*_{wo}, where $\mathcal{M} = (M, R^{\mathcal{M}}) \in T^*_{wo}$ if and only if $R^{\mathcal{M}}$ is a weak order on M. Let T^*_v denote the vacuous theory, consisting of all structures with binary relations. We claim that $\mathrm{ec}(T^*_{wo}) = T^*_v$. This means, in particular, that the theory of weak order has no empirical content unless one can reasonably observe absence of preference.

To see why this is the case, let $\mathcal{D} = (D, R^{\mathcal{D}})$ be a data set, and let $\mathcal{M} = (D, R^{\mathcal{M}})$, where $R^{\mathcal{M}}$ is the binary relation which ranks all pairs. Then $\mathcal{D} \subseteq \mathcal{M}$, and $\mathcal{M} \in T^*_{wo}$.

The assertion that $\mathrm{ec}(T^*_{wo}) = T^*_v$ seems surprising, but it says nothing more than the well-known fact that the preference which is indifferent between all alternatives can rationalize any choices whatsoever when choices are not fully observable.

8.1. **The theorem of Łoś-Tarski.**

Theorem (Łoś-Tarski). *A first order theory is closed under substructures if and only if it admits a universal axiomatization.*

We now turn to give an analogue of Łoś-Tarski's theorem for the case of UNCAF axiomatizations. Let $\mathcal{L}$ be a language. Let $\mathcal{M}$ and $\mathcal{N}$ be structures of $\mathcal{L}$ with domains M and N respectively. Recall that $\mathcal{M}$ is a *weak substructure* of $\mathcal{N}$ if there exists an embedding $\eta : M \to N$ such that

(1) $\eta\left(f^{\mathcal{M}}(a_1, \ldots, a_n)\right) = f^{\mathcal{N}}(\eta(a_1), \ldots, \eta(a_n))$ for every n-ary function symbol f
(2) $(a_1, \ldots, a_n) \in R^{\mathcal{M}}$ only if $(\eta(a_1), \ldots, \eta(a_n)) \in R^{\mathcal{N}}$ for every n-ary relation symbol R
(3) $\eta(c^{\mathcal{M}}) = c^{\mathcal{N}}$ for every constant symbol c.

Theorem 34. *A first order theory is closed under weak substructures if and only if it admits an UNCAF axiomatization.*

The proof is similar to the proof of Łoś-Tarski's Theorem and is omitted.

9. Applications

9.1. **Application: Afriat's theorem.**

Afriat's theorem (Afriat, 1967; Varian, 1982) states that consumption data are rationalizable by a monotonic, continuous, and concave utility if and only if they are rationalizable by a locally nonsatiated preference. Similarly, for demand data satisfying Walras' Law, data which are rationalizable at all are rationalizable by a monotonic, continuous, and concave utility. We shall recast his theorem, using our results, as a statement about the empirical content (the falsifiable closure) of the theory of concave utility maximization.

The language and definitions are similar to those of Example 4, but we need to make some changes to model that preferences are revealed by demand choices at competitive budgets.

Let $\Pi \subseteq \mathbf{R}^n_{++} \times \mathbf{R}_+$. A function $d : \Pi \to \mathbf{R}^n_+$ that satisfies

(1) $p \cdot d(p, I) = I$, and
(2) $d(p, I) = d(\lambda p, \lambda I)$ for all $\lambda > 0$ such that $(\lambda p, \lambda I) \in \Pi$

is a *demand function.*

Let $\mathcal{L}$ be a language with two binary relations, $\succeq$ and $\succ$. The language should also include a constant symbol for every element of $\mathbf{R}^n_+$ and $\mathbf{R}$.[7] We shall introduce three theories: the theory T' of classical demand theory, the subtheory T'_{wo} of weak-order maximization, and the subtheory T_c of concave utility maximization.

First, T' is the class of all structures isomorphic to some $\mathcal{M}$ of $\mathcal{L}$ with $M = \mathbf{R}^n_+$, all constant symbols refer to their named objects, and for which there is a demand function d and $\Pi \subseteq \mathbf{R}^N_{++} \times \mathbf{R}_+$, such that

- $(x, y) \in \succeq$ if and only if there is $(p, I) \in \Pi$ such that $x = d(p, I)$ and $p \cdot y \leq I$;
- $(x, y) \in \succ$ if and only if there is $(p, I) \in \Pi$ such that $x = d(p, I)$ and $p \cdot y < I$.

