Revealed Preference in Game Theory

Ádám Galambos [1]

Department of Managerial Economics and Decision Sciences, Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL, 60208

Abstract

I characterize joint choice behavior generated by the pure strategy Nash equilibrium solution concept by an extension of the Congruence Axiom of Richter(1966) to multiple agents. At the same time, I relax the "complete domain" assumption of Yanovskaya(1980) and Sprumont(2000) to "closed domain." Without any restrictions on the domain of the choice correspondence, determining pure strategy Nash rationalizability is computationally very complex. Specifically, it is NP–complete even if there are only two players. In contrast, the analogous problem with a single decision maker can be determined in polynomial time.

Key words: Nash equilibrium; Revealed preference; Complexity

Email address: a-galambos@kellogg.northwestern.edu (Ádám Galambos).

[1] I am grateful to Professor Marcel K. Richter for many inspiring and stimulating discussions on these topics, as well as many suggestions. I wish to thank Professors Beth Allen, Andrew McLennan and Jan Werner, and participants of the Micro/Finance and Micro/Game theory workshops at the University of Minnesota for their comments. 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: Professor Jan Werner).

27 September 2005

1 Introduction

What are the testable implications of the Nash equilibrium solution? If we observed a group of agents play different games, could we tell, without knowing their preferences, whether they are playing according to Nash equilibrium? Such questions could be of interest to a regulatory agency, wanting to know if some firms they observe in the market are behaving in a competitive or in a collusive way. A manager might ask the same question about her employees. A mechanism designer might want to test if a certain group of agents behave according to the Nash equilibrium solution, to see if he can realistically assume that the agents will behave that way when faced with his mechanism. But these questions are also of interest in themselves from a theoretical point of view.

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, when there is an individual decision maker (Richter, 1966, 1971, 1975). The analogous revealed preference questions in the multi-person decision making problem were addressed only recently.[2] 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"

[2] 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.

conditions of Peleg and Tijs (1996).

I generalize existing results by relaxing the requirement that the domain of the choice correspondence be complete, and by allowing for infinfite action sets. My characterization is a straightforward extension of the revealed preference result of Richter (1966) to a multi-agent setting.[3] Then I consider the question of "Nash rationalizability" when no restrictions are imposed on the domain of the choice correspondence, and show that this problem is computationally very complex. Specifically, it is NP-complete even if there are only two players. In contrast, the analogous problem with one decision maker can be solved in polynomial time.

2 Characterization of Nash rationalizability

Suppose we observe a finite set I of players play different games, with *universal strategy spaces* $\mathcal{S}_i$. Let $\mathcal{S} := \prod_{i \in I} \mathcal{S}_i$. Let Λ be a finite set of game forms, i.e. a set of Cartesian product subsets of $\mathcal{S}$ (sometimes called *subforms* below). For each subform S in Λ, 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 non-empty valued *choice correspondence* $\mathfrak{C} : \Lambda \rightrightarrows \mathcal{S}$ with the property that $\mathfrak{C}(S) \subseteq S$ for all $S \in \Lambda$.

[3] 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.

Question 1 When can we find total, transitive and reflexive preferences $(\succsim_i)_{i\in I}$ 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 $(S, (\succsim_i)_{i\in I})$, i.e.

$$s^* \in \mathfrak{C}(S) \iff \forall i \in I, \forall s_i \in S_i, s^* \succsim_i (s_i, s^*_{-i})?$$

If we can find such preferences, we say that $\mathfrak{C}$ is *(pure strategy Nash equilibrium) rationalizable.*

To answer this question, define a *revealed preference relation* (Samuelson, 1938; Richter, 1966) for the individual choice problem where, for any $i \in I$, the allowable "budgets" are sets of the form $\{s \in S | \forall_{j\neq i} s_j = s'_j\}$ with $s' \in S$.

