

# The Complexity of Nash Rationalizability

Ádám Galambos<sup>1</sup>

*Department of Economics, Lawrence University, 711 E Boldt Way, Appleton, WI, 54911*

---

## Abstract

The observed choices of a set of players interacting in various related games are said to be *Nash rationalizable* if there exist preferences for the players that make those choices the pure strategy Nash equilibria for each game. We provide an intuitive characterization of Nash rationalizability under less restrictive conditions than those in the existing literature, and show that this restricted Nash rationalizability problem is computationally tractable. Then we show that the general Nash rationalizability problem is NP-complete, i.e. it is considered computationally intractable. We use results from descriptive complexity theory to explain the implications of our theorems for revealed preference theory for games.

*Key words:* Nash equilibrium; Revealed preference; Complexity.

*JEL classifications:* C72, D70.

---

## 1. Introduction

The importance of questions of computational complexity has been increasingly recognized in the game theory literature (Halpern, 2008). Previous authors have examined the complexity of computing Nash equilibria

---

*Email address:* adam.galambos@lawrence.edu (Ádám Galambos)

<sup>1</sup>I am grateful to Marcel K. Richter for many inspiring and stimulating discussions on these topics, as well as many suggestions. I wish to thank Beth Allen, Jan Werner, and participants of the Micro/Finance and Micro/Game theory workshops at the University of Minnesota for their comments, and especially Andrew McLennan for suggesting that I investigate the computational complexity literature. This paper is based on my doctoral dissertation at the University of Minnesota. I gratefully acknowledge the financial support of the NSF through grant SES-0099206 (principal investigator: Jan Werner). I am grateful for the detailed comments of two anonymous referees and an Associate Editor.

(e.g. Daskalakis et al., 2009), and the complexity of mechanism design (e.g. Conitzer and Sandholm, 2002). The present paper contributes to this growing literature by investigating the complexity of determining whether the observed behavior of players can be described using the Nash equilibrium solution. That is, we study the complexity of testing whether the choices of a group of players are *Nash rationalizable* (see below for a formal definition).

Thus we also contribute to the literature on the testable implications of the Nash equilibrium solution concept. In the theory of individual decision making, the revealed preference literature asks questions of the form: “What conditions characterize choice behavior that is generated by maximization of a preference relation with certain properties?” Starting with Samuelson (1938) and Houthakker (1950), answers to such questions have been obtained in quite general settings (Richter, 1966, 1971, 1975). The analogous revealed preference questions in collective choice theory were addressed only recently.<sup>2</sup> Yanovskaya (1980) and Sprumont (2000) characterize joint choice behavior that is consistent with the pure strategy Nash equilibrium solution for normal form games. Their conditions are very similar to the “Consistency” and “Converse Consistency” conditions of Peleg and Tijs (1996).

These existing results on the testable implications of Nash equilibrium assume that observations are complete in a very strong sense. We relax these assumptions and provide a characterization that is a straightforward extension of the revealed preference result of Richter (1966) to a multi-agent setting.<sup>3</sup> Our condition is plainly analogous with the strong axiom of revealed preference. Even with our weakened assumptions on the structure of observations, testing Nash rationalizability is a polynomial problem. Then we consider the general problem of testing Nash rationalizability when no

---

<sup>2</sup>See Carvajal et al. (2004) for a survey. Yanovskaya (1980) and Sprumont (2000) are most closely related to our work, because they formulate their questions for normal form games, as we do. A complementary literature (Ray and Zhou, 2001; Ray and Snyder, 2003) considers analogous questions for extensive form games.

<sup>3</sup>In the context of extensive form games and subgame perfect Nash equilibrium, Ray and Zhou (2001) also use a revealed preference approach and allow for infinite action sets. However, they impose the “complete domain” assumption, and use a “subgame consistency” and an “internal consistency” condition in addition to the revealed preference condition (“acyclicity of the revealed base relation”). Demuynck and Lauwers (2009) consider Nash rationalizability with choices over lotteries, and independently obtain a slightly different generalization of Richter’s Congruence Axiom. For a discussion of the relationship to this paper, see (Demuynck and Lauwers, 2009, p. 12).

restrictions are imposed on the structure of observations. We show that this problem is NP-complete even if there are only two players.

Our result has important practical and theoretical implications. From an applied perspective, it is computationally difficult to determine whether agents involved in some interaction can be described as behaving noncooperatively, according to the Nash equilibrium solution. To understand the relevance of our computational complexity result for the literature on the testable implications of Nash equilibrium, we turn to the notion of descriptive complexity. Using known connections between computational complexity and descriptive complexity, we show that the NP-completeness of the general Nash rationalizability problem means that there exists no attractive, general characterization of Nash rationalizability analogous to the strong axiom of revealed preference. We believe that this paper is the first to point out the relevance of descriptive complexity notions to the game theory or economics literature.

## 2. Characterization of Nash rationalizability

In this section we formulate the testable implications question for Nash equilibrium, so that in the next section we can consider the complexity of that problem. We also present a simple characterization theorem that relaxes assumptions made in the previous literature.

We consider the testable implications of *pure strategy* Nash equilibrium, and so it is natural to take choices of pure strategies (by all players) as the observed data. We suppose that we observe a set of games played, and for each game we see the players, the game form, as well as the chosen pure strategy of each player. Because we do not observe the payoffs, we will say that we observe these players as they *play game forms*.<sup>4</sup> Each observed game form is considered a one-shot interaction, even though some players might encounter each other in several game forms. In our analysis, the sequence in which the observed game forms are played has therefore no significance. Though this is not a fully general formulation of the problem, it has proven to be a useful starting point in the literature (Sprumont, 2000).

