THE AXIOMATIC STRUCTURE OF EMPIRICAL CONTENT

CHRISTOPHER P. CHAMBERS, FEDERICO ECHENIQUE,
AND ERAN SHMAYA

Abstract. In this paper, we provide a formal framework for studying the empirical content of a given theory. We define the **falsifiable closure** of a theory to be the least weakening of the theory that makes only falsifiable claims. The falsifiable closure is our notion of empirical content. We prove that the empirical content of a theory can be exactly captured by a certain kind of axiomatization, one that uses axioms which are universal negations of conjunctions of atomic formulas. Our results establish an explicit connection between the data one assumes one may observe, and the empirical content of the theory.

We present applications to standard revealed preference theory, and to recent theories of multiple selves from behavioral economics.

Date: February 23, 2010.

Chambers and Echenique are 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, and Leeat Yariv for comments. Chambers and Echenique acknowledge support from the NSF through grant SES-0751980. Emails: `chambers@hss.caltech.edu` (Chambers), `fede@caltech.edu` (Echenique), and `e-shmaya@kellogg.northwestern.edu` (Shmaya).

Contents

1. Introduction 1
1.1. The concepts. 1
1.2. Falsifiability, empirical content and falsifiable closure. 3
1.3. Axiomatizations. 4
1.4. Partial observability. 6
1.5. Illustration of falsifiable closure. 6
1.6. Relative theories. 7
1.7. Joint hypotheses. 7
1.8. Applications. 7
1.9. Previous literature 8
1.10. Outline 9
2. Theories and structures 10
3. Falsifiable Closure: Semantics 12
4. Syntax 15
4.1. Joint hypotheses 20
4.2. Relative notions 20
4.3. A result on axiomatizations using unobservables 23
5. Relation to Tarski 23
5.1. Data sets vs. substructures 23
5.2. Tarski's result on relational systems 23
5.3. The theorem of Łoś-Tarski 26
6. Applications 27
6.1. Application: Multiple selves preferences 27
6.2. Application: Afriat's theorem 33
7. Other notions of refutability 35
8. Conclusion 38
Appendix A. The dual of falsifiable completeness 39
Appendix B. Basic definitions from Model Theory 41
References 44

> *Die Hauptquelle unserer Unwissenheit liegt darin, dass unser Wissen nur begrenzt sein kann, während unsere Unwissenheit notwendigerweise grenzenlos ist.*
> (The main source of our ignorance lies in the fact that our knowledge can only be finite, while our ignorance must necessarily be infinite.)
>
> Karl Popper

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 has an informal understanding of the testable implications of his theory. We study the common formal structure of all particular instances of empirical content.

The contribution of the paper is to formalize and characterize the concept of empirical content. In the introduction, we describe our main ideas using a simple example.

1.1. 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 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 *universes* representing possible relationships between the observables. The economist claims that one of these universes 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.

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 binary relation, but if we want to test the theory we shall have to be more formal. 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 binary comparisons between alternatives: we observe *revealed weak preference* and *revealed strict preference.* Revealed strict preference is not in principle the strict part of revealed weak 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 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 $\succeq^{\mathcal{D}^1}$ 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. Denote this data set by $\mathcal{D}^1$. In words, we observe that a is weakly revealed preferred to b, and b to c, but we do not observe any strict comparisons. 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. 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.*

In our study, data sets are always finite and observations are partial. A data set consists of finitely many objects that are observed, and a finite number of observed relations among the objects. This is a basic assumption underlying our results. Partial observability is an important feature of actual economic data sets. Partial observability is, for example, unavoidable in revealed preference analysis in consumption theory, where we may observe that a bundle A is revealed preferred to B and B to C but we do not observe the relation between A and C.

A theory is an abstraction that can generate some kinds of data but not others. In our revealed preference example, a theory is a collection of triples $(X, \succeq^X, \succ^X)$, where X is a set of alternatives (not necessarily finite), and $\succeq^X$ and $\succ^X$ are binary relations on X. For example, the theory of rational preference maximization is the class of triples $(X, \succeq^X, \succ^X)$ such that $\succeq^X$ is complete and transitive and $\succ^X$ is the strict part of $\succeq^X$. We denote the theory of rational preference maximization by T_R. A second 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 previous two theories relate to "abstract" environments. We can also discuss theories which are related to "concrete" environments; an example of this is the set of all triples $(X, \succeq^X, \succ^X)$ in T_R where X is a subset of some Euclidean vector space, and where $x \succeq^X y$ whenever x is a larger vector than y (in the usual Euclidean order), and $x \succ^X y$ whenever x is a strictly larger vector than y. We call the elements $(X, \succeq^X, \succ^X)$ of a theory *models of the theory.* When a researcher claims the theory T_R she says that our universe is a model of the theory.

1.2. **Falsifiability, empirical content and falsifiable closure.** We can think of a theory as being consistent with certain data sets and not with others. For example, the theory T_R of rational preference is consistent with the data set $\mathcal{D}_1$ described above. As we remarked above, $\mathcal{D}_1$ 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 falsification of the theory. On the other hand, the data set $\mathcal{D}_2$, where $D_2 = \{a, b, c\}$, $\succeq^{\mathcal{D}_2}$ is empty and $\succ^{\mathcal{D}_2}$ is

$$a \succ^{\mathcal{D}_2} b \succ^{\mathcal{D}_2} c \succ^{\mathcal{D}_2} a$$

is a falsification of the theory. Formally, a data set *falsifies* a theory if it is not contained in any instance (or member) of the theory. It is clear that no member $(X, \succeq^X, \succ^X)$ of T_R can contain $(D_2, \succeq^{\mathcal{D}_2}, \succ^{\mathcal{D}_2})$. A theory is *falsifiable* if there is a data set that falsifies it.

We now come to the most important concept in our paper: empirical content. 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. However, *every data set that falsifies T_U also falsifies T_R.* The example of lexicographic preferences in a Euclidean space, the canonical example of a rational preference with no utility representation, helps us see why: restricted to any finite set, the lexicographic relation is representable by a utility function. Generally, it follows from 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.

