WINPEC Working Paper Series No. E2603
April 2026

Epistemological Bases for Nash and Rationalizability Theories of Prediction/Decision-Making in Games

Tai-Wei Hu Mamoru Kaneko

Waseda INstitute of Political EConomy
Waseda University
Tokyo, Japan

Epistemological Bases for Nash and Rationalizability Theories of Prediction/Decision-Making in Games*

Tai-Wei Hu† and Mamoru Kaneko‡

April 10, 2026

Abstract

We explore the conceptual bases of the Nash (noncooperative) theory. To emphasize the conceptual issues, we compare it to the rationalizability theory. Both theories target individual *ex ante* prediction/decision making in a game. We pinpoint their difference in the treatment of an inside player's prediction-making, with respect to the quantifiers "for all" and "for some", about the other's prediction/decision. We argue that the Nash theory follows the fundamental postulate that each player respects the other as an independent decision-maker, and that this independence is captured in the corresponding prediction/decision criterion. The rationalizability theory is based on a different conceptual base; the quantifier "for some" is highly speculative. We will discuss other conceptual bases of the Nash theory from various points of view, which suggest further developments along the line of the Nash theory.

Keywords: Nash equilibrium, Nash solution, Subsolution, Rationalizable Strategy, Prediction/decision Making, Theory of Mind

1 Introduction

Game theory studies human decision-making and resulting behavior in social situations, where individual decisions affect others. The literature have studied these interactions (cf., Maschler *et al.* [12]), but has not addressed the reasoning process from basic individual beliefs to behavior. Since the reasoning process takes place in people's minds, a study of this process encounters the basic problem of how a person treats another person's mind. We consider this problem from both epistemological and ontological perspectives. We restrict our consideration to the context of a 2-person game as it already contains the central characteristics of the problem.

We explore conceptual bases of two leading theories of how people make predictions and decisions in a game: the Nash noncooperative theory, abbreviated as the *Nash theory*, due to

*We thank for supports by JSPS Grant-in-Aids for Scientific Research 23K20588, 23K21869.

†Department of Economics, University of Bristol, UK, taiwei.hu@bristol.ac.uk

‡Waseda University, Tokyo, (mkanekoepi@waseda.jp), University of Tsukuba, Tsukuba, Japan

[15] and the theory of rationalizable strategies, *the rationalizabilty theory,* due to Bernheim [1] and Pearce [16]. These describe how each player makes a prediction/decision from the *one-shot* and *ex ante* view. We pinpoint their epistemological difference in the criteria of a player's prediction about the other's decision. The key distinction is on respect for the other player's independence as a decision-maker. In the former, each treats the other as independent with respect when forming predictions, while the latter is incompatible with this independence.

An equivalent formulation of rationalizable strategies known as "*iterated elimination of strictly dominated strategies (IESDS)*" is based on the negative form of individual predictions (cf., [1] and [16]). This clarifies the pinpointed difference from a different conceptual point of view. We focus mainly on the positive form of predictions, while giving some remarks on the negative form in suitable places.

1.1 Backgrounds

In the literature, a game theoretical solution is treated as a black box that maps a game to a set of resulting outcomes satisfying "plausible" properties on it; "rationality" is included in "plausible" properties. The rationality concept involved here is classified to "*substantive rationality*", i.e., an attribute of an outcome, due to Simon [19]. He also proposed another notion, "*procedural rationality*", i.e., an attribute of a process or procedure. The Nash theory and rationalizability theory include not only the resulting outcomes but also the interactive decision-making process. Rationality involved in such processes could be classified as "procedural rationality". Because we study the logical reasoning process from basic beliefs conducted by a player, our target is more specific than "procedural rationality"; we call it "*individual cognitive rationality*", illustrated in Figure 1. This will be discussed in Section 4.[1, 2]

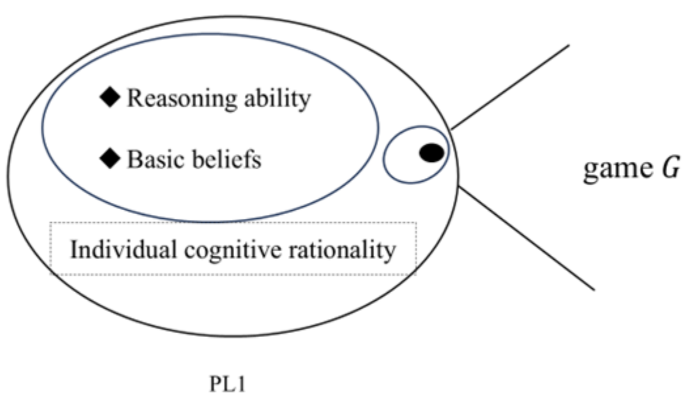

Figure 1:

[1]A similar idea is discussed in Grimm and Richter [6], from the experimental point of view.

[2]See Sent [17] for a survey of papers on rationality in the game theory literature.

The concept of Nash equilibrium, abbreviated as NE, permits different interpretations. For example, it may describe a stationary stable state in a recurrent situation, which is typical in general equilibrium theory and or Cournot equilibrium theory in economics.[3] In contrast, despite the fact that Nash himself gave no interpretational discussions in [15], his various mathematical developments on NE such as solvability, solutions, and subsolutions, lead to its interpretation in terms of individual *ex ante* prediction/decision making in one-shot situation.

The rationalizability theory is typically regarded as a faithful description of individual *ex ante* decision making in games; this view is common in standard game theory/micro-economics textbooks. For example, Mas-Colell *et al.* [13], p.243 wrote "*The set of rationalizable strategies consists precisely of those strategies that may be played in a game where the structure of the game and the player's rationality are common knowledge among the players.*" This theory is regarded as a good reference point to be compared with the Nash theory.

We argue that the key difference between the Nash theory and rationalizabilty theory is in the quantifiers, "for all" and "for some", on predictions about the other's decision. This difference implies that the Nash theory faithfully follows the principle of *methodological individualism* which claims that an individual person is the true base for action-taking (cf., von Mises [22], Chapter II, Section 5), while the rationalizabilty theory deviates from it. After all, we take the rationalizability theory as an item to be compared, because these considerations manifest conceptual bases of the Nash theory.