Formally, we observe a finite set  $I$  of players play a finite number of game

---

<sup>4</sup>But we assume that players know their own (and possibly other players') preferences. Only us, the observers, are ignorant of players' preferences.

forms. Each player has a *universal strategy space*  $\mathcal{S}_i$ .<sup>5</sup> Let  $\mathcal{S} := \prod_{i \in I} \mathcal{S}_i$ . The strategy space of a player  $i$  in any observed game form will always be a subset of his universal strategy space  $\mathcal{S}_i$ . Because we assume that in each observed game form we observe each player's choice of strategy, it is without loss of generality to identify the outcome space with the Cartesian product of players' strategy spaces.<sup>6</sup> Thus the set of observed game forms, denoted by  $\Lambda$ , is a set of Cartesian product subsets of  $\mathcal{S}$ . For each game form  $S$  in  $\Lambda$ , we observe the strategy profiles played. We assume that if there are several strategy profiles which players would be willing to choose, then we observe all of these as chosen. Formally, we are given a *choice correspondence*  $\mathfrak{C} : \Lambda \rightrightarrows \mathcal{S}$  with the property that  $\mathfrak{C}(S) \subseteq S$  for all  $S \in \Lambda$ . Though  $\mathfrak{C}$  is not required to be non-empty valued in general, we do assume that  $\mathfrak{C}(S)$  is non-empty for all effectively *one-player* game forms  $S$ , i.e. all game forms in which at most one player has more than one strategy available to her. We first ask whether there exist preferences that make the observed choice correspondence the pure strategy Nash equilibrium correspondence.

**Definition 1.** A choice correspondence  $\mathfrak{C} : \Lambda \rightrightarrows \mathcal{S}$  is Nash rationalizable if there exist total, transitive and reflexive preferences  $(\precsim_i)_{i \in I}$  on  $\mathcal{S}$  such that for all  $S \in \Lambda$ , the chosen set  $\mathfrak{C}(S)$  is the set of pure strategy Nash equilibria of  $(S, (\precsim_i)_{i \in I})$ , i.e.

$$s^* \in \mathfrak{C}(S) \iff \forall i \in I, \forall s_i \in \mathcal{S}_i, (s_i, s_{-i}^*) \precsim_i s^*.$$

As in revealed preference theory, it is useful to define a *revealed preference relation* (Samuelson, 1938; Richter, 1966) for each player. These revealed preference relations formalize the notion that if an observed strategy profile is to be rationalized as an equilibrium, then it must be preferred by each player to all other strategy profiles that could have been reached by him via unilateral deviation. Formally, define for each  $i \in I$  a relation  $V_i$  on  $\mathcal{S}$ :

$$s' V_i s \iff \exists S \in \Lambda [s, s' \in S \text{ and } \forall j \in I \setminus \{i\} s_j = s'_j \text{ and } s \in \mathfrak{C}(S)]. \quad (1)$$

---

<sup>5</sup>Note that the spaces  $\mathcal{S}_i$  are *not* assumed to be finite.

<sup>6</sup>Our results would easily generalize to the case where the outcome space is not necessarily identical with the product of strategy spaces. However, making the less restrictive assumption that only outcomes are observed and players' strategy choices are not would complicate the analysis. The NP-completeness result of the next section would, a fortiori, still hold.

In words,  $s$  is *directly revealed preferred* to  $s'$  if, and only if, in some observed game form  $S$ , the strategy profile  $s$  was chosen, and the strategy profile  $s'$  was not chosen but was reachable to  $i$  through a unilateral deviation from  $s$ . Let  $W_i$  be the transitive closure of  $V_i$ , i.e. the *indirectly revealed preferred* relation.

**Definition 2** (I-Congruence). *A choice correspondence  $\mathfrak{C}$  satisfies I-Congruence if the following condition holds:*

$$\forall_{S \in \Lambda} \forall_{s \in S} \left[ \left[ \forall_{i \in I} \forall_{s'_i \in S_i} (s'_i, s_{-i}) W_i s \right] \Rightarrow s \in \mathfrak{C}(S) \right]. \quad (2)$$

*In words, if a strategy profile is indirectly revealed preferred by each player to all strategy profiles that he could unilaterally deviate to, then this strategy profile should be observed as chosen.*

This condition is a direct generalization of that in Richter (1966) to games. Previous authors have made the assumption that observations are complete in the sense that the domain  $\Lambda$  contains *all* Cartesian product subsets of  $\mathcal{S}$ . For the theorem below, this assumption of “complete domain” is relaxed to “closed domain:”

**Definition 3.** *A class  $\Lambda$  of game forms is closed if  $S \in \Lambda$  implies that for any  $s \in S$  and any  $i \in I$ , the reduced game form  $s_{-i} \times S_i$  is also in  $\Lambda$ , where  $s_{-i} \times S_i$  denotes the game form with singleton strategy sets  $\{s_j\}$  for all players  $j \neq i$  and with strategy set  $S_i$  for  $i$ .*

This definition of closedness is essentially the same as the one used in Peleg and Tijs (1996) in their axiomatization of Nash equilibrium.

**Theorem 1.** *Suppose  $\Lambda$  is closed. A choice correspondence  $\mathfrak{C} : \Lambda \rightrightarrows \mathcal{S}$  is (pure strategy Nash equilibrium) rationalizable if and only if it satisfies the I-Congruence axiom.*

This characterization relaxes the complete domain assumption to closed domain. However, the theorem does not hold if  $\Lambda$  is allowed to be any arbitrary set of game forms, as the example below illustrates.

**Example 1.** *Suppose that two players are observed playing four game forms. The universal strategy spaces for the players are  $\mathcal{S}_1 = \{U, D\}$  and  $\mathcal{S}_2 = \{L, M, R\}$ . We display the four observed game forms below, with a square marking the (only) strategy profile that’s observed as chosen.*

	L	M					
U							
D	■						
$(U, L)V_1(D, L)$			$(D, R)V_1(U, R)$			$(U, L)V_2(U, M)$	
$(D, M)V_2(D, L)$			$(U, M)V_2(U, R)$			$(D, R)V_2(D, M)$	

The domain of this choice correspondence is not closed; for example, the one-player game form  $\{U, D\} \times \{M\}$  is not observed, even though the larger game form  $\{U, D\} \times \{M, R\}$  is.<sup>7</sup> The example does, however, satisfy *I*-Congruence. Under each observation, we show the directly revealed preferred relations derived from that observation. The only additional indirectly revealed preferred relations are:  $(D, R)W_2(D, L)$  and  $(U, L)W_2(U, R)$ . It is straightforward to check that in each game form above, every strategy profile that is directly (*V*) or indirectly (*W*) revealed preferred by each player to the strategy profiles that he could reach by unilaterally deviating is, in fact, chosen. But this choice correspondence is not Nash rationalizable. To see this, suppose we found rationalizing preferences for each player. This preferences will have to be, by definition, extensions of the revealed preferred relations above. How do player 1's preferences rank  $(U, M)$  and  $(D, M)$ ? If they rank  $(U, M)$  higher, then we should have observed  $(U, M)$  as chosen in the first game form above, because  $(U, L)V_2(U, M)$ , and 2's preferences will have to respect this fact. But if 1's preferences rank  $(D, M)$  higher, then we should have observed  $(D, M)$  as chosen in the second game form above, because  $(D, R)V_2(D, M)$ , and 2's preferences will have to respect this fact. Thus it is impossible to find preferences that rationalize this choice correspondence.

