

THE AXIOMATIC STRUCTURE OF EMPIRICAL CONTENT

CHRISTOPHER P. CHAMBERS, FEDERICO ECHENIQUE, AND ERAN SHMAYA

ABSTRACT. We define the **empirical content** of an economic theory as the least restrictive observationally equivalent theory. We show that the empirical content of a theory is captured by a certain kind of axiomatization, with axioms that are universal negations of conjunctions of atomic formulae.

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. We¹ introduce a framework for studying the empirical content of an economic theory. Our main result is a description of the types of axioms that characterize empirical content.

We define theories as hypothetical “extensions” of data sets. For example, if one can observe revealed preference demand data, then a theory consists of a class of preferences whose consistency with the data is to be tested. If there is a preference in the theory that could rationalize the data, then the data would be called *rationalizable*; otherwise, the data *falsify* the theory.

Some theories make non-testable claims, in that there are less restrictive, observationally equivalent theories. The theory of utility maximization is an example.

Date: November 9, 2013.

Chambers is affiliated with the Department of Economics, University of California, San Diego. Echenique is affiliated with the Division of the Humanities and Social Sciences, California Institute of Technology. Shmaya is affiliated with the Kellogg School of Management at Northwestern University. We thank Nageeb Ali, Yaron Azrieli, Bettina Klaus, Matias Iaryczow, Kota Saito, Tomasz Strzalecki, and Leeat Yariv for comments. We also thank Larry Samuelson and four anonymous referees. Chambers and Echenique acknowledge support from the NSF through grant SES-0751980. Email addresses: cpchambers@ucsd.edu (Chambers), fede@caltech.edu (Echenique), and e-shmaya@kellogg.northwestern.edu (Shmaya).

¹Throughout the paper, “we” refers to the authors.

It is observationally equivalent to the theory of weak order (complete and transitive preference) maximization, because every finite data set rationalizable by a weak order is also rationalizable by a utility function. We define the empirical content of a theory T as the *weakest* theory observationally equivalent to T .

The goal of this paper is to establish that the falsifiability of a theory is related to the *form* (or syntax) of the axioms that characterize the theory. Our result ties empirical content to a certain kind of axiom, called “UNCAF.” An UNCAF axiom precludes certain simultaneous observations.

Two familiar examples of axioms are the weak axiom of revealed preference and the completeness axiom (stating that any two alternatives must be comparable). The weak axiom is UNCAF: for any pair x and y , it precludes the simultaneous observation of x being revealed weakly preferred to y and y being revealed strictly preferred to x . Therefore, it expresses a falsifiable property. The completeness axiom is not UNCAF, and not falsifiable.

In Section 2, we present our framework for the well-known case of revealed preference theory: the empirical content of the theories of utility maximization and weak order maximization are captured by the strong axiom of revealed preference, which is UNCAF. The remainder of the paper contains a rigorous exposition of our results, together with some examples that are more involved than the basic examples in Section 2. The mathematics used are simple but unfamiliar to many economists. The ideas are borrowed from mathematical logic and model theory, which are areas of mathematics studying the relationship between the structure of mathematical statements and the mathematical objects to which they apply.

2. ILLUSTRATION OF EMPIRICAL CONTENT

Two economists, Woland and Ulysses, present their theories about data on agents’ choices. Woland’s theory, denoted T_{wo} , is the theory of weak order maximization. According to T_{wo} , agents have some complete and transitive preference

relation over alternatives. Woland claims that weak and strong revealed preferences are instances of some theoretical weak order preference relation. Ulysses's theory, denoted T_u , is the theory of utility maximization. According to T_u , agents have some utility function over alternatives, and revealed preferences are compatible with this utility function.

Both Ulysses and Woland agree on which situations represent 'weak revealed preference' and 'strict revealed preference.' For example, they both agree that an individual reveals a weak preference for alternative x to alternative y when the individual is willing to choose x over y in direct comparison, and that the individual reveals a strict preference for x over y when the individual is willing to choose x over y and some amount of money. Thus, both economists translate behavior to revealed preference in the same way. However, they present different theories about preferences.

Ulysses makes a stronger claim than Woland, because any utility function represents a weak order, while there are weak orders that have no representation by a utility function. A well-known example is the lexicographic order on \mathbf{R}^2 , which is a weak order without a utility representation. So, it may seem as though Ulysses takes more risks than Woland, since if one could demonstrate that an individual has lexicographic preferences, Ulysses's theory would be falsified while Woland's would not. It is easy to see, however, that this can never be demonstrated. Data sets are finite, so even if a lexicographic order governs an individual's choices, any data set taken from her observed behavior can be rationalized by a utility function.

So some claims made by Ulysses are not falsifiable. Is Woland guilty of the same delinquency? The answer is yes. For example, Woland's claim that the agent's weak preference relation is complete is not falsifiable. Under Woland's theory, Mark Twain must either weakly prefer coffee to tea or strictly prefer tea to coffee. All the same, even if Twain holds neither preference, no data can reveal

this fact. Woland can always claim that non-observation of these choices does not imply their non-existence.

The non-falsifiability of Woland’s claims is a consequence of a phenomenon we call *partial observability*. Data sets can affirm the existence of the entities and relationships we observe, but cannot imply that relationships we do not observe do not exist. Partial observability reflects a fundamental feature of scientific data. In the words of Carl Sagan, “Absence of evidence is not evidence of absence.” We shall see that partial observability has important consequences for the form of the axioms that capture empirical content.

Of course, T_{wo} and T_u also make falsifiable claims. If Merlin is observed to strictly prefer Sangrail to Excalibur, strictly prefer Excalibur to Ring of Dispel and strictly prefer Ring of Dispel to Sangrail, then both T_{wo} and T_u are proven false.