1.2 Prediction/decision-criteria

To understand the difference mentioned above, consider a 1-person game and a 2-person game. The left of Figure 2 describes the decision-making in the 1-person game G^1, where player 1's decision-making is determined by his objective to maximize his payoffs. As long as the components of G^1 are known to him, the decision-making process is straightforward. Here, a basic postulate is that the *OT (outside theorist)* is objective and treats/respects player 1 as an independent decision-maker, and the OT is not involved in 1's decision-making.

On the other hand, the game G^2 in the right includes two players. Conceptually, the game G^2 differs in that each player notices the presence of the other. A basic methodological postulate is that each player, say i, treats/respects the other, j, as an independent player reciprocally treating/respecting i. The OT recognizes this symmetry regarding the two players' independence and their interaction, but he himself is not involved in either player's prediction/decision process.

The 1-person and 2-person games differ only in the presence of player 2 in G^2, but this makes the situation critically differ. In G^2, because each treats the other as an independent player

[3] Johansen [8] gave discussions on various interpretations of Nash equilibrium in the economics literature. Since the publication of his paper in 1982, a lot of mathematical developments have occurred, but the foundational/conceptual issues about belief-formation in one-shot games have remained more or less the same until now, 2026.

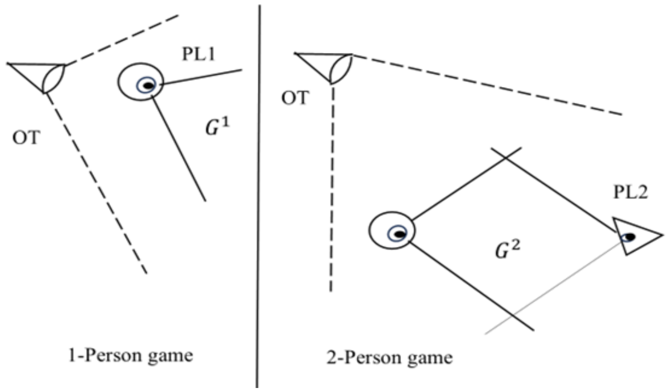

Figure 2:

and because the situation is interactive, he needs to *predict* what the other would choose. We postulate that one predicts what the other would choose by simulating the other's mind and then he maximizes his payoff against the prediction. Now, the question is how one makes prediction about the other. In this paper, we take and study the Nash theory and rationalizability theory to explore possible answers to this question.

We formulate these theories in terms of the prediction/decision criteria in a unified manner. For the Nash theory, the criterion is given as follows:

N1^o: player 1 chooses his best action for all of his predictions of player 2's possible decisions based on N2^o;

N2^o: player 2 chooses his best action for all of his predictions of player 1's possible decisions based on N1^o.

In N1^o, 1's choice takes all of his predictions, and N2^o is symmetric. This means that each respects the other's independent choice in his prediction-making. Here, N1^o and N2^o form circularity; for player 1 to make a decision based on N1^o, it requires a prediction about 2's possible decisions to be determined by N2^o, which requires predictions based on N1^o. Because of this circularity, these requirements form an infinite regress depicted in Table 1.1:

N1^o		N1^o		$\cdots$
$\downarrow$	$\nearrow$	$\downarrow$	$\nearrow$	$\downarrow$
N2^o		N2^o		

N1^o
$\downarrow\uparrow$
N2^o

Table 1: Table 1.1 Table 1.2

Here, N1^o occurs in 1's mind to make his decision, while for his prediction, he needs to take 2's perspective and N2^o. Here, N1^o and N2^o occur in different epistemic scopes. However, this distinction does not appear in the above formulation, because N1^o-N2^o are formulated in a purely set-theoretical framework. Thus, the above infinite regress is reduced to circularity, i.e., a form of fixed-point, as in Table 1.2. Nevertheless, the distinction can be explicitly captured in an epistemic logic framework such as in Hu *et al.* [5]. This will be discussed in Section 4.

The rationalizability theory is characterized by another prediction/decision criterion:

R1^o: player 1 chooses his best action for some of his predictions of player 2's choice based on R2^o;

R2^o: player 2 chooses his best action for some of his predictions of player 1's choice based on R1^o.

Thus, R1^o-R2^o are parallel except for the quantifiers before how a player treats his predictions; the quantifier "for all" for N1^o-N2^o in substituted by "for some" in R1^o-R2^o.[4] The point is how this difference together with the resulting outcomes should be interpreted.

Both N1^o-N1^o and R1^o-R2^o show circularity. The mathematical treatment in a formal epistemic framework only requires a straightforward extension of criterion R1^o-R2^o for the rationalizability theory, while the Nash theory requires much more, to be specified in Section 2. Nevertheless, the conceptual base for R1^o-R2^o is less secure than that for N1^o-N2^o.

Finally, consider the conceptual bases of these theories from the perspective of the *individual cognitive rationality*. In both theories, the players are assumed to understand their own payoff functions and to have sufficient cognitive abilities for logical inferences. The Nash theory is founded on mutual beliefs about these and the assumption that the players can perform interpersonal inferences.

When a game has a unique NE (or satisfies *interchangeability* condition to be mentioned in Section 2), those basic beliefs are sufficient for players to reach a definite set of decisions and predictions, which consist of all NE strategies. For unsolvable games, players who follow the Nash theory would have difficulty in reaching definite decisions. However, we do not regard this difficulty as a failure of the Nash theory. Our characterization gives a boundary of strategic situations where basic beliefs about payoffs and logical abilities are sufficient for strategic decisions; outside this boundary, additional ingredients such as social norms are needed.

For the rationalizability theory, on the contrary, players can always reach a set of possible decisions and predictions. However, we meet some difficulty in interpreting this set of decisions due to the quantifier "for some". This means that when predicting the other player's decision, the player is assumed to pick an arbitrary decision among all possible ones without any reasoning. Also, he assumes that the other player does the same, and this arbitrariness is preserved in all epistemic layers. We argue that this is incoherent with our notion of individual cognitive rationality,

Our conclusion is that we provide a perspective regarding the two theories with respect to their connections to rationality, which differs from the typical attitude quoted above from Mas-Colell *et al.* [13].

---

[4] R1^o-R2^o are close to the "best-response property" in Bernheim [1] and Pearce [16].

The paper is written as follows: Section 2 introduces the concepts of NE and rationalizable strategies. Section 3 formulates N1^o-N2^o as the Nash theory and R1^o-R2^o as the rationalizability theory. We give various theorems characterizing these theories. In Section 4, we discuss further foundational issues from N1^o-N2^o. Section 5 concludes the paper with a summary of these discussions.

