Descriptive Complexity and Revealed Preference Theory

Adam Galambos
Department of Economics, Lawrence University, Appleton, Wisconsin
adam.galambos@lawrence.edu

June 2018

Abstract

This paper formalizes revealed preference theory using the notion of Ramsey eliminability in logic, and shows how the language required to state a revealed preference axiom for some choice theory is closely connected with the computational tractability of testing that theory. The connection is made through results in descriptive complexity theory, a relatively new field in finite model theory. It is shown that checking whether observed choices of players in normal form games are Nash rationalizable is an NP-complete problem. This also means that there does not exist an analogue (in a precise sense) of the Strong Axiom of Revealed Preference for Nash equilibrium. Keywords: Revealed preference, Descriptive complexity, Choice theory, Ramsey eliminability

1 Introduction

The extensive literature on revealed preference theory goes back to Paul Samuelson's seminal work in the 1930s. In Samuelson's words, "[i]n its narrow version the theory of "revealed preference" confines itself to a finite set of observable price-quantity competitive demand data, and attempts to discover the full empirical implications of the hypothesis that the individual is consistent." (Samuelson, 1953) This consistency is embodied in some version of the Strong Axiom of Revealed Preference. The quote highlights that for Samuelson, the revealed preference approach was motivated by his goal of deriving *operationally meaningful* theorems; that is, results that can be translated to a set of operations that were empirically meaningful. (Hands, 2001, pp. 60-69)[1] In fact, as Hands points out, Samuelson did not use the words "revealed preference" in his early work, presumably because he was intent on putting consumer theory on the more solid foundations of choice data. In principle, neither utility nor preference need appear as part of such a theory. The first goal of what became revealed

[1] See also Backhouse (2017).

preference theory was, therefore, to put consumer theory on foundations that are observable and empirically meaningful.

In practical terms, this meant that the theory had to be restated without reference to utility or preference, which were purely theoretical notions. Instead, the empirically meaningful notion of observed choice had to suffice. In the philosophy of science literature, this process of restating theory in terms of observables only is called *Ramsey elimination*. A revealed preference axiom characterizing a choice theory is equivalent to the *Ramsey sentence* for that theory. A particularly simple equivalent of the Ramsey sentence, such as the Strong Axiom of Revealed Preference, tells us what the "core" of the theory is in observational terms, and can give us some insight into and intuition about the meaning of the theory. The first contribution of this paper is to identify the notion of "simplicity" of a revealed preference axiom, or any Ramsey sentence of a theory, with its *descriptive complexity*, i.e. the logical complexity of the language necessary to state it. This refinement of Ramsey eliminability opens up the possibility of characterizing the complexity of a revealed preference axioms, and also of comparing the complexity of different revealed preference axioms.

Empirically meaningful foundations make it possible (or easier) to test the theory as well. Do observations corroborate or contradict the theory? The revealed preference approach thus also came to be viewed as a way of making the foundations of microeconomics testable as a scientific theory. Therefore a desirable property for a revealed preference axiom is tractability: it should be possible to check in finite and "reasonable" time whether choice data satisfy the axiom. (Gradwohl and Shmaya, 2015) The second contribution of this paper is to point out that the simplicity and the computational tractability of revealed preference axioms are very tightly connected. This follows from important results in descriptive complexity theory. The implication is that there is a simple, elegant connection between the simplicity of the language of a revealed preference axiom, and the computational tractability of using the axiom to test data. Both of these issues have received attention in the revealed preference literature, but the connection between the two has not been made. The first three theorems in this paper draw conclusions about the computational complexity of revealed preference questions based on the descriptive complexity of revealed preference conditions; the last theorem draws conclusions about the descriptive complexity of any revealed preference condition for Nash equilibrium based on the computational complexity of Nash rationalizability.

The main contributions of this paper are thus in showing how results from descriptive complexity theory can shed light on important issues that the revealed preference literature has grappled with, and in characterizing the descriptive complexity of revealed preference axioms of various choice theories by introducing a suitable refinement of Ramsey eliminability. Finally, we illustrate this approach for the Nash equilibrium solution concept. We first show that the Nash rationalizability problem is *NP*-complete, i.e., computationally intractable. Then we use this result and a fundamental theorem from descriptive complexity theory to show that the descriptive complexity of any revealed preference axiom for Nash equilibrium must be higher than first order logic extended by transitive closure.

In this paper, only finite structures are considered. Understanding the finite case is generally important in choice theories, and assuming finiteness makes it possible to ask questions about computational tractability as well. For finite structures, some of the approaches to falsifiability of choice theories in the literature are not particularly meaningful (Chambers et al., 2017). Section 4 will address this and other questions regarding related literature. Section 2 introduces the required formal logic framework. Section 3 provides a formalization of the notion of simplicity for axioms, which is a refinement of Ramsey eliminability (to be defined in that section), and shows that this notion of simplicity is equivalent to the tractability of testing whether observed data satisfy the axiom.

2 Ramsey elimination

The purpose of this section is to show how revealed preference conditions can be stated using formal language. This is necessary in order to be able to characterize the complexity of the language used in a revealed preference condition.

Philosophers of science have turned to model theory to formalize the general notion of a theory. A theory is described as a set of sentences in a formal language $\mathcal{T}$, which particular structures (models) stated in the language $\mathcal{T}$ may or may not satisfy.[2] In order to formalize the notion of observation, the language has two parts: a more restrictive language of observation $\mathcal{O}$, and the full language $\mathcal{T}$ that also includes theoretical terms in addition to those used to describe observations. *Ramsey elimination* entails stating the theory entirely in the language of observation, eliminating all theoretical terms. (Benthem, 1978) This helps one articulate the empirical content of a theory, and it corresponds exactly to the revealed preference approach in economics, where individual and collective choice theories theories are reframed in terms of the language of choice alone, without making use of the language of preference or utility.

The use of formal logic in this literature is probably more extensive than what most economists are used to, and the details of modeling choice theory may seem tedious. However, just as in other areas in economics, using formal reasoning helps make the arguments more transparent and rigorous. The payoff is the new insight that descriptive complexity theory brings to the connection between the language and the computational tractability of revealed preference axioms.

2.1 Definitions

The two definitions below are standard—see, for example, Immerman (1999). We do not spell out standard notions in detail, such as the meaning of “truth” or “implication” in formal logic, but we follow standard usage that the reader may find in any introduction to logic.

[2]See below for formal definitions.

Definition 1. A *relational language* is a tuple

$$\mathcal{L} = \langle R_1^{a_1}, \ldots, R_r^{a_r}, c_1, \ldots, c_s \rangle \tag{1}$$

together with second-order predicate logic, where each $R_i^{a_i}$ is a relation of arity a_i, and the c_i are constant symbols.

We use the implication "$\Rightarrow$" and bi-implication " $\iff$ " symbols in the conventional sense. The second-order quantifiers will be denoted by "$\forall_2$" and "$\exists_2$."

Definition 2. An $\mathcal{L}$-*structure* is a tuple

$$M_{\mathcal{L}} = \langle U, R_1^M, \ldots, R_r^M, c_1^M, \ldots, c_s^M \rangle, \tag{2}$$

where the finite nonempty set U is the *universe*, and each R_i^M is a relation of arity a_i on U. For each constant symbol c_i of $\mathcal{L}$, there is a constant $c_i^M \in U$.[3]

Sometimes we will refer to R_i^M as the $M_{\mathcal{L}}$-interpretation of $R_i^{a_i}$, and we extend this usage to sentences in $\mathcal{L}$ as well. We will occasionally omit the language subscript and refer to a structure M when the language used is clear from the context.

2.2 Formalizing choice theories

The theories considered here will be (individual or collective) choice theories. In general, one can describe observations in choice theories by describing a variety of choice situations for agents, and the choices made by those agents in each situation. Therefore it is useful to introduce a general language $\mathcal{O}^n$ for describing observations in choice theories for some set of agents $N = \{1, \ldots, n\}$:

$$\mathcal{O}^n = \langle S_1^1, \ldots, S_n^1, X^1, E^{n+1}, I^1, A^2, C^2 \rangle, \tag{3}$$

where each S_i^1 is a unary relation that describes the "universal strategy set" for i, the unary relation X^1 is the set of outcomes, the relation E^{n+1} is an outcome function[4] assigning an outcome to every strategy profile in $S_1^1 \times \cdots \times S_n^1$ (see example 2 below), the unary relation I^1 is the index set of observations, the binary relation A^2 describes the possible strategy profiles available for each observed choice situation, and the binary relation C^2 describes choices for each observed choice situation. The language does not have any constant symbols. The two examples below will further clarify the interpretations of the various components of $\mathcal{O}^n$. In order to describe a choice theory, the language of observations must be expanded to include preference relations. Thus the language of choice theories is

$$\mathcal{T}^n = \langle S_1^1, \ldots, S_n^1, X^1, E^{n+1}, I^1, A^2, C^2, R_1^2, \ldots, R_n^2 \rangle, \tag{4}$$

[3] Sometimes an $\mathcal{L}$-structure is called a model for $\mathcal{L}$.