Second, the theory of weak order maximization is the subtheory T'_{wo} of T' defined as structures isomorphic to some $(\mathbf{R}^n_+, \succeq^*, \succ^*)$ in T' for which there is a complete, reflexive, and transitive binary relation $\succeq$ on X such that

$$(x, y) \in \succeq^* \Rightarrow (x, y) \in \succeq$$
$$(x, y) \in \succ^* \Rightarrow (x, y) \in \succ .$$

The theory of concave utility maximization is the subtheory T_c of T' that is the class of all structures isomorphic to some $(\mathbf{R}^n_+, \succeq^*, \succ^*)$ in T' for which there is a monotonic and concave function $u : \mathbf{R}^n_+ \to \mathbf{R}$ such that

$$(x, y) \in \succeq^* \Rightarrow u(x) \geq u(y)$$
$$(x, y) \in \succ^* \Rightarrow u(x) > u(y).$$

We obtain the following expression of Afriat's (1967) theorem:

Theorem 35. *T'_{wo} is the empirical content of T_c with respect to T'.*

Proof. Consider the set $\Sigma = \{\phi_n, : n = 2, \ldots\}$ of UNCAF formulas, where ϕ_n is

$$\forall v_1, \ldots, \forall v_n (\neg(v_1, v_2) \in \succeq \vee \neg(v_2, v_3) \in \succeq \vee, \ldots, \vee \neg(v_n, v_1) \in \succ).$$

By a well-known theorem (see Richter (1966) and Suzumura (1976)), if a structure $(X, \succeq^*, \succ^*)$ satisfies these sentences, then it is in T'_{wo}. And if a structure $(X, \succeq^*, \succ^*)$ is in T'_{wo}, it is clear to see it satisfies these sentences. So

[7]We introduce constant symbols for each element of $\mathbf{R}^n_+$ so that we do not need to worry about describing consumption space and the relation $\geq$, the function $\cdot$, etc. as part of the problem. The technique of introducing a constant to represent every element in some concrete set is very useful in a variety of contexts in which the underlying set is something whose behavior is well-understood, but whose defining symbols are not meant to be taken as data. Otherwise, we would need to take $\geq$ and the values of the function $\cdot$ as "observable data."

$(X, \succeq^*, \succ^*) \in T'_{wo}$ if and only if it is in T' and satisfies the formulas in Σ. Then, by Theorem 25 $T'_{wo} = T' \cap \mathcal{T}(\Sigma)$ implies that T'_{wo} is totally falsifiable with respect to T', as the formulas in Σ are all UNCAF.

Note that for $(X, \succeq^*, \succ^*)$ in T', the interpretation of the sentences in Σ is that the strong axiom of revealed preference holds.[8] Note that it is meaningful to talk about a finite data set as "satisfying" a collection of sentences in this case, so long as the sentences do not refer to any constants. This is because there are no function symbols in our language. A data set in this context is a structure for our language ignoring constants. Formally, Afriat's theorem then states that if a finite data set $(D, \succeq^{\mathcal{D}}, \succ^{\mathcal{D}})$ satisfies the sentences in Σ, there is a structure $(X, \succeq^*, \succ^*)$ in T_c containing it.

Let $(X, \succeq^*, \succ^*)$ be a structure in $T'_{wo} \backslash T_c$, and let $\mathcal{D}$ be a finite data set contained in $(X, \succeq^*, \succ^*)$. It is easy to verify that each of the axioms in Σ are true for $\mathcal{D}$. So, there exists $\mathcal{M} \in T_c$ containing $\mathcal{D}$ by the argument implied by Afriat's theorem.