Define for each $i \in I$ a relation V_i on $\mathcal{S}$:

$$sV_is' \iff \exists_{S\in\Lambda} \left[s, s' \in S \text{ and } \forall_{j\in I\setminus\{i\}} s_j = s'_j \text{ and } s \in \mathfrak{C}(S)\right] \tag{1}$$

Let W_i be the transitive closure V_i, i.e. the *indirectly revealed preferred* relation.

Axiom 1 **[I-Congruence]**

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

This condition generalizes that in Richter (1966) to situations with several agents.

Previous authors have made the assumption that observations are complete in the sense that Λ contains *all* Cartesian product subsets of $\mathcal{S}$. For the theorem below, this assumption of "complete domain" is relaxed to "closed domain:"

<u>Definition 1</u> A class Λ 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 $s_{-i} \times S_i \in \Lambda$, 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.

<u>Theorem 1</u> Suppose Λ 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.

The theorem does not hold if Λ is allowed to be any arbitrary set of game forms.

<u>Proof</u>

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: $\mathfrak{C}(S) = \{s \in S | \forall_{i \in I} \forall_{s'_i \in S_i} s \succsim_i (s'_i, s_{-i})\}$. 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: $s^* W_i (s'_i, s_{-i})$ for all $s'_i \in S^*_i$, and yet $s^* \notin \mathfrak{C}(S^*)$. Since, under our initial supposition, for any $s, s' \in \mathcal{S}$, the relation $s V_i s'$ implies that $s \succsim_i s'$ and since $\succsim_i$ is transitive, $s^* W_i (s'_i, s_{-i})$ implies that $s^* \succsim_i (s'_i, s_{-i})$ 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 subform 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 Λ_i. For all $\hat{S} \in \Lambda_i$, let

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

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 Λ_i. Since I-Congruence restricted to the one-player games Λ_i is the same as the Congruence axiom of Richter (1966), for each $i \in I$ there exists a total, transitive, reflexive binary relation $\succsim_i$ 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 $\succsim_i$-maximal elements. We will show that these preferences (pure strategy Nash equilibrium) rationalize $\mathfrak{C}$ on Λ.

<u>$\mathfrak{C}(S)$ are Nash equilibria</u>: 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 $\succsim_i$ extends W_i (see Richter (1966)), this implies that for all $i \in I$, $s' \succsim_i s''$ for all $s'' \in S_i^{s'}$, i.e. that s' is a Nash equilibrium.

<u>all Nash equilibria are chosen by $\mathfrak{C}$</u>: Given $(S) \in \Lambda$, suppose that $s' \in S$ is a Nash equilibrium, i.e. for all $i \in I$, it holds that $s' \succsim_i s''$ for all $s'' \in S_i^{s'}$. Since $\succsim_i$ rationalizes $\mathfrak{C}_i = \mathfrak{C}$ on Λ_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, $s' \in \mathfrak{C}(S)$. *Q.E.D.*

3 Nash equilibrium rationalizability under arbitrary domains

In the previous section we substantially weakened the "complete domain" assumption of the previous literature to "closedness." In a revealed preference context even this assumption seems restrictive, and one would like to avoid any restrictions on the set of game forms observed. Formally this means that Λ could be an arbitrary set of game forms. While it is possible, even under arbitrary domains, to characterize Nash rationalizability using revealed preference conditions, we show in this section that this problem is NP-complete. That is, any algorithm that decides Nash rationalizability for choice correspondences with arbitrary domains is necessarily very complex computationally. This holds even if there are only two players. It is natural to ask whether the same complexity arises if there is only a single player. We show that it does not — the appropriate one player analogue of the Nash rationalizability problem can be decided in polynomial time.

In this section we assume that the universal strategy spaces $\mathcal{S}_i$ are finite. Let Λ be an arbitrary finite set of game forms, i.e. a set of Cartesian product subsets of[4] $\mathcal{S}$. For each subform $S \in \Lambda$, 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.[5] That is, we ask

[4] Recall that $\mathcal{S} := \prod_{i \in I} \mathcal{S}_i$.

[5] All the results below continue to hold if we ask for rationalizability by weak preferences.