2 Preliminaries

We begin with basic concepts in a finite 2-person game. Let $G = (N, \{S_i\}_{i\in N}, \{h_i\}_{i\in N})$ be a finite 2-person game, where $N = \{1, 2\}$ is the set of players, S_i is the finite set of pure strategies and $h_i : S_1 \times S_2 \to R$ is the payoff function for player $i \in N$. We assume $S_1 \cap S_2 = \emptyset$. When we take one player $i \in N$, the remaining player is denoted by j. Also, we write $h_i(s_i; s_j)$ for $h_i(s_1, s_2)$. The property that s_i is a *best-response* to s_j, i.e.,

$$h_i(s_i; s_j) \geq h_i(s_i'; s_j) \text{ for all } s_i' \in S_i, \tag{1}$$

is denoted by $\text{Bst}_i(s_i; s_j)$. Since $S_1 \cap S_2 = \emptyset$, the expression $\text{Bst}_i(s_i; s_j)$ has no ambiguity. A pair of strategies (s_1, s_2) is a *Nash equilibrium,* abbreviated as NE*,* in G iff $\text{Bst}_i(s_i; s_j)$ holds for $i = 1, 2$. We use $E(G)$ to denote the set of all NE's in G. The set $E(G)$ may be empty; it is empty in Table 2.2 (Matching Pennies). We call $s_i \in S_i$ as a *Nash strategy* iff $(s_i; s_j)$ is a NE for some $s_j \in S_j$.

Nash (noncooperative) solutions: The main part of Nash's [15] theory starts with the concept of interchangeability.[5] A subset E of $S_1 \times S_2$ is *interchangeable* iff

$$(s_1, s_2), (s_1', s_2') \in E \text{ imply } (s_1, s_2') \in E. \tag{2}$$

This is known to be equivalent for E to have the product form.

Lemma 2.1 Let $E \subseteq S_1 \times S_2$ and let $E_i = \{s_i : (s_i; s_j) \in E \text{ for some } s_j \in S_j\}$ for $i = 1, 2$. Then, E satisfies (2) if and only if $E = E_1 \times E_2$.

We say that E is the *Nash solution* iff E is a nonempty set of NE's and is the greatest set satisfying (2), i.e., E satisfies (2) and $E' \subseteq E$ for any E' with (2). When the Nash solution exists for game G, G is called *solvable.* In a solvable game G, the Nash solution E is the the set $E(G)$ of all NE's. Table 2.1 (Prisoner's dilemma) has the unique NE (s_{12}, s_{22}), which forms the Nash solution $E = \{(s_{12}, s_{22})\}$. Since Table 2.2 has no NE's, it has no Nash solution.

A game G may be *unsolvable* for two cases: either $E(G) = \emptyset$ or $E(G)$ is nonempty but violates (2). Nash [15] gave the concept of a subsolution: For a game G with $E(G) \neq \emptyset$, a subset E^o of $E(G)$ is a *subsolution* iff E^o is a maximal subset of $E(G)$ and satisfies condition (2), where E^o is maximal iff there is no subsolution E' with $E^o \subsetneq E'$. When $(s_1, s_2) \in E(G)$, there is a

[5] Jansen [7] studied the mathematical structure of the solution and subsolutions.

	s_{21}	s_{22}
s_{11}	$(5,5)$	$(1,6)$
s_{12}	$(6,1)$	$(3,3)^{NE}$

	s_{21}	s_{22}
s_{11}	$(1,-1)$	$(-1,1)$
s_{12}	$(-1,1)$	$(1,-1)$

	s_{21}	s_{22}
s_{11}	$(2,2)^{NE}$	$(0,0)$
s_{12}	$(0,0)$	$(1,1)^{NE}$

Table 2: Table 2.1 Table 2.2 Table 2.3

subsolution E^o containing (s_1, s_2). This E^o may not be unique: Table 2.3 is unsolvable and has two subsolutions: $\{(s_{11}, s_{21})\}$ and $\{(s_{12}, s_{22})\}$.

In Section 3, it will be argued that the Nash solution can be regarded as a theory of *ex ante* prediction/decision-making in games.

Rationalizable Strategies: The pure strategy version is known as *point-rationalizability*. Following Bernheim [1], we give the iterative definition of rationalizability. A sequence of pairs of strategy sets, $\{(R_1^\nu(G), R_2^\nu(G))\}_{\nu=0}^\infty$, is inductively defined as follows: for $i = 1, 2$, and

$$\begin{aligned} &R_i^0(G) = S_i; \\ &R_i^\nu(G) = \{s_i \in S_i : \mathrm{Bst}_i(s_i; s_j) \text{ holds for some } s_j \in R_j^{\nu-1}(G)\} \text{ for any } \nu \geq 1. \end{aligned} \tag{3}$$

We obtain rationalizable strategies by taking the intersection of these sets, i.e., $R_i(G) = \bigcap_{\nu=0}^\infty R_i^\nu(G)$ for $i = 1, 2$; a pure strategy $s_i \in S_i$ is *rationalizable* iff $s_i \in R_i(G)$.

It is an immediate result that for any 2-person finite game G and $i \in N$,

$$\text{if } (s_1, s_2) \text{ is a NE , then each strategy } s_i \text{ is rationalizable.} \tag{4}$$

Although a NE does not necessarily exist, a rationalizable strategy always does. We report two more properties related to rationalizability. Let $\{(R_1^\nu(G), R_2^\nu(G))\}_{\nu=0}^\infty$ be defined by (3). Then, these sequences are monotonically decreasing: for all $i = 1, 2$ and $\nu \geq 0$,

$$R_i^\nu(G) \supseteq R_i^{\nu+1}(G). \tag{5}$$

The existence of a rationalizable strategy is one consequence; for any finite game G and $i \in N$,

$$\text{player } i \text{ has at least one rationalizable strategy.} \tag{6}$$

When game G has a NE, (6) follows from (4). Proofs of (5) and (6) are given in Section 5 to be self-contained.