Since T'_{wo} is totally falsifiable, we conclude that T'_{wo} is the empirical content of T_c with respect to T'. □

10. Other notions of refutability

We are not the first to formalize the notions of falsification and Popper's logical positivism. We discussed the work of Adams, Fagot, and Robinson (1970), Adams (1992) and Pfanzagl, Baumann, and Huber (1971) in the introduction. The excellent book by Luce, Krantz, Suppes, and Tversky (1990) discusses these contributions. Here, we discuss an approach whose formalism is more similar to ours. In a series of papers, Herbert Simon and coauthors (Simon and Groen, 1973; Simon, 1979, 1983, 1985; Rynasiewicz, 1983; Shen and Simon, 1993) discuss a notion of falsifiability, and the formal structure of falsifiable theories. The focus of this work, as we mentioned in the introduction, is on the elimination of theoretical terms.

This literature has based the idea of falsification on the notion of data as a substructure. We now discuss their notion of falsification, and argue that substructures are inadequate as a notion of data. The definition of falsifiability was proposed by Simon and Groen (1973).[9] They intend their definition to capture the theories that can be axiomatized using only universal quantifiers.

[8] In first-order logic, the strong axiom is an infinite number of axioms, as we make evident here.

[9] Rynasiewicz (1983) proposes a different notion, which he calls "finitely strongly falsifiable." One can show that example 39 presents a theory that is totally falsifiable, and closed under substructures, but is not finitely strongly falsifiable.

A structure $\mathcal{M}$ is *finite* if its domain M is finite.

Definition 36. A theory T is *finitely testable* if there is a structure $\mathcal{M}$ that is not a model of T, and if, for every structure $\mathcal{M}$ that is not a model of T, $\mathcal{M}$ has a finite substructure that is not a model of T.

Definition 37. A theory T is *irrevocably testable* if no model of T has a finite substructure that is not a model of T.

Thus T is finitely and irrevocably testable (FIT) if there is a structure that is not a model of T, and if for every structure $\mathcal{M}$, $\mathcal{M}$ is not a model of T if and only if $\mathcal{M}$ contains a finite substructure that is not a model of T. That is, $\mathcal{M}$ is a model of T if and only if every finite substructure of $\mathcal{M}$ is a model of T. Note that this latter condition also appears in Theorem 29, on relational systems. FIT is the notion of falsifiability used by Simon and Groen. It build on substructures as a notion of data. Note that a relative definition exists: for $T \subseteq T'$, T is FIT with respect to T' if there exists a structure in T' that is not a model of T, and if for every structure $\mathcal{M} \in T'$, $\mathcal{M}$ is not a model of T if and only if $\mathcal{M}$ contains a finite substructure that is not a model of T.

Proposition 38. *If a theory satisfies FIT then it is closed under substructures.*

Proof. Let T satisfy FIT. Let $\mathcal{M}$ be a structure in T. If $\mathcal{M}$ has a substructure that is not in T then this substructure has a finite substructure $\mathcal{B}$ that is not in T. But $\mathcal{B}$ is also a substructure of $\mathcal{M}$, so FIT implies that $\mathcal{M}$ is not in T. It follows that $\mathcal{M}$ cannot have any substructure that is not a model of T. □

By Proposition 38 and the Łoś-Tarski Theorem, FIT implies a universal axiomatization whenever T is a first order theory. The relation between falsifiability and the Łoś-Tarski Theorem is, we hope, clear from our results in Section 8.1.

The following example shows that a theory T may be totally falsifiable with respect to another theory T' [Definition 9], but fail to be FIT (with respect to T'). The example points out that FIT-ness may fail simply because there are no finite substructures of a theory. This can occur for technical reasons related to the definition of substructure.

Example 39. Consider the language $L = \langle 0, q, <, f \rangle$ where q is an unary relation symbol, $<$ is a binary relation symbol, f is a one-place function symbol, and 0 is a constant symbol. Let T' be the class of structures isomorphic to some $\mathcal{M} = (\mathbb{Z}, 0^{\mathcal{M}}, q^{\mathcal{M}}, <^{\mathcal{M}}, f^{\mathcal{M}})$ where $0^{\mathcal{M}}$ is 0 in $\mathbb{Z}$, $<^{\mathcal{M}}$ is a linear order and $x <^{\mathcal{M}} f^{\mathcal{M}}(x)$.

Let T be the class of structures in T' where the formula

$$\forall x \neg q(x)$$

is true. Then by Theorem 25, T is totally falsifiable with respect to T'.