Question 2 When can we find total, transitive and asymmetric preferences $(\prec_i)_{i\in I}$ 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 $(S, (\prec_i)_{i\in I})$, i.e.

$$s^* \in \mathfrak{C}(S) \iff \forall i \in I, \forall s_i \in S_i, (s_i, s^*_{-i}) \prec_i s^*?$$

If we can find such preferences, we say that $\mathfrak{C}$ is *(pure strategy Nash equilibrium) rationalizable.*

It is possible to characterize pure strategy Nash rationalizability for arbitrary domains (Galambos, 2004). 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 a 3×3 subform in Λ — involving only two players — is $\{s_1, s_2, s_2\} \times \{z_1, z_2, z_3\}$, and the strategy profile chosen by $\mathfrak{C}$ is (s_1, z_1). As in section 2, we then infer that if $\mathfrak{C}$ is to be Nash rationalized, it must be that

$$(s_2, z_1) \prec_1 (s_1, z_1) \text{ and } (s_3, z_1) \prec_1 (s_1, z_1) \tag{4}$$
$$(s_1, z_2) \prec_2 (s_1, z_1) \text{ and } (s_1, z_3) \prec_2 (s_1, z_1). \tag{5}$$

In addition, since (s_3, z_3) was not chosen, it must also be that

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

It is not surprising that deciding Nash rationalizability from a set of such statements 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.

<u>Theorem 2</u> The (pure strategy Nash equilibrum) rationalizability problem is NP-complete.

In fact, we prove a stronger statement: The (pure strategy Nash equilibrum) rationalizability problem is NP-complete *even if* we have only two players. The proof (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.[6]

3.1 Supra-semirationalizability

With only one player, the Nash rationalizability problem (as formulated in section 3) specializes to the classical revealed preference problem with finite budgets. It is not difficult to show then that rationalizability of a non-empty valued choice correspondence $\mathfrak{C} : \Lambda \rightrightarrows \mathcal{S}$ can be decided in polynomial time. One might conjecture that the complexity of the two-player problem does not arise with one decision maker because rationalizability can be characterized without having to use the disjunctive revealed preferred relations (like (6) above) that are derived from non-chosen points in the multi-player problem. Somewhat surprisingly, this is only partly true. The statement in (6) has disjunctions in two roles. The capitalized "OR" separates two statements, one of which relates to player 1 only, and the other to player 2 only. This disjunction corresponds, intuitively, to "player 1 would deviate at (s_3, z_3) **OR** player 2 would deviate at (s_3, z_3)." In each of those two statements, the occurrences of

[6] Specifically, we use 3SAT, a version of the satisfiability problem that was shown to be NP-complete in Cook (1971).

the small case "or" separate the different possible deviations of a particular player. While disjunctions of the first kind ("OR") clearly cannot occur with one player, it is possible to formulate the problem for one decision maker so that disjunctions of the second kind ("or") can occur. We present such a formulation below, and show that even this more general and intuitively more complex problem can be decided in polynomial time.

Suppose that one decision maker chooses from finite budgets $S \in \Lambda$. However, our observations of her choices are imperfect in the sense that her *actual* chosen set is known to be only a subset of the *observed* set $\mathfrak{C}(S)$.[7] To determine whether her choices could be generated by maximizing a preference relation, we must ask the question: Does there exist a preference relation $\prec$ on $\mathcal{S}$ such that for each budget $S \in \Lambda$ the $\prec$-maximal elements are *contained in* $\mathfrak{C}(S)$? This notion of rationality was labeled *supra-semirationality* by Matzkin and Richter (1991). Since we do not know whether an element $s \in \mathfrak{C}(S)$ is actually chosen or not, we can derive revealed preferred relations only from elements *not* in $\mathfrak{C}(S)$. We know that if $s \notin \mathfrak{C}(S)$, some element of $\mathfrak{C}(S)$ must be preferred to it. For example, if $S = \{s_1, s_2, s_3, s_4\}$ and $\mathfrak{C}(S) = \{s_2, s_4\}$, then supra-semirationalizability of $\mathfrak{C}$ requires

$$\begin{aligned} &s_1 \prec s_2 \ \textbf{or} \ s_1 \prec s_4, \\ &s_3 \prec s_2 \ \textbf{or} \ s_3 \prec s_4. \end{aligned} \tag{7}$$