Figure 3 describes the set-theoretical inclusions of the Nash strategies by the rationalizable strategies. The leftmost case is a typical case, as stated in (4). In the middle case, there is a rationalizable strategy as stated by (6) but there is no Nash strategy, with Table 2.2 as an example. The rightmost case of Figure 3 illustrates the Nash and rationalizable strategies in Table 3; the set of rationalizable strategies coincides with the entire set $S_1 \cup S_2$, but this game has the unique NE. In this game, (s_{11}, s_{21}) is the NE; so, s_{11}, s_{21} are rationalizable by (4). The matrix determined by $\{s_{12}, s_{13}\}$ and $\{s_{22}, s_{23}\}$ is the same as Table 2.2, which form the cycle

	s_{21}	s_{22}	s_{23}
s_{11}	$(5,5)^{NE}$	$(-2,-2)$	$(-2,-2)$
s_{12}	$(-2,-2)$	$(1,-1) \longrightarrow$	$\downarrow (-1,1)$
s_{13}	$(-2,-2)$	$\uparrow (-1,1)$	$\longleftarrow (1,-1)$

Table 3:

described by the small arrows with respect to best responses. For this reason, this part has no NE's, but it implies that s_{12}, s_{13} and s_{22}, s_{23} are all rationalizable.

Economics has a tradition to regard predictability as a virtue of a theory. Neither Nash nor rationalizable strategies grants good predictability. However, conceptual bases of a theory is prior to predictability. Once we find some solution concepts to be well accepted with its conceptual bases, we consider predictability as a further possible criterion.

3 Comparisons of the Nash and Rationalizabilty Theories

Here, we give a precise formulation of prediction/decision-making in each of those theories, and pinpoint the difference between them. Then, we argue that the Nash theory is compatible with methodological individualism with respect for the other player's independence as a decision-maker. However, the rationalizabilty theory suffers from a serious difficulty with the treatment of the other's mind.

3.1 The Nash noncooperative theory

The decision criterion N1^o-N2^o was given informally in Section 1. Here, we formulate it in a set-theoretical manner, by introducing E_1 and E_2 as the candidate sets of possible decisions for

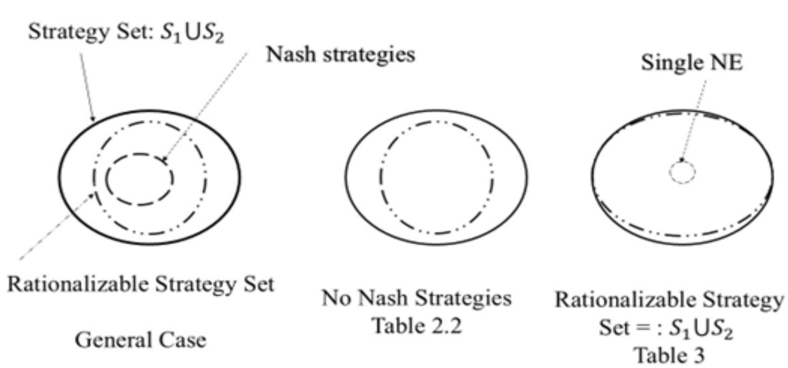

Figure 3:

players 1 and 2, as follows:

N1: for each $s_1 \in E_1$, $\mathrm{Bst}_1(s_1; s_2)$ holds for all $s_2 \in E_2$;

N2: for each $s_2 \in E_2$, $\mathrm{Bst}_2(s_2; s_1)$ holds for all $s_1 \in E_1$.

The candidate sets E_1 and E_2 include *all* possible decisions that satisfy N1-N2. In N1, from 1's perspective, E_1 describes his possible decisions, while E_2 does his predictions. To respect 2's independent choice, 1 needs to consider the (set-theoretically) greatest set E_2. We consider the greatest sets E_1 and E_2 satisfying N1-N2. For a solvable game G, we show that the players can reach definite sets of possible decisions using N1-N2. For an unsolvable game, we encounter some difficulty and need some other considerations.

The criterion N1-N2 is viewed from the perspective of the OT. The quantifier "for each $s_1 \in E_1$" in the front of N1 means that the OT considers all strategies satisfying N1 as a possible decision for player 1. Thus, player 1 takes the same perspective as the OT regarding player 2's mind, that is, 1 respects 2's independent decision-making. N2 is described from the OT's viewpoint as well as from player 2's viewpoint. This argument generates an infinite regress, described in Table 1.1. As mentioned in Section 1, the infinite regress is reduced into the circularity in Table 1.2, since "decision" and "prediction" are distinguished only in interpretation. In Section 4, we will briefly mention this distinction in a formal manner in the epistemic logic approach in order to treat the infinite regress without the reduction.

First, we give the following lemma.

Lemma 3.1. Let E_i be a nonempty subset of S_i for $i = 1, 2$. Then, (E_1, E_2) satisfies N1-N2 if and only if any $(s_1, s_2) \in E_1 \times E_2$ is a NE in G.

It is an immediate consequence that G has a NE if and only if there is a nonempty pair (E_1, E_2) satisfying N1-N2. Lemma 3.1 is regarded as a statement about an equation with unknown E_1 and E_2. For example, for any NE (s_1, s_2), the pair $(E_1, E_2) = (\{s_1\}, \{s_2\})$ is a solution of the equation. However, the intention of N1-N2 is to describe each player's logical thinking treating the other as an independent thinker, as the right in Figure 2. Thus, we think about the greatest sets E_1 and E_2 satisfying N1-N2.

Theorem 3.1 states that the requirement is fulfilled when the game G is a solvable game. A proof is given in Section 5.

Theorem 3.1 (N1-N2 for a Solvable Game): Let G be a solvable game. Then, the greatest pair (E_1, E_2) satisfying N1-N2 exists and $E = E_1 \times E_2$ is the Nash solution $E(G)$.

The *individual cognitive rationality* in Figure 1 implies that the player can make the same inferences as the OT and can derive Theorem 3.1. This is interpreted as three steps. Step (i) finds the set of all NE's; this needs payoff comparisons, together with recording the calculation results. Step (ii) verifies the set to satisfy interchangeability (2). Step (iii) establishes that $E(G)$ is the greatest set satisfying N1-N2. The steps may take places simultaneously; (ii) and (iii) require comparison of sets of NE's with the inclusion relation $\supseteq$. The resulting greatest (E_1, E_2) involves the entirety of the targets required in the *individual cognitive rationality.* Since G is a finite game, these steps can be done with the finite set theory. This remark is relevant in comparisons with Johansen's [8] approach to be discussed in Section 4.

Theorem 3.1 is extended into the following form for an unsolvable game G.

Theorem 3.2 Suppose that G is unsolvable but $E(G) \neq \emptyset$. Then E is a subsolution if and only if (E_1, E_2) is a maximal pair satisfying N1-N2.