We define the empirical content of a theory T to be the largest theory $T' \supseteq T$ with the property that any data set that falsifies T also falsifies T'. We call the mapping from a theory to its empirical content the *falsifiable closure*, denoted as $T \mapsto \mathrm{fc}(T)$. Our notion of empirical content captures the sense in which T_R and T_U are observationally equivalent, and applies to any arbitrary theory. A theory $T' \supseteq T$ is a weakening of T; hence it is less restrictive. But T may include some non-falsifiable restrictions, as is the case, for example, with T_U. The empirical content of T is the weakening that exactly retains all empirical implications of T.

A theory is *falsifiably complete* if it is identical to its falsifiable closure ($T = \mathrm{fc}(T)$). In other words, T is falsifiably complete if all its restrictions are falsifiable.

1.3. Axiomatizations.

The main result of our paper is that empirical content is described by certain kinds of axioms. To study the structure of axioms, we need to have a way of talking about axioms as formal mathematical concepts. The mathematical field of model theory provides us with tools for such an analysis. Our paper uses definitions and basic ideas from model theory. We model the data we can observe by a first order language, involving relation and function symbols. The relation and function symbols should correspond to primitive observables.

Consider our examples of T_R and T_U. The following axiom is "true" in these theories:

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

The axiom expresses a simple consequence of T_R. It is true in T_R and T_U in the sense, which we make precise later, that the axiom is true in every model of T_R and T_U.

Model theory provides a formal language to express assertions about the universe. We call such assertion 'axioms'. Theoretical economists will recognize these axioms even without being familiar with the underlying formal model-theoretic framework. Here is another example of an axiom, which expresses the property of nonsatiation:

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

Axiom 1 has a very specific structure: It starts with universal quantifiers, then a negation symbol, and then a conjunction of statements of the form '$y \succeq z$' or $y \succ x$, which in the terminology of model theory are called atomic formulas. We call axioms with these features "universal negation of conjunctions of atomic formulas," or UNCAF for short. Note that axiom 2 is not UNCAF. Our main result is that the falsifiable closure (the empirical content) of a theory T is the theory axiomatized by all of the UNCAF axioms that are true in T. A falsifiably complete theory is one that has an UNCAF axiomatization. We stress that we do not assume theories can be axiomatized at all; instead, this is derived as a consequence of falsifiable completeness.

UNCAF axioms have two important features. The first feature is that the axiom is a *universal sentence*. It is a statement that is required to hold for all instances of its variables. Popper (1959) famously regarded universality to be a basic property of scientific theories. Contrast the theory that claims "all swans are white" with the theory "there is a non-white 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 non-white swan. The second feature is that the axiom is the negation of the conjunction of certain possible observations.

1.4. **Partial observability.** We have emphasized the partial observability of data sets. Partial observability has important consequences for UNCAF axiomatizations. For example, consider the axiom

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

which expresses that $\succ$ is the strict part of $\succeq$. Axiom (3) is not an UNCAF axiom. In fact, given our primitive assumption that we can only observe revealed weak and revealed strict preference, the theory T_R makes non-testable claims. Suppose the universe is represented by $(\mathbf{R}, \geq, >')$, so that the set of alternatives is $\mathbf{R}$, $\geq$ is the usual order on $\mathbf{R}$ and where $>'$ is defined as $x >' y$ if $x > y + 5$. Then axiom (3) does not hold for our universe, and the universe is not consistent with our theory T_R. However, no data set will ever falsify the theory since every data set that is contained in $(\mathbf{R}, \geq, >')$ is compatible with T_R. For example, the data set with alternatives $\{3, 5\}$, $5 \geq 3$ and $>' = \emptyset$ is not incompatible 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. As a result, the theory T_R is not falsifiably complete.

1.5. **Illustration of falsifiable closure.** To illustrate our main result, it should be familiar to most readers that a data set falsifies T_R if and only if it contains a preference cycle containing at least one strict comparison. Intuitively, we can conclude that $\mathrm{fc}(T_R)$ is the exactly the theory precluding these cycles. According to our results, $\mathrm{fc}(T_R)$ should be axiomatized by the UNCAF axioms satisfied by T_R. These axioms are: for all $n \geq 1$,

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

where for all i, R_i is either $\succ$ or $\succeq$. Note that these axioms correspond exactly to our intuition: each such axiom rules out certain types of cycles.

1.6. **Relative theories.** The reason T_R is not falsifiably complete 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 with respect to this theory, the theory of rational preference maximization becomes falsifiably complete. 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,

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

1.7. **Joint hypotheses.** Falsifiably complete 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 $\mathrm{fc}(T_0 \cap T_1)$ be a proper subset of $\mathrm{fc}(T_0) \cap \mathrm{fc}(T_1)$), *but only if T_0 and T_1 are not falsifiably complete.* 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.

1.8. **Applications.** We present two applications in the paper, one to the recent theories on multiple selves in behavioral economics, and one where we recreate the classical Afriat's theorem using our framework. Here we describe the multiple-selves application.

A "multiple-selves theory" postulates that individual behavior is guided by several preference relations (see e.g. Ambrus and Rozen (2008), Green and Hojman (2008), Manzini and Mariotti (2007), O'Donoghue and Rabin (1999) or Fudenberg and Levine (2006)). We summarize the theory using a preference aggregation rule, a mapping from a list of n preferences to a single individual's preference.[1] We assume that we can observe the individual's revealed preference and absence of revealed preference, as above, and we investigate the empirical content of these theories. The proposers of multiple selves theory were clearly motivated by empirical phenomena, but the full empirical content of these theories is not known. We prove that multiple selves theories are falsifiably complete, and thus possess an UNCAF axiomatization, as long as the preference aggregation rule satisfies some hypotheses.

1.9. **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.

---

[1] Many modern works take choice functions as primitive, instead of revealed preference. Our general framework can be applied to these theories, but in the result we prove about multiple selves we have assumed revealed preference as primitive.

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, wich 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. 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 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.