*Proof of Theorem 1. Necessity:* Suppose there are total, transitive and reflexive preferences  $(\succsim_i)_{i \in I}$  on  $\mathcal{S}$  such that for any  $S \in \Lambda$ , the choice set is the set of Nash equilibria:  $\mathfrak{C}(S) = \{s \in S \mid \forall_{i \in I} \forall_{s'_i \in S_i} (s'_i, s_{-i}) \succsim_i s\}$ . Suppose that for some  $S^*$ , there is a  $s^* \in S^*$  such that for all  $i \in I$ , it is revealed preferred to all others available:  $(s'_i, s_{-i})W_i s^*$  for all  $s'_i \in S_i^*$ , and yet  $s^* \notin \mathfrak{C}(S^*)$ .

---

<sup>7</sup>The reader can use this example to compare the complete domain assumption with our closed domain assumption. This domain could be made closed by adding all one-player game forms to it that are contained in one of the four game forms. These would be  $\{U, D\} \times \{M\}$ ,  $\{U, D\} \times \{R\}$ ,  $\{U, D\} \times \{L\}$ ,  $\{D\} \times \{L, M\}$ ,  $\{U\} \times \{M, R\}$ ,  $\{U\} \times \{L, R\}$ , and  $\{D\} \times \{L, R\}$ . However, the domain would still not be complete, for the game forms  $\{U, D\} \times \{L, R\}$  and  $\{U, D\} \times \{L, M, R\}$  would not be included in it.

Since, under our initial supposition, for any  $s, s' \in \mathcal{S}$ , the relation  $s'V_i s$  implies that  $s' \precsim_i s$  and since  $\precsim_i$  is transitive,  $(s'_i, s_{-i})W_i s^*$  implies that  $(s'_i, s_{-i}) \precsim_i s^*$  for all  $s'_i \in S_i^*$ , for all  $i \in I$ . Thus  $s^*$  is a Nash equilibrium and so  $s^* \in \mathfrak{C}(S^*)$ , contradicting our initial supposition and proving the necessity of I-Congruence.

*Sufficiency:* Assume that I-Congruence holds. For  $S \in \Lambda$  and  $s \in S$ , let  $S_i^s$  denote the one-player game form with strategy sets  $\{s_j\}$  for all  $j \neq i$ , and strategy set  $S_i$  for player  $i$ . For each  $i \in I$ , let  $\Lambda_i := \{S_i^s | S \in \Lambda, s \in S\}$ . (Note that by closedness  $\emptyset \neq \Lambda_i \subseteq \Lambda$ .) We derive, for each  $i \in I$ , an “individual choice correspondence”  $\mathfrak{C}_i$  on  $\Lambda_i$ . For all  $\hat{S} \in \Lambda_i$ , let

$$\mathfrak{C}_i(\hat{S}) := \{x \in \hat{S} | x \in \mathfrak{C}(S) \text{ for some } S \text{ with } \hat{S} = S_i^x\} \quad (3)$$

Note that  $\mathfrak{C}_i$  is non-empty valued because we assumed that  $\mathfrak{C}$  is non-empty valued for one-player game forms. By definition, the revealed preferred relation derived from  $\mathfrak{C}_i$  coincides with  $W_i$ . Therefore, by I-Congruence,  $\mathfrak{C}_i$  coincides with  $\mathfrak{C}$  on  $\Lambda_i$ . Since I-Congruence restricted to the one-player games  $\Lambda_i$  is the same as the Congruence axiom of Richter (1966), for each  $i \in I$  there exists a total, transitive, reflexive binary relation  $\precsim_i$  on  $\mathcal{S}$  such that for each  $\hat{S} \in \Lambda_i$ , the set  $\mathfrak{C}_i(\hat{S}) = \mathfrak{C}(\hat{S})$  is the set of  $\precsim_i$ -maximal elements. We will show that these preferences (pure strategy Nash equilibrium) rationalize  $\mathfrak{C}$  on  $\Lambda$ .

$\mathfrak{C}(S)$  are Nash equilibria: Given  $S \in \Lambda$ , suppose  $s' \in \mathfrak{C}(S)$ . Then, for all  $i \in I$ , by the definition of  $W_i$  it must be that  $s''W_i s'$  for all  $s'' \in S_i^{s'}$ . Since  $\precsim_i$  extends  $W_i$  (see Richter (1966)), this implies that for all  $i \in I$ ,  $s'' \precsim_i s'$  for all  $s'' \in S_i^{s'}$ , i.e. that  $s'$  is a Nash equilibrium.

All Nash equilibria are chosen by  $\mathfrak{C}$ : Given  $(S) \in \Lambda$ , suppose that  $s' \in S$  is a Nash equilibrium, i.e. for all  $i \in I$ , it holds that  $s'' \precsim_i s'$  for all  $s'' \in S_i^{s'}$ . Since  $\precsim_i$  rationalizes  $\mathfrak{C}_i = \mathfrak{C}$  on  $\Lambda_i$ , we have  $s' \in \mathfrak{C}_i(S_i^{s'}) = \mathfrak{C}(S_i^{s'})$ . Then  $s''W_i s'$  for all  $s'' \in S_i^{s'}$ , and, by I-Congruence,  $s' \in \mathfrak{C}(S)$ .  $\square$

### 3. The complexity of Nash rationalizability

The characterization theorem in the previous section shows that under a closed domain, the characterization of rational behavior in individual decision theory can be naturally extended to characterize Nash rationalizable behavior in games. Two questions arise. First, is there a similarly intuitive and simple characterization for general (unrestricted) domains? In other words, can we

drop the “closed domain” assumption in Theorem 1? And second, would it be computationally tractable in practice to determine whether a set of observations is Nash rationalizable? It turns out that the answer to the second question helps us answer the first. Determining Nash rationalizability under a closed domain is computationally tractable, but in general it is an NP-complete problem. The *computational* complexity of the general problem translates into a *descriptive* complexity that precludes a characterization of the kind we stated for closed domains. Similarly, the simple characterization of Nash rationalizability under closed domains implies the computational tractability of that problem.