To sum up, T_{wo} and T_u make some falsifiable and some non-falsifiable claims. There is a well-known way to separate the wheat from the chaff. Denote weak and strict preference by \succeq and \succ . The (common) falsifiable implications of T_{wo} and T_u are captured by the strong axiom of revealed preference: For all $n \geq 1$,

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

where for all i , R_i is either \succ or \succeq . Note that the “strong axiom of revealed preference” is actually a collection of axioms. These axioms include $\forall x_1 \forall x_2 \neg ((x_1 \succeq x_2) \wedge (x_2 \succ x_1))$ (the weak axiom of revealed preference) and $\forall x_1 \forall x_2 \forall x_3 \neg ((x_1 \succeq x_2) \wedge (x_2 \succeq x_3) \wedge (x_3 \succ x_1))$ (ruling out cycles of length 3). After purging all non-falsifiable assertions made by Ulysses and Woland, their theories claim that agents’ behavior does not exhibit cycles, as described by Axiom (1).

The structure of Axiom (1) is responsible for its falsifiability. It starts with universal (\forall) quantification, then negation (\neg), and then a conjunction (\wedge) of

statements of the form ‘ $y \succeq z$ ’ or ‘ $z \succ x$ ’. The latter statements are called atomic formulae in mathematics. In our framework, they are statements about the observations that comprise a data set. Thus, axioms like (1) are falsifiable because they are universal statements that negate a conjunction of basic statements about observables: “universal negation of conjunctions of atomic formulae” (UNCAF).

The main result of the paper is that the empirical content of a theory is axiomatized by the UNCAF axioms that are true in the theory. A theory with an UNCAF axiomatization makes only falsifiable claims.

Much of modern decision theory is motivated by the idea that axioms provide the testable implications of a theory. Our main result provides a guide as to which kinds of axioms serve the stated purpose of decision theorists. Not all axioms used in economics or decision theory are UNCAF. For example, the completeness axiom $\forall x \forall y (x \succeq y) \vee (y \succeq x)$ and the non-satiation axiom $\forall x \exists y (y \succ x)$ are not UNCAF. We present additional examples in Section 4.1.

3. A FORMAL MODEL FOR EMPIRICAL CONTENT

In Section 2 we discussed empirical content in the context of classical revealed preference. It was enough then to consider weak and strong revealed preference, and our axioms used only the symbols \succeq and \succ . But now these symbols may no longer be enough. In order to accommodate other theories and frameworks, we need to allow for other symbols. For example, to discuss some modern models in decision theory, we need to introduce the union and intersection of sets (for example, to accommodate the models of Kreps (1979) and Gul and Pesendorfer (2001); see Section 4.1). Yet other models may need symbols to signify that an object is a strategy, or that an element is chosen from a set of available actions.

We collect the symbols that we need into a *language*. A language comprises symbols for the relations that can potentially be observed and that one can talk

about in axioms. After formally introducing languages, we proceed to define data and theories, and the empirical content of a theory.

A *language* \mathcal{L} is given by a finite set of *relation symbols* and, for each relation symbol R , a positive integer n_R , the *arity* of R . For the revealed preference example discussed in Section 2, the language is $\mathcal{L}_{\text{RP}} = \langle \succeq, \succ \rangle$. It has two binary relation symbols (i.e., relation symbols of arity 2): \succeq , which is intended to express revealed weak preference, and \succ , which is intended to express revealed strict preference.

Let \mathcal{L} be a language. An \mathcal{L} -*data set* \mathcal{D} is given by:

- (1) A finite non-empty set D (the *domain* of \mathcal{D}).
- (2) An n -ary relation $R^{\mathcal{D}}$ over D for every n -ary relation symbol R of \mathcal{L} .²

For each R , one should think of an element $(x_1, \dots, x_{n_R}) \in R^{\mathcal{D}}$ as an *observation*. We have observed that the elements x_1, \dots, x_{n_R} of D stand in relation $R^{\mathcal{D}}$. A data set collects a finite number of observations.

Consider the language \mathcal{L}_{RP} discussed in Section 2. An example of an \mathcal{L}_{RP} -data set is

$$D = \{\text{Sangrail}, \text{Excalibur}, \text{Ring of Dispel}\},$$

with

$$\text{Sangrail} \succ^{\mathcal{D}} \text{Excalibur},$$

$$\text{Excalibur} \succ^{\mathcal{D}} \text{Ring of Dispel},$$

$$\text{Ring of Dispel} \succ^{\mathcal{D}} \text{Sangrail}.$$

Less colorful examples abound in empirical studies of consumption.

We now turn to the definition of theory. The main building block is the concept of structure. An \mathcal{L} -*structure* \mathcal{M} is given by the following:

²An n -ary relation on D is a subset of D^n . By convention, we write $R^{\mathcal{D}}(x_1, \dots, x_n)$ if $(x_1, \dots, x_n) \in R^{\mathcal{D}}$, and if $R^{\mathcal{D}}$ is a binary relation, we use the more familiar $x R^{\mathcal{D}} y$ instead of $R^{\mathcal{D}}(x, y)$.

- (1) A nonempty set M (the *domain* of \mathcal{M}).
- (2) An n -ary relation $R^{\mathcal{M}}$ over M for every n -ary relation symbol R of \mathcal{L} .

A structure is a hypothetical construct used to extend, or rationalize, data. A structure describes objects and relations that were not necessarily observed, but that might have been observed. In contrast, a data set describes actual observations.

Definition 1. An \mathcal{L} -structure \mathcal{M} *rationalizes* an \mathcal{L} -data set \mathcal{D} if the following conditions are satisfied:

- (1) $D \subseteq M$, where D and M are the domains of \mathcal{D} and \mathcal{M} .
- (2) $R^{\mathcal{D}} \subseteq R^{\mathcal{M}}$.