1.10. **Outline.** The presentation of the paper is as follows. Section 2 discusses our general notion of theory, building from concepts in model theory. Section 3 discusses our semantic notions of data, falsifiability, falsifiable completeness, and falsifiable closure. Section 4 contains our main results: syntactic characterizations of the notions presented in Section 3. The culmination of this section is Section 4.2, where we present our general results relating to relative notions of falsifiability. In Section 5, we present some related works involving Tarski. Section 6 is devoted to two applications: one involves unknown results on multiple selves; the other is a presentation of Afriat's theorem in our context. Section 7 discusses the relation of our work to the work of Herbert Simon, mainly as expressed in Simon and Groen (1973). Lastly, Section 8 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.

2. 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.

The language we choose should correspond to those objects which we believe to be observable as data in our theory. There are important and subtle issues involved in the formulation of a language. 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.

1. *Remark.* 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).

2. **Definition.** 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.

3. **Example.** Consider the language $\mathcal{L} = \langle R, \tilde{R} \rangle$ with two binary relations:

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

A structure of L is a triple $\mathcal{M} = (M, R^{\mathcal{M}}, \tilde{R}^{\mathcal{M}})$, where M is a set, and $R^{\mathcal{M}}$ and $\tilde{R}^{\mathcal{M}}$ 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 $(M, R^{\mathcal{M}}, \tilde{R}^{\mathcal{M}})$ for which $R^{\mathcal{M}}$ is complete and transitive, and for all $x, y \in M$, $x\, \tilde{R}^{\mathcal{M}}\, y$ if and only if $x\, R^{\mathcal{M}}\, y$ is false. That is, $R^{\mathcal{M}}$ expresses weak preference, while $\tilde{R}^{\mathcal{M}}$ expresses the absence of weak preference.

We can write this more carefully as follows: T_{wo} is the class of all $\mathcal{L}$-structures for which the following axioms are true:

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

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.

For future reference, we denote the class of all structures for which axioms 2,3, and 4 are true by T_w.

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 $x\, R\, y \leftrightarrow u(x) \geq u(y)$ and $x\, \tilde{R}\, y \leftrightarrow u(x) < u(y)$.

Finally, we can define the "vacuous theory" T_v of all the structures of L. 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.

4. *Remark.* Marker and other model theory textbooks only study first-order theories (See Definition 17 below). In our definition of theory we follow Tarski (1954).

3. FALSIFIABLE CLOSURE: SEMANTICS

5. **Definition.** 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 $f^{\mathcal{D}} : \mathrm{Dom}(f^{\mathcal{D}}) \subseteq D^n \to D$ 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 $\{P | P^{\mathcal{D}} \neq \emptyset\}$, $\{f | \mathrm{Dom}(f^{\mathcal{D}}) \neq \emptyset\}$, and $C(\mathcal{D})$ of, respectively, relation symbols, function symbols and constant symbols that *appear in* $\mathcal{D}$ are finite.

There are some subtle issues in the definition of data set. In particular, as we explain in detail in Section 5.1, 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. We model data sets in this way in order to capture the idea of partial observability.

6. **Definition.** 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 $\mathrm{Dom}(f^{\mathcal{D}})$ 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 3, 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 relation $R^{\mathcal{D}} = \{(2, 1)\}$, is contained in $\mathcal{M}$.

7. **Definition.** 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 is incompatible with data that could be observed if $\mathcal{M}$ represented the universe.

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

8. **Lemma.** *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 $\mathcal{M}$.*

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

10. **Lemma.** *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 8. Let $\mathcal{D} \subseteq \mathcal{M}$ be a finite data set that falsifies T_2. Then $\mathcal{D}$ falsifies T_1. $\square$

Proof of Lemma 9. 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 $T_1 \cup T_2$. $\square$

Proof of Lemma 10. 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. $\square$

11. **Definition.** 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 3. The theory T_u of utility maximization is falsifiable: the data set $\mathcal{D} = (D, R^{\mathcal{D}}, \tilde{R}^{\mathcal{D}})$ with domain $D = \{a, b\}$ and where $R^{\mathcal{D}} = \emptyset$ and $\tilde{R}^{\mathcal{D}} = \{(a, b), (b, a)\}$ 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.

12. **Definition.** A theory T is *falsifiably complete* if T is falsifiable at every structure which is not a model of T.

13. **Definition.** Let T be a theory. The *falsifiable closure* of T, denoted $\mathrm{fc}(T)$ is the class of all structures $\mathcal{M}$ such that T is not falsifiable at $\mathcal{M}$.

From Lemma 10 it follows that $\mathrm{fc}(T)$ is a theory (*i.e.* closed under isomorphism). The theory $\mathrm{fc}(T)$ captures our idea of empirical content. In particular, T is falsifiably complete if and only if $\mathrm{fc}(T) = T$.

14. **Example.** Consider again Example 3. Then $\mathrm{fc}(T_u) = \mathrm{fc}(T_{wo}) = T_w$. 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 2-4 of Example 3. Axiom 1 expresses a non-falsifiable property, and the additional hypotheses implicit in T_u are also non-testable.

15. **Lemma.** *If a theory T is falsifiable at a structure $\mathcal{M}$ then $fc(T)$ 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 13 no model of $\mathrm{fc}(T)$ contains $\mathcal{D}$ (since $\mathcal{D}$ falsifies T). By Definition 7 this means that $\mathcal{D}$ falsifies $\mathrm{fc}(T)$. Since $\mathcal{M}$ contains $\mathcal{D}$ it follows that $\mathrm{fc}(T)$ is falsifiable at $\mathcal{M}$. $\square$

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

16. **Proposition.** *The falsifiable closure has the following properties.*

Extensiveness: *$T \subseteq fc(T)$ for every theory T.*
Idempotence: *$fc(fc(T)) = fc(T)$ for every theory T.*
Preservation of Nullary Union: *$fc(\varnothing) = \varnothing$.*
Preservation of Binary Union: *$fc(T_1 \cup T_2) = fc(T_1) \cup fc(T_2)$ for all theories T_1, T_2.*

Proof. Extensiveness follows from the fact that T is not falsifiable at its own models. Idempotence from Lemma 15: If $\mathcal{M} \notin \mathrm{fc}(T)$ then T is falsifiable at $\mathcal{M}$ and therefore $\mathrm{fc}(T)$ is falsifiable at $\mathcal{M}$, *i.e.* $\mathcal{M} \notin \mathrm{fc}(\mathrm{fc}(T))$. Preservation of nullary union follows as every model contains a data set falsifying $\varnothing$. Preservation of binary union follows from Lemma 9. $\square$

4. 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 term, formula, sentence, and truth. The changes necessary should be obvious to those familiar with mathematical logic; again, details are presented in Appendix B.