### 3.1. Computational complexity

In this section we assume that the universal strategy spaces  $\mathcal{S}_i$  are finite. Let  $\Lambda$  be an arbitrary finite set of game forms, i.e. a set of Cartesian product subsets of  $\mathcal{S}$ .<sup>8</sup> For each game form  $S \in \Lambda$ , we observe the strategy profiles played. We assume that if there are several strategy profiles which players would be willing to choose, then we observe all of these as chosen. As before, we ask whether a given choice correspondence  $\mathfrak{C}$  is pure strategy Nash rationalizable. To simplify notation, we now require Nash rationalizability by *strict* preferences.<sup>9</sup>

It is possible to characterize pure strategy Nash rationalizability for arbitrary domains (Galambos, 2004), though that characterization involves a statement using second-order logic. In addition to deriving revealed preference relations from chosen strategy profiles (as in section 2), one must also derive revealed preference relations from non-chosen strategy profiles. Suppose, for example, that  $\{s_1, s_2, s_3\} \times \{z_1, z_2, z_3\}$  is a  $3 \times 3$  game form in  $\Lambda$  (involving only two players), and the strategy profile chosen by  $\mathfrak{C}$  is  $(s_1, z_1)$  (see Figure 1). As in section 2, we then infer that if  $\mathfrak{C}$  is to be Nash rationalized, it must be that

$$(s_2, z_1) \prec_1 (s_1, z_1) \quad \text{and} \quad (s_3, z_1) \prec_1 (s_1, z_1) \quad \text{and} \quad (4)$$

$$(s_1, z_2) \prec_2 (s_1, z_1) \quad \text{and} \quad (s_1, z_3) \prec_2 (s_1, z_1). \quad (5)$$

These relations are all derived from the observation that  $(s_1, z_1)$  is observed as chosen. But it turns out to be also necessary to derive relations from the

---

<sup>8</sup>Recall that  $\mathcal{S} := \prod_{i \in I} \mathcal{S}_i$ .

<sup>9</sup>All the results below continue to hold if we ask for rationalizability by weak preferences.



observation that certain other strategy profiles were *not* chosen. For example, since  $(s_3, z_3)$  was not chosen, it must also be that

$$\begin{aligned} (s_3, z_3) \prec_1 (s_2, z_3) \quad \textbf{or} \quad (s_3, z_3) \prec_1 (s_1, z_3) \quad \textbf{or} \\ (s_3, z_3) \prec_2 (s_3, z_2) \quad \textbf{or} \quad (s_3, z_3) \prec_2 (s_3, z_1). \end{aligned} \tag{6}$$

It is not surprising that deciding Nash rationalizability from a set of such disjunctions could be computationally very complex. It is natural to ask whether there exists an alternative, not so complex method for deciding rationalizability. The main result of this section answers that question in the negative.

**Theorem 2.** *The (pure strategy Nash equilibrium) rationalizability problem is NP-complete.*

In fact, we prove a stronger statement: The (pure strategy Nash equilibrium) rationalizability problem is NP-complete *even if* there are only two players. As Fortnow and Homer state in their lucid review of complexity theory, “A proof of NP-completeness has come to signify the (worst case) intractability of a problem.” (Fortnow and Homer, 2003) The proof of Theorem 2 (in Appendix A) is based on a standard technique in the theory of computational complexity: “polynomially reducing” a problem that is known to be NP-complete to the given problem.<sup>10</sup>

### 3.2. Descriptive complexity

The computational intractability of the general Nash rationalizability problem could, in principle, still leave open the possibility that an intuitive characterization of Nash rationalizable choice correspondences exists. Some might argue that from the point of view of the testable implications literature, it is the insightfulness and elegance of the characterizing property that matters more than the practical issue of ease of computability. However, an important and mathematically satisfying insight of the complexity literature says that the two are equivalent in a very specific sense.

---

<sup>10</sup>Specifically, we use 3SAT, a version of the satisfiability problem that was shown to be NP-complete in Cook (1971).

While computational complexity theory focuses on the (worst case) length of time<sup>11</sup> that computing the answer to a query could take, descriptive complexity focuses on the logical language necessary to characterize the answer to that query. Thus descriptive complexity theory is strictly about logical language.<sup>12</sup> We refer the interested reader to the authoritative book by Immerman (1999) for an in-depth exposition. To appreciate how the two notions of complexity are connected, consider the problem of Nash rationalizability. On the one hand, one can formulate the (boolean) query: Is a particular choice correspondence among those that can be (pure strategy) Nash rationalized? Our result above says that answering this question is computationally very complex. On the other hand, one can ask: What kind of logical language must one use to characterize the set of choice correspondences that can be Nash rationalized? The first connection between these two kinds of questions is provided by a famous theorem in the seminal paper of Fagin (1973), as stated by Immerman (1999):

**Theorem 3** (Fagin’s Theorem). *NP is equal to the set of existential, second-order boolean queries,  $NP = SO\exists$ . Furthermore, this equality remains true when the first-order part of the second-order formulas is restricted to be universal.*

We have shown that Nash rationalizability is in the class NP, or nondeterministically polynomial problems. Now we have a different way to show this: the very definition of Nash rationalizability (Definition 1) is a second-order existential formula (of the form “there exist binary (preference) relations such that...”). But more importantly, we showed that Nash rationalizability is NP-complete, that is, it is at least as complex computationally as any other problem in the class NP. This result gains added significance in light of the following:

**Theorem 4** (Theorem 4.10 in (Immerman, 1999)). *Over finite, ordered*

---

<sup>11</sup>Though we focus on time complexity here, analyzing space complexity is also common in the literature. Also, computing time is measured using a standardized model of a universal computer, a *Turing machine*. See Fortnow and Homer (2003) for a brief but broad review.

<sup>12</sup>Descriptive complexity is not to be confused with Kolmogorov complexity, which is sometimes also labeled “descriptive.” The latter measures the length of the shortest algorithm that produces a given string. See Chapter 14 in the comprehensive book by Cover and Thomas (2006) for more.

structures,  $FO(LFP) = P$ .