T is also closed under substructures because, if $(\mathbb{Z}, 0^{\mathcal{M}}, q^{\mathcal{M}}, <^{\mathcal{M}}, f^{\mathcal{M}})$ is isomorphic to a model of T and $\mathcal{B}$ is a substructure of $\mathcal{M}$, then $q_{\mathcal{B}}$ coincides with the $q^{\mathcal{M}}$ on $|\mathcal{B}|$.

On the other hand, no model of T' contains any finite substructures. Suppose, to the contrary, that $\mathcal{B}$ is a substructure of $\mathcal{M} \in T'$ and that $|\mathcal{B}|$ is finite. Then $|\mathcal{B}|$ has a largest element $\bar{z}$ according to $<_{\mathcal{B}}$. Note that $f_{\mathcal{B}} = f^{\mathcal{M}}|_{|\mathcal{B}|}$ and $\bar{z} <^{\mathcal{M}} f^{\mathcal{M}}(\bar{z}) = f_{\mathcal{B}}(\bar{z}) \in |\mathcal{B}|$. But $\bar{z}, f_{\mathcal{B}}(\bar{z}) \in |\mathcal{B}|$ and $\bar{z} <^{\mathcal{M}} f_{\mathcal{B}}(\bar{z})$ imply that $\bar{z} <_{\mathcal{B}} f_{\mathcal{B}}(\bar{z})$, which contradicts that $\bar{z}$ was the largest element of $|\mathcal{B}|$.

Consequently, if T were to satisfy FIT with respect to T', it must contain every model of T', which is false. It follows that T does not satisfy FIT with respect to T'.

A theory may satisfy FIT but fail to be totally falsifiable; a simple example involves one unary relation R and theory T axiomatized by $\forall R(x)$.

11. Conclusion

We have developed a theory of the empirical content of an economic theory. The leading examples, throughout the paper, are borrowed from revealed-preference theory; they should be familiar to most economists. We have also shown that the results are applicable to less well-understood theories, and can give new substantive results.

Many papers in decision theory are motivated as providing the testable implications, in the form of testable axioms, of mathematical specifications of individual behavior. We investigate when such axioms are indeed testable, and argue that our analysis is useful for understanding and advancing modern decision theory.

We discuss briefly examples from the recent literature in decision theory. The working paper version of our paper has conditions under which theories of multiple-selves in behavioral economics, and theories of preference aggregation in social choice, are totally falsifiable. That is, all its claims are fully testable.

A recurring methodological issue in economics is the argument over unreal assumptions. There is an early literature, sparked by Milton Friedman's 1953 position that the truth of assumptions does not matter. Recent methodological discussions by Rubinstein (2006), Gul and Pesendorfer (2008), Dekel and Lipman

(2009), and Gilboa (2009), deal with (among other issues) whether the truth of the "story" behind a theory is relevant. In our results, assumptions and stories do not appear explicitly. They appear implicitly in the specification of concrete theories (see for example the theories in Example 4, and Sections 9.1). This is because we have focused on the testable implications of a theory: an UNCAF axiomatization can be seen as a test for the theory.

The framework we have laid out is, however, applicable to discussions of realism as well. An illustration lies in Paul Samuelson's (see Archibald, Simon, and Samuelson (1963)) response to Friedman's position on assumptions. Samuelson effectively counters Friedman by using ideas that we have formalized in our paper. Samuelson makes the point that assumptions matter because either a theory T (described by its "assumptions") is totally falsifiable and thus equivalent to its empirical content, in which case Friedman's point is moot; or it makes non-falsifiable claims, in which case the failure to refute the theory is uninformative about the theory's non-falsifiable claims. In fact, Samuelson argues, by Occam's Razor one should choose the weaker theory, consisting of the empirical content of T (what we have formally termed $\mathrm{ec}(T)$), rather than unnecessary claims in T. Regardless of one's position on the question of realism, this example shows how our notions may be useful.