17. **Definition.** 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* $\mathcal{T}(\Gamma)$.

18. **Example.** In Example 3, 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.

19. **Definition.** 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 $v_1, \ldots, v_n$.

The following result provides the syntactic characterization of the semantic concept of falsifiable completeness. Falsifiably complete theories are exactly those which have an UNCAF axiomatization. This is our main result.

20. **Theorem.** *A theory T is falsifiably complete if and only if it admits an UNCAF axiomatization.*

The following corollary is an immediate consequence of Theorem 20 and Definition 12. It will be of interest to us later, in comparing our work with that of Tarski (1954).

21. **Corollary.** *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.*

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 6.2), 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 22 from Vaught (1954); the proof follows from the proof of Theorem 20 and it is omitted.

22. **Corollary.** *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.*

For a theory T denote by uncaf(T) the set of UNCAF formulas that are true in all models of T.

Theorem 20 is an immediate consequence of Proposition 23

23. **Proposition.** *For every theory T one has $fc(T) = \mathcal{T}\left(uncaf(T)\right)$.*

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 $\mathcal{L}$:

(4)

$$\phi_{\mathcal{D}} = \forall \bar{v} \neg \bar{\phi}_{\mathcal{D}}(\bar{v}), \text{ where}$$

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

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 $(d_1, \ldots, d_n) \in \mathrm{Dom}(f^{\mathcal{D}})$.

24. **Lemma.** *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 23. We divide the proof into two steps:

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

Let $\mathcal{D}$ be a data set that falsifies $\mathrm{fc}(T)$. Then from Lemma 24, 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 24 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.* $\mathcal{M} \in \mathrm{fc}(T)$ as desired.

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

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,

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

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

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

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 $\bar{\phi}[d_1,\dots,d_k]$ 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 $\mathcal{M} \notin \mathrm{fc}(T)$. □

Proof of Lemma 24. 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 (4) are true. In particular, $(d_1,\dots,d_n) \in P_{\mathcal{M}'}$ for every relation symbol P that appears in $\mathcal{D}$ and every $(d_1,\dots,d_n) \in P^{\mathcal{D}}$. 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 $\mathcal{D}$. □

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

25. **Example.** 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 $\forall x, R(x) \to \neg S(x)$. Note that the falsifiable closure of T_1 is T', while the falsifiable closure of T_2 is T_2 itself. Consequently, the intersection of the falsifiable closures 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 falsifiable closure of $T_1 \cap T_2$ is a proper subtheory of the intersection of the individual falsifiable closures.

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 falsifiably complete.

4.2. **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 3. 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 falsifiably complete. 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."[2]

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 *falsifiably complete with respect to* T' if T is falsifiable at every model of T' which is not a model of T. The *falsifiable closure of* T *in* T', denoted $\mathrm{fc}_{T'}(T)$, is given by $\mathrm{fc}_{T'}(T) = T' \cap \mathrm{fc}(T)$, the class of all models $\mathcal{M}$ of T' such that T is not falsifiable at $\mathcal{M}$. Note that T is falsifiably complete with respect to T' if and only if $\mathrm{fc}_{T'}(T) = T$. We have the following theorem:

26. **Theorem.** *Suppose $T \subseteq T'$. Then T is falsifiably complete 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 26. If T is falsifiably complete with respect to T', then by Proposition 23

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

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

---

[2] We may also decide to take some mathematical objects as given, so that our axiomatization only needs to characterize economically meaningful hypotheses.

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 $\Sigma \subseteq \text{uncaf}(T)$. It follows that

$$T' \cap \text{fc}(T) = T' \cap \mathcal{T}\,(\text{uncaf}(T)) \subseteq T' \cap \mathcal{T}(\Sigma) = T,$$

where the first equality follows from Proposition 23 and the inclusion from the fact that $\Sigma \subseteq \text{uncaf}(T)$. Since in addition $T \subseteq T' \cap \text{fc}(T)$, we get that $T = T' \cap \text{fc}(T)$, so that T is falsifiably complete with respect to T'. $\square$

27. **Proposition.** *Let $T \subseteq T'$ be theories. Then $fc_{T'}(T)$ is the smallest theory that contains T and is falsifiably complete with respect to T'.*

Proof. From the fact that fc is idempotent and monotone (Proposition 16), we conclude that

$$\text{fc}_{T'}(\text{fc}_{T'}(T)) = \text{fc}(\text{fc}(T) \cap T') \cap T' \subseteq \text{fc}(\text{fc}(T)) \cap T' = \text{fc}(T) \cap T' = \text{fc}_{T'}(T).$$

Therefore $\text{fc}_{T'}(T)$ is falsifiably complete with respect to T'. Assume now that $T \subseteq \tilde{T} \subseteq T'$ and $\tilde{T}$ is falsifiably complete with respect to T'. Then

$$\text{fc}_{T'}(T) \subseteq \text{fc}_{T'}(\tilde{T}) = \tilde{T},$$

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 falsifiably complete with respect to T'. $\square$

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

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

Then $\text{fc}_{T_o}(T_u) = \text{fc}_{T_o}(T_{wo}) = T_{wo}$. That is, if we assume that every pair is either ranked or unranked (in fact, this assumption would usually be implicit), then the theory of weak order is falsifiably complete. The theory of weak order is the falsifiable closure 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.

4.3. **A result on axiomatizations using unobservables.** Often, a theory has an axiomatization involving unobservables. Obviously, such an axiomatization cannot directly lead to empirical falsification. We find conditions under which a theoretical axiomatization can be "projected" on observables to yield a falsifiably complete theory.