More generally, our "data set" looks like

[7] For example, she might choose bundles consisting of three different goods, but we observe her choice only in two of those goods.

$$s_0^1 \notin \mathfrak{C}(\{s_0^1, s_1^1, \ldots, s_{k_1}^1\}) \tag{8}$$
$$s_0^2 \notin \mathfrak{C}(\{s_0^2, s_1^2, \ldots, s_{k_2}^2\}) \tag{9}$$
$$\vdots \tag{10}$$
$$s_0^m \notin \mathfrak{C}(\{s_0^m, s_1^m, \ldots, s_{k_m}^m\}), \tag{11}$$

where $s_r^q \in \mathcal{S}$ for $q = 1, \ldots, m$ and $r = 0, \ldots, k_q$. We write the set of direct revelations as

$$s_0^1 \lhd \{s_1^1, \ldots, s_{k_1}^1\} \tag{12}$$
$$s_0^2 \lhd \{s_1^2, \ldots, s_{k_2}^2\} \tag{13}$$
$$\vdots$$
$$s_0^m \lhd \{s_1^m, \ldots, s_{k_m}^m\}. \tag{14}$$

The interpretation is that the decision maker strictly prefers some element of $\{s_1^1, \ldots, s_{k_1}^1\}$ to s_0^1, and strictly prefers some element of $\{s_1^2, \ldots, s_{k_2}^2\}$ to s_0^2, etc. It is clear that supra-semirationalizability is equivalent to (12 – 14).

If every alternative appearing in one of the sets on the right hand side in (12–14) also appeared on the left hand side on some other line, it is clear that there would exist no rationalization. If there were one, the alternatives that appear in (12–14) would contain a preference cycle: s_0^1 would be worse than some alternative on the right in (12), which in turn would appear on the left on another line and so would be worse than another alternative, which in turn would appear on the left on another line, and so would be worse than . . . . With a finite number of alternatives, this would result in a preference cycle. For this reason, if a set of revelations has the property that all alternatives appearing on the right also appear on the left, we say that it is an *implicit cycle*. Thus it is a necessary condition for rationalizability that the set of revelations contain no implicit cycle. In separate work on supra-semirationalizability we show that this condition is also sufficient for rationalizability.

Remark 1 In an equivalent formulation using a payoff function u, the relations (12),(13), and (14) make a system of inequalities:

$$u(s_0^1) < u(s_1^1) \vee \cdots \vee u(s_{k_1}^1) \quad (15)$$
$$u(s_0^2) < u(s_1^2) \vee \cdots \vee u(s_{k_2}^2) \quad (16)$$
$$\vdots$$
$$u(s_0^m) < u(s_1^m) \vee \cdots \vee u(s_{k_m}^m), \quad (17)$$

where "$\vee$" denotes "supremum." Thus supra-semirationalizability is equivalent to the solvability of (15 – 17). If all alternatives that appear on the right in (15 – 17) also appear on the left, i.e.

$$\bigcup_{j=1}^{m} \{s_1^j, \ldots, s_{k_j}^j\} \subseteq \{s_0^1, s_0^2, \ldots, s_0^m\}, \quad (18)$$

it follows that

$$\bigvee_{j=1}^{m} u(s_1^j) \vee \cdots \vee u(s_{k_j}^j) < u(s_0^1) \vee u(s_0^2) \vee \cdots \vee u(s_0^m). \quad (19)$$

On the other hand, from (15),(16), and (17) it follows that

$$u(s_0^1) \vee u(s_0^2) \vee \cdots \vee u(s_0^m) < \bigvee_{j=1}^{m} u(s_1^j) \vee \cdots \vee u(s_{k_j}^j), \quad (20)$$

which is a contradiction, proving that implicit cycles contradict supra-semirationalizability.