In words, the set of queries that can be characterized using first-order logic extended by the *least fixed point* (LFP) operator is exactly the set of queries that can be computed in deterministic polynomial time. The LFP operator adds the power of inductive definitions to first-order logic, and the best way to understand it is through its most relevant example, the transitive closure operator. Given a binary relation  $R_0$ , applying the single-step transitive closure rule “if  $aR_0b$  and  $bR_0c$  then let  $aR'_0c$ ” and taking the union  $R_1 = R_0 \cup R'_0$  gives us a new binary relation,  $R_1$ . Iterating this procedure will eventually cease to yield any new pairs in the relation, i.e. for some (smallest)  $k$  we will have  $R_k = R_{k+1}$ . This relation  $R_k$  is then the least fixed point, the relation that the LFP operator would yield when applied to the original relation  $R_0$  and the single-step transitive closure rule. Though the operator is far more general than this example, in revealed preference theory the transitive closure operator is of most interest.<sup>13</sup> We note that the direction of the theorem we will be using,  $FO(LFP) \subseteq P$ , does not need the assumption of an order on structures.<sup>14</sup> We address the finiteness assumption in the proof Corollary 2.

The following corollary of Theorem 1 uses descriptive complexity theory to characterize the computational complexity of Nash rationalizability under a closed domain.

**Corollary 1.** *Under the assumption of a closed domain and a finite strategy space  $\mathcal{S}$ , the Nash rationalizability problem is decidable in deterministic polynomial time, i.e. it is in the class  $P$ .*

As a corollary of Theorem 2, we have the following result on the descriptive complexity of the general Nash rationalizability problem.

**Corollary 2.** *For general, unrestricted domains, there exists no characterization of Nash rationalizability using only  $FO(LFP)$ , unless  $P = NP$ .*

### 3.2.1. Discussion of corollaries

First, Corollary 1 is immediate from Theorems 1 and 4. Since Theorem 1 provides a simple characterization of Nash rationalizability using only first-order logic and transitive closure, we can apply the  $FO(LFP) \subseteq P$  part

---

<sup>13</sup>For a detailed discussion, see section 4.1 in Immerman (1999).

<sup>14</sup>That assumption would do no harm to our results, but we don’t need to worry about it. See discussion after the theorem in Immerman (1999).

of Theorem 4. We thus see that the simple, intuitive description of Nash rationalizability under closed domains implies the computational tractability of that problem.<sup>15</sup> We note that this result highlights both the beauty and the utility of the classic revealed preference results, which offer first-order characterizations (with transitive closure) in place of definitions of rationality that use second-order logic (viz. of the form “there exist binary relations with certain properties such that...”).

Though Corollary 2 follows easily from Theorems 2 and 4, we provide a proof below to make some important points. The qualifier at the end of this result is a reminder that the  $P \stackrel{?}{=} NP$  problem is one of the most important open problems in mathematics today, even though it is widely believed that  $P \subsetneq NP$ . We follow the literature in assuming that this is very likely the case. The implication of Corollary 2 is then that the feat of reducing definitions using second-order logic to characterizations using transitive closure (or some other LFP relation) and first-order logic is *not* possible for the general Nash rationalizability problem. Thus a characterization that is as simple and intuitive as the Strong Axiom of Revealed Preference or our I-Congruence Axiom cannot be found for the general Nash rationalizability problem.

**Proof of Corollary 2.** If  $P \subsetneq NP$ , then  $NP$ -complete problems are not in the class  $P$ , and therefore Nash rationalizability is not in the class  $P$  by our Theorem 2. By Immerman’s Theorem 4.10 (Theorem 4 above), Nash rationalizability then cannot be formulated using first order logic extended by the LFP operator.

A brief remark regarding our application of Immerman’s Theorem 4.10 is in order. Throughout his book, Immerman works with finite structures. In our application, structures are choice correspondences, and the finiteness assumption includes for Immerman that all structures are defined on the same vocabulary that has a finite number of relation variables, each with a specific and finite arity. At first this might seem to be at odds with our setting in that we do not limit a priori the sizes of game forms observed, and if these observations are to be represented as relations, that might mean that relations of arbitrary arity could be needed. But in the proof of our Theorem 2, we use only game forms that are of very limited size (at most

---

<sup>15</sup>The finiteness assumption in Corollary 1 is necessary so that it makes sense even to ask the computational complexity question. Theorem 1 applies to not necessarily finite structures as well.

five strategies for each player). This means that our Theorem 2, in fact, holds even if restricted only to choice correspondences that only have game forms with at most 5 strategies per player. Thus Corollary 2 holds even for this limited class of choice correspondences, and, a fortiori, for the general Nash rationalizability problem.  $\square$

#### 4. Conclusion

This paper brings together two growing literatures: that on the testable implications of collective choice theories and that on complexity and game theory. We considered the Nash rationalizability problem under a closed domain and also under unrestricted domains. In the first, more restricted setting, we presented a natural characterization of Nash rationalizability. In the second, more general setting, we showed that the problem of Nash rationalizability is computationally very complex. Then we explained that the simplicity of the characterization in the first setting is equivalent to the computational tractability of that problem, while the computational intractability of Nash rationalizability in the unrestricted setting implies that no characterization exists that is as simple as the classical revealed preference conditions. We relied on results from descriptive complexity theory, a subfield that may be of further interest to game theorists and economists in general.

Our work establishes that the assumption of a closed domain makes Nash rationalizability computationally tractable, while unrestricted domains lead to untractability. One direction for further work in this area would be to explore what level of complexity other assumptions on the structure of observations would imply.

An interesting question for future research is the role of beliefs in multi-agent decision making. Since the literature so far has addressed only the Nash equilibrium solution concept, the role of beliefs has been hidden by the implicit assumption that agents' beliefs correspond exactly to the actions taken. If one were to study behavior generated by non-equilibrium solution concepts, such as Pearce-Bernheim rationalizability, the prominent role of beliefs would become apparent.

Another interesting aspect of this problem is the relationship between the analyst or observer and the decision making process. In rationalizability for individual choice problems, it seems clear that the observer and the decision making process are entirely separate. That is, the analyst is outside

the decision making problem, observing the behavior of the decision-maker. In collective decision making situations, it is conceivable that the analyst is himself one of the decision-makers. For example, a player in a game, not knowing the preferences of the other players, might attempt to draw conclusions concerning the plausibility of certain possible outcomes, based on some previous experiences of games played by the same agents. Analyzing situations of this kind might lead to interesting applications.

## A. Appendix

Here we prove Theorem 2 of section 3, which states that the Nash rationalizability problem (NR) is NP-complete. Our proof involves two additional problems: Nash rationalizability with only two players<sup>16</sup> (NR2), and the classic problem of determining the satisfiability of a Boolean formula in conjunctive normal form with three disjuncts in each conjunct (3SAT).