Theorem 3.2 differs from Theorem 3.1 in that a player meets multiplicity of the maximal pairs (E_1, E_2) satisfying N1-N2. In Table 2.3, both $\{(s_{11}, s_{21})\}$ and $\{(s_{12}, s_{22})\}$ are maximal pairs. Either player meets difficulty in choosing one maximal pair from $\{(s_{11}, s_{21})\}$ and $\{(s_{12}, s_{22})\}$; it is a failure of individual *ex ante* prediction/decision-making.

One possible way out is to allow the player's belief to include some additional information to indicate a specific subsolution. Let $s^* = (s_1^*, s_2^*)$ be a pair of strategies in game G, which we call the *identifier.* We modify N1-N2 by adding the identifier s^* as follows:

N1(s^*): [for each $s_1 \in E_1$, $\mathrm{Bst}_1(s_1; s_2)$ holds for all $s_2 \in E_2$] $\wedge$ ($s_2^* \in E_2$);

N2(s^*): [for each $s_2 \in E_2$, $\mathrm{Bst}_2(s_2; s_1)$ holds for all $s_1 \in E_1$] $\wedge$ ($s_1^* \in E_1$).

In N1(s^*), the square bracket is the same as N1 but the additional ($s_2^* \in E_2$) requires the prediction to contain the identifier s^*; the subsolution is suggested by the identifier s^*. N2(s^*) is symmetric. The identifier $s^* = (s_1^*, s_2^*)$ is the additional information to each player to determine which subsolution could be adopted. The following theorem states that the identifier $s^* = (s_1^*, s_2^*)$ determines a unique subsolution $E_1 \times E_2$ if and only if N1(s^*)-N2(s^*) work in the same way as Theorem 3.1.

Theorem 3.3 (Unique subsolution determined by N1(s^*)-N2(s^*)): Suppose that G is unsolvable with $s^* \in E(G)$. Then, (E_1, E_2) is the greatest pair satisfying N1(s^*)-N2(s^*) if and only if $E_1 \times E_2$ is a subsolution with $s^* \in E_1 \times E_2$ but $s^* \notin E_1' \times E_2'$ for any other subsolution $E_1' \times E_2'$ in G.

In order for the pair (E_1, E_2) in the above theorem to work in the same way as Theorem 3.1, we need $s^* = (s_1^*, s_2^*)$ to identify the subsolution $E_1 \times E_2$ uniquely. Otherwise, the players meet multiplicity of candidate subsolutions. When a 2×2 game G has a NE, it has at least one s^* to identify each subsolution uniquely. Unfortunately, this does not necessarily hold in an arbitrarily given game. Table 4 is a counterexample; it has 6 subsolutions, and each NE is contained in two subsolutions.

Theorem 3.3 permits two interpretations of additional information. The first is: the latter statement may be derived from some additional property of a given game, such as the Pareto dominance among the NE's. For example, a finite game of strategic complementarity (or super modularity) has a NE in pure strategies, and under some condition, that if it has multiple

	s_{21}	s_{22}	s_{23}
s_{11}	$(1,1)^{NE}$	$(1,1)^{NE}$	$(0,0)$
s_{12}	$(0,0)$	$(1,1)^{NE}$	$(1,1)^{NE}$
s_{13}	$(1,1)^{NE}$	$(0,0)$	$(1,1)^{NE}$

Table 4:

equilibria, they are Pareto-ranked (see Vives [21] for a survey of this theory and its applications). Here, the identifier s^* is the strategy pair which Pareto-dominates the other NE's. The additional ($s_2^* \in E_2$) of N1(s^*) means that 1 predicts this dominating judgement.

The second interpretation is that the situation is occurring in a broader recurrent context, and the players learn the additional information s^* as saliently appearing to the players, and they take the identifier s^* for his prediction/decision-making. This interpretation resembles the focal point due to Schelling [18] with the basic idea that "people can often concert their intentions or expectations with others if each knows that the other is trying to do the same".

3.2 Rationalizable Strategies

The criterion, R1^o-R2^o, for rationalizability is formulated in a more accurate manner: for nonempty $E_1 \times E_2 \subseteq S_1 \times S_2$,

R1^A: for all $s_1 \in E_1$, $\mathrm{Bst}_1(s_1; s_2)$ holds for some $s_2 \in E_2$;

R2^A: for all $s_2 \in E_2$, $\mathrm{Bst}_2(s_2; s_1)$ holds for some $s_1 \in E_1$.

This differs from N1-N2 only in that the two occurrences "for all" in the ends of N1 and N2 are replaced by "for some". Nevertheless, we keep the "for all" in the front of each of R1^A-R2^A. In fact, R1^A-R2^A is the pure-strategy version of the BP-property given by Bernheim [1] and Pearce [16]. The greatest pair (E_1, E_2) satisfying R1^A-R2^A coincides with the sets of rationalizable strategies $(R_1(G), R_2(G))$. A more general version of the following theorem is reported in Bernheim [1], Proposition 3.1; we include the proof for self-containment in Section 5.

Theorem 3.4 (Rationalizability): $(R_1(G), R_2(G))$ is the greatest pair (E_1, E_2) satisfying R1^A-R2^A.

Theorem 3.4 and Lemma 2.1 imply that the greatest pair satisfying R1^A-R2^A exists and consists of the sets of rationalizable strategies. Interchangeability is automatically satisfied by construction. In this respect, the rationalizability theory may appear advantageous to the Nash theory in that it avoids the identification of each subsolution raised by Theorem 3.3. In fact, we reverse this view for the following reasons.

Consider the symmetric treatment between the OT and each of the players. In the case, R1^A-R2^A are replaced by the following R1^S-R2^S, that is, for nonempty $E_1 \times E_2 \subseteq S_1 \times S_2$,

R1^S: for some $s_1 \in E_1$, $\mathrm{Bst}_1(s_1; s_2)$ holds for some $s_2 \in E_2$;

$R2^S$: for some $s_2 \in E_2$, $\mathrm{Bst}_2(s_2; s_1)$ holds for some $s_1 \in E_1$.