Let $\mathcal{F} \subseteq \mathcal{L}$ be languages, such that $\mathcal{L}$ contains all the symbols of $\mathcal{F}$ and possibly additional relation symbols. The idea is that the additional symbols in $\mathcal{L}$ are meant to signify theoretical and unobservable terms. For every $\mathcal{L}$-structure $\mathcal{M}$, we denote by $F(\mathcal{M})$ the $\mathcal{F}$-structure induced from $\mathcal{M}$ by forgetting the relations that corresponds relation symbols not in $\mathcal{F}$. For every $\mathcal{L}$-theory T we denote by $F(T)$ the theory of all structures of the form $F(\mathcal{M})$ for some model $\mathcal{M}$ of T.

29. **Proposition.** *If T is a falsifiably complete $\mathcal{L}$-theory then $F(T)$ is a falsifiably complete $\mathcal{F}$-theory.*

So a theory that is falsifiably complete when we say that theoretical objects are observable is automatically falsifiably complete in the correct observable form—as long as the observable structures are obtained by "projection" from unobservables as in the proposition.

5. Relation to Tarski

5.1. **Data sets vs. 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. The competing notion of substructures as data sets (see the discussion in Section 7) does not.

5.2. **Tarski's result on relational systems.** An UNCAF sentence is a special case of a *universal sentence*, i.e. a sentence of the form

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

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:

30. **Theorem.** *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 21 and Tarski's second condition is clear: In our framework data sets replace substructures. Indeed; 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.

31. **Lemma.** *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 $\Lambda_{\mathcal{L}} \vdash \phi \leftrightarrow \phi_1 \wedge \cdots \wedge \phi_n$.*

32. **Corollary.** *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

$$\bigwedge_{i=1}^{m} \forall \bar{v} \neg \bigwedge_{j=1}^{n} \phi_{i,j}. \tag{5}$$

Finally, under $\Lambda_{\mathcal{L}}$ every literal is equivalent to an atomic formula: for every term $t_0, t_1, \dots, t_k$, $\neg f(t_1, \dots, t_k) = t_0$ is equivalent to $f(t_1, \dots, t_k) \neq t_0$, and $\neg P(t_1, \dots, t_k)$ is equivalent to $\tilde{P}(t_1, \dots, t_k)$. Therefore we can change the formulas $\phi_{i,j}$ in (5) to atomic formulas and so we arrive at a conjunction of UNCAFs, as desired. $\square$

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.

33. **Example.** 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 $\text{fc}(T^*_{wo}) = T^*_v$. 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 result 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.

5.3. 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.

34. **Theorem.** *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.

6. Applications

6.1. Application: Multiple selves preferences.

We apply our concepts to a popular model without a known axiomatization, the model of multiple selves. The purpose of this exercise is to demonstrate that the concepts we introduce are useful for studying theories which have no known axiomatizations (and hence whose empirical content is not completely understood). Models of multiple selves are motivated by empirical observations (see e.g. Ambrus and Rozen (2008), Green and Hojman (2008), Manzini and Mariotti (2007), O'Donoghue and Rabin (1999) or Fudenberg and Levine (2006)), but often they lack an axiomatization in terms of observables. Here we exhibit a broad class of such models which are falsifiably complete.

In our framework, given is a fixed and finite set of agents, the "selves." Given is also a rule for aggregating agents' preferences into a single preference. The interpretation is that an individual has conflicting preferences (perhaps different preferences for different motivations) and reconciles these preferences with a preference aggregation rule. We observe an aggregate preference (a revealed preference), and we would like to know whether it could be generated by the rule for *some* profile of agents' preferences.[3] We want to test whether or not a specific group of selves uses a particular preference aggregation rule in making decisions, only having observed the aggregate ranking. This question is the correct formulation of the standard revealed-preference exercise for the multiple selves model.

---

[3] In this paper, we focus on preferences which are linear orders; however the results apply more broadly.

Multiple selves theories are an excellent example of how hard it can be to show falsifiability. The theories have a trivial existential (second-order) axiomatization: Given a preference aggregation rule, the theory is the collection of observables for which *there exists* preferences for individual selves generating the observable behavior. Leaving aside the second-order nature of this axiomatization, the problem with an existential axiomatization is that we cannot conclude that the theory is falsifiable. Recall the example of Popper (1959): the theory that there is a non-white swan is not testable because we would need to examine all the swans in the universe. Here, for a given observed behavior, we would need to check all possible preferences that the selves might have; for an infinite set of alternatives, this set of preferences is vast. The fact that the axiomatization is second order means we have to search over preference profiles–themselves extremely complicated objects. We present a class of aggregation rules that lead to falsifiably complete theories; theories with an UNCAF axiomatization.

We require a *finite* cardinality of agents, and any preference aggregation rule which is neutral and satisfies independence of irrelevant alternatives. We show that the theory is falsifiably complete, given that we can observe both aggregate preference and absence of aggregate preference (and that these relations behave in the proper way).

The previous literature relates to the theory of social choice, where, given some preference aggregation rule, there have been efforts to axiomatize relations which are rationalized by some society of agents. When the society can be arbitrarily large, it is known that *any* transitive antisymmetric relation is the Pareto relation for *some* society (which may be large)–this is essentially the Szpilrajn theorem. Because of this, any complete binary relation with a transitive asymmetric part is the result of the Pareto extension rule for some society (we identify indifferent alternatives for the Pareto extension rule with unranked alternatives for the Pareto ordering–see Sen (1969)). Results for majority rule are even weaker: McGarvey (1953) showed that any complete binary relation is the majority rule relation for some society of agents (which again may be large). Deb (1976) extends the result to more social choice rules; Kalai (2004) generalizes this result to an even broader class. Shelah (2009) establishes results on different domains of preferences of individuals.