In contrast to Nash rationalizability, the supra-semirationalizability (**SSR**) problem (involving one decision maker) is polynomial. Let $\mathcal{S}$ denote the set of alternatives, and suppose we observe for each "budget" S^i, $i = 1, \ldots k$ a set $\mathfrak{C}(S^i)$. Recall that this means that the *actual* chosen set is a subset of $\mathfrak{C}(S^i)$. Define an instance of SSR as a set of pairs of sets

$$\{(\mathfrak{C}(S^1), S^1 \backslash \mathfrak{C}(S^1)), (\mathfrak{C}(S^2), S^2 \backslash \mathfrak{C}(S^2)), \ldots, (\mathfrak{C}(S^k), S^k \backslash \mathfrak{C}(S^k))\}. \quad (21)$$

Such a list is a *yes*-instance if there exists a preference relation on $\mathcal{S}$ such that for each S^i, the set $\mathfrak{C}(S^i)$ contains the preference-maximal elements. Otherwise it is a *no*-instance.

THEOREM 3 The supra-semirationalizability problem can be decided in polynomial time.

Proof

The following algorithm determines in polynomial time whether an instance of SSR is a *yes*-instance or a *no*-instance. By "polynomial time" we mean, intuitively, that the number of steps in the algorithm is polynomial in the length of the input string ((21) above, with the finite sets $\mathfrak{C}(S^i)$ and $S^i \backslash \mathfrak{C}(S^i)$ written out element by element).

Algorithm:

1. Let $I := \{1, \ldots, k\}$. Let $Q = \emptyset$.
2. Let $q = 1$ and $r = 1$.
3. Scan the sets $S^i \backslash \mathfrak{C}(S^i)$, $i \in I$, to check if the qth element of $\mathfrak{C}(S^r)$ (denote it by x^*) appears in any of them.
   - 3.1. If it does:
     - 3.1.1. If $\mathfrak{C}(S^r)$ has q elements and r is the highest index in I, STOP.
     - 3.1.2. If $\mathfrak{C}(S^r)$ has q elements, let $q = 1$ and increase r by 1. Go to step 3.
     - 3.1.3. Increase q by 1 and go to step 3.
   - 3.2. If x^* does not appear in any $S^i \backslash \mathfrak{C}(S^i)$, add it to Q and set $x \prec x^*$ for all $x \in \mathcal{S} \backslash Q$.
   - 3.3. Scan the sets $\mathfrak{C}(S^i)$, $i \in I$ to check if x^* appears in any of them,

and let $I' := \{i \in I : x^* \notin \mathfrak{C}(S^i)\}$ (note that by the definition of x^* in step 3, $I' \subsetneq I$). If $I' = \emptyset$, set $I = \emptyset$ and STOP. (Note that the observations with labels in $I \backslash I'$ are now supra-semirationalized by $\prec$.)

3.4. Relabel the pairs $(\mathfrak{C}(S^i), S^i \backslash \mathfrak{C}(S^i)),\ \ i \in I'$, with the labels $1, \dots, |I'|$. Let $I := \{1, \dots, |I'|\}$. Go to step 2.

At every iteration the algorithm either returns to step 2 or 3 or it stops. The algorithm stops after at most $\sum_{i=1}^{k} |\mathfrak{C}(S_i)|$ iterations. If $I \neq \emptyset$ when the algorithm stops, the input choice correspondence is a *no-instance*. In this case $\{(\mathfrak{C}(S^i), S^i \backslash \mathfrak{C}(S^i))|\ i \in I\}$ is an implicit cycle, and by the argument on page 11 the input choice correspondence is not rationalizable. If $I = \emptyset$ when the algorithm stops, the input choice correspondence is a *yes-instance*. In this case $\prec$ is a partial order on $\mathcal{S}$ that supra-semirationalizes the input choice correspondence.[8] It is clear that each step is polynomial in the length of the input, and so is the number of iterations. *Q.E.D.*

4 Conclusion

We characterized behavior generated by the pure strategy Nash equilibrium concept for normal form games under a "closed domain" assumption. Our characterization is a straightforward extension of classical revealed preference theory to a multi-agent setting.