Here, the quantifier "for all" in the front of $R1^A$-$R2^A$ is replaced by "for some". This version is entirely different form the theory formulated by $R1^A$-$R2^A$. In fact, each of $R1^S$ and $R2^S$ is equivalent to

$$\mathrm{Bst}_i(s_i; s_j) \text{ holds for some } s_i \in E_i \text{ and } s_j \in E_j, \tag{7}$$

for $i = 1, 2$. These give no direct connection between $R1^S$ and $R2^S$, which does not generate the regress in Table 1.1. In fact, the situation is much simpler; when $E_i = S_i$ and $E_j = S_j$, (7) holds automatically for each $i = 1, 2$, since the best response property itself is "for any $s_j \in E_j$, $\mathrm{Bst}_i(s_i; s_j)$ holds for some $s_i \in E_i$". Thus, the rationalizability theory does not follow the symmetric treatment between the OT and each player.

Finally, consider the formulation IESDS of the rationalizability theory. It treats the negative form of prediction, as said in Section 1. A player eliminates what could not be a decision for the other player, provided that the other takes the parallel process focusing on the remaining strategies and repeats the process until no further elimination is possible. The remaining strategies are rationalizable strategies. It is based on the negative concept of "prediction" and also the candidates are changing in the process. The basic concept totally differs from the Nash theory N1-N2; it deviates from the requirement that prediction is based on the same principle of one's decision-making. The difficulty is applied to the positive form of the rationalizability theory, too, because the positive and negative forms are mathematically equivalent with respect to the resulting outcomes.

4 Further Discussions on the Nash Theory

In the above study of the conceptual bases of the Nash theory and the rationalizability theory, we focus on the treatment of the prediction/decision criteria using set-theoretical arguments to pinpoint their differences. However, this formulation, although convenient for readers familiar with typical mathematical models in economics/game theory, is not suitable to discuss two important issues. First, it does not allow for an explicit distinction between decision and prediction. Second, the issue of rationality can only be implicitly touched without an explicit treatment of its ontological nature. We discuss these issues based on Johansen's [8] argument for NE. Finally, we touch a different but related issue, that the Nash theory was treated as an isolated theory in the form of Section 3.1; we discuss this from the methodological individualism and singularism, based on von Mises [22].

4.1 Johansen's consideration of the Nash equilibrium concept

Johansen [8] regarded NE as natural in economic research and wanted to give suitable bases. He formulated five postulates, J1 to J5, for prediction/decision-making in a game G with a unique

NE. He claimed, in p.435, that those postulates determine the NE to be a pair of one's decision and prediction about the other's decision in G. He treated his postulates and claims as if they constitute a mathematical argument, but they are not rigorous according to the present standard of mathematics/logic. Nevertheless, his intentions and arguments contain valuable insights for research in social sciences and deserve to be scrutinized.

Indeed, if the term "*rational*" in his postulates is interpreted in terms of the *individual cognitive rationality* mentioned in Section 1.1, his postulates J1 to J4 can be well interpreted along the line from N1-N2 to Theorem 3.1. His last postulate J5 is an *ex post* evaluation of the outcome from J1 to J4, and we come back to J5 in a later subsection. Now, his four postulates are as follows, provided that the game has a unique NE.[6]

Postulate J1 A player makes his decision $s_i \in S_i$ on the basis of, and only on the basis of information concerning the action possibility sets of two players S_1, S_2 and their payoff functions h_1, h_2.

Postulate J2 In choosing his own decision, a player assumes that the other is rational in the same way as he himself is rational.

Postulate J3 If some decision is a rational decision to make for an individual player, then this decision can be correctly predicted by the other player.

Postulate J4 Being able to predict the actions to be taken by the other player, a player's own decision maximizes his payoff function corresponding to the predicted actions of the other player.

Postulate J1 is basic to the other postulates. The difficulty is that J2 and J3 contain the key term "*rational*" but it is not explicitly defined. In J2, it is an attribute of the players, which, as we will argue below, corresponds to the *individual cognitive rationality.* On the other hand, it becomes an attribute of a decision in J3; it is Simon's [19] "*substantive rationality"*. Johansen explained his intention by the term "*rational*" only in paragraphs around the postulates. J4 is connected to J3 through "correctly predicted" with payoff maximization to "*rationality*". Thus, these postulates as a whole have various ambiguous components. Incidentally, "If some decision" in the beginning of J3 must be corrected to "If any decision", following the standard mathematics/logic.

Johansen wrote suggestively in p.433 that rationality is used to analyze the game situation, which means that each player has logical ability together with individual basic beliefs. Also, he wrote that these abilities are assumed equally good for the players. It is relevant here that Johansen makes a clear distinction between one's decision and his prediction on the other's decision, which is the key for a player to think about the other's mind. Nevertheless, we remark that the formulation of N1-N2 and Theorem 3.1 has no explicit distinction either; so, Table 1.1 is reduced to the circularity in Table 1.2.

[6]He pointed out (p.437) that interchangeability is sufficient for his assertion.

4.2 Individual cognitive rationality

Here, we look at the "*individual cognitive rationality*" in the context of N1-N2 to Theorem 3.1. Its main part is the logical ability of a player formulated as the classical logic. Then, the basic beliefs describe what he adopts as the source for his deduction. In this treatment, decision and prediction are distinguished still only in the contextual interpretations. To make explicit distinctions, we need to introduce individual belief operators and discuss briefly an epistemic logic.

Logical treatments of N1-N2: The concept of "*Individual cognitive rationality*" consists of (i) *reasoning ability* and (ii) *basic beliefs*, which are in the mind of each player.

***(i)* Reasoning ability**: It starts with the game theoretical primitives with logical connectives, $\wedge, \Longrightarrow, \neg$, and the conjunctions of finite sets of actions and of pairs of actions are allowed to be formulae. The set of all possible formulae is defined by mathematical induction. The classical logic axioms (schemata), inference rules, and the concepts of proofs are given (see, e.g., Chellas [3], Kaneko [9]).[7] These form a description of the *reasoning ability* of each player.