[4] Even though we model the outcome function as a relation, we will refer to it as the outcome function for consistency with standard game theoretic usage.

where the R_i^2 are binary relations representing preferences. For these relations, we will use the notation xR_i^2y instead of $(x, y) \in R_i^2$.

The following examples show how the theory of individual preference maximization and that of Nash equilibrium can be described using the languages of observation and choice theory as defined above. The main points made in this paper can be understood without the level of detail shown in these examples, but the examples are included for those who are interested in the precise application of the abstract concepts used.

Example 1 (Individual choice)**.** Consider the following individual choice data. A decision maker chooses from three-member subsets of the set $\{a, b, c, d\}$, and her choice is always the letter that comes first in the alphabet of those available. An $\mathcal{O}^n$-structure $\hat{O}$ describing this situation would have: the universe

$$\hat{U} = \{1, 2, 3, 4, 5, 6, 7, 8, a, b, c, d\};$$

the observation index set

$$I^{\hat{O}} = \{1, 2, 3, 4\},$$

the strategy set

$$S_1^{\hat{O}} = \{5, 6, 7, 8\},$$

outcomes $X^{\hat{O}} = \{a, b, c, d\}$, outcome function

$$E^{\hat{O}} = \{(5, a), (6, b), (7, c), (8, d)\},$$

the available alternatives relation

$$A^{\hat{O}} = \{(1, a), (1, b), (1, c), (2, a), (2, b), (2, d), (3, a), (3, c), (3, d), (4, b), (4, c), (4, d)\}$$

identifying the four choice situations; and the choice relation that identifies the chosen element for each observation:

$$C^{\hat{O}} = \{(1, a), (2, a), (3, a), (4, b)\}.$$

Example 2 (Two-player game)**.** We observe two players who play normal form games. One has universal strategy set $\{a, b\}$, the other has universal strategy set $\{A, B\}$. (Their strategy sets in any particular game observed will be subsets of their universal strategy sets.) We observe them choose (a, A) from the full 2×2 game, and (b, B) from the game where the first player is restricted to choosing b. An $\mathcal{O}^2$-structure $\hat{O}$ describing this situation would have: the universe

$$\hat{U} = \{1, 2, a, b, A, B, (a, A), (a, B), (b, A), (b, B)\},$$

the observations index set

$$I^{\hat{O}} = \{1, 2\},$$

the universal strategy sets

$$S_1^{\hat{O}} = \{a, b\},\ S_2^{\hat{O}} = \{A, B\},$$

the outcomes $X^{\hat{O}} = \{(a, A), (a, B), (b, A), (b, B)\}$the outcome function

$$E^{\hat{O}} = \{(a, A, (a, A)), (a, B, (a, B)), (b, A, (b, A)), (b, B, (b, B))\},$$

the accessible alternatives relation

$$A^{\hat{O}} = \{(1, (a, A)), (1, (a, B)), (1, (b, A)), (1, (b, B)), (2, (b, A)), (2, (b, B))\}$$

identifying the two game forms; and the choice relation that identifies the observed outcome for each of the two game forms observed:

$$C^{\hat{O}} = \{(1, (a, A)), (2, (b, B))\}.$$

Note that in this formulation it is assumed that strategy choices are directly observable (i.e., the set of outcomes is identified with the set of strategy profiles).

This example demonstrates how observations of two players playing normal form games can be described in the language defined in (3), and it will be of particular interest below in section 3.2.2 when we consider the Nash rationalizability problem. We note here that observations of two-player normal form games can be summarized more succinctly. The example above could be described as

$$(\{a, b\} \times \{A, B\}, aA), (\{b\} \times \{A, B\}, bB), \tag{5}$$

where each of the two items shows the game form observed together with the outcome observed. This shorthand notation will be used in the proof of Theorem 3. All of the relations that are explicit in the language $\mathcal{O}^2$ used above are implicit in this shorthand. However, in order to use results from descriptive complexity theory, we needed to show that choice theories (such as Nash equilibrium) can be formulated using the relational language of observations. While this shorthand simplifies the description significantly, it shortens the description by a polynomial factor (and this will be important for the proof of Theorem 3).

The strategy sets allow for describing strategic interaction, but non-strategic collective choice theories can also be described using this language. In this case, only one relation S_1 would be used, which would specify a universal set of "strategies," and the outcome function E would simply create a one-to-one map between these "strategies" and the set of outcomes X, just as in the case of individual choice.

The preceding examples illustrate how a structure (or model) uses a given language. In this framework, it is possible to define a *theory* as a collection of statements (or simply a sentence) in a language. Given a theory, a particular structure in the language may or may not satisfy the theory.

Definition 3. An $\mathcal{L}$-*theory* is a set of sentences in the language $\mathcal{L}$. An $\mathcal{L}$-structure M satisfies the $\mathcal{L}$-theory Σ if the M-interpretation of every sentence in Σ is true in M.

Every theory considered below will be assumed to include the standard logical axioms, though these will not be shown explicitly (in keeping with common practice). In addition, since only choice theories will be considered, the common properties of all choice theories will be assumed throughout. These properties specify that the interpretations of the components of the formal language of observations $\mathcal{O}^n$ are as intended; for example, the choice relation actually chooses an element of accessible alternatives in each choice situation. The assumption below formalizes these properties.

Assumption 1. Every theory we consider will include the logical axioms. In addition, every theory we consider will include the following (first-order universal) sentence CT, which states the axioms defining a choice-theoretic structure:

$$CT := \Big[\forall d, y\big[(d, y) \in C^2 \;\Rightarrow\; [d \in I^1 \;\wedge\; y \in X^1]\big]\Big] \wedge \tag{6a}$$

$$\Big[\forall d, y\big[(d, y) \in A^2 \;\Rightarrow\; [d \in I^1 \;\wedge\; y \in X^1]\big]\Big] \wedge \tag{6b}$$

$$\Big[\forall i_{i \in \{1,\dots,n\}} S_i^1 \cap I^1 = \emptyset\Big] \wedge \tag{6c}$$

$$\Big[\forall x_1, \dots, x_{n+1}\big[(x_1, \dots, x_{n+1}) \in E^{n+1}\big] \;\Rightarrow\; \big[[\forall i_{i \in \{1,\dots,n\}} x_i \in S_i^1] \wedge [x_{n+1} \in X^1]\big]\Big]. \tag{6d}$$

(6a) says that the binary relation C^2 describing choice has an element of the index set for observations as its first component, and as its second component it has an element of the universe that can be used for outcomes. (6b) says the same thing for the binary relation defining budgets. (6c) says that the strategy sets are disjoint from the index set for observations, and (6d) says that the first n arguments of the relation defining the outcome function are from the appropriate universal strategy sets, and the last argument is from the set of outcomes.

2.2.1 Formalizing solutions in choice theories

In our applications, an $\mathcal{O}^n$-structure will describe a set of choices, and a $\mathcal{T}^n$-structure will describe a set of choices together with preferences. The language $\mathcal{T}^n$ can then be used to express a particular choice theory. The following is a general formulation of choice theory in the language $\mathcal{T}^n$. In the definition below, the first-order sentence $\varphi(y, z, S_1^1, \dots, S_n^1, X^1, E^{n+1}, R_1^2, \dots, R_n^2)$ defines the solution concept, and contains occurrences of the universal strategy sets S_i^1, the set of outcomes X^1, the outcome function E^{n+1}, the preference relations R_i^2, unbound occurrences of y and z, as well as bound occurrences of other variables. Note that φ is not allowed to depend on I^1, the index set of observations, so the solution for a particular observed choice situation cannot depend on other choice situations.

$$\Psi^n(\varphi) := \big[\forall d, y\, \big[[(d,y) \in C^2] \Rightarrow [\forall z_{(d,z)\in A^2}\ \varphi]\big]\,\big] \wedge \tag{7a}$$

$$\big[\forall d, y\, \big[[(d,y) \in A^2] \wedge [(d,y) \notin C^2] \Rightarrow [\exists z_{(d,z)\in A^2}\ \neg\varphi]\big]\,\big] \wedge \tag{7b}$$

$$\big[\forall i_{i\in\{1,\ldots,n\}} \forall x, y_{x,y\in X^1} [x R_i^2 y] \vee [y R_i^2 x]\big] \wedge \tag{7c}$$

$$\big[\forall i_{i\in\{1,\ldots,n\}} \forall x, y, z_{x,y,z\in X^1} [x R_i^2 y \wedge y R_i^2 z] \Rightarrow x R_i^2 z\big]. \tag{7d}$$

(7a) says that for every observation, elements that are observed to be chosen are solutions, as defined by φ. (7b) says that for every observation, elements that are not observed as chosen are not solutions, as defined by φ. (7c) says that each agent's preference relation is total, and (7d) that each preference relation is transitive. Thus the theory $\Psi^n(\varphi)$ defines φ-rationality for n decision makers. Note that $\Psi^n(\varphi)$ is, in fact, a first-order sentence.