[8] This can be seen by noticing that indicies are ommitted from I in step 3.3 only if the corresponding observation is rationalized by the partial order defined so far.

Without restrictions on the domain of the choice correspondence, pure strategy Nash rationalizability is a computationally very complex problem. Specifically, it is an NP-complete problem *even if* the number of players is held fixed (and is two or more). In contrast, the one-player analogue of the problem, supra-semirationalizability, can be decided in polynomial time. This implies that the classical revealed preference problem (with finite budgets) and the Nash rationalizability problem under a “closed domain” can also be decided in polynomial time.

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 other solution concepts that are not “Nash-like,” 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[9] (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 [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

[9] I.e. the *same* two players are involved in every observed subform.

player 1 (in any subform 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 subform 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 subforms of $S \times Z$. For example, the following instance of NR2 encodes a choice function on two subforms.

$$(\{s_0, s_1, s_2\} \times \{z_0, z_1\}, s_2 z_1), (\{s_0, s_4, s_5\} \times \{z_0, z_2\}, s_4 z_0) \tag{A.1}$$

The first subform 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 subform–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.[10]

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 $T : X \to \{\mathfrak{t}, \mathfrak{f}\}$, 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 $\{C_1, \dots, C_k\}$ of clauses is given. A truth assignment $T : X \to \{\mathfrak{t}, \mathfrak{f}\}$ *satisfies* $\{C_1, \dots, C_k\}$ if for every clause C_i there exists

[10] 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.

$x \in C_i$ 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:

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

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 subform–outcome pair $(\{s_0, s_1, s_3, s^*\} \times \{z_0, z_2, z^*\}, s^*z^*)$. We will add two additional subform–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 subform–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 A.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 subform–outcome pairs:

$$\begin{aligned}
&(\{s_0, s_h, s_*\} \times \{z_0, z_h, z_*\},\ s_* z_*) \\
&(\{s_0\} \times \{z_h, z_*\},\ s_0 z_h) \\
&(\{s_0\} \times \{z_0, z_*\},\ s_0 z_0)\ ^{11} \\
&(\{s_h\} \times \{z_h, z_*\},\ s_h z_h) \\
&(\{s_h\} \times \{z_0, z_*\},\ s_h z_0) \\
&(\{s_h, s_*\} \times \{z_0\},\ s_h z_0) \\
&(\{s_0, s_*\} \times \{z_0\},\ s_0 z_0)\ ^{11} \\
&(\{s_h, s_*\} \times \{z_h\},\ s_h z_h) \\
&(\{s_0, s_*\} \times \{z_h\},\ s_0 z_h)
\end{aligned} \tag{A.3}$$

Figure A.1 illustrates this set of subform–outcome pairs. For transparency, the first pair in (A.3) is not shown (and, given the other eight subform–outcome pairs in the list, its rationalizability will depend only on orienting the edge cycle in Figure A.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 A.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

$$(\{s_0, s_e, s_f, s_g, s^*\} \times \{z_0, z^*\}, s^* z^*). \tag{A.4}$$

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

$$(\{s_0, s_e, s_f, s^*\} \times \{z_0, z_g, z^*\}, s^* z^*) \tag{A.5}$$

and Γ_g (see (A.3) for the definition of the nine subform–outcome pairs in

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

Γ_h for $h = 1, \ldots, k$).

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

$$(\{s_0, s_e, s^*\} \times \{z_0, z_f, z_g, z^*\}, s^* z^*) \tag{A.6}$$

and Γ_g and Γ_f.

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

$$(\{s_0, s^*\} \times \{z_0, z_e, z_f, z_g, z^*\}, s^* z^*) \tag{A.7}$$

and Γ_g, Γ_f and Γ_e.