***(ii)* Basic beliefs**: In the case of Theorem 3.1, we adopt two *basic beliefs*, $H_{1,2}$ and $Gr(E^*)$. The first is about the players' *preference relations*, $p_i(s;s')$, determined by the payoff functions. Symbolically, it is formulated as

$$H_{1,2} = H_1 \wedge H_2, \text{ where } H_i = \wedge\{p_i(s;s') : h_i(s) \geq h_i(s')\} \text{ for } i = 1,2. \tag{8}$$

Then, we define the second basic belief $Gr(E^*)$. Let $E = (E_1, E_2)$ be a pair of sets $E_i \subseteq S_i, i = 1,2$, which are intended to be the set of possible decisions for $i = 1,2$. For simplicity, let us focus on the perspective of player 1, since 2's perspective is symmetric. In N1, player 1 utilizes the set E_2 as his predictions on 2's possible decisions. We define two formulae for $s_1 \in S_1$ and $s_2 \in S$:

$$N1(s_1;E_2) = \wedge_{s_2 \in E_2}\mathrm{Bst}_1(s_1;s_2) \text{ and } N2(s_2;E_1) = \wedge_{s_1 \in E_1}\mathrm{Bst}_2(s_2;s_1). \tag{9}$$

In the language of Section 3.1, $N1(s_1;E_2)$ is "$\mathrm{Bst}_1(s_1;s_2)$ holds for all $s_2 \in E_2$", which is the part of N1 after "for each $s_1 \in E_1$". Hence, adding $\wedge_{s_1 \in E_1}$to $N1(s_1;E_2)$, we have the representation of N1 in the present language; these are denoted as $N1(E)$ and $N2(E)$, i.e.,

$$N1(E) = \wedge_{s_1 \in E_1}[\wedge_{s_2 \in E_2}\mathrm{Bst}_1(s_1;s_2)] \text{ and } N2(E) = \wedge_{s_2 \in E_2}[\wedge_{s_1 \in E_1}\mathrm{Bst}_2(s_2;s_1)]. \tag{10}$$

Further, we denote $N12(E) = N1(E) \wedge N2(E)$. Now, let $E^* = (E_1^*, E_2^*)$ be intended to be the pair of the greatest sets, which is expressed as

$$Gr(E^*) = [N12(E^*)] \wedge [\wedge_{E \subseteq S} N12(E) \Longrightarrow \wedge_{s \in S}(s \in E \Longrightarrow s \in E^*)]. \tag{11}$$

That is, E^* satisfies $N12(\cdot)$ and is the greatest among E's satisfying $N12(\cdot)$. In sum, we adopt $H_{1,2}$ and $Gr(E^*)$ as the basic beliefs.

---

[7] It is possible to adopt a different logic than the classical logic such as the intuitionistic logic (cf., Suzuki [20]).

In the above language, NE is expressed as $\text{Bst}_1(s_1; s_2) \wedge \text{Bst}_2(s_2; s_1)$, denoted by $\text{Nash}(s_1, s_2)$. Theorem 3.1 is described in the above language as follows.

Theorem 3.1* Let G be a solvable game. Then,

$$H_{1,2}, Gr(E^*) \vdash \wedge_{(s_1,s_2)\in E^*}[(\wedge_{t_2 \in E_2^*}\text{Bst}_1(s_1; t_2)) \wedge (\wedge_{t_1 \in E_1^*}\text{Bst}_2(s_2; t_1)) \Longleftrightarrow \text{Nash}(s_1, s_2)], \quad (12)$$

where $\vdash$ means that the right side is logically derived from $H_{1,2}, Gr(E^*)$.

The above theorem is regarded as a faithful translation of Theorem 3.1. The exact proof needs step-by-step verifications and we omit it.

Consider Theorem 3.1* from the viewpoint of Johansen's J1 to J4. First, J1 implies that the language consists of formulae based on primitive symbols from game G and players' inferences are all based on such formulae. Postulate J2 corresponds to the above (i) and (ii) for $i = 1, 2$. Let us interpret the meaning of *"rational"* decision and "*correctly predicted*" in J3. Action $s_1 \in E_1$ in $N1(E)$ of (10) is 1's possible decision, and E_2 is the set of player 2's possible decisions, interpreted as "*correctly prediction*" by player 1. Finally, J4 defines i's decision as a best response to his prediction on j's decision. Thus, Johansen's J1 to J4 are well interpreted in terms of the statement of Theorem 3.1*, and *vice versa*.

Yet, the same objects, E_1 and E_2, would occur in different scopes of thinking, depending on the perspective. We now consider how to distinguish them explicitly.

Distinction between of one's mind and the other's: Although we derived an infinite regress from N1-N2 as well as R1^A-R2^A, illustrated in Table 1.1, they are reduced into circularity in Table 1.2, because the depths of interpersonal beliefs for the players are not explicitly formulated. Nevertheless, Diagram 1 becomes the proper form by introducing individual beliefs to distinguish the players' epistemic scopes. For the rigorous formulation of this epistemological argument, we need a system of epistemic logic. Such a logical system has been developed in the literature; we refer to only works directly related to the present paper, i.e., Kaneko [9], Hu, *et al.* [5], and Hu and Kaneko [4].[8] Here, we explain the idea of how a belief hierarchy in Table 1.1 is derived, sacrificing rigor in the logic.

We use the belief operator symbols $\mathbf{B}_1(\cdot)$, $\mathbf{B}_2(\cdot)$ of players 1, 2; these are symbolic and their meanings are all determined by symbolical operations in the logical system. First, the interpersonal understanding of H_1,H_2 is described from 1's perspective as

$$\mathbf{B}_1(H_1), \mathbf{B}_1\mathbf{B}_2(H_2), \mathbf{B}_1\mathbf{B}_2\mathbf{B}_1(H_1), \cdots. \quad (13)$$

This is the infinite regress described in Table 1.1. To complete the argument with explicit epistemic layers, we need to modify $Gr(E^*)$ accordingly. However, this will be a large undertaking

[8] It may be asked what differs from the field called "epistemic game theory" (cf., Brandenburger and Dekel [2]), which follows the tradition of the Bayesian game theory with subjective probability (cf., Maschler, *et al.* [12]). Our epistemic logic approach follows the tradition of symbolic logic in the purely symbolic manner, while the semantical development is secondary (cf., Mendelson [14], Kaneko [9]).

because the components, say $N(E^*)$ and $N(E^*)$, of $Gr(E^*)$ should be modified in a similar manner to (13) with some nested epistemic structures. Here, we have infinite candidates of formulae again. Direct treatments of these infinite candidates can be constructed using the *small infinitary formulae* in Hu, *et al.* [5], which permits conjunctions and disjunctions of constructively generated infinite sets of formulae. We remark that the modified formula of $Gr(E^*)$ as a whole needs to be extended once again in the way of (13). The formalized theory with explicit epistemic layers for the rationalizability theory is treated in [5], corresponding to the detailed arguments in Theorem 3.4,[9] but the case of the Nash theory remains open.