To define *strict* φ-rationality, simply replace (7c) by

$$\big[\forall i_{i\in\{1,\ldots,n\}} \forall x, y_{x,y\in X^1, x\neq y} [x R_i^2 y] \vee [y R_i^2 x]\big] \tag{6c'}$$

and add another conjunct expressing asymmetry and non-reflexivity of preferences:

$$\big[\forall i_{i\in\{1,\ldots,n\}} \forall x, y_{x,y\in X^1} \neg\big[[x R_i^2 y] \wedge [y R_i^2 x]\big]\big]. \tag{6e}$$

The resulting theory (the conjunction of (7a), (7b), (6c'), (7d), and (6e)) will be denoted by $\Psi^n(\text{str-}\varphi)$.

In some applications, *partial* rationalizability is of interest. This theory can be obtained from $\Psi^n(\varphi)$ simply by omitting (7b), and will be denoted by $\Psi^n_{sub}(\varphi)$. In this case, what is observed as chosen is understood to be a subset of the φ-solutions, not the entire set of φ-solutions. (This notion has been called *sub-semirationality* by Matzkin and Richter (1991).) The strict version $\Psi^n_{sub}(\text{str-}\varphi)$ is defined analogously.

Example 3. To obtain the theory of individual preference maximizing choice, let $n = 1$ and let φ be "$y R_1^2 z$." In the case of individual preference maximizing choice, there is no role for the strategy set, and so we may define rationality without reference to S_1^1. This theory is therefore $\Psi^1(y R_1^2 z)$.

Example 4. To obtain the theory of Nash equilibrium, let φ be the sentence φ_{Nash} defined by

$$\forall i_{i\in\{1,\ldots,n\}} \Big[\Big[\forall j_{j\neq i} \exists x_{j\,x_j\in S_j^1} \exists x_i, x_i'{}_{\,x_i, x_i'\in S_i^1} \big[(x_1,\ldots,x_n,y) \in E^{n+1}\ \wedge\ (x_1,\ldots,x_i',\ldots,x_n,z) \in E^{n+1}\big]\Big]\ \Rightarrow\ y R_i^2 z\Big].$$

Example 5. Let $\mathcal{C}$ denote a set of subsets of $\{1,\ldots,n\}$. In a $\mathcal{C}$-strong Nash equilibrium, no coalition in $\mathcal{C}$ can jointly deviate to make each member of the coalition better off. To obtain the theory of $\mathcal{C}$-strong Nash equilibrium, let φ be

$$\forall G_{G\in\mathcal{C}} \Big[\Big[\forall j_{j\notin G} \exists x_{j\,x_j\in S_j^1} \forall i_{i\in G} \exists x_i, x_i'{}_{\,x_i, x_i'\in S_i^1} \big[(x_1,\ldots,x_n,y) \in E^{n+1}\ \wedge\ ((x_j)_{j\notin G}, (x_i')_{i\in G}, z) \in E^{n+1}\big]\Big]\ \Rightarrow\ \exists i_{i\in G} \neg[z R_i^2 y]\Big].$$

Note that $\mathcal{C}$ is not part of the language $\mathcal{T}^n$, and it is not necessary to describe it, or to quantify over subsets of it. In any particular case of $\mathcal{C}$, the sentence φ involves conjunctions and disjunctions that could be written without reference to $\mathcal{C}$ or its subsets. The quantifiers used on $\mathcal{C}$ or its members in the sentence above are shorthand for conjunctions or disjunctions, used only for the sake of better comprehensibility. Because we state all theories for a fixed number n of individuals, in the sentence above quantification over $\mathcal{C}$ or G could be substituted with conjunctions or disjunctions.

2.3 Eliminability

Suppose that $\mathcal{T} = \langle O_1, \ldots, O_k, T_1, \ldots, T_m \rangle$ is a relational language, with the O_i describing observable relationships, and the relations T_i being theoretical terms. Suppose $\psi(O_1, \ldots, O_k, T_1, \ldots, T_m)$ is a first-order sentence in $\mathcal{T}$, i.e., a (finitely axiomatizable) $\mathcal{T}$-theory. Ramsey (1931) introduced the idea of rewriting ψ without the theoretical terms using what is now called a *Ramsey sentence*:

$$\exists_2 X_1, \ldots, X_m \psi(O_1, \ldots, O_k, X_1, \ldots, X_m). \tag{10}$$

This sentence is meant as an observationally equivalent way to write ψ, but without the theoretical terms. Instead, second-order existential quantification is introduced, with predicate variables X_i. Rewriting a theory without the theoretical terms is called *Ramsey elimination*. Hintikka (1998) provides a review of the relevant notions, and points out that the mere introduction of a Ramsey sentence does nothing to eliminate the effect that theoretical terms have:

> In order to do their job, theoretical concepts merely need to exist in certain relations to the observational concepts of the theory, but it does not matter how or where they exist. In such a model-theoretical perspective, the Ramsey reduction does not change anything at all. It merely means that the left hand (predicate constants) lends money (condition-imposing power) to the right hand (second-order quantifiers).

In fact, Hintikka (1998) argues that the notion of Ramsey eliminability merely points to the structural role that quantifiers play in a theory.

A stronger requirement is to eliminate theoretical terms without introducing second order existential quantification. This is called *strong Ramsey eliminability*: there exists a first-order sentence $\rho(O_1, \ldots, O_k)$ that is equivalent to (10). (Sneed, 1971) While every theory clearly has a Ramsey sentence, not every theory has a strong Ramsey sentence. Therefore strong Ramsey eliminability imposes real restrictions on the role that theoretical terms play in a theory.

3 Descriptive complexity, computational complexity, and Ramsey eliminability

In this section we apply Ramsey eliminability to choice theories and connect it with notions of computational and descriptive complexity. In the process, we find that strong Ramsey eliminability is too restrictive to be of interest for choice theories, and define natural refinements of Ramsey eliminability that are less restrictive.

3.1 Strong Ramsey eliminability and descriptive complexity

Descriptive complexity theory[5] relates the logical language required to describe a property (its *descriptive complexity*) to the computational complexity of checking that property. It is a subfield of finite model theory, and applies to finite structures only—indeed, questions of computational complexity would not make sense for infinite structures. The following analysis uses results from descriptive complexity theory to connect the computational complexity of testing a choice theory with the language required to describe its revealed preference axiom.

To formalize this, note that revealed preference axioms describe those observations that are consistent with a particular choice theory. This is exactly what the following mapping $\mathbf{Q}(\cdot)$ does; that is, it identifies those observation structures that are consistent with a given theory.

Definition 4. Let $\mathcal{T} = \langle O_1, \ldots, O_k, T_1, \ldots, T_m \rangle$ be a relational language, with the O_i describing observable relationships, the relations T_i being theoretical terms, and let ψ be a $\mathcal{T}$-sentence. Let $\mathcal{O} = \langle O_1, \ldots, O_k \rangle$ be the relational language for observations, and let $STRUC[\mathcal{O}]$ be the set of $\mathcal{O}$-structures. Define[6] the *consistency query* $\mathbf{Q}_\psi : STRUC[\mathcal{O}] \to \{0, 1\}$ as: for all $\mathcal{O}$-structures A,

$$\mathbf{Q}_\psi(A) = 1 \iff \text{there exists a } \mathcal{T}\text{-structure } A' \text{ that extends } A \text{ to satisfy } \psi. \tag{11}$$

A $\mathcal{T}$-structure A' extends an $\mathcal{O}$-structure A if the interpretation of every $\mathcal{O}$-relation is the same in A' as in A, and in addition it includes interpretations of the theoretical terms T_i in $\mathcal{T}$.

Fundamental results in descriptive complexity theory relate the *computational* complexity of the query $\mathbf{Q}$ to the *descriptive* complexity of the logical language necessary to define the set of $\mathcal{O}$-structures A such that $\mathbf{Q}(A) = 1$. Fagin's Theorem (Fagin, 1973) was the first result establishing a close correspondence between computational and descriptive complexity. It shows that the class *NP* (non-deterministically polynomial queries) is equal to the class $SO\exists$ (queries that can be defined using existential second-order logic).[7] The

[5] See Immerman (1995) for a short introduction.