In the context of our recurring example of revealed preference theory, to rationalize data is to “complete” the relations by adding all unobserved relations to render the agents’ full preference relation consistent with a given theory.

In Definition 1, note that we require $R^{\mathcal{D}} \subseteq R^{\mathcal{M}}$ and not that $R^{\mathcal{D}}$ is the restriction of $R^{\mathcal{M}}$ to D . This requirement reflects our assumptions that relations are only partially observed in the data set. For example, a data set $D = \{\text{coffee}, \text{tea}\}$ in which we only observe the weak preference $\text{tea} \succeq^D \text{coffee}$ is rationalized by a structure in which tea is both weakly and strictly preferred to coffee. This is the nature of partial observability.

We now turn to the definition of theory. We define a theory as a collection of isomorphic structures in some language. Two structures are isomorphic if we can identify the objects across the two structures so that all relations are preserved: Let \mathcal{M} and \mathcal{N} be \mathcal{L} -structures with domains M and N respectively. Formally, an *isomorphism from \mathcal{M} to \mathcal{N}* is a bijective map $\eta : M \rightarrow N$ that preserves the interpretations of all symbols of \mathcal{L} : $(a_1, \dots, a_{n_R}) \in R^{\mathcal{M}}$ if and only if $(\eta(a_1), \dots, \eta(a_{n_R})) \in R^{\mathcal{N}}$ for every relation symbol R of \mathcal{L} and $a_1, \dots, a_{n_R} \in M$.

Definition 2. An \mathcal{L} -theory T is a class of structures that is closed under isomorphism. A data set \mathcal{D} is T -rationalizable if there is a structure in T that rationalizes \mathcal{D} . Otherwise, \mathcal{D} falsifies T .

We note that the term ‘theory’ is used in model theory for a different purpose than ours. For a similar use of the term in economics, see Schipper (2009).

Consider the example in Section 2. Woland’s theory and Ulysses’s theory are both expressed in the same language \mathcal{L}_{RP} . As we emphasized in Section 2, we assume that Woland and Ulysses interpret the symbols \succeq and \succ of the language in the same way. Woland’s theory T_{wo} consists of all structures $\mathcal{M} = (M, \succeq^{\mathcal{M}}, \succ^{\mathcal{M}})$, where M is a set, $\succeq^{\mathcal{M}}$ is a weak order on M , and $\succ^{\mathcal{M}}$ is the strict preference associated to $\succeq^{\mathcal{M}}$. Ulysses’s theory T_u consists of all structures $\mathcal{M} = (M, \succeq^{\mathcal{M}}, \succ^{\mathcal{M}})$, where M is a set, and there is a utility function $u : M \rightarrow \mathbf{R}$, such that $x \succeq^{\mathcal{M}} y$ if and only if $u(x) \geq u(y)$ and $x \succ^{\mathcal{M}} y$ if and only if $u(x) > u(y)$. The data set $D = \{\text{Sangrail}, \text{Excalibur}, \text{Ring of Dispel}\}$, whereby $\text{Sangrail} \succ^{\mathcal{D}} \text{Excalibur}$, $\text{Excalibur} \succ^{\mathcal{D}} \text{Ring of Dispel}$, and $\text{Ring of Dispel} \succ^{\mathcal{D}} \text{Sangrail}$ falsifies both T_{wo} and T_u , as it cannot be rationalized by any weak order or utility function.

We are now in a position to introduce the main new concept in the paper, the concept of *empirical content*. Recall that Ulysses’s theory is strictly more restrictive than Woland’s: $T_u \subsetneq T_{wo}$. The two theories are, however, observationally equivalent. Any data set that falsifies the weaker theory T_{wo} also falsifies T_u . According to the definition we propose below, these theories have the same empirical content.

The theory T_{wo} can also be weakened without changing its observable implications. Consider the structure (X, \succeq^X, \succ^X) , in which X is a set of packages of salt. Suppose that \succeq^X is the weak order in which $x \succeq^X y$ whenever x has more salt than y . Suppose, however, that $x \succ^X y$ only when x has at least 50 grams more salt than y . The structure (X, \succeq^X, \succ^X) is not in T_{wo} , as \succ^X is not the

strict preference relation associated to \succeq^X (such objects are studied by decision theorists in the literature on semiorders). However, the choices generated by an agent with preferences \succeq^X and \succ^X over X do not falsify T_{wo} .

For example, consider the data set with alternatives $\{x_{350}, x_{375}\}$, where x_k is a packet with k grams of salt. The observations consist of $x_{375} \succeq^X x_{350}$ and $\succ^X = \emptyset$, so there is no observed strict comparison. Such data is rationalizable by T_{wo} , because the data are silent on whether x_{375} is strictly preferred to x_{350} or x_{350} is weakly preferred to x_{375} . Note that partial observability plays a crucial role in this example.

The idea that a theory can be weakened without observable consequences motivates the notion of empirical content. The empirical content of a theory is the most permissive observationally equivalent weakening of the theory. Formally:

Definition 3. The *empirical content* of a theory T , denoted $ec(T)$, is the class of all structures \mathcal{M} that do not rationalize any data set that falsifies T .

4. MAIN RESULT

We proceed to define UNCAF axioms and state our main result.

Given a language \mathcal{L} , we can write formulae using the symbols in \mathcal{L} . In addition to the relation symbols specified by \mathcal{L} , we shall use certain logical symbols. These symbols are fixed, and we are allowed to use them regardless of the language under consideration. The logical symbols are the *quantifiers* “exists” (\exists) and “for all” (\forall); “not” (\neg); the *logical connectives* “and” (\wedge) and “or” (\vee); a countable set of *variable symbols* x, y, z, u, v, w, \dots ; parentheses ‘(’ and ‘)’; and equality and inequality symbols ‘=’ and ‘ \neq ’. The use of an inequality symbol is not standard in mathematical logic (see Remark 1 for why we added it).