*Proof of Theorem 2.* 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), which is a special case of the Nash rationalizability problem (henceforth denoted by NR). That is, we will 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 if and only if the 3SAT instance is satisfiable. 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:** The Nash rationalizability problem with two players can be described as follows. 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

---

<sup>16</sup>I.e. the *same* two players are involved in every observed game form.

finite game forms of  $S \times Z$ . For example, the following instance of NR2 encodes a choice function on two game forms.

$$(\{s_0, s_1, s_2\} \times \{z_0, z_1\}, s_2 z_1), (\{s_0, s_4, s_5\} \times \{z_0, z_2\}, s_4 z_0) \quad (7)$$

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.<sup>17</sup>

**3SAT:** Suppose that  $X = \{x_1, x_2, \dots, x_m\}$  is a set of Boolean variables and  $\bar{X} = \{\bar{x}_1, \dots, \bar{x}_m\}$  is the set of their negations. For any truth assignment  $T : X \rightarrow \{\mathbf{t}, \mathbf{f}\}$ , we define for  $\bar{x} \in \bar{X}$  the extension of  $T$  by  $T(\bar{x}) = \mathbf{t}$  if, and only if  $T(x) = \mathbf{f}$ . The set  $X^* := X \cup \bar{X}$  is the set of *literals*. A subset  $C$  of  $X^*$  is a *clause*. Suppose a set  $\{C_1, \dots, C_k\}$  of clauses is given. A truth assignment  $T : X \rightarrow \{\mathbf{t}, \mathbf{f}\}$  *satisfies*  $\{C_1, \dots, C_k\}$  if for every clause  $C_i$  there exists  $x \in C_i$  with  $T(x) = \mathbf{t}$ . 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  $\rightarrow$  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:

$$V = \{\{v_1^1, v_1^2, v_1^3\}, \{v_2^1, v_2^2, v_2^3\}, \dots, \{v_l^1, v_l^2, v_l^3\}\}, \quad (8)$$

where  $v_j^i \in X^*$ . Suppose w.l.o.g. that the set of variables that appear in  $V$  is  $\{x_1, \dots, x_k\}$ . We will construct an instance of NR2 for  $V$ , using the actions

---

<sup>17</sup>It is clear that NR2 is in the class NP: given an instance of NR2 and preference relations for every player, it can be checked in polynomial time whether the preferences Nash rationalize the given choice function.

$s_*, s^*, s_0, s_1, \dots, s_k$  for player 1, and the actions  $z_*, z^*, z_0, z_1, \dots, z_k$  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 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  $\Gamma_2$ ) 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 2).

**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  $\Gamma_h$ , which consists of the following game form–outcome pairs:

$$\begin{aligned}
& (\{s_0, s_h, s_*\} \times \{z_0, z_h, z_*\}, s_* z_*) \\
& (\{s_0\} \times \{z_h, z_*\}, s_0 z_h) \\
& (\{s_0\} \times \{z_0, z_*\}, s_0 z_0) \quad ^{18} \\
& (\{s_h\} \times \{z_h, z_*\}, s_h z_h) \\
& (\{s_h\} \times \{z_0, z_*\}, s_h z_0) \\
& (\{s_h, s_*\} \times \{z_0\}, s_h z_0) \\
& (\{s_0, s_*\} \times \{z_0\}, s_0 z_0) \quad ^{18} \\
& (\{s_h, s_*\} \times \{z_h\}, s_h z_h) \\
& (\{s_0, s_*\} \times \{z_h\}, s_0 z_h)
\end{aligned} \tag{9}$$

Figure 2 illustrates this set of game form–outcome pairs. For transparency, the first pair in (9) 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 2 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 2 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

$$(\{s_0, s_e, s_f, s_g, s^*\} \times \{z_0, z^*\}, s^* z^*). \tag{10}$$

2. Replace every clause of the form  $\{x_e, x_f, \bar{x}_g\}$  with

$$(\{s_0, s_e, s_f, s^*\} \times \{z_0, z_g, z^*\}, s^* z^*) \tag{11}$$

and  $\Gamma_g$  (see (9) for the definition of the nine game form–outcome pairs in  $\Gamma_h$  for  $h = 1, \dots, k$ ).

---

<sup>18</sup>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.

3. Replace every clause of the form  $\{x_e, \bar{x}_f, \bar{x}_g\}$  with

$$(\{s_0, s_e, s^*\} \times \{z_0, z_f, z_g, z^*\}, s^* z^*) \quad (12)$$

and  $\Gamma_g$  and  $\Gamma_f$ .

4. Replace every clause of the form  $\{\bar{x}_e, \bar{x}_f, \bar{x}_g\}$  with

$$(\{s_0, s^*\} \times \{z_0, z_e, z_f, z_g, z^*\}, s^* z^*) \quad (13)$$

and  $\Gamma_g, \Gamma_f$  and  $\Gamma_e$ .

5. Add the following game form–outcome pairs:

$$(\{s_0\} \times \{z_0, z^*\}, s_0 z_0), (\{s_0, s^*\} \times \{z_0\}, s_0 z_0). \quad (14)$$

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<sup>19</sup>  $S_k := \{s_*, s^*, s_0, s_1, \dots, s_k\}$  and  $Z_k := \{z_*, z^*, z_0, z_1, \dots, z_k\}$ , and denote the players' rationalizing preferences on  $S_k \times Z_k$  by  $\prec_1$ , and  $\prec_2$ . Define, for each variable  $x_i$  with  $i \in \{1, 2, \dots, k\}$  (recall that these are exactly the variables that appear in  $V$ ) a truth assignment:

$$T_{\prec}(x_i) = \mathbf{t} \iff s_0 z_0 \prec_1 s_i z_0. \quad (15)$$

Consider a clause of the form  $\{x_e, x_f, x_g\}$ . Since  $NR_V$  contains (see (10) and (14))

$$\begin{aligned} &(\{s_0, s_e, s_f, s_g, s^*\} \times \{z_0, z^*\}, s^* z^*), \\ &(\{s_0\} \times \{z_0, z^*\}, s_0 z_0), \\ &(\{s_0, s^*\} \times \{z_0\}, s_0 z_0), \end{aligned} \quad (16)$$

---

<sup>19</sup>Recall that  $V$  involves the variables  $x_1, \dots, x_k$ .