The current behavioral literature borrows from the social choice literature, and interprets the society as a group of conflicting tendencies within an individual decision maker: multiple selves. This literature attempts to understand the empirical content of such assumptions. In particular, Green and Hojman (2008) generalize McGarvey's program to choice functions. Ambrus and Rozen (2008) give sufficient conditions (stated in terms of number of "violations" of classical rationality) for a choice function to be rationalizable by conflicting selves for a fixed number of agents. DeClippel and Eliaz (2009) provide a full characterization of choice rules which can result from a specific social choice rule–the fallback solution on a *fixed pair* of agents. We only consider preference relations and not choice functions here; however, we show that the predictions of nearly every such model can be empirically falsified even in the case where we hypothesize a finite and known cardinality of "selves."[4]

There are very few results like ours, assuming a fixed and finite population of selves. Dushnik and Miller (1941) give necessary and sufficient conditions for a binary relation to be the intersection of a pair of linear orders; this can dually be seen as an axiomatization for binary relations which are the image of the Pareto extension rule for two agents. This characterization theorem both relies on existential quantification, and is not a first order characterization.[5] Dushnik and Miller (1941)'s existential axiomatization cannot be the basis for falsification.[6] Sprumont (2001) provides a similar characterization in a restricted case. Both of these results are of interest as those relations which are the intersection of a pair of linear orders are exactly those relations which can be rationalized by the Pareto-extension rule.

---

[4] It is surprisingly more difficult to axiomatize such models for a fixed and known set of selves, than for an arbitrary set of agents.

[5] That is, it involves quantification over relations.

[6] In particular, the theory of Pareto relations for n agents was not known to be falsifiably complete. Our Theorem 36 demonstrates that it is.

We work with neutral preference aggregation rules which satisfy independence of irrelevant alternatives. By working with such preference aggregation rules, we need not specify what the global set of alternatives is in advance. A set of agents N is fixed and finite. A *preference aggregation rule* is therefore defined to be a mapping carrying any set of alternatives X and any N vector of linear orders[7] (termed a *preference profile*) over those alternatives $(R^1, ..., R^n)$ to a complete binary relation over X. We write $R_{f(R^1,...,R^n)}$ for the binary relation which results (suppressing notation for dependence on X). We assume the following property:

35. **Definition.** (Neutrality and Independence of irrelevant alternatives): For all sets X and Y, for all $x, y \in X$ and all $w, z \in Y$ and all preference profiles $(R^1, ..., R^n)$ over X and $(R'^1, ..., R'^n)$ over Y, if for all $i \in N$, $x\ R^i\ y \Leftrightarrow w\ R'^i\ z$, then $x\ R_{f(R^1,...,R^n)}\ y \Leftrightarrow w\ R_{f(R'^1,...,R'^n)}\ z$.[8]

This hypothesis embeds both the neutrality and independence of irrelevant alternatives assumptions. These assumptions seem to be the minimal assumptions needed to apply Theorem 30.

Given f, we will say that a binary relation R on a set X is *f-rationalizable* if there exists a profile of linear orders $(R_1, ..., R_n)$ for which $R = R_{f(R^1,...,R^n)}$.

Denote by $\mathcal{L} = \langle R, \tilde{R} \rangle$ the language involving two binary relations, and let $\mathcal{T}(\Lambda_{\mathcal{L}})$ be the theory of all structures satisfying the axiom

$$\forall x \forall y, x R^{\mathcal{M}} y \leftrightarrow \neg x \tilde{R}^{\mathcal{M}} y.$$

A structure is *f-rationalizable* if $R^{\mathcal{M}}$ is f-rationalizable and $x R^{\mathcal{M}} y \leftrightarrow \neg x \tilde{R}^{\mathcal{M}} y$. The class of f-rationalizable structures is denoted $\mathcal{T}_f$. Note that $\mathcal{T}_f$ is in fact a theory, as it is closed under isomorphism (this is the content of neutrality).

---

[7] A linear order is complete, transitive, and anti-symmetric

[8] Formally, neutrality means that social rankings should be independent of the names of alternatives, and independence of irrelevant alternatives means that the social preference between a pair of alternatives should depend only on the individual preferences between that pair. We have collapsed these two hypotheses into one larger condition.

36. **Theorem.** *For every f, $\mathcal{T}_f$ is falsifiably complete with respect to $\mathcal{T}(\Lambda_{\mathcal{L}})$.*

37. *Remark.* If absence of ranking is unobservable, that is if we consider the language that include only a single relation symbol R, then the theory of all structures in which R is an aggregation of n linear orders is not falsifiably complete. The easiest example is when the aggregation rule is such that $xR_{f(R_1,\ldots,R_n)}y$ for every $x, y, R_1, \ldots, R_n$. Then the theory is axiomatized by $\forall x \forall y xRy$, which is not falsifiably complete.

Proof. We first show that $\mathcal{T}_f$ has a universal axiomatization; the result then follows immediately from Corollary 32.

We use Theorem 30 to show that $\mathcal{T}_f$ is universally axiomatizable.

We must verify that T_f satisfies the following two properties:

1. (1) Closure under substructures: If $\mathcal{A} \in \mathcal{T}$, and $\mathcal{A}'$ is a substructure of $\mathcal{A}$, then $\mathcal{A}' \in \mathcal{T}$.[9]
2. (2) Finite substructure property: If for all finite substructures $\mathcal{A}'$ of $\mathcal{A}$, $\mathcal{A}' \in \mathcal{T}$, then $\mathcal{A} \in \mathcal{T}$.[10]

The first property is obviously satisfied; it follows from the neutrality and IIA assumption.

To prove that the second is satisfied, let $\mathcal{A}$ be an arbitrary structure for the language $\mathcal{L}$, and suppose that for all finite substructures $\mathcal{A}'$ of $\mathcal{A}$, $\mathcal{A}' \in \mathcal{T}$. A structure consists of a set X and a complete binary relation $R^{\mathcal{M}}$ on X, where $\tilde{R}^{\mathcal{M}}$ is a binary relation which is the complement of $R^{\mathcal{M}}$. The assumption that for all finite substructures $\mathcal{A}'$, $\mathcal{A}' \in \mathcal{T}$ means that for all finite subsets $Y \subseteq X$, $R^{\mathcal{M}}|_Y$ is f-rationalizable. We need to show that $R^{\mathcal{M}}$ on X is also f-rationalizable.