[6] The *consistency query* is a boolean query—see Immerman (1999) for more on boolean and more general types of queries.

[7] For an in-depth exposition, see the authoritative book by Immerman (1999) .

following result follows immediately from Fagin's Theorem.[8] We use the same notation as in Definition 4 above.

Theorem 1. *Suppose ψ is a first-order $\mathcal{T}$-sentence. Then the computational problem of determining whether an $\mathcal{O}$-structure is extendable to satisfy ψ is in the class NP.*

Proof. The Ramsey sentence (10) of ψ defines the computational query in the theorem. The Ramsey sentence is in $SO\exists$, so by the easy direction of Fagin's Theorem, the query is in *NP*. $\square$

The consistency query being in *NP* means that it can be *verified* in polynomial time that a particular $\mathcal{T}$-structure extends a given observation structure and satisfies ψ. Problems that are not polynomial are usually considered to be computationally intractable. The Ramsey sentence being in $SO\exists$ implies the tractability of *verification*, but it would be desirable if actually determining *whether* an $\mathcal{O}$-structure is extendable to satisfy ψ were polynomial. In fact, the requirement of strong Ramsey eliminability guarantees that. Recall that the language $\mathcal{T}$ is an extension of the language of observations $\mathcal{O}$ by the addition of theoretical terms.

Theorem 2. *Suppose ψ is a $\mathcal{T}$-theory (a sentence in the language $\mathcal{T}$) that satisfies strong Ramsey eliminability. Then for any $\mathcal{O}$-structure $\hat{O}$ it can be determined in polynomial time whether $\hat{O}$ can be extended to a $\mathcal{T}$-structure that satisfies ψ.*

This theorem means that strong Ramsey eliminability of a theory implies the computational tractability of testing whether observations are consistent with the theory.

Proof. Strong Ramsey eliminability of ψ means that the Ramsey sentence (10) for ψ is equivalent to a first-order sentence ρ. The descriptive complexity class *FO* (first-order logic) corresponds to the *logarithmic-time hierarchy* in computational complexity (Immerman, 1999, Theorem 5.30), so the computational complexity of testing a theory that is strongly Ramsey eliminable is actually even lower than polynomial. $\square$

For finite structures, Theorems 1 and 2 give meaning to the notion of (strong) Ramsey eliminability in terms of the tractability of the computational problem of testing a theory. Though testability is, in principle, always possible for finite structures, the computational tractability of determining whether a particular structure is extendable to a theory is often of great interest. Strong Ramsey eliminability guarantees the computational tractability of testing a theory.

[8] This result was also stated in a working paper version of Chambers et al. (2017).

3.2 Eliminability in choice theories

In choice theories, the theoretical terms are the preference relations of the decision makers. The Ramsey sentence for a choice theory therefore states that there exist preferences such that the alternatives observed as chosen are solutions (for sub-semirationality), or that the alternatives observed as chosen are the only solutions (for exact rationality). Identifying the required preferences is referred to as "rationalizing the observed choices." Sometimes the solution concept is also stated explicitly, as in "Nash rationalizing the observed choices."

3.2.1 Individual preference maximization

Ramsey elimination in itself is simply a rewriting of the theory in the form of the Ramsey sentence; as noted earlier, "[i]t merely means that the left hand (predicate constants) lends money (condition-imposing power) to the right hand (second-order quantifiers)." (Hintikka, 1998) Strong Ramsey eliminability, however, places real restrictions on theories and guarantees the computational tractability of testing those theories, as we saw earlier. Do theories of individual preference maximization satisfy strong Ramsey eliminability? It turns out that not even the weakest theory of individual preference maximization, partial rationalizability, satisfies this requirement. The results in this subsection use the notation and assumptions established in section 2.2 (see (6) and (7)). To avoid the trivial cases of rationalization by complete indifference, we focus on the strict preference versions of these theories.

Proposition 1. *The theories of individual strict preference maximization do not satisfy strong Ramsey eliminability, whether exact rationality or sub-semirationality (i.e. "partial" rationalizability) is required. That is,* $\Psi^1(\text{str-}yR_1^2z)$ *and* $\Psi^1_{sub}(\text{str-}yR_1^2z)$ *are not strongly Ramsey eliminable.*

Proof. This result follows from the fact that acyclicity is not first-order expressible (see Immerman, 1999, Proposition 6.24). Any revealed preference condition for individual preference maximization must entail that rationalizing preferences be acyclic. That condition cannot be expressed in first-order logic alone. □

The preceding result implies that strong Ramsey eliminability is too restrictive of a requirement for choice theories. In order to characterize the descriptive complexity of revealed preference conditions for individual choice, we would need a property that is intermediate between the extremes of Ramsey eliminability and strong Ramsey eliminability. This suggests that it would be helpful to define various degrees of Ramsey eliminability that correspond to languages of various descriptive complexities.

Definition 5. Suppose $\psi(O_1, \ldots, O_k, T_1, \ldots, T_m)$ is a first-order sentence in $\mathcal{T}$, i.e., a (finitely axiomatizable) $\mathcal{T}$-theory. If the *Ramsey sentence*

$$\exists_2 X_1, \ldots, X_m \psi(O_1, \ldots, O_k, X_1, \ldots, X_m)$$

of ψ is equivalent to a sentence $\rho(O_1, \ldots, O_k)$ in the descriptive complexity class $\mathbb{D}$, we say that ψ is $\mathbb{D}$ *Ramsey eliminable.*

Using this terminology, strong Ramsey eliminability is simply *FO* Ramsey eliminability (where *FO* stands for first-order). We can now characterize the descriptive complexity of individual preference maximization using the descriptive complexity class *FO(TC)*, which is first-order logic extended by the transitive closure operator.

Proposition 2. *The theories of individual strict preference maximization are FO(TC) Ramsey eliminable, when exact rationality or sub-semirationality is required. That is,* $\Psi^1(\text{str-}yR_1^2z)$ *and* $\Psi^1_{sub}(\text{str-}yR_1^2z)$ *are FO(TC) Ramsey eliminable.*

Proof. The Strong Axiom of Revealed Preference (SARP) is the *FO(TC)* sentence that is equivalent to the Ramsey sentence of these theories (see Houthakker, 1950; Richter, 1966). □

We include the following known result (Varian, 1982) to highlight the fact that it follows from the previous theorem on the descriptive complexity of SARP, and the Immerman-Vardi Theorem. The Immerman-Vardi Theorem[9] shows that the computational complexity class *P* of *polynomial* problems is equal to the descriptive complexity class *FO(LFP)*, the class of queries that can be stated using first-order logic extended by the *least fixed point* (*LFP*) operator. The *LFP* operator adds the power of inductive definitions to first-order logic. Transitive closure is one such inductive definition.

Proposition 3. *The computational complexity of determining the (exact or partial) strict preference rationalizability of individual choices is polynomial. That is, the computational complexity of the consistency query* **Q** *for each of the theories* $\Psi^1(\text{str-}yR_1^2z)$ *and* $\Psi^1_{sub}(\text{str-}yR_1^2z)$ *is polynomial.*

Proof. Since SARP is in *FO(TC)*, it is also in *FO(LFP)*, and so by the easy direction of the Immerman-Vardi Theorem the consistency queries for these theories are polynomial. Note that even though the Immerman-Vardi Theorem requires an ordering on the universe, the direction used here does not require an ordering (see the discussion after Theorem 4.10 in Immerman (1999)). □

3.2.2 Complexity of Nash equilibrium rationalizability

There is an extensive literature on questions of testability in game theory (see, e.g., Yanovskaya, 1980; Sprumont, 2000; Galambos, 2004; Carvajal et al., 2004). This literature asks versions of the revealed preference question for collective choice theories. In the case of Nash equilibrium, the question is: given observations of players' choices in a set of games, is it possible to find preferences for the players on the set of outcomes that make the observed choices (a subset of) the Nash equilibria of for each observed game? If it is possible to find such preferences, we say that the observations are (partially) Nash rationalizable.

Here we show that Nash rationalizability is an *NP*-complete problem, and then use this result to show that the theory of Nash equilibrium is not *FO(TC)*

[9] Immerman (1982); Vardi (1982); stated as Theorem 4.10 in Immerman (1999)

Ramsey eliminable. These two results are interesting in their own right: the first characterizes the computational complexity of the revealed preference question for Nash equilibrium, and the second states that any revealed preference axiom for Nash equilibrium has to have higher descriptive complexity than first order logic extended by transitive closure. In particular, this means that there is no analogue of SARP for Nash equilibrium. Beyond the individual contributions of these two results, they also illustrate how the approach outlined in this paper can be fruitfully applied: the second result about descriptive complexity follows easily from the first result on computational complexity and the Immerman-Vardi Theorem.