Finally, we have studied basic ideas from philosophical positivism. They are seen as naive by some philosophers because researchers may have complicated agendas, and be motivated by their environment, in ways that makes falsification not the focus of their research: Philosophy of science since Popper has therefore focused on the sociology of what drives actual research. We are not expert on these matters, of course, but it seems to us that most economists still find the problem of falsification interesting.[10] In fact, the recent methodological discussions in Gul and Pesendorfer (2008), Dekel and Lipman (2009), and Gilboa (2009), all take for granted that one wants to understand a theory's empirical content (possible exceptions are Hicks (1983) and Rubinstein (2006)). We believe that a formal understanding of empirical content is useful, independently of the complexities involved in the actual production of research.[11]

Appendix A. The dual of falsifiable completeness

We have so far discussed falsifiability as a primitive notion, but falsifiability has a dual concept: verifiability. The simplest way to explain these concepts using

[10]Olszewski and Sandroni (2009) study a falsifiability problem for non-deterministic theories.
[11]Gilboa (2009; Chapter 7.3) presents this viewpoint very convincingly.

those we already have is as follows. We can say that a theory T is *verifiably complete* with respect to T' if $T' \backslash T$ is totally falsifiable with respect to T'. Hence, just as falsifiable completeness specifies that all claims of a theory should be falsifiable, verifiable completeness specifies that all claims should be verifiable. Falsifying the complement of a theory is the same as verifying the theory itself–in this sense, falsification and verification are dual.

We can then define the *verifiable interior* of a theory T with respect to T', $\mathrm{vi}_{T'}(T) = T' \backslash \mathrm{ec}_{T'}(T' \backslash T)$. Thus, the verifiable interior of a theory T with respect to T' is the largest subtheory of T which is verifiably complete. It corresponds to the weakest strengthening of the hypotheses for which the theory becomes verifiably complete. Unsurprisingly, the verifiable interior operation is a topological interior, corresponding to the same topology as the empirical content.

Lastly, we can define a sentence to be an ECAF (existential conjunction of atomic formulas) if it is a sentence of the form

$$\exists v_1 \exists v_2 ... \exists v_n (\phi_1 \wedge \phi_2 ... \wedge \phi_n)$$

where each ϕ_i is an atomic formula.

The following result is a trivial consequence of Theorem 25.

Theorem 40. *A theory T is verifiably complete with respect to T' if and only if there exists a set of ECAF sentences, Λ, for which $T = (\bigcup_{\lambda \in \Lambda} \mathcal{T}(\lambda)) \cap T'$.*

We present here a simple example of a theory which is verifiably complete.

Example 41. The example here is one in which we study a private-goods economy, where each individual has her own consumption. We will thus speak of *allocations*. The theory of egalitarian equivalence of some specified allocation, described by Pazner and Schmeidler (1978), asks whether there is some fixed consumption bundle for which each individual is indifferent between her private consumption and the fixed consumption.

To model this, we will suppose that each individual has a preference, and we will consider some fixed allocation; this fixed allocation will be specified in our language by constant symbols.

The language $\mathcal{L}$ involves n binary predicates $R_1, ..., R_n$ and n constant symbols, $c_1, ..., c_n$. The theory that $(c_1, ..., c_n)$ is an egalitarian equivalent allocation is axiomatized by the following sentence:

$$\exists x \bigwedge_{i=1}^{n} (x R_i c_i \wedge c_i R_i x)$$

.

This axiom is immediately seen to be of the ECAF form; hence the theory that $(c_1, ..., c_n)$ is egalitarian equivalent is a verifiably complete theory. This is intuitive, as to verify that the theory holds, one must simply demonstrate the existence of x to which each individual is indifferent.

Appendix B. Definitions from Model Theory (not for publication)

The following definitions are taken, for the most part, quite literally from (Marker, 2002), pp. 8-12. We refer readers to this excellent text for more details; but present the basics here to keep the analysis self-contained. The $\overline{x}$ notation is here used to denote a list, or vector, or elements $(x_1, ..., x_m)$.

We first must specify our language $\mathcal{L}$. The language is a primitive and specifies the *syntax*, or the things we can say.

Definition 42. A *language* $\mathcal{L}$ is given by specifying the following:

(1) a set of function symbols $\mathcal{F}$ and positive integers n_f for each $f \in \mathcal{F}$
(2) a set of relation symbols $\mathcal{R}$ and positive integers n_R for each $R \in \mathcal{R}$
(3) a set of constant symbols $\mathcal{C}$.