To this end, consider $\{0, 1\}$ endowed with the discrete topology. Identify the set of binary relations on X with $\mathcal{B} = \{0, 1\}^{X \times X}$ and topologize with the product topology. Then $\mathcal{B}$ is a compact topological space. Denote the set of f-rationalizable binary relations on X by $\mathcal{B}_f$. For $x, y \in X$, let $B_{x,y} = \{B \in \mathcal{B}_f : B(x, y) = R^{\mathcal{M}}(x, y)\}$. Note that for

---

[9] A structure $\mathcal{A}' = (X', R')$ is a substructure of $\mathcal{A} = (X, R)$ if $X' \subseteq X$ and $R|_{X'} = R'$.

[10] A structure $\mathcal{A} = (X, R)$ is finite if X is finite.

all (x, y), $B_{x,y}$ is nonempty.[11] We now seek to show that it is a closed subset of $\mathcal{B}$. To see this, note that for all $B \in B_{x,y}$, by definition, there exists a preference profile $(R^1, ..., R^n)$ for which $B = R_{f(R^1,...,R^n)}$. For each $B \in B_{x,y}$, choose one such profile. Suppose $\{B_\lambda\}_{\lambda\in\Lambda} \subseteq B_{x,y}$ is a net converging to some $\overline{B}$. By compactness of $\mathcal{B}$, we may without loss of generality assume that $(R^1_\lambda, ..., R^n_\lambda) \to (\overline{R^1}, ..., \overline{R^n})$ (see Kelley (1955), p. 71). In particular, it is easy to verify that each $\overline{R^i}$ is a linear order (by definition of product topology convergence). Also by definition $\lim_{\lambda\in\Lambda} R_{f(R^1_\lambda,...,R^n_\lambda)} = B$. The limit can be passed through f.[12] Conclude that $R_{f(\overline{R^1},...,\overline{R^n})} = B$. Clearly, $B(x, y) = R^{\mathcal{M}}(x, y)$. Conclude that $B \in B_{x,y}$, so that $B_{x,y}$ is closed.

Now, we claim that $\bigcap_{(x,y)\in X\times X} B_{x,y} \neq \varnothing$. To show this, we will show that for every finite set $Z \subseteq X \times X$, $\bigcap_{(x,y)\in Z} B_{x,y} \neq \varnothing$ and appeal to the finite intersection property. So, let $Z \subseteq X \times X$ be finite. Let $Y = Z_1 \times Z_2$, where Z_i denotes the projection of Z on the ith coordinate. Note that Y is finite; so by hypothesis, $R^{\mathcal{M}}|_Y$ is f-Pareto rationalizable. Let $(R^1, ..., R^n)$ be linear orders on Y for which $R_{f(R^1,...,R^n)} = R^{\mathcal{M}}|_Y$. Each of these can be extended to linear orders on all of X by the Szpilrajn theorem, say to R^{i*}. Then $R_{f(R^{1*},...,R^{n*})}|_Y = R^{\mathcal{M}}|_Y$ (this follows from the neutrality and independence of irrelevant alternatives hypothesis). In particular, for all $(x, y) \in Z$, $R_{f(R^{1*},...,R^{n*})}(x, y) = R^{\mathcal{M}}(x, y)$, so that $\bigcap_{(x,y)\in Z} B_{x,y} \neq \varnothing$. This verifies the finite intersection property, and as each $B_{x,y}$ is closed and $\mathcal{B}$ is compact, we conclude that $\bigcap_{(x,y)\in X\times X} B_{x,y} \neq \varnothing$. This establishes that $R^{\mathcal{M}} \in \mathcal{B}_f$.

□

---

[11] This follows from the fact that $R^{\mathcal{M}}|_{\{x,y\}}$ is f-rationalizable. This implies that there exist linear orders $(R^1, ..., R^n)$ on $\{x, y\}$ for which $R_{f(R^1,...,R^n)} = R^{\mathcal{M}}|_{\{x,y\}}$. The argument now follows from the Szpilrajn theorem, by taking appropriate extensions of R^i for all i and appealing to independence of irrelevant alternatives.

[12] To see this, note that for all $x, y \in X$ by definition of convergence, there exists $\lambda^* \in \Lambda$ for which for all $\lambda \geq \lambda^*$ and for all $i \in N$, $R^i_\lambda(x, y) = \overline{R^i}(x, y)$. Recall that for a pair $x, y \in X$ for which $x \neq y$ and a linear order R over X, $R|_{\{x,y\}}$ is determined by $R(x, y)$. As $R^i_\lambda(x, y) = \overline{R^i}(x, y)$ for all $i \in N$ and $\lambda \geq \lambda^*$, we may conclude that $R_{f(R^1_\lambda,...,R^n_\lambda)}|_{\{x,y\}} = R_{f(\overline{R^1},...,\overline{R^n})}|_{\{x,y\}}$ for all such λ. Therefore, $R_{f(R^1_\lambda,...,R^n_\lambda)} \to R_{f(\overline{R^1},...,\overline{R^n})}$.

The above discussion assumes that preferences are linear orders, but many of the multiple-selves papers put different restrictions on the selves' preferences. While the proof above does not directly apply, it is easy to see that the theorem is true on different domains of preference profiles: Any domain of preference profiles which is closed in the product topology, and closed under restriction and arbitrary permutation, will work.

6.2. **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 3, 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, R and P. The language should also include a constant symbol for every element of $\mathbf{R}^n_+$ and $\mathbf{R}$.[13] 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.

---

[13] 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."