Using the language of observations and theory defined earlier, asking whether observations are Nash rationalizable is a consistency query (Definition (4)). Example (2) showed how a set of observed game forms can be represented in a language of observations $\mathcal{O}^n$. If we add to $\mathcal{O}^n$ binary relations representing preferences for each player, we get the language $\mathcal{T}^n$ of theory. Example (4) shows how Nash equilibrium can be formalized in the language $\mathcal{T}^n$. The following theorem shows that the consistency query for Nash equilibrium is *NP*-complete, even with only two players. Recall that $\mathcal{O}^2$ is the language of observations for a choice theory with two individuals, and that the theory of Nash equilibrium is $\Psi^n(\varphi_{\text{Nash}})$, with φ_{Nash} the first-order sentence in Example (4).

Theorem 3. *Let* $\mathbf{Q}_{\varphi_{Nash}} : STRUC[\mathcal{O}^2] \to \{0,1\}$ *be the consistency query for Nash equilibrium with only two players; i.e.,* $\mathbf{Q}_{\varphi_{Nash}}(A) = 1$ *if, and only if, the structure A describes observations that are Nash rationalizable. This query is NP-complete.*

Proof. See the Appendix. The proof uses a common technique in the computational complexity literature: it shows that a problem known to be *NP*-complete can be polynomially reduced to the query $\mathbf{Q}_{\varphi_{\text{Nash}}}$. □

Theorem 3 tells us that determining whether observations are Nash rationalizable is computationally more complex than rationalizability for individual choice. The connection between computational and descriptive complexity means that the descriptive complexity of any revealed preference axiom for Nash equilibrium must therefore also be higher than the descriptive complexity of SARP.

Theorem 4. *The theory of Nash equilibrium is not FO(TC) Ramsey eliminable, unless* $P = NP$.[10]

Proof. If the theory of Nash equilibrium $\Psi^n(\varphi_{\text{Nash}})$ (with φ_{Nash} the first-order sentence in (4)) were *FO(TC)* Ramsey eliminable, then the consistency query for Nash equilibrium would be polynomial by the Immerman-Vardi Theorem

[10] The qualification in the theorem is a reminder that the $P \stackrel{?}{=} NP$ is one of the most important open questions in mathematics today. It is widely believed that $P \subsetneq NP$, and many results in complexity theory are qualified in this way.

(see the discussion of this theorem before Proposition 3). Since the Nash rationalizability query is *NP*-complete by Theorem 3, Nash equilibrium can't be *FO(TC)* Ramsey eliminable, unless $P = NP$. □

This contrasts with the *FO(TC)* Ramsey eliminability of the theory of individual preference maximization (Proposition 2 above). The theorem tells us that any revealed preference characterization of Nash equilibrium will use language more complex than first order logic with transitive closure.[11] The works cited above do find less complex revealed preference axioms for Nash rationalizability, but only in restricted settings (such as "complete domains;" see also Chambers and Echenique (2016)). Theorem 4 makes it clear that in general, the descriptive complexity of revealed preference axioms for Nash equilibrium is inherently greater than *FO(TC)*, and therefore there is no hope of finding a SARP-like revealed preference axiom for the theory of Nash equilibrium. This result differs from those of Chambers et al. (2017), who are not able to differentiate the degree of testability of individual preference maximization and Nash equilibrium. The main reason for this difference is that they consider only *partial* Nash rationalizability, that is, the theory $\Psi^n_{sub}(\varphi_{\text{Nash}})$. It is not difficult to see that this partial rationalizability version of the theory of Nash equilibrium is, in fact, *FO(TC)* Ramsey eliminable, and the *FO(TC)* equivalent of its Ramsey sentence is a straightforward generalization of the Strong Axiom of Revealed Preference (called "I-Congruence" in Galambos (2004); see also (Chambers and Echenique, 2016, Chapter 10)).

4 Comments on related literature

This paper connects the literatures on revealed preference theory, on the computational complexity on rationalizing choices, on descriptive complexity, and on Ramsey eliminability. The relavent references to these literatures were indicated above at the appropriate points. This section offers additional comments on those papers that are most closely related to this one.

Gradwohl and Shmaya (2015) propose "tractable falsifiability" as a criterion for assessing theories. A theory is tractably falsifiable if there is an algorithm V that determines in polynomial time whether a set of observations is consistent with the theory or not. Their algorithm V is analogous to the consistency query (Definition 4) in this paper. The formalism of their paper is very different, as they do not explicitly model theories or language, but instead identify a theory with a set of observations encoded in $0-1$ strings. However, tractable falsifiability is analogous with *FO(LFP)* Ramsey eliminability, and the Immerman-Vardi Theorem says that when ordering is part of the language of observations, the two are equivalent.

[11] In fact, as the proof suggests, first order logic extended by the power of any inductive definitions will not suffice, either. In choice theories, however, the inductively defined operator of interest is transitive closure.

Another paper that is closely related to this one is by Chambers et al. (2017), who use a formal logic approach similar to the present one to prove general results about falsifiability of choice theories. There are some important differences between their approach and the one taken here. First, they identify a "practical" test of a theory with an axiomatization that is effectively enumerable. As they point out in their discussion, all finite axiomatizations are effectively enumerable, and all the applications from choice theory they consider have finite axiomatizations. In contrast, in this paper a "practical" test is one that is computationally tractable. This also means that finiteness is assumed throughout in this paper–an assumption Chambers et al. (2017) do not impose. A second major difference between this paper and theirs is that Chambers et al. (2017) study only partial rationalizability (or sub-semirationality, using language from this paper) when they consider theories of strategic group choice. They call attention to this assumption, though they do not explain why they consider partial rationalizability for strategic choice but exact rationalizability for non-strategic choice theories.

5 Conclusion

The Ramsey sentence (10) for a choice theory could, in principle, serve as its revealed preference axiom, because it axiomatizes the theory in terms of observational terms only. However, it would be an utterly useless and uninformative revealed preference formulation of the theory. In contrast, a good revealed preference axiom restates the Ramsey sentence in simpler language, and in a way that gives insight and intuition about the theory. In addition, it is helpful if checking whether observations satisfy the axiom is computationally tractable. A great deal of work in the revealed preference literature has been devoted to these two goals: simple, intuitive axioms, and tractable axioms. The contribution of this paper is to connect these two desiderata for revealed preference axioms using fundamental results in descriptive complexity theory. This connection means that the language used by a revealed preference axiom has implications about the computational tractability of the axiom, and vice versa. This combination of descriptive complexity theory and Ramsey eliminability provide a new perspective for understanding and working with revealed preference theory.

While computational complexity considerations have been studied in connection with a number of questions in economic theory, descriptive complexity considerations have not. As this paper demonstrates, results from descriptive complexity theory can bring a useful perspective to revealed preference theory, and possibly to other areas in economic theory as well.

Acknowledgement

I would like to thank the Editor and an anonymous referee for very helpful comments that greatly improved the organization and exposition of the paper.

A Proof of Theorem 3

Here we prove Theorem 3, which states that the Nash rationalizability query $\mathbf{Q}_{\varphi_{\text{Nash}}}$ is *NP*-complete, even if there are only two players.

Proof of Theorem 3. We will prove the theorem using *polynomial-time reduction*, a standard technique in the theory of computational complexity. We will show that the 3SAT problem, known to be NP-complete (see Cook (1971) and Garey and Johnson (1979)), polynomially transforms into the Nash rationalizability problem with two players (henceforth denoted by NR2). 3SAT is the classic problem of determining the satisfiability of a Boolean formula in conjunctive normal form with three disjuncts in each conjunct. We will refer to a particular input into the NR2 or the 3SAT problem as an instance; thus an instance of NR2 is a set of game forms observed together with the strategy combinations that are observed as chosen, and an instance of 3SAT is a Boolean formula in conjunctive normal form with three disjuncts in each conjunct. Below we construct an algorithm that runs in polynomial time, and, given any instance of 3SAT, produces an instance of NR2 with the property that the NR2 instance is rationalizable (by preferences for the two players) if, and only if, the 3SAT instance is satisfiable (by assigning "true" or "false" to every Boolean variable in the formula). This will imply that if there exists a polynomial-time algorithm for deciding NR2, then any instance of 3SAT can be decided in polynomial time by first polynomially transforming it into an instance of NR2 and then deciding that in polynomial time. Since 3SAT is NP-complete, this argument will establish that NR2 is NP-complete.