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

$$[s_0 z_0 \prec_1 s_e z_0] \text{ or } [s_0 z_0 \prec_1 s_f z_0] \text{ or } [s_0 z_0 \prec_1 s_e z_0]. \quad (17)$$

Under  $T_{\prec}$  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  $\prec_1$ , and  $\prec_2$  rationalize  $NR_V$ , then it follows from the construction of  $\Gamma_g$  that either  $s_0 z_g \prec_2 s_0 z_0$  holds, or  $s_g z_0 \prec_1 s_0 z_0$  holds, *but not both*.<sup>20</sup>

If  $s_g z_0 \prec_1 s_0 z_0$ , then by definition  $T_{\prec}(x_g) = \mathbf{f}$ , so  $\{x_e, x_f, \bar{x}_g\}$  is satisfied. If, on the other hand,  $s_0 z_0 \prec_1 s_g z_0$ , then  $s_0 z_g \prec_2 s_0 z_0$  holds (the edge cycle in  $\Gamma_g$  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 (11)), it must be that either  $s_0 z_0 \prec_1 s_e z_0$  or  $s_0 z_0 \prec_1 s_f z_0$ . Then, by the definition of  $T_{\prec}$ , either  $T_{\prec}(x_e) = \mathbf{t}$  or  $T_{\prec}(x_f) = \mathbf{t}$ , 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  $T_{\prec}$ . Thus the truth assignment  $T_{\prec}$  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  $\prec_1$  on  $S_k$  and  $\prec_2$  on  $Z_k$ , and we will show that they are acyclic.<sup>21</sup> Then extensions of these orders to total orders will also rationalize  $NR_V$ . First we define player 1's preferences. The example in Figure 3 illustrates the construction of rationalizing preferences (for both players).

1. For  $z \in Z_k \setminus \{z_0\}$ , let  $(s^*, z)$  be the best element in the row  $S_k \times \{z\}$  under  $\prec_1$ . (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 3.)
2. In the row  $S_k \times \{z_0\}$  let  $(s^*, z_0)$  be the worst element under  $\prec_1$ .
3. For  $z \in Z_k \setminus \{z_*, z^*, z_0\}$ , let  $(s_*, z)$  be the worst element in the row  $S_k \times \{z\}$  under  $\prec_1$ .
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).

<sup>20</sup>In fact,  $\Gamma_g$  is constructed so that it is rationalizable if and only if the “edge cycle” indicated by a dashed line in Figure 2 b) is oriented in one direction or the other.

<sup>21</sup>Recall that  $S_k := \{s_*, s^*, s_0, s_1, \dots, s_k\}$  and  $Z_k := \{z_*, z^*, z_0, z_1, \dots, z_k\}$ .

5. In the row  $S_k \times \{z_*\}$  let  $(s_*, z_*)$  be the second best element under  $\prec_1$  (in step 1. we defined  $(s^*, z_*)$  as the best element in this row).
6. For all  $i \in \{1, 2, \dots, k\}$  such that  $T(x_i) = \mathbf{t}$ , let  $s_0 z_0 \prec_1 s_i z_0$  and

$$(s_k, z_i) \prec_1 (s_{k-1}, z_i) \prec_1 \cdots \prec_1 (s_1, z_i) \prec_1 (s_0, z_i), \quad (18)$$

and for all  $i \in \{1, 2, \dots, k\}$  such that  $T(x_i) = \mathbf{f}$ , let  $s_i z_0 \prec_1 s_0 z_0$  and

$$(s_0, z_i) \prec_1 (s_1, z_i) \prec_1 \cdots \prec_1 (s_{k-1}, z_i) \prec_1 (s_k, z_i). \quad (19)$$

The preferences  $\prec_2$  for player 2 are defined symmetrically — one can just exchange the roles of “s” and “z” in the preceding definition, and substitute  $\prec_2$  for  $\prec_1$  and “column” for “row” — except for the crucial step 6., which becomes:

- 6'. For all  $i \in \{1, 2, \dots, k\}$  such that  $T(x_i) = \mathbf{t}$ , let  $s_0 z_i \prec_2 s_0 z_0$  and

$$(s_i, z_0) \prec_2 (s_i, z_1) \prec_2 \cdots \prec_2 (s_i, z_{k-1}) \prec_2 (s_i, z_k), \quad (20)$$

and for all  $i \in \{1, 2, \dots, k\}$  such that  $T(x_i) = \mathbf{f}$ , let  $s_0 z_0 \prec_2 s_0 z_i$  and

$$(s_i, z_k) \prec_2 (s_i, z_{k-1}) \prec_2 \cdots \prec_2 (s_i, z_1) \prec_2 (s_i, z_0). \quad (21)$$

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  $\prec_1$  and  $\prec_2$  have no cycles; in fact, we can define preferences on these rows and columns as shown in Figure 3. 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  $\Gamma_i$  (for  $i = 1, \dots, k$ ) are rationalized by these preferences (that is, the outcome  $(s_*, z_*)$  is a Nash equilibrium, and at any other profile either player 1 prefers to deviate under  $\prec_1$  or player 2 prefers to deviate

under  $\prec_2$ ). Checking that the other game form–outcome pairs (10–14) are also rationalized by  $\prec_1$  and  $\prec_2$  is also routine. For example, consider one of the type defined in (11):  $(\{s_0, s_e, s_f, s^*\} \times \{z_0, z_g, z^*\}, s^*z^*)$ . Under  $\prec_1$  and  $\prec_2$ , 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  $\prec_1, \prec_2$  were defined) is satisfied, and thus either  $(s_0, z_0) \prec_1 (s_e, z_0)$  or  $(s_0, z_0) \prec_1 (s_f, z_0)$  holds (by step 6. in the definition of  $\prec_1$ ), or  $(s_0, z_0) \prec_2 (s_0, z_g)$  holds (by step 6'. in the definition of  $\prec_2$ ). 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 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.  $\square$

## References

- 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.
- Conitzer, V., Sandholm, T., 2002. Complexity of mechanism design. In: Dawid, A., Friedman, N. (Eds.), *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*. 18th Annual Conference, Morgan Kaufmann Publishers, San Francisco, California, pp. 103–110.  
URL <http://arxiv.org/abs/cs/0205075v1>
- 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.
- Cover, T. M., Thomas, J. A., 2006. *Elements of Information Theory*, 2nd Edition. Wiley–Interscience.