Certain strings of symbols can be put together to form *axioms*. We shall not spell out the formal rules for forming axioms because we mostly consider a special kind of axiom in this paper, UNCAF axioms, defined below. Rules for forming

axioms are in, for example, Marker (2002). Such rules are intuitive and immediately recognizable: The string ‘ $\forall x \exists y x \succ y$ ’ is a legitimate axiom. The string ‘ $\forall y \exists \succ x$ ’ is not.

UNCAF axioms are built out of simple blocks. The first notion is that of an *atomic formula*.

An *atomic formula* ϕ of a language \mathcal{L} is either

- (1) $t_1 = t_2$ or $t_1 \neq t_2$, where t_1, t_2 are variable symbols.
- (2) $R(t_1, \dots, t_{n_R})$ where R is a relation symbol of \mathcal{L} and t_1, \dots, t_{n_R} are variable symbols.

In the language \mathcal{L}_{RP} of revealed preference, all atomic formulae use at most two variable symbols. The atomic formulae using one variable symbol, say x , are $x \succeq x$, $x \succ x$, $x = x$, and $x \neq x$. The atomic formulae using two variable symbols, say x and y , are $x \succeq y$, $x \succ y$, $x = y$, and $x \neq y$. Atomic formulae represent the types of observations that data sets contain.

Definition 4. Let \mathcal{L} be a language. A *universal negation of a conjunction of atomic formulae (UNCAF)* axiom is a string of the form

$$\forall v_1 \forall v_2 \dots \forall v_n \neg (\phi_1 \wedge \phi_2 \cdots \wedge \phi_m)$$

where $\phi_1, \phi_2, \dots, \phi_m$ are atomic formulae with variables from v_1, \dots, v_n .

For a set Γ of axioms of \mathcal{L} , let $\mathcal{T}(\Gamma)$ be the theory consisting of the structures for which all axioms in Γ are true. If $T = \mathcal{T}(\Gamma)$ for some set Γ of axioms, we say that Γ is an *axiomatization* of T . If all axioms in Γ are UNCAF, we say that Γ is an UNCAF axiomatization of T .

For a theory T , denote by $\text{uncaf}(T)$ the set of UNCAF axioms that are true in all members of T . The following theorem is the main result of the paper. It asserts that the empirical content of a theory has an axiomatization (whether or

not the theory itself does), and that it can be axiomatized by the UNCAF axioms true for every structure in the theory.

Theorem 1. *For every theory T , $ec(T)$ is the theory axiomatized by the UNCAF axioms that are true in T : $ec(T) = \mathcal{T}(\text{uncaf}(T))$.*

In the revealed preference example of Section 2, $ec(T_u) = ec(T_{wo})$. The empirical content of these theories is the class of structures satisfying (1), the strong axiom of revealed preference, which constitutes an UNCAF axiomatization of $ec(T_{wo})$ and of $ec(T_u)$.

Remark 1. The use of an inequality symbol is not standard in mathematical logic. Adding this symbol as a primitive, while not affecting the expressional power of the language, increases the set of UNCAF axioms. To see why, consider a language with a single unary predicate R representing the property of being red, and the theory that says that there exists exactly one red object. The empirical content of the theory is that there exists at most one red object. It is axiomatized by the UNCAF axiom

$$\forall x \forall y \neg((x \neq y) \wedge R(x) \wedge R(y)).$$

We can write an equivalent axiom without using the inequality symbol

$$\forall x \forall y \neg(\neg(x = y) \wedge R(x) \wedge R(y)),$$

however, the latter axiom is not UNCAF.

Remark 2. In the working version of the paper (Chambers, Echenique, and Shmaya, 2011), we establish results for languages with function and constant symbols, as well as an infinite list of relation symbols. We dismiss these concepts here in the interest of exposition.

4.1. **Examples.** We review examples of axioms taken from Kreps (1979), Gul and Pesendorfer (2001), Dekel, Lipman, and Rustichini (2009), as well as two classical axioms.

The following examples of axioms provide illustrations of Theorem 1. The first two are in the language \mathcal{L}_{RP} . The remaining examples are in the language $\mathcal{L}_K = \langle \in, \cup, \cap, \succeq, \succ \rangle$ (for \mathcal{L}_K we write \cup and \cap in the usual ways, even though they are relation symbols; see below).

- (1) $\forall x \forall y \neg((x \succeq y) \wedge (y \succ x))$
- (2) $\forall x \forall y((x \succeq y) \vee (y \succeq x))$
- (3) $\forall x \forall y(x \succ y \rightarrow x \succeq x \cup y \succeq y)$
- (4) $\forall x \exists y((y \in x) \wedge (y \succeq x))$
- (5) $\forall x \forall y \forall z(x \sim x \cup y \rightarrow x \cup z \sim x \cup y \cup z)$

Axiom (1) is UNCAF; it is a version of the weak axiom of revealed preference.

Axiom (2) is the completeness axiom discussed in Section 2; it is not UNCAF since it uses the disjunction symbol \vee .