NR2: We will describe the Nash rationalizability problem with two players using the shorthand notation introduced at the end of Example 2. Let $S := \{s_*, s^*, s_0, s_1, s_2, s_3, \dots\}$ be the set of potential actions of player 1 (in any game form a finite subset of this will be player 1's action space). Let $Z := \{z_*, z^*, z_0, z_1, z_2, \dots\}$ be the set of potential actions of player 2 (in any game form a finite subset of this will be player 2's action space). An *instance of* NR2 consists of a choice function on a finite set of finite game forms of $S \times Z$. For example, the following instance of NR2 encodes a choice function on two game forms.

$$(\{s_0, s_1, s_2\} \times \{z_0, z_1\}, s_2 z_1), (\{s_0, s_4, s_5\} \times \{z_0, z_2\}, s_4 z_0) \tag{12}$$

The first game form is $\{s_0, s_1, s_2\} \times \{z_0, z_1\}$, and the (*only*) observed outcome is (s_2, z_1). In general, an instance of NR2 consists of a list of game form–outcome pairs of the form $(A \times B, ab)$, where $A \subset S$, $B \subset Z$ and $a \in A, b \in B$. An instance of NR2 is a *yes-instance* if the corresponding choice function is (pure strategy Nash equilibrium) rationalizable, and it is a *no-instance* if it is not. A polynomial-time algorithm for NR2 is a polynomial-time algorithm that returns, for any given instance of NR2, a *yes* if and only if it is a yes-instance. Below we will show that if there exists a polynomial-time algorithm for NR2, then there exists a polynomial-time algorithm for 3SAT, which proves that NR2 is NP-complete.[12]

3SAT: Suppose that $X = \{x_1, x_2, \ldots, x_m\}$ is a set of Boolean variables and $\bar{X} = \{\bar{x}_1, \ldots, \bar{x}_m\}$ is the set of their negations. For any truth assignment $T : X \to \{\mathfrak{t}, \mathfrak{f}\}$, we define for $\bar{x} \in \bar{X}$ the extension of T by $T(\bar{x}) = \mathfrak{t}$ if, and only if $T(x) = \mathfrak{f}$. The set $X^* := X \cup \bar{X}$ is the set of *literals*. A subset C of X^* is a *clause*. Suppose a set $\{C_1, \ldots, C_k\}$ of clauses is given. A truth assignment $T : X \to \{\mathfrak{t}, \mathfrak{f}\}$ *satisfies* $\{C_1, \ldots, C_k\}$ if for every clause C_i there exists $x \in C_i$ with $T(x) = \mathfrak{t}$. A set of clauses is *satisfiable* if there exists a truth assignment that satisfies it. We can now state 3SAT: *Given an arbitrary finite set of clauses with exactly three elements in every clause, does there exist a satisfying truth assignment?* 3SAT is known to be NP-complete (see Garey and Johnson (1979)).

3SAT $\to$ NR2: We now define the polynomial-time transformation mentioned at the beginning of the proof. That is, we define a polynomial-time algorithm that takes any instance of 3SAT as its input, and produces an instance of NR2 that is rationalizable if and only if the input 3SAT instance is satisfiable. Suppose we are given an arbitrary instance of 3SAT:

$$V = \left\{\{v_1^1, v_1^2, v_1^3\}, \{v_2^1, v_2^2, v_2^3\}, \cdots, \{v_l^1, v_l^2, v_l^3\}\right\}, \tag{13}$$

where $v_j^i \in X^*$. Suppose w.l.o.g. that the set of variables that appear in V is $\{x_1, \ldots, x_k\}$. We will construct an instance of NR2 for V, using the actions $s_*, s^*, s_0, s_1, \ldots, s_k$ for player 1, and the actions $z_*, z^*, z_0, z_1, \ldots, z_k$ for player 2.

Informal description of the construction: For every clause, we construct a game form where player 1's action set is s_0, s^*, and all s_i such that x_i appears in the clause and is not negated; player 2's action set is z_0, z^*, and all z_i such that x_i appears in the clause and *is* negated. The (unique) outcome for this game form is (s^*, z^*). We will construct these game forms in such a way that rationalizing (s^*, z^*) as a Nash equilibrium will always be possible (and very simple), and it will also be possible (and simple) to rationalize all other points *except* (s_0, z_0) as not Nash equilibria. Thus rationalizability will boil down to being able to assign preferences in such a way that (s_0, z_0) is not a Nash equilibrium, and this will be possible if, and only if, the clause on which the game form was based is satisfied. Satisfying all clauses simultaneously will be possible if, and only if, the set of games constructed according to the above description can be simultaneously rationalized. Using an example, I will present further details of the construction, and then I will proceed to a general description. Suppose the variables appearing in an instance of 3SAT are x_1, x_2, x_3, x_4, x_5, and one particular clause is $\{x_1, \bar{x}_2, x_3\}$. Following the above described construction, we have a game form–outcome pair $(\{s_0, s_1, s_3, s^*\} \times \{z_0, z_2, z^*\}, s^* z^*)$. We will add two additional game form–outcome pairs that will imply that player 1 prefers (s_0, z_0) to (s^*, z_0) and that player 2 prefers (s_0, z_0) to (s_0, z^*). Rationalizability will boil down to finding preferences for the players such that either player 1 prefers (s_1, z_0) to (s_0, z_0), or player 1 prefers (s_3, z_0) to (s_0, z_0), or player 2

[12] Theorem 1 shows that NR2 is in the class *NP*.

prefers (s_0, z_2) to (s_0, z_0). The first of these will correspond to setting x_1 *true*, the second will correspond to setting x_3 *true*, and the third will correspond to setting x_2 *false*. This procedure, however, may lead us to assign preferences implying both that a variable x_i is *true* and that it is *false*. In the example just described, we might rationalize (s_0, z_0) not being a Nash equilibrium by assigning player 2 a preference of (s_0, z_2) over (s_0, z_0), which would correspond to setting x_2 *false*. At the same time, we might rationalize (s_0, z_0) not being a Nash equilibrium in another game form by assigning player 1 a preference of (s_2, z_0) over (s_0, z_0), which would correspond to setting x_2 *true*. To prevent this, we construct a "module" of game form–outcome pairs (denoted below by Γ_2) that will be rationalizable, but only if exactly one of the above two possibilities hold: either player 2 prefers (s_0, z_2) to (s_0, z_0), or player 1 prefers (s_2, z_0) to (s_0, z_0), but not both (see Figure 1).

Detailed description of the construction: First we construct a set of games for every variable that is negated in some clause in V. That is, suppose $\{v_j^1, v_j^2, \bar{x}_h\} \in V$. Then we construct Γ_h, which consists of the following game form–outcome pairs:

$$\begin{array}{rll}
(\{s_0, s_h, s_*\} \times \{z_0, z_h, z_*\}, & s_* z_*) & \\
(\{s_0\} \times \{z_h, z_*\}, & s_0 z_h) & \\
(\{s_0\} \times \{z_0, z_*\}, & s_0 z_0) & ^{13} \\
(\{s_h\} \times \{z_h, z_*\}, & s_h z_h) & \\
(\{s_h\} \times \{z_0, z_*\}, & s_h z_0) & \\
(\{s_h, s_*\} \times \{z_0\}, & s_h z_0) & \\
(\{s_0, s_*\} \times \{z_0\}, & s_0 z_0) & ^{13} \\
(\{s_h, s_*\} \times \{z_h\}, & s_h z_h) & \\
(\{s_0, s_*\} \times \{z_h\}, & s_0 z_h) &
\end{array} \tag{14}$$

Figure 1 illustrates this set of game form–outcome pairs. For transparency, the first pair in (14) is not shown (and, given the other eight game form–outcome pairs in the list, its rationalizability will depend only on orienting the edge cycle in Figure 1 b)). Each of the remaining eight involve only one player, and only two points, and so each has one revealed preference implication: the point chosen is preferred to the one not chosen. Figure 1 a) shows the resulting eight such implications, with the arrows pointing to the preferred point. For example, $(\{s_0\} \times \{z_h, z_*\}, s_0 z_h)$ is shown as an arrow pointing from $(s_0 z_*)$ to $(s_0 z_h)$.

Now we transform the 3SAT instance V into an instance of NR2 as follows.

1. Replace every clause of the form $\{x_e, x_f, x_g\}$ with
$$(\{s_0, s_e, s_f, s_g, s^*\} \times \{z_0, z^*\}, s^* z^*). \tag{15}$$

2. Replace every clause of the form $\{x_e, x_f, \bar{x}_g\}$ with
$$(\{s_0, s_e, s_f, s^*\} \times \{z_0, z_g, z^*\}, s^* z^*) \tag{16}$$

[13] Note that this is independent of h, so this game form–outcome pair could be included only once, not for every variable x_h that is negated in some clause.