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

- $(x, y) \in R$ if and only if there is $(p, I) \in \Pi$ such that $x = d(p, I)$ and $p \cdot y \leq I$;
- $(x, y) \in P$ 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 $(\mathbf{R}^n_+, R^*, P^*)$ in T' for which there is a complete, reflexive, and transitive binary relation $\succeq$ on X such that

$$\begin{aligned}(x, y) \in R^* &\Rightarrow (x, y) \in\succeq\\ (x, y) \in P^* &\Rightarrow (x, y) \in\succ .\end{aligned}$$

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

$$\begin{aligned}(x, y) \in R^* &\Rightarrow u(x) \geq u(y)\\ (x, y) \in P^* &\Rightarrow u(x) > u(y).\end{aligned}$$

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

38. **Theorem.** *T'_{wo} is the falsifiable closure of T_c with respect to T'.*

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

$$\forall v_1, \ldots, \forall v_n(\neg(v_1, v_2) \in R \vee \neg(v_2, v_3) \in R \vee, \ldots, \vee\neg(v_n, v_1) \in P).$$

By a well-known theorem (see Richter (1966) and Suzumura (1976)), if a structure (X, R^*, P^*) satisfies these sentences, then it is in T'_{wo}. And if a structure (X, R^*, P^*) is in T'_{wo}, it is clear to see it satisfies these sentences. So $(X, R^*, P^*) \in T'_{wo}$ if and only if it is in T' and satisfies the formulas in Σ. Then, by Theorem 26 $T'_{wo} = T' \cap \mathcal{T}(\Sigma)$ implies that

T'_{wo} is falsifiably complete with respect to T', as the formulas in Σ are all UNCAF.

Note that for (X, R^*, P^*) in T', the interpretation of the sentences in Σ is that the strong axiom of revealed preference holds.[14] 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 $(D, R^{\mathcal{D}}, P^{\mathcal{D}})$ satisfies the sentences in Σ, there is a structure (X, R^*, P^*) in T_c containing it.

Let (X, R^*, P^*) be a structure in $T'_{wo} \setminus T_c$, and let $\mathcal{D}$ be a finite data set contained in (X, R^*, P^*). 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 falsifiably complete, we conclude that T'_{wo} is the falsifiable closure of T_c with respect to T'. $\square$

7. 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,

---

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

and argue that substructures are inadequate as a notion of data. The definition of falsifiability was proposed by Simon and Groen (1973).[15] They intend their definition to capture the theories that can be axiomatized using only universal quantifiers.

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

39. **Definition.** 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.

40. **Definition.** 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 30, 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.

41. **Proposition.** *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. $\square$

---

[15] Rynasiewicz (1983) proposes a different notion, which he calls "finitely strongly falsifiable." One can show that example 42 presents a theory that is falsifiably complete, and closed under substructures, but is not finitely strongly falsifiable.

By Proposition 41 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 5.3.

The following example shows that a theory T may be falsifiably complete with respect to another theory T' [Definition 12], 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.

42. **Example.** 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 26, T is falsifiably complete 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 falsifiably complete; a simple example involves one unary relation R and theory T axiomatized by $\forall R(x)$.

8. 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. In particular, we have illustrated the usefulness of our results by presenting conditions under which theories of multiple-selves in behavioral economics, and theories of preference aggregation in social choice, are falsifiably complete. 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 3, and Sections 6.1 and 6.2). 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 falsifiably complete 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 $\mathrm{fc}(T)$), 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. 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.[16]

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 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 falsifiably complete 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.

---

[16] Gilboa (2009; Chapter 7.3) presents this viewpoint very convincingly.

We can then define the *verifiable interior* of a theory T with respect to T', $\mathrm{vi}_{T'}(T) = T' \backslash \mathrm{fc}_{T'}(T' \backslash T)$. 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 falsifiable closure.

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 26.

43. **Theorem.** *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.

44. **Example.** 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:

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

.

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. Basic definitions from Model Theory

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.

45. **Definition.** 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.

46. **Definition.** 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.

47. **Definition.** 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 $c \in \mathcal{C}$.

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

49. **Definition.** $\mathcal{M}$ 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.

50. **Definition.** 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.

51. **Definition.** 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

52. **Definition.** 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 $\mathcal{W}$.

53. **Definition.** 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.

54. **Definition.** 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).

55. **Definition.** 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 $b \in M$, $\mathcal{M} \models \psi(\overline{a}, b)$.

56. **Definition.** $\mathcal{M}$ *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.

57. **Definition.** 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.

Ambrus, A., and K. Rozen (2008): "Revealed conflicting preferences," working paper.

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

Deb, R. (1976): "On constructing generalized voting paradoxes," *The Review of Economic Studies*, 43(2), 347–351.

DeClippel, G., and K. Eliaz (2009): "Reason-Based Choice: A Bargaining Rationale for the Attraction and Compromise Effects," working paper.

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

Dushnik, B., and E. Miller (1941): "Partially ordered sets," *American Journal of Mathematics*, 63(3), 600–610.

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

Fudenberg, D., and D. Levine (2006): "A dual-self model of impulse control," *American Economic Review*, 96(5), 1449–1476.

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

Green, J. R., and D. Hojman (2008): "Choice, rationality, and welfare measurement," working paper.

Gul, F., and W. Pesendorfer (2008): "The case for mindless economics," *The foundations of Positive and normative Economics: A handbook*, pp. 3–42.

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

Kalai, G. (2004): "Social indeterminacy," *Econometrica*, 72, 1565–1581.

Kelley, J. L. (1955): *General Topology.* D. Van Nostrand Company, Inc.

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.

Manzini, P., and M. Mariotti (2007): "Sequentially rationalizable choice," *American Economic Review*, 97(5), 1824–1839.

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

McGarvey, D. C. (1953): "A theorem on the construction of voting paradoxes," *Econometrica*, 21(4), 608–610.

O'Donoghue, T., and M. Rabin (1999): "Doing It Now or Later," *The American Economic Review*, 89(1), 103–124.

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.

Sen, A. (1969): "Quasi-transitivity, rational choice and collective decisions," *The Review of Economic Studies*, 36(3), 381–393.

Shelah, S. (2009): "What majority decisions are possible," *Discrete Mathematics*, 309(8), 2349–2364.

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.

SPRUMONT, Y. (2001): "Paretian quasi-orders: the regular two-agent case," *Journal of Economic Theory*, 101(2), 437–456.

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.