The semantics are specified by concrete mathematical objects, called *structures*. Structures provide the appropriate framework for interpreting our syntax.

Definition 43. An $\mathcal{L}$-*structure* $\mathcal{M}$ is given by the following:

(1) a nonempty set M called the *domain* of M
(2) a function $f^{\mathcal{M}} : M^{n_f} \to M$ for each $f \in \mathcal{F}$
(3) a set $R^{\mathcal{M}} \subseteq M^{n_R}$ for each $R \in \mathcal{R}$
(4) an element $c^{\mathcal{M}} \in M$ for each $c \in \mathcal{C}$.

When the language $\mathcal{L}$ is understood, we refer to an $\mathcal{L}$-structure simply as a *structure*. The elements $f^{\mathcal{M}}$, $R^{\mathcal{M}}$, and $c^{\mathcal{M}}$ are called *interpretations* of the corresponding symbols in the language $\mathcal{L}$.

It is useful to be able to give a meaning to certain relations *across* structures. For example, in our case, we have reason to study both the notion of *substructure* and *isomorphism*. The following makes these precise.

Definition 44. Suppose that $\mathcal{M}$ and $\mathcal{N}$ are $\mathcal{L}$-structures with universes M and N respectively. An $\mathcal{L}$-*embedding* $\eta : \mathcal{M} \to \mathcal{N}$ is a one-to-one map $\eta : M \to N$ that preserves the interpretations of all symbols of $\mathcal{L}$: specifically,

(1) $\eta(f^{\mathcal{M}}(a_1, ..., a_{n_f})) = f^{\mathcal{N}}(\eta(a_1), ..., \eta(a_{n_f}))$ for all $f \in \mathcal{F}$ and $a_1, ..., a_{n_f} \in M$

(2) $(a_1, ..., a_{m_R}) \in R^{\mathcal{M}}$ if and only if $(\eta(a_1), ..., \eta(a_{m_R})) \in R^{\mathcal{N}}$ for all $R \in \mathcal{R}$ and $a_1, ..., a_{m_R} \in M$

(3) $\eta(c^{\mathcal{M}}) = c^{\mathcal{N}}$ for $c \in \mathcal{C}$.

Definition 45. An *isomorphism* is a bijective $\mathcal{L}$-embedding.

Definition 46. $\mathcal{M}$ is a *substructure* of $\mathcal{N}$ if $M \subseteq N$ and the inclusion map $\iota : M \to N$ defined by $\iota(m) = m$ for all $m \in M$ is an $\mathcal{L}$-embedding.

The following definition gives us the basic building blocks of our syntax. Note that we include a countable list of "variables" to be used in this definition; these are not part of the language *per se*, but rather part of a "meta language" in that they are present in all languages.

Definition 47. The set of $\mathcal{L}$*-terms* is the smallest set $\mathcal{TE}$ such that

(1) $c \in \mathcal{TE}$ for each constant symbol $c \in \mathcal{C}$

(2) each variable symbol $v_i \in \mathcal{TE}$ for $i = 1, 2, ...,$

(3) if $t_1, ..., t_{n_f} \in \mathcal{TE}$ and $f \in \mathcal{F}$, then $f(t_1, ..., t_{n_f}) \in \mathcal{TE}$.

The following definitions mark our departure from Marker. Specifically, we want to allow atomic formulas to include expressions involving the $\neq$ sign–and we want to include this symbol as part of our meta-language, in the sense that it is present in every language.

Definition 48. Say that ϕ is an *atomic* $\mathcal{L}$*-formula* if ϕ is one of the following

(1) $t_1 = t_2$, where t_1 and t_2 are terms

(2) $t_1 \neq t_2$, where t_1 and t_2 are terms

(3) $R(t_1, ..., t_{n_R})$, where $R \in \mathcal{R}$ and $t_1, ..., t_{n_R}$ are terms

Definition 49. The set of $\mathcal{L}$*-formulas* is the smallest set $\mathcal{W}$ containing the atomic formulas such that

(1) if ϕ is in $\mathcal{W}$, then $\neg\phi$ is in $\mathcal{W}$

(2) if ϕ and ψ, then $(\phi \wedge \psi)$ and $(\phi \vee \psi)$ are in $\mathcal{W}$

(3) if ϕ is in $\mathcal{W}$, then $\exists v_i \phi$ and $\forall v_i \phi$ are in $\mathcal{W}$.