Figure 1: **a)** The game forms in Γ_h **b)** The "edge cycle" must be oriented for rationalizability (recall that these four points are not chosen in the first game form in (14))

and Γ_g (see (14) for the definition of the nine game form–outcome pairs in Γ_h for $h = 1, \ldots, k$).

3. Replace every clause of the form $\{x_e, \bar{x}_f, \bar{x}_g\}$ with
$$(\{s_0, s_e, s^*\} \times \{z_0, z_f, z_g, z^*\}, s^* z^*) \tag{17}$$
and Γ_g and Γ_f.

4. Replace every clause of the form $\{\bar{x}_e, \bar{x}_f, \bar{x}_g\}$ with
$$(\{s_0, s^*\} \times \{z_0, z_e, z_f, z_g, z^*\}, s^* z^*) \tag{18}$$
and Γ_g, Γ_f and Γ_e.

5. Add the following game form–outcome pairs:
$$(\{s_0\} \times \{z_0, z^*\}, s_0 z_0),\ (\{s_0, s^*\} \times \{z_0\}, s_0 z_0). \tag{19}$$

The resulting instance of NR2 will be denoted by NR_V.

In the worst case, all variables that appear in V are distinct and are negated, which gives $l \cdot 30$ game form–outcome pairs, i.e. the input size is increased by a multiplicative factor. The transformation involves only replacing each clause by at most 30 game form–outcome pairs, as described above, and so it runs in polynomial time (in fact in linear time).

V **satisfiable** $\iff$ NR_V **Nash rationalizable:** Now we must show that the polynomial transformation $V \mapsto NR_V$ constructed above has the property mentioned at the beginning of the proof: V is satisfiable if and only if NR_V is Nash rationalizable.

$\Leftarrow$ First, suppose NR_V is Nash rationalizable. Let[14] $S_k := \{s_*, s^*, s_0, s_1, \ldots, s_k\}$

[14] Recall that V involves the variables $x_1, \ldots, x_k$.

and $Z_k := \{z_*, z^*, z_0, z_1, \ldots, z_k\}$, and denote the players' rationalizing preferences on $S_k \times Z_k$ by $\succ_1$, and $\succ_2$. Define, for each variable x_i with $i \in \{1, 2, \ldots, k\}$ (recall that these are exactly the variables that appear in V) a truth assignment:

$$T_\succ(x_i) = \mathfrak{t} \iff s_i z_0 \succ_1 s_0 z_0. \tag{20}$$

Consider a clause of the form $\{x_e, x_f, x_g\}$. Since NR_V contains (see (15) and (19))

$$\begin{aligned} (\{s_0, s_e, s_f, s_g, s^*\} \times \{z_0, z^*\}, &\qquad s^* z^*), \\ (\{s_0\} \times \{z_0, z^*\}, &\qquad s_0 z_0), \\ (\{s_0, s^*\} \times \{z_0\}, &\qquad s_0 z_0), \end{aligned} \tag{21}$$

and since $s_0 z_0$ is not a Nash equilibrium in the first game form, but it is an equilibrium in the second and the third, it must be that

$$[s_e z_0 \succ_1 s_0 z_0] \textbf{ or } [s_f z_0 \succ_1 s_0 z_0] \textbf{ or } [s_g z_0 \succ_1 s_0 z_0]. \tag{22}$$

Under $T_\succ$ this means that $\{x_e, x_f, x_g\}$ is satisfied.

Now consider a clause of the form $\{x_e, x_f, \bar{x}_g\}$. It is easy to see that if $\succ_1$, and $\succ_2$ rationalize NR_V, then it follows from the construction of Γ_g that either $s_0 z_0 \succ_2 s_0 z_g$ holds, or $s_0 z_0 \succ_1 s_g z_0$ holds, *but not both.*[15]

If $s_0 z_0 \succ_1 s_g z_0$, then by definition $T_\succ(x_g) = \mathfrak{f}$, so $\{x_e, x_f, \bar{x}_g\}$ is satisfied. If, on the other hand, $s_g z_0 \succ_1 s_0 z_0$, then $s_0 z_0 \succ_2 s_0 z_g$ holds (the edge cycle in Γ_g must be oriented), and since $s_0 z_0$ is not a Nash equilibrium in $(\{s_0, s_e, s_f, s^*\} \times \{z_0, z^g, z^*\}, s^* z^*)$ (see (16)), it must be that either $s_e z_0 \succ_1 s_0 z_0$ or $s_f z_0 \succ_1 s_0 z_0$. Then, by the definition of $T_\succ$, either $T_\succ(x_e) = \mathfrak{t}$ or $T_\succ(x_f) = \mathfrak{t}$, and so $\{x_e, x_f, \bar{x}_g\}$ is satisfied.

The situation for clauses of the type $\{x_e, \bar{x}_f, \bar{x}_g\}$ and $\{\bar{x}_e, \bar{x}_f, \bar{x}_g\}$ is analogous, and these clauses will also be satisfied by $T_\succ$. Thus the truth assignment $T_\succ$ satisfies V.

$\Rightarrow$ To prove the converse, suppose that V is satisfied by a truth assignment T. We will describe rationalizing (non-total) preference relations $\succ_1$ on S_k and $\succ_2$ on Z_k, and we will show that they are acyclic.[16] Then extensions of these orders to total orders will also rationalize NR_V. First we define player 1's preferences. The example in Figure 2 illustrates the construction of rationalizing preferences (for both players).

1. For $z \in Z_k \backslash \{z_0\}$, let (s^*, z) be the best element in the row $S_k \times \{z\}$ under $\succ_1$. (In fact, for simplicity, we may order the points in the rows $S_k \times \{z^*\}$ and $S_k \times \{z_*\}$ as shown in figure 2.)

2. In the row $S_k \times \{z_0\}$ let (s^*, z_0) be the worst element under $\succ_1$.

[15] In fact, Γ_g is constructed so that it is rationalizable if and only if the "edge cycle" indicated by a dashed line in Figure 1 b) is oriented in one direction or the other.

[16] Recall that $S_k := \{s_*, s^*, s_0, s_1, \ldots, s_k\}$ and $Z_k := \{z_*, z^*, z_0, z_1, \ldots, z_k\}$.

3. For $z \in Z_k \backslash \{z_*, z^*, z_0\}$, let (s_*, z) be the worst element in the row $S_k \times \{z\}$ under $\succ_1$.
4. In the row $S_k \times \{z_0\}$ let (s_*, z_0) be worse than any other point except (s^*, z_0) (which we have already defined to be the bottom element in that row).
5. In the row $S_k \times \{z_*\}$ let (s_*, z_*) be the second best element under $\succ_1$ (in step 1. we defined (s^*, z_*) as the best element in this row).
6. For all $i \in \{1, 2, \ldots, k\}$ such that $T(x_i) = \mathfrak{t}$, let $s_i z_0 \succ_1 s_0 z_0$ and

$$(s_0, z_i) \succ_1 (s_1, z_i) \succ_1 \cdots \succ_1 (s_{k-1}, z_i) \succ_1 (s_k, z_i), \tag{23}$$

and for all $i \in \{1, 2, \ldots, k\}$ such that $T(x_i) = \mathfrak{f}$, let $s_0 z_0 \succ_1 s_i z_0$ and

$$(s_k, z_i) \succ_1 (s_{k-1}, z_i) \succ_1 \cdots \succ_1 (s_1, z_i) \succ_1 (s_0, z_i). \tag{24}$$

The preferences $\succ_2$ for player 2 are defined symmetrically — one can just exchange the roles of "s" and "z" in the preceding definition, and substitute $\succ_2$ for $\succ_1$ and "column" for "row" — except for the crucial step 6., which becomes:

6'. For all $i \in \{1, 2, \ldots, k\}$ such that $T(x_i) = \mathfrak{t}$, let $s_0 z_0 \succ_2 s_0 z_i$ and

$$(s_i, z_k) \succ_2 (s_i, z_{k-1}) \succ_2 \cdots \succ_2 (s_i, z_1) \succ_2 (s_i, z_0), \tag{25}$$

and for all $i \in \{1, 2, \ldots, k\}$ such that $T(x_i) = \mathfrak{f}$, let $s_0 z_i \succ_2 s_0 z_0$ and

$$(s_i, z_0) \succ_2 (s_i, z_1) \succ_2 \cdots \succ_2 (s_i, z_{k-1}) \succ_2 (s_i, z_k). \tag{26}$$

One can easily verify that the above defined preferences are acyclic. Since we defined relations only on rows and columns, we can check acyclicity for each row and for each column separately. In the row $S_k \times \{z_0\}$ and in the column $\{s_0\} \times Z_k$ all relations involve the point (s_0, z_0), and so there is no possibility of a cycle. In the rows $S_k \times \{z_*\}$ and $S_k \times \{z^*\}$ and in the columns $\{s_*\} \times Z_k$ and $\{s^*\} \times Z_k$ it is again clear that $\succ_1$ and $\succ_2$ have no cycles; in fact, we can define preferences on these rows and columns as shown in Figure 2. As to the remaining rows and columns, we will verify acyclicity on just one — preferences on the others are defined very similarly. Consider the row $S_k \times \{z_i\}$ (where $0 < i \leq k$). The point (s^*, z_i) is the best element in that row, (s_*, z_i) is the worst, and the remaining are ordered linearly — i.e., the entire row is ordered linearly.