Axiom (3) is the Set Betweenness axiom from Gul and Pesendorfer (2001). Strictly speaking, this axiom should read $\forall x \forall y \forall z((x \succ y) \wedge (\cup(z, x, y))) \rightarrow x \succeq z \succeq y$, where the relation \cup of arity 3 is meant to say that z is the union of x and y . The reason is that, in our language, \cup is a relation symbol, not a function symbol. Under the assumption of completeness, Set Betweenness is equivalent to the pair of UNCAF axioms:

$$\forall x \forall y \neg((x \succeq y) \wedge (x \cup y \succ x))$$

$$\forall x \forall y \neg((x \succeq y) \wedge (y \succ x \cup y)).$$

(In every structure \mathcal{M} such that $\succeq^{\mathcal{M}}$ is complete and $\succ^{\mathcal{M}}$ is its strict part, axiom (3) is true if and only if its equivalent UNCAF form is true. We often need to

use such “relative” notions, and we explain how they work in Section 4.2.) We develop the implications for Gul and Pesendorfer’s theory in Section 5.1 below.

Axiom (4) is Axiom 3 in Dekel, Lipman, and Rustichini (2009). It is not UNCAF because it requires an existential quantifier. Hence the theory described by this axiom makes non-falsifiable claims.

Axiom (5) is the main axiom in Kreps (1979), where we have used the notation \sim to express the two directions of weak preference in the usual way. Similarly to (3), Axiom (5) is equivalent to an UNCAF axiom under completeness.

4.2. Relative theories. Often a researcher wants to take certain assumptions as given, and find the empirical content of his theory *relative to* such given assumptions. For example, decision theorists often regard axioms such as continuity or completeness as “technical”; they want to build on such axioms by studying other, more substantive, axioms.

Consider two theories, T and T' , where $T \subseteq T'$. We can define the *empirical content of T relative to T'* , written $ec_{T'}(T)$, as the class of all structures $\mathcal{M} \in T'$ that do not rationalize any data set that falsifies T , i.e.,

$$(2) \quad ec_{T'}(T) = ec(T) \cap T'.$$

The following is an immediate consequence of Theorem 1.

Corollary 1. *For any theories T and T' such that $T \subseteq T'$, $ec_{T'}(T) = \mathcal{T}(\text{uncaf}(T)) \cap T'$*

We say that a collection of UNCAF axioms Λ is a *UNCAF axiomatization of T relative to T'* if $T = \mathcal{T}(\Lambda) \cap T'$. Then Corollary 1 implies that the empirical content of T relative to T' admits an UNCAF axiomatization relative to T' .

4.3. Observing absence of relations. The previous discussion has emphasized the role of partial observability: the phenomenon that data affirms the existence

of certain relations, but cannot certify the non-existence of the relations one does not observe. We shall now expand on this issue by introducing the possibility that absence of relations may be observed.

To do so, we add, for every relation symbol R , a relation \tilde{R} that represents absence of R .

Say that a language \mathcal{L} *supports negation of relations* if its relation symbols are divided into pairs (R, \tilde{R}) with the same arity. The idea is that \tilde{R} should represent the relation ‘ R does not hold’. If \mathcal{L} supports negation of relations, we denote by ON (observed negation) the theory of all structures \mathcal{M} of \mathcal{L} such that $\tilde{R}^{\mathcal{M}}$ is the complement of $R^{\mathcal{M}}$ for every relation symbol R of \mathcal{L} ; so that for any x_1, \dots, x_n , $\tilde{R}^{\mathcal{M}}(x_1, \dots, x_n)$ if and only if it is not the case that $R^{\mathcal{M}}(x_1, \dots, x_n)$. Say that T *respects negation of relations* if $T \subseteq \text{ON}$.

We denote by $\text{universal}(T)$ the set of universal axioms that are true in every structure of T . Universal axioms are more general than UNCAF axioms. They are axioms in which all the quantifiers over all the variables are “for all”; they have no existential quantifiers. The completeness axiom $\forall x \forall y ((x \succeq y) \vee (y \succeq x))$ is an example of a universal axiom that is not UNCAF.

The following corollary of Theorem 1 is proved in the appendix. It is based on the fact that, for theories that respect negation of relations, every universal axiom is equivalent to an UNCAF axiom.

Corollary 2. *Let \mathcal{L} be a language that supports negation of relations and let T be a theory that respects negation of relations. Then $ec_{\text{ON}}(T) = \mathcal{T}(\text{universal}(T))$.*

Corollary 2 illustrates the relation between our results and the discussion in Popper (1959). Popper advocated universal theories. The distinction between UNCAF and universal axiomatizations is rooted in partial observability.

5. APPLICATIONS

We develop two simple applications of our framework to recent theories in behavioral economics and decision theory. In each case, we show that the theories in question have UNCAF axiomatizations. This result is important because it implies that the theories only make falsifiable claims.

The first application is to the theory of temptation and self-control, a model in decision theory dealing with the choice of menus of alternatives. Our second application (Section 5.2) is to models of multiple selves, which are used frequently in behavioral economics. No axiomatization is known for this application, but we can still show that the theory has an UNCAF axiomatization, and therefore that all the claims made by the theory are testable.

5.1. Temptation and self-control (Gul and Pesendorfer, 2001). Consider the language $\mathcal{L}_K = \langle \in, \cup, \cap, \succeq, \succ \rangle$, which we used in Section 4.1. We consider the theory of Gul and Pesendorfer's set-betweenness axiom, relative to a basic theory that we take as given, using the framework of relative theories from Section 4.2.

Let T' be the theory of all the structures $(X, \in^X, \cup^X, \cap^X, \succeq^X, \succ^X)$ in which:

- $X = 2^A$ for some set A ;
- \cup^X and \cap^X represent union and intersection of subsets of A ;
- \succeq^X is a complete binary relation on 2^A , and \succ^X is its strict part;

and all structures that are isomorphic to a structure as just described. Thus T' collects basic properties of the primitives of the model, and it would not be very interesting to test for these primitives.