5. Add the following subform–outcome pairs:

$$(\{s_0\} \times \{z_0, z^*\}, s_0 z_0), \ (\{s_0, s^*\} \times \{z_0\}, s_0 z_0). \tag{A.8}$$

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 $l \cdot 30$ subform–outcome pairs, i.e. the input size is increased by a multiplicative factor. The transformation involves only replacing each clause by at most 30 subform–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[12] $S_k := \{s_*, s^*, s_0, s_1, \ldots, s_k\}$ 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 $\prec_1$, and $\prec_2$. Define, for each variable x_i with $i \in$

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

$\{1, 2, \ldots, k\}$ (recall that these are exactly the variables that appear in V) a truth assignment:

$$T_{\prec}(x_i) = \mathfrak{t} \iff s_0 z_0 \prec_1 s_i z_0. \tag{A.9}$$

Consider a clause of the form $\{x_e, x_f, x_g\}$. Since NR_V contains (see (A.4) and (A.8))

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

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

$$[s_0 z_0 \prec_1 s_e z_0] \textbf{ or } [s_0 z_0 \prec_1 s_f z_0] \textbf{ or } [s_0 z_0 \prec_1 s_e z_0]. \tag{A.11}$$

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 Γ_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.*[13]

If $s_g z_0 \prec_1 s_0 z_0$, then by definition $T_{\prec}(x_g) = \mathfrak{f}$, 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 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 (A.5)), 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 $T_{\prec}(x_e) = \mathfrak{t}$ or $T_{\prec}(x_f) = \mathfrak{t}$, and so $\{x_e, x_f, \bar{x}_g\}$ is satisfied.

[13] 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 A.1 b) is oriented in one direction or the other.

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.[14] Then extensions of these orders to total orders will also rationalize NR_V. First we define player 1's preferences. The example in Figure A.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 $\prec_1$. (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 A.2.)
2. In the row $S_k \times \{z_0\}$ let (s^*, z_0) be the worst element under $\prec_1$.
3. For $z \in Z_k \backslash \{z_*, z^*, z_0\}$, 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).
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 $i \in \{1, 2, \ldots, k\}$ such that $T(x_i) = \mathfrak{t}$, let $s_0 z_0 \prec_1 s_i z_0$ and

$$(s_k, z_i) \prec_1 (s_{k-1}, z_i) \prec_1 \cdots \prec_1 (s_1, z_i) \prec_1 (s_0, z_i), \qquad \text{(A.12)}$$

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

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

$$(s_0, z_i) \prec_1 (s_1, z_i) \prec_1 \cdots \prec_1 (s_{k-1}, z_i) \prec_1 (s_k, z_i). \tag{A.13}$$

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 $i \in \{1, 2, \ldots, k\}$ such that $T(x_i) = \mathfrak{t}$, let $s_0 z_i \prec_2 s_0 z_0$ and

$$(s_i, z_0) \prec_2 (s_i, z_1) \prec_2 \cdots \prec_2 (s_i, z_{k-1}) \prec_2 (s_i, z_k), \tag{A.14}$$

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

$$(s_i, z_k) \prec_2 (s_i, z_{k-1}) \prec_2 \cdots \prec_2 (s_i, z_1) \prec_2 (s_i, z_0). \tag{A.15}$$

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 A.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 subform–

outcome pairs in NR_V. It is immediate that the sets of subform–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 $\prec_1$ or player 2 prefers to deviate under $\prec_2$). Checking that the other subform–outcome pairs (A.4–A.8) are also rationalized by $\prec_1$ and $\prec_2$ is also routine. For example, consider one of the type defined in (A.5): $(\{s_0, s_e, s_f, s^*\} \times \{z_0, z_g, z^*\}, s^*z^*)$. 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 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. *Q.E.D.*

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.

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.

Du, D.-Z., Ko, K.-I., 2000. Theory of Computational Complexity. Discrete Mathematics and Optimization. John Wiley & Sons, Inc.

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.

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

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

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.