- Daskalakis, C., Goldberg, P. W., Papadimitriou, C. H., 2009. The complexity of computing a nash equilibrium. *Commun. ACM* 52 (2), 89–97.  
URL <http://doi.acm.org/10.1145/1461928.1461951>
- Demuynck, T., Lauwers, L., 2009. Nash rationalization of collective choice over lotteries. *Mathematical Social Sciences* 57, 1–15.
- Fagin, R., 1973. Contributions to the model theory of finite structures. Ph.D. thesis, U.C. Berkeley.
- Fortnow, L., Homer, S., 2003. A short history of computational complexity. *Bulletin of the European Association for Theoretical Computer Science* 80.
- 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.
- Halpern, J. Y., 2008. Computer science and game theory. In: Durlauf, S. N., Blume, L. E. (Eds.), *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, Basingstoke.
- Houthakker, H. S., May 1950. Revealed preference and the utility function. *Economica* NS 17 (66), 159–174.
- Immerman, N., 1999. *Descriptive Complexity*. Graduate Texts in Computer Science. Springer.
- Peleg, B., Tijs, S. H., 1996. The consistency principle for games in strategic form. *International Journal of Game Theory* 25, 13–34.
- Ray, I., Snyder, S., 2003. Observable implications of Nash and subgame-perfect behavior in extensive games. Tech. Rep. 2, Department of Economics, Brown University.
- Ray, I., Zhou, L., 2001. Game theory via revealed preferences. *Games and Economic Behavior* 37, 415–424.
- Richter, M. K., 1966. Revealed preference theory. *Econometrica* 34, 635–645.

- Richter, M. K., 1971. Rational choice. In: Chipman, J. S., Hurwicz, L., Richter, M. K., Sonnenschein, H. (Eds.), *Preferences, Utility, and Demand*. Harcourt Brace Jovanovich, pp. 635–645.
- Richter, M. K., 1975. Rational choice and polynomial measurement theory. *Journal of Mathematical Psychology* 12, 99–113.
- Samuelson, P. A., 1938. A note on the pure theory of consumer's behaviour. *Economica* 5, 61–71.
- Sprumont, Y., 2000. On the testable implications of collective choice theories. *Journal of Economic Theory* 93, 205–232.
- Yanovskaya, E., 1980. Revealed preference in non-cooperative games. *Mathematical Methods in the Social Sciences* 24 (13), in Russian.

	$z_1$	$z_2$	$z_3$
$s_1$	■		
$s_2$			
$s_3$			

Figure 1: The strategy profile observed chosen is  $(s_1, z_1)$ .

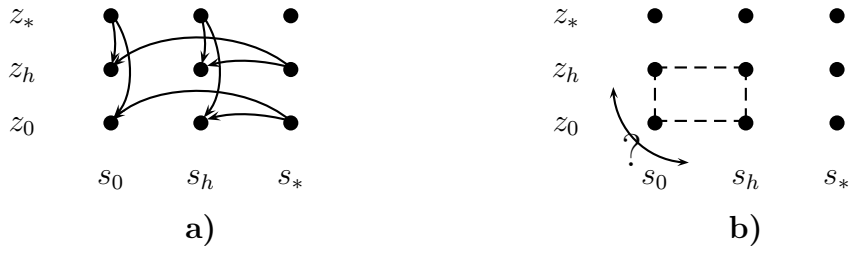


Figure 2: **a)** The game forms in  $\Gamma_h$  **b)** The “edge cycle” must be oriented for rationalizability (recall that these four points are not chosen in the first game form in (9))



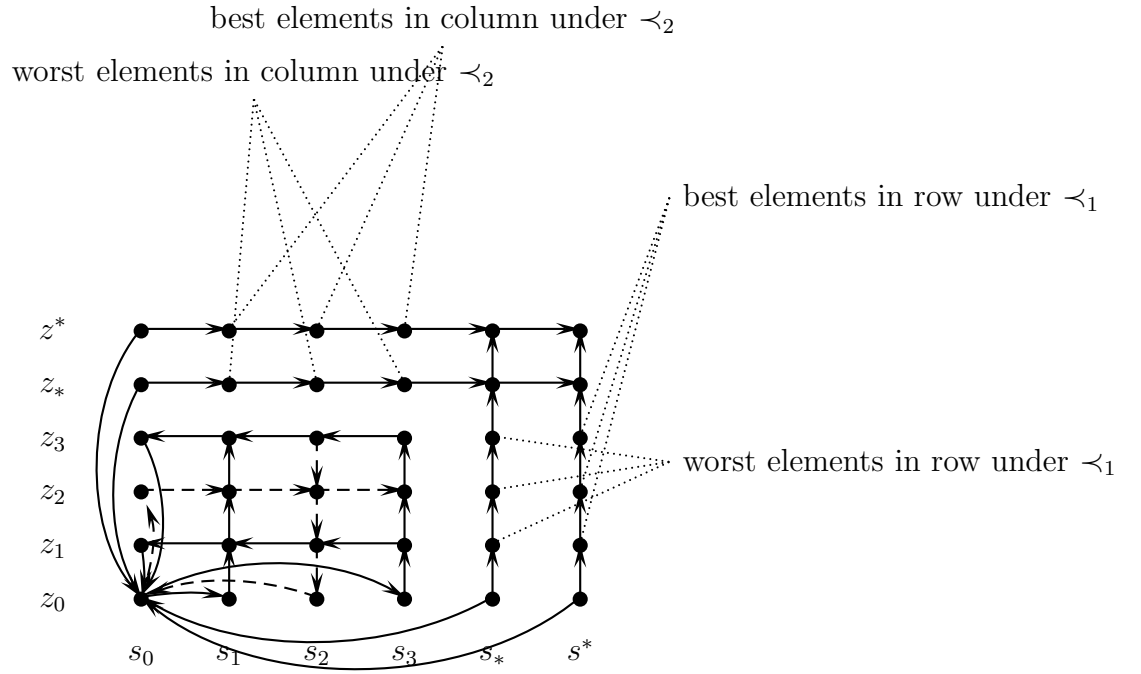


Figure 3: Rationalizing preferences for  $T(x_1) = \mathbf{t}, T(x_2) = \mathbf{f}, T(x_3) = \mathbf{t}$ . The dashed line indicates the relations that arise from  $T(x_2) = \mathbf{f}$ .