It remains to show that these preferences do, in fact, rationalize all the game form–outcome pairs in NR_V. It is immediate that the sets of game form–outcome pairs Γ_i (for $i = 1, \ldots, k$) are rationalized by these preferences (that is, the outcome (s_*, z_*) is a Nash equilibirum, and at any other profile either player 1 prefers to deviate under $\succ_1$ or player 2 prefers to deviate under $\succ_2$). Checking that the other game form–outcome pairs (15–19) are also rationalized

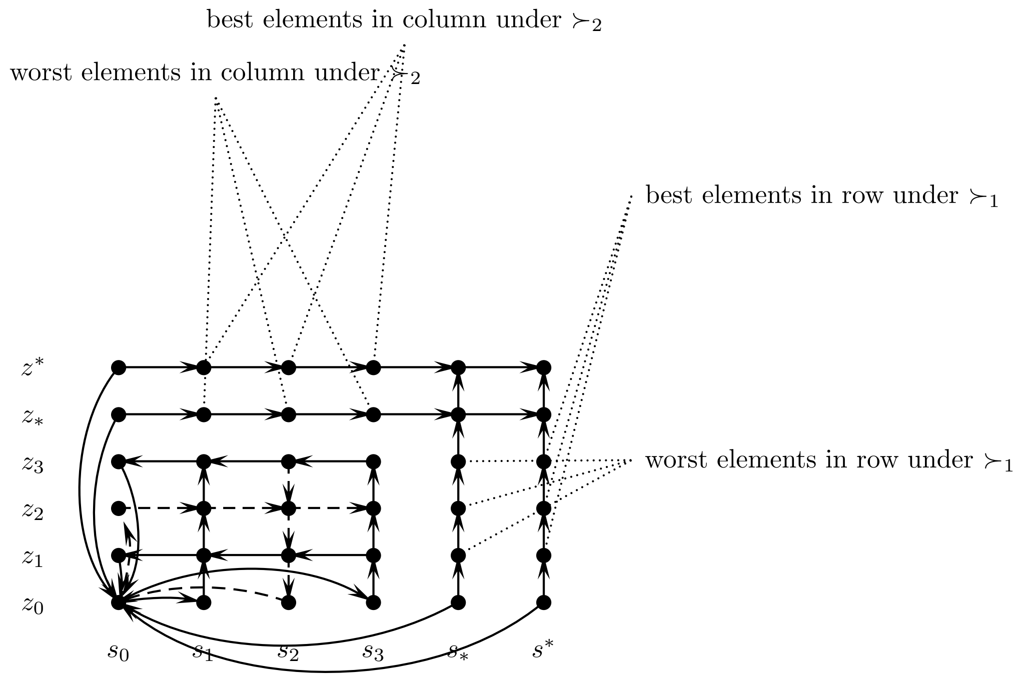

Figure 2: Rationalizing preferences for $T(x_1) = \mathfrak{t}, T(x_2) = \mathfrak{f}, T(x_3) = \mathfrak{t}$. The dashed line indicates the relations that arise from $T(x_2) = \mathfrak{f}$.

by $\succ_1$ and $\succ_2$ is also routine. For example, consider one of the type defined in (16): $(\{s_0, s_e, s_f, s^*\} \times \{z_0, z_g, z^*\}, s^*z^*)$. Under $\succ_1$ and $\succ_2$, the profile (s^*, z^*) is clearly a Nash equilibrium. The profiles on the same row or column as (s^*, z^*) are not Nash equilibria, because they are dominated by (s^*, z^*). The profile (s_0, z_0) is not a Nash equilibrium because the truth assignment T (based on which $\succ_1, \succ_2$ were defined) is satisfied, and thus either $(s_e, z_0) \succ_1 (s_0, z_0)$ or $(s_f, z_0) \succ_1 (s_0, z_0)$ holds (by step 6. in the definition of $\succ_1$), or $(s_0, z_g) \succ_2 (s_0, z_0)$ holds (by step 6'. in the definition of $\succ_2$). The remaining points are not Nash equilibria because either player 1 would deviate to his s^* strategy, or player 2 would deviate to her z^* strategy (or both).

We have shown that our polynomial transformation produces a Nash rationalizable instance of NR2 if and only if the input 3SAT instance is satisfiable. Thus if an algorithm could decide any instance of NR2 in polynomial time, then any instance V of 3SAT could be be decided in polynomial time by first using our algorithm to produce NR_V in polynomial time, and then deciding NR_V in polynomial time. Since 3SAT is NP-complete, this proves that NR2 is NP-complete. □

References

Backhouse, R., 2017. Founder of Modern Economics: Paul A. Samuelson: Volume 1: Becoming Samuelson, 1915-1948. Oxford Studies in History of Economics. Oxford University Press.
URL https://books.google.com/books?id=65-PDgAAQBAJ

Benthem, J. F. A. K. v., 1978. Ramsey eliminability. Studia Logica: An International Journal for Symbolic Logic 37 (4), pp. 321–336.
URL http://www.jstor.org/stable/20014913

Carvajal, A., Ray, I., Snyder, S., February 2004. Equilibrium behavior in markets and games: testable restrictions and identification. Journal of Mathematical Economics 40 (1-2), 1–40.

Chambers, C., Echenique, F., 2016. Revealed Preference Theory. Econometric Society Monographs. Cambridge University Press.
URL https://books.google.com/books?id=4fSTCwAAQBAJ

Chambers, C. P., Echenique, F., Shmaya, E., 2017. General revealed preference theory. Theoretical Economics 12 (2), 493–511.

Cook, S. A., 1971. The complexity of theorem-proving procedures. In: Proceedings of the third annual ACM symposium on Theory of computing. Association for Computing Machinery, pp. 151–158.

Fagin, R., 1973. Contributions to the model theory of finite structures. Ph.D. thesis, U.C. Berkeley.

Galambos, A., 2004. Revealed preference in game theory. Ph.D. thesis, University of Minnesota.

Garey, M. R., Johnson, D. S., 1979. Computers and Intractability. W.H. Freeman and Company.

Gradwohl, R., Shmaya, E., 2015. Tractable falsifiability. Economics & Philosophy 31 (2), 259–274.

Hands, D. W., 2001. Reflection without rules: economic methodology and contemporary science theory. Cambridge University Press.

Hintikka, J., 1998. Ramsey sentences and the meaning of quantifiers. Philosophy of Science 65 (2), pp. 289–305.
URL http://www.jstor.org/stable/188262

Houthakker, H. S., May 1950. Revealed preference and the utility function. Economica NS 17 (66), 159–174.

Immerman, N., 1982. Relational queries computable in polynomial time. In: 14th Symposium on Theory of Computation. ACM, pp. 147–152.

Immerman, N., 1995. Descriptive complexity: a logician's approach to computation. Notices of the American Mathematical Society 42.

Immerman, N., 1999. Descriptive Complexity. Graduate Texts in Computer Science. Springer.

Matzkin, R. L., Richter, M. K., 1991. Testing strictly concave rationality. Journal of Economic Theory 53, 287–303.

Ramsey, F., 1931. Theories. In: Braithwaite, R. (Ed.), The foundations of mathematics and other logical essays. No. v. 5 in The international library of philosophy: Philosophy of logic and mathematics. Routledge and Kegan Paul.

Richter, M. K., 1966. Revealed preference theory. Econometrica 34, 635–645.

Samuelson, P. A., 1953. Consumption theorems in terms of overcompensation rather than indifference comparisons. Economica 20 (77), pp. 1–9.
URL `http://www.jstor.org/stable/2550984`

Sneed, J., 1971. The logical structure of mathematical physics. Synthese library. Reidel.

Sprumont, Y., 2000. On the testable implications of collective choice theories. Journal of Economic Theory 93, 205–232.

Vardi, M. Y., 1982. Complexity of relational query languages. In: 14th Symposium on Theory of Computation. ACM, pp. 137–146.

Varian, H. R., 1982. The nonparametric approach to demand analysis. Econometrica: Journal of the Econometric Society, 945–973.

Yanovskaya, E., 1980. Revealed preference in non-cooperative games. Mathematical Methods in the Social Sciences 24 (13), in Russian.