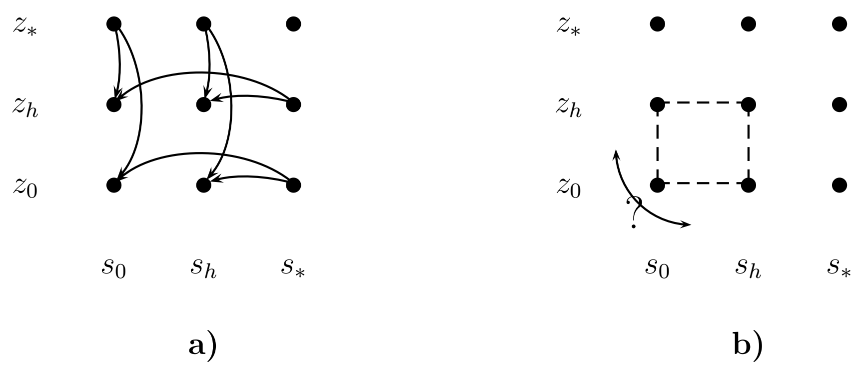

Fig. A.1. **a)** The subforms in Γ_h **b)** The "edge cycle" must be oriented for rationalizability (recall that these four points are not chosen in the first subform in (A.3))

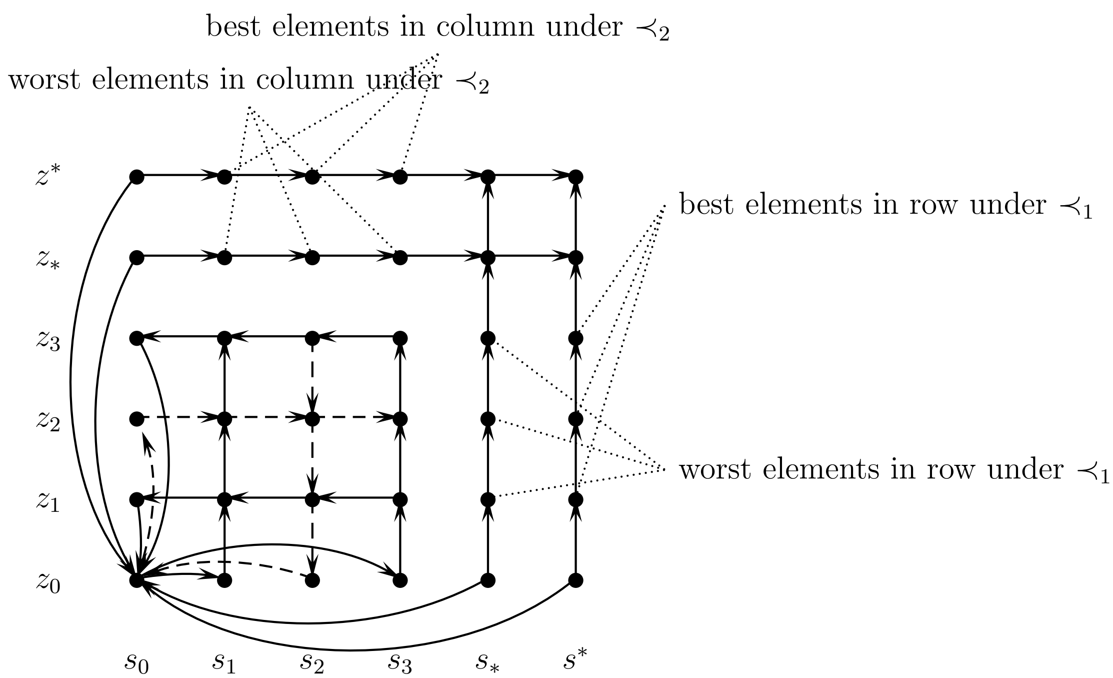

Fig. A.2. Rationalizing preferences for $T(x_1) = \mathfrak{t}, T(x_2) = \mathfrak{f}, T(x_3) = \mathfrak{t}$. The dashed line indicates the relations that arise from $T(x_2) = \mathfrak{f}$.