Let T_{GP} be the set of all structures in T' that satisfy Gul and Pesendorfer's Set Betweenness axiom, equivalent (under completeness) to:

$$\forall x \forall y \neg((x \succeq y) \wedge (x \cup y \succ x))$$

and

$$\forall x \forall y \neg((x \succeq y) \wedge (y \succ x \cup y))$$

written as two axioms. Each of these axiom(s) is UNCAF, so that according to (2) and Theorem 1, $ec_{T'}(T_{GP}) = T_{GP}$.

5.2. 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 that have no known axiomatizations (and whose empirical content is therefore 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 that make only falsifiable claims.

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 and reconciles these preferences with a preference aggregation rule. Suppose an aggregate preference (a revealed preference) is observed. We ask whether it could have been generated by the rule for *some* profile of agents’ preferences.

Let n , the number of agents, be fixed. A *preference aggregation rule* maps every set X of *alternatives* and every *preference profile* (R^1, \dots, R^n) of linear orders³ over X to a complete binary relation over X . We write $\succeq_{f(R^1, \dots, R^n)}$ for the binary relation that results (suppressing notation for dependence on X). We assume the following property:

Definition 5 (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

³A linear order is complete, transitive, and anti-symmetric

(R^1, \dots, R^n) over X and (R'^1, \dots, R'^n) over Y , if for all $i \in N$, $x R^i y \Leftrightarrow w R'^i z$, then $x \succeq_{f(R^1, \dots, R^n)} y \Leftrightarrow w \succeq_{f(R'^1, \dots, R'^n)} z$.

Fix a preference aggregation rule f . Consider the language with two binary relation symbols \succeq and $\tilde{\succeq}$. A structure \mathcal{M} is *f-rationalizable* if $\succeq^{\mathcal{M}} = \succeq_{f(R^1, \dots, R^n)}$ for some profile of linear orders (R_1, \dots, R_n) and $\tilde{\succeq}^{\mathcal{M}}$ is the complement of $\succeq^{\mathcal{M}}$. The class of *f-rationalizable* structures is denoted T_f . Note that T_f is in fact a theory, as it is closed under isomorphism (this is the content of neutrality).

Proposition 1. *For every f , $ec_{ON}(T_f) = T_f$.*

Proposition 1 follows from Corollary 2 and results in Chambers, Echenique, and Shmaya (2012) that imply that T_f admits a universal axiomatization. We omit the details.

6. RELATED LITERATURE

We are not the first to investigate empirical content; results exist in the literatures of philosophy of science and mathematical psychology. Popper (1959) is a seminal reference, viewing falsifiable theories as those that admit universal axiomatizations (see Section 4.3 on how universal and UNCAF axiomatizations differ).

Much early literature in philosophy of science was concerned with whether the restrictions on observable relations imposed by axioms involving unobservable relations could be expressed in terms of observable relations alone. Craig (1956) provides a seminal result in this direction.

Adams, Fagot, and Robinson (1970) seem to be the first social scientists to discuss 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 axioms (of a certain type) consistent with one theory is equivalent to the set of all axioms consistent with the other. 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.

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)). Their notion of data does not allow for partial observability.

Finally, some of our formal arguments are close to results by Tarski (1954). Tarski characterizes those theories that have a universal axiomatization. As we demonstrated in Section 4.3, the issue of universal axiomatization is related to falsification, but Tarski never explored this aspect of his result. We discuss Tarski's Theorem and its relation to our theorem in Appendix B.

The discussion in Brown and Kubler (2008) also provides a general framework for falsification in economic theories. The focus is on mathematical environments that admit quantifier elimination, and on economic theories that can be expressed using these environments.

7. DISCUSSION

We develop a theory of empirical content. Many studies in decision and choice theory aim to provide testable axioms characterizing individual behavior. We investigate when such axioms are indeed testable, and argue that our analysis is useful for understanding and advancing modern decision and choice theory. Our leading examples are borrowed from revealed preference theory and should be familiar to most economists, but we have also shown that the results are applicable to less well-understood theories, and can give new substantive results.

There is a recurring methodological debate in economics over the importance of the falsifiability of a theory. Early literature was sparked by Milton Friedman's 1953 position that the truth of assumptions does not matter. In our terminology, a non-UNCAF axiomatization may be preferable to an UNCAF axiomatization because it is a more tractable description of the theory. (We thank an anonymous

referee for phrasing the issue in this way.) We agree that this may be the case, but to verify that the two descriptions are observationally equivalent, we must understand their empirical content in the first place. This is why our results are useful. The two axioms of decision theory that seem to have generated the greatest amount of research are the independence axiom and the sure thing principle, each of which have been experimentally falsified and each of which can be specified in UNCAF form.

Paul Samuelson (see Archibald, Simon, and Samuelson (1963)) counters Friedman's position with ideas that we have formalized. Samuelson claims that assumptions matter because either a theory T (described by its "assumptions") only makes falsifiable claims and is 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 $ec(T)$), rather than unnecessary claims in T . Regardless of one's position on the question of realism, this example shows how our notions may be useful.

Finally, we have studied basic ideas from philosophical positivism. As scientists may have agendas other than falsification, philosophy of science since Popper has focused on the sociology of actual research. It seems that most economists still find the problem of falsification important. In fact, 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)). In addition, Olszewski and Sandroni (2009) study a falsifiability problem for non-deterministic theories. We believe that a formal understanding of empirical content is useful, independently of the complexities involved in the actual production of research.