Definition 50. A variable v *occurs freely* in a formula ϕ if it is not inside a $\exists v$ or $\forall v$ quantifier. It is *bound* in ϕ if it does not occur freely in ϕ.

Definition 51. A *sentence* is a formula ϕ with no free variables.

We are now prepared to define a concept of "truth" relating syntax and semantics. We want to define what it means for a sentence to be true in a given structure. The notion we define here is slightly different than Marker, as it again relies on the correct interpretation of the $\neq$ symbol, which is not a primitive there (nor in any other standard text).

Definition 52. Let ϕ be a formula with free variables from $\overline{v} = (v_{i_1}, ..., v_{i_m})$, and let $\overline{a} = (a_{i_1}, ..., a_{i_m}) \in M^m$. We inductively define $M \models \phi(\overline{a})$ as follows. The notation $M \not\models \psi(\overline{a})$ means that $M \models \phi(\overline{a})$ is not true.

(1) If ϕ is $t_1 = t_2$, then $\mathcal{M} \models \phi(\overline{a})$ if $t_1^{\mathcal{M}}(\overline{a}) = t_2^{\mathcal{M}}(\overline{a})$
(2) If ϕ is $t_1 \neq t_2$, then $\mathcal{M} \models \phi(\overline{a})$ if $t_1^{\mathcal{M}}(\overline{a}) \neq t_2^{\mathcal{M}}(\overline{a})$
(3) If ϕ is $R(t_1, ..., t_{n_R})$, then $\mathcal{M} \models \phi(\overline{a})$ if $(t_1^{\mathcal{M}}(\overline{a}), ..., t_{n_R}^{\mathcal{M}}(\overline{a})) \in R^{\mathcal{M}}$
(4) If ϕ is $\neg\psi$, then $\mathcal{M} \models \phi(\overline{a})$ if $\mathcal{M} \not\models \psi(\overline{a})$
(5) If ϕ is $(\psi \wedge \theta)$, then $\mathcal{M} \models \phi(\overline{a})$ if $\mathcal{M} \models \psi(\overline{a})$ and $\mathcal{M} \models \theta(\overline{a})$
(6) If ϕ is $(\psi \vee \theta)$, then $\mathcal{M} \models \phi(\overline{a})$ if $\mathcal{M} \models \psi(\overline{a})$ or $\mathcal{M} \models \theta(\overline{a})$
(7) If ϕ is $\exists v_j \psi(\overline{v}, v_j)$, then $\mathcal{M} \models \phi(\overline{a})$ if there is $b \in M$ such that $\mathcal{M} \models \psi(\overline{a}, b)$
(8) If ϕ is $\forall v_j \psi(\overline{v}, v_j)$, then $\mathcal{M} \models \phi(\overline{a})$ if for all $b \in M$, $\mathcal{M} \models \psi(\overline{a}, b)$.

Definition 53. $\mathcal{M}$ *satisfies* $\phi(\overline{a})$ or $\phi(\overline{a})$ *is true* in $\mathcal{M}$ if $\mathcal{M} \models \phi(\overline{a})$.

Lastly, for our purposes, it is useful to have a notion of a *universal* sentence.

Definition 54. A *universal sentence* or *universal formula* is a sentence of the form $\forall\overline{v}\phi(\overline{v})$, where ϕ is quantifier free.

References

Adams, E. (1992): "On the empirical status of measurement axioms: The case of subjective probability," in *Philosophical and foundational issues in measurement theory*, pp. 53–73. Lawrence Erlbaum Associates.

Adams, E., R. Fagot, and R. Robinson (1970): "On the Empirical Status of Axioms in Theories of Fundamental Measurement," *Journal of Mathematical Psychology*, 7(3), 379–409.

Afriat, S. (1967): "The Construction of Utility Functions from Expenditure Data," *International Economic Review*, pp. 67–77.

Archibald, G. C., H. A. Simon, and P. A. Samuelson (1963): "Discussion," *The American Economic Review*, 53(2), 227–236.

Brown, D., and F. Kubler (2008a): *Computational Aspects of General Equilibrium Theory: Refutable Theories of Value*. Springer Publishing Company, Incorporated.

——— (2008b): "Refutable Theories of Value," in *Computational Aspects of General Equilibrium Theory: Refutable Theories of Value*. Springer Publishing Company, Incorporated.

Chang, C., and H. Keisler (1990): *Model theory*. North Holland.

Dekel, E., and B. Lipman (2009): "How (not) to do decision theory," working paper.

Dekel, E., B. L. Lipman, and A. Rustichini (2009): "Temptation-Driven Preferences," *The Review of Economic Studies*, 76(3), pp. 937–971.

Friedman, M. (1953): “The Methodology of Positive Economics,” in *Essays in positive economics*. The University of Chicago Press, Chicago Illinois.

Gilboa, I. (2009): *Theory of decision under uncertainty*. Cambridge University Press.

Gul, F., and W. Pesendorfer (2001): “Temptation and Self-Control,” *Econometrica*, 69(6), pp. 1403–1435.

Gul, F., and W. Pesendorfer (2008): “The case for mindless economics,” in *The foundations of Positive and normative Economics: A handbook*, pp. 3–42. Oxford University Press, USA.

Hicks, J. (1983): “A discipline not a science,” in *Classics and Moderns: Collected Essays on Economic Theory Volume 3*, pp. 365–375.

Kreps, D. M. (1979): “A Representation Theorem for ”Preference for Flexibility”,” *Econometrica*, 47(3), pp. 565–577.

Levy, A. (2002): *Basic set theory*. Dover.

Luce, R., D. Krantz, P. Suppes, and A. Tversky (1990): *Foundations of measurement, Vol. 3: Representation, axiomatization, and invariance*. Academic.

Marker, D. (2002): *Model theory: an introduction*. Springer Verlag.

Mas-Colell, A., M. D. Whinston, and J. R. Green (1995): *Microeconomic Theory*. Oxford University Press.

Olszewski, W., and A. Sandroni (2009): “Falsifiability,” working paper.

Pazner, E., and D. Schmeidler (1978): “Egalitarian equivalent allocations: A new concept of economic equity,” *The Quarterly Journal of Economics*, 92(4), 671–687.

Pfanzagl, J., V. Baumann, and H. Huber (1971): *Theory of Measurement*. Physica-Verl.

Popper, K. R. (1959): *The Logic of Scientific Discovery*. Hutchinson.

Richter, M. K. (1966): “Revealed Preference Theory,” *Econometrica*, 34(3), 635–645.

Rubinstein, A. (2006): “Dilemmas of an economic theorist,” *Econometrica*, 74(4), 865–883.

Rynasiewicz, R. A. (1983): “Falsifiability and the Semantic Eliminability of Theoretical Languages,” *The British Journal for the Philosophy of Science*, 34(3), 225–241.

Shen, W.-M., and H. A. Simon (1993): “Fitness Requirements for Scientific Theories Containing Recursive Theoretical Terms,” *The British Journal for the Philosophy of Science*, 44(4), 641–652.

Simon, H. A. (1979): “Fit, Finite, and Universal Axiomatization of Theories,” *Philosophy of Science*, 46(2), 295–301.

——— (1983): “Fitness Requirements for Scientific Theories,” *The British Journal for the Philosophy of Science*, 34(4), 355–365.

——— (1985): “Quantification of Theoretical Terms and the Falsifiability of Theories,” *The British Journal for the Philosophy of Science*, 36(3), 291–298.

Simon, H. A., and G. J. Groen (1973): “Ramsey Eliminability and the Testability of Scientific Theories,” *The British Journal for the Philosophy of Science*, 24(4), 367–380.

Suzumura, K. (1976): “Remarks on the Theory of Collective Choice,” *Economica*, 43(172), 381–390.

Tarski, A. (1954): “Contributions to the theory of models I,” *Indagationes Mathematicae*, 16, 572–581.

Varian, H. R. (1982): “The Nonparametric Approach to Demand Analysis,” *Econometrica*, 50(4), 945–974.

Vaught, R. (1954): “Remarks on universal classes of relational systems,” *Indagationes Mathematicae*, 16, 589–591.