APPENDIX A. PROOFS

A.1. **Proof of Theorem 1.** Let \mathcal{L} be a language and \mathcal{D} a \mathcal{L} -data set. For every $d \in D$ let v_d be a variable. Let $\phi_{\mathcal{D}}$ be the following UNCAF formula of \mathcal{L} :

$$(3) \quad \begin{aligned} \phi_{\mathcal{D}} &= \forall_d v_d \neg \bar{\phi}_{\mathcal{D}}, \text{ where} \\ \bar{\phi}_{\mathcal{D}} &= \bigwedge (v_d \neq v_{d'}) \bigwedge R(v_{d_1}, \dots, v_{d_n}) \end{aligned}$$

where $\forall_d v_d$ stands for concatenation of universal quantifiers over all the variables v_d for $d \in D$. In the definition of $\bar{\phi}_{\mathcal{D}}$, the first conjunction ranges over all pairs $d \neq d' \in D$, and the second conjunction ranges over all relation symbols R and every $(d_1, \dots, d_n) \in R^{\mathcal{D}}$.

Lemma 1. *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 rationalized by some isomorphic copy of \mathcal{M} .*

Proof of Lemma 1. If a structure \mathcal{M} rationalizes \mathcal{D} then substituting d for v_d we get that $\bar{\phi}_{\mathcal{D}}$ is true in \mathcal{M} under this substitution, and therefore $\phi_{\mathcal{D}}$ is not true in \mathcal{M} . Since truth is preserved under isomorphisms, it follows that if an isomorphic copy of \mathcal{M} rationalizes \mathcal{D} then $\phi_{\mathcal{D}}$ is not true in \mathcal{M} .

Assume now that \mathcal{M} is an \mathcal{L} -structure 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 and use the fact that truth is preserved under isomorphism). Since $\phi_{\mathcal{D}}$ is not true in \mathcal{M} there exist elements $\bar{m} = (m_d)_{d \in D}$ of M such that $\bar{\phi}_{\mathcal{D}}$ is true in \mathcal{M} under the substitution s given by $s(v_d) = m_d$. Consider the isomorphic structure of \mathcal{M}' which is obtained by replacing every element m_d with d . (Note that this replacement is possible since $m_d \neq m_{d'}$ whenever $d \neq d'$, which follows from the fact that $\bar{\phi}_{\mathcal{D}}$ is true in \mathcal{M} under s .) Then $\bar{\phi}_{\mathcal{D}}$ is true in \mathcal{M}' under substitution s' given by $s'(v_d) = d$. It follows that $(d_1, \dots, d_n) \in R^{\mathcal{M}'}$ for every relation symbol R and every $(d_1, \dots, d_n) \in R^{\mathcal{D}}$.

Thus, $R^{\mathcal{D}} \subseteq R^{\mathcal{M}'}$ for every relation symbol R , and so property (2) in Definition 1 is satisfied. Therefore \mathcal{M}' is an isomorphic copy of \mathcal{M} that rationalizes \mathcal{D} . \square

Proof of Theorem 1. We divide the proof into two steps:

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

Let \mathcal{D} be a data set that falsifies T . Then from Lemma 1, and the fact that T is closed under isomorphism it follows that $\phi_{\mathcal{D}}$ is true in all structures of T , so that that $\phi_{\mathcal{D}} \in \text{uncaf}(T)$. Therefore $\phi_{\mathcal{D}}$ is true in \mathcal{M} , as by hypothesis $\mathcal{M} \in \mathcal{T}(\text{uncaf}(T))$. By Lemma 1 again it follows that \mathcal{M} does not rationalize \mathcal{D} .

Thus, we proved that \mathcal{M} does not rationalize any data set that falsifies T . By Definition 3, it follows that $\mathcal{M} \in \text{ec}(T)$ as desired.

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

Let $\phi \in \mathcal{T}(\text{uncaf}(T))$ be not true in \mathcal{M} . Let V be the finite set of variables of ϕ so that $\phi = \forall_{v \in V} v \neg \bar{\phi}$ where $\bar{\phi}$ is a conjunction of atomic formulae with variables in V and $\forall_{v \in V} v$ stands for a concatenation of universal quantifiers over all variables $v \in V$.

Since ϕ is not true in \mathcal{M} , it follows that there exists some assignment $s : V \rightarrow M$ of elements in the universe of \mathcal{M} to variables such that $\bar{\phi}$ becomes true in \mathcal{M} under the substitution $s(v)$ for every variable v of $\bar{\phi}$.

Let \mathcal{D} be a finite data set defined as follows: The domain D of \mathcal{D} is given by $D = s(V)$. For every relation symbol R ,

$$R^{\mathcal{D}} = \{(s(v_1), \dots, s(v_k)) \mid R(v_1, \dots, v_k) \text{ appears in } \bar{\phi}\},$$

Then \mathcal{D} is a data set that is rationalized by \mathcal{M} and ϕ is not true in any structure that rationalizes \mathcal{D} . But ϕ is true in every structure of T , and therefore \mathcal{D} falsifies

T . Thus, we proved that \mathcal{M} contains a data set \mathcal{D} that falsifies T . Therefore $\mathcal{M} \notin \text{ec}(T)$ by Definition 3. \square

A.2. Proof of Corollary 2. The corollary follows immediately from Theorem 1 and the following Lemma.

Lemma 2. *Let \mathcal{L} be a language that supports negation of relations. Then for every universal axiom ϕ in \mathcal{L} there exist UNCAF axioms ϕ_1, \dots, ϕ_n such that $\phi \leftrightarrow \phi_1 \wedge \dots \wedge \phi_n$ is true for ON.*

Proof of Lemma 2. We provide a purely syntactic proof. Consider the universal axiom $\forall \bar{v} \bar{\phi}(\bar{v})$, where ϕ is quantifier free and \bar{v} are the variables that appear in ϕ . Writing $\bar{\phi}$ in its conjunctive normal form,⁴ 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

$$(4) \quad \bigwedge_{i=1}^m \forall \bar{v} \neg \bigwedge_{j=1}^n \phi_{i,j}.$$

Finally, under ON every literal is equivalent to an atomic formula since $\neg R(v_1, \dots, v_k)$ is equivalent to $\tilde{R}(v_1, \dots, v_k)$ for every variables v_1, \dots, v_k . Therefore we can change the formulae $\phi_{i,j}$ in (4) to atomic formulae and arrive at a conjunction of UNCAFs, as desired. \square

⁴The conjunctive normal form of a quantifier free formula is a logically equivalent formula in which all conjunctions and disjunctions occur at the beginning of the formula, the conjunction coming first.

APPENDIX B. COMPARISON WITH TARSKI'S THEOREM

As discussed in Section 6, Karl Popper regarded *universality*, rather than UNCAF, as the defining characteristic of falsifiability. Popper's theory "all swans are white" is not an UNCAF theory, yet Popper regarded it as falsifiable. Tarski's Theorem characterizes those theories that admit a universal axiomatization. The aim of this section is to explain how our theorem relates to Tarski's Theorem.

We first reformulate Tarski's Theorem to highlight the relationship with our theorem. Let \mathcal{L} be a language without constant symbols. Recall that, given two \mathcal{L} -structures \mathcal{M} and \mathcal{N} with universes M and N , we say \mathcal{N} is a *substructure* of \mathcal{M} if $N \subseteq M$, and for all $x, \dots, x_n \in N$ and all relations R , $R(x_1, \dots, x_n)$ holds in \mathcal{N} if and only if it holds in \mathcal{M} .

Finally, we say that an axiom is *universal* if it has the form $\forall v_1 \dots \forall v_n \phi(v_1, \dots, v_n)$, where ϕ is quantifier free, and a theory *admits a universal axiomatization* if there is a collection of universal axioms Λ such that $T = \mathcal{T}(\Lambda)$.

Tarski (1954) proved the following theorem:

Tarski's Theorem. *Let \mathcal{L} be a language without constant symbols and let T be an \mathcal{L} -theory. Then T admits a universal axiomatization if and only if the following condition holds: Every structure \mathcal{M} that is not in T contains a finite substructure \mathcal{N} such that \mathcal{N} is not sub-structure of any member of T .*

We note that the condition in Tarski's Theorem is very similar to the condition that Simon and Groen (1973) call finitely and irrevocably testable.⁵ The similarity between Tarski's Theorem and our theorem is now transparent: We replace Tarski's finite substructure with data sets and get an UNCAF axiomatization instead of a universal axiomatization.

⁵Under two qualifications: Simon and Groen (1973) added the restriction that there exists some structure \mathcal{M} such that $\mathcal{M} \notin T$. In addition, instead of using our definition of theory as a class of structures closed under isomorphism, they assume that the class of structures all share the same domain.

Falsifiability was not Tarski’s motivation. Indeed, substructures are unsatisfactory as mathematical models for observed data since they correspond to a situation in which the scientist observes the presence or absence of every possible relation among the elements in his data and therefore cannot accommodate partial observability. In contrast, our definition of data set assumes that, when a relation is present, the scientist can *potentially* observe it, but will not necessarily observe it.

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.
- 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.
- BROWN, D., AND F. KUBLER (2008): “Refutable Theories of Value,” in *Computational Aspects of General Equilibrium Theory: Refutable Theories of Value*. Springer Publishing Company, Incorporated.
- CHAMBERS, C., F. ECHENIQUE, AND E. SHMAYA (2011): “The axiomatic structure of empirical content,” Caltech SSWP 1317.
- (2012): “General revealed preference theory,” working paper.
- CRAIG, W. (1956): “Replacement of auxiliary expressions,” *Philosophical Review*, 65, 38–55.
- DEKEL, E., AND B. LIPMAN (2009): “How (not) to do decision theory,” working paper.
- DEKEL, E., B. L. LIPMAN, AND A. RUSTICHINI (2009): “Temptation-Driven Preferences,” *The Review of Economic Studies*, 76(3), pp. 937–971.
- FRIEDMAN, M. (1953): “The Methodology of Positive Economics,” in *Essays in positive economics*. The University of Chicago Press, Chicago Illinois.
- 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 (2001): “Temptation and Self-Control,” *Econometrica*, 69(6), pp. 1403–1435.
- GUL, F., AND W. PESENDORFER (2008): “The case for mindless economics,” in *The foundations of Positive and normative Economics: A handbook*, pp. 3–42. Oxford University Press, USA.
- HICKS, J. (1983): “A discipline not a science,” in *Classics and Moderns: Collected Essays on Economic Theory Volume 3*, pp. 365–375.

- KREPS, D. M. (1979): "A Representation Theorem for "Preference for Flexibility"," *Econometrica*, 47(3), pp. 565–577.
- MANZINI, P., AND M. MARIOTTI (2007): "Sequentially rationalizable choice," *American Economic Review*, 97(5), 1824–1839.
- MARKER, D. (2002): *Model theory: An Introduction*. Springer.
- O'DONOGHUE, T., AND M. RABIN (1999): "Doing It Now or Later," *The American Economic Review*, 89(1), 103–124.
- OLSZEWSKI, W., AND A. SANDRONI (2009): "Falsifiability," working paper.
- PFANZAGL, J., V. BAUMANN, AND H. HUBER (1971): *Theory of Measurement*. Physica-Verl.
- POPPER, K. R. (1959): *The Logic of Scientific Discovery*. Hutchinson.
- 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.
- SCHIPPER, B. (2009): "How mindless is standard economics really?," SSRN Working Paper 1492970.
- 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.
- TARSKI, A. (1954): "Contributions to the theory of models I," *Indagationes Mathematicae*, 16, 572–581.