Individualistic one-shot play postulate: We have studied a one-shot play situation of a 2-person game G from the perspective of the individual *ex ante* prediction/decision-making, which we call the *literal interpretation*. However, other sources of information/beliefs may be available to players, such as N1(s^*)-N2(s^*) and Theorem 3.3. Should we regard such an additional structure as granted, too? If the answer is positive without considering any source, the theory could be simply a playground for mathematical manipulations without relations to society. Instead, we like to have a broader interpretation that is connected to social environments.

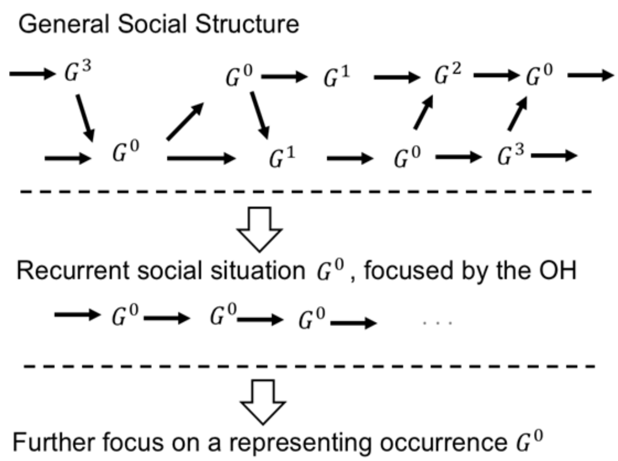

Figure 4:

We take the view of the *principles of methodological individualism* (MI) and *methodological singularism* (MS), due to von Mises [22], pp.41-44, which are two phases of one large ontological principle. The top of Figure 4 illustrates a large society consisting of many small social situations from the past to the future.[10] In the middle, we take a recurrent situation of G^0, which is supposed to be the 2-person game discussed in this paper. Then, in the bottom, we focus on a

[9] In fact, this infinite regress can be avoided by focusing on the fixed-point property. See Hu and Kaneko [4] for an explicit treatment of this property.

[10] The idea of Figure 4 is borrowed from Kaneko and Kline [10].

single occurrence as a representative. Behind the bottom case, G^0 has many occurrences, and players' experiences from those plays would be the source for beliefs of G^0 such as the additional beliefs on s^* in N1(s^*)-N2(s^*). Thus, the situations described in the top and middle can be the sources for individual beliefs. One possible route to study such sources further is to follow Kaneko and Matsui [11] and Kaneko and Kline [10], the *inductive game theory.*

Johansen's [8] last postulate J5 is about the *ex post* evaluation of one's decision and prediction about the other's revealed choice. It holds as long as J1 to J4 are assumed for both players. Hence, its status is a corollary from J1 to J4 rather than an additional postulate. Nevertheless, it is on the boundary to inductive game theory.

Two remark on strategy: von Neumann and Morgenstern's [23] concept of a strategy is a behavior plan for a player in an extensive form game; a strategy is a function that maps a piece of information received to a specific action. An action without such a structure is a special case of a strategy. In this paper, we did not specify which interpretation we adopt. Now, an action has no conceptual difficulties, but a strategy in the sense of [23] requires careful considerations, since it includes a time concept related to actual behavior "*cause and effect*" in the underlying extensive game. In the perspective of an individual *ex ante* prediction/decision-making, each strategy can be treated just as one complete plan with the base structure hidden. As long as we follow this stipulation, a strategy can be treated as if it is an action without an underlying structure.

Finally, we give a remark on the use of a mixed strategy. Mathematically, a mixed strategy is defined as a probability distribution over the set of pure strategies. In the literature, its introduction is regarded as important because it guarantees the existence of a NE, and is typically motivated by the necessity to hide a planned and patterned behavior. Consistent with this motivation, a mixed strategy should be implemented with a random device, e.g., a coin-toss. Under this interpretation, our arguments may accommodate mixed strategies at the conceptual level. The literature typically include all possible probability distributions, which is an infinite object. Under the postulate that our players have only finite logical abilities, it requires a mixed strategy to have a simple probability distribution such as represented by rational number probability values.

5 Conclusions

Both Nash theory and rationalizability theory aim at the problem of "individual *ex ante* prediction/ decision-making in a one-shot game", while they yield very different outcomes. In this paper, we questioned what conceptual bases differentiate these theories. We presented the unified framework so that we pinpointed the difference in their logical bases; it is in the choices of the quantifier "for all" or "for some" for predictions about the other's decision. It tells how the theories treat of a player's own mind vs. the other's mind.

In Section 3, we argued that the Nash theory expresses each player's respect and independence of the other. On the other hand, the rationalizability theory faces serious difficulty according to our notion of "*individual cognitive rationality*", and the common attitude that the rationalizability theory is faithful to rational prediction/decision-making in games seems scientifically inadequate.

In Section 4, we discussed conceptual bases of the Nash theory, taking Johansen's [8] postulates for NE. We reconciled his key term "*rational*" in the postulates with our notion of "*individual cognitive rationality*". We argued that the Nash theory succeeds to separate one's decision from his prediction on the other's mind, which can become formally distinguished if one introduces individual belief operators.

Finally, we returned to more foundational questions about the perspective of individual *ex ante* approach in one-shot game from von Mises's [22] principles of methodological individualism and singularism. According his principles, the perspective of the approach is abstracted from the complicated social situation continuing from the past to the future and is a representative of one occurrence of the partial situation. One important problem is to connect the Nash theory to recurrent situations; the Nash theory is based on merely deductive reasoning, while the new theory involves inductive reasoning from experiences to decision and prediction. This paper suggests to study this broader view; it would give better understanding of society.

Appendix

Proof of (5) and (6): We show by induction over ν that the two sequences $\{R_i^\nu(G)\}_\nu$, $i = 1, 2$, are decreasing with respect to the set-inclusion relation and each $R_i^\nu(G)$ is nonempty. Once this is shown, since S_i is finite, we have $R_i(G) = \bigcap_{\nu=0}^{\infty} R_i^\nu(G) \neq \emptyset$, i.e., (6). Let us show (5). For the base case of $\nu = 0$, we have $R_i^0(G) = S_i \supseteq R_i^1(G)$ and $R_i^1(G) \neq \emptyset$ by (3) for $i = 1, 2$. Now, suppose the hypothesis that this result holds up to ν and $i = 1, 2$. Since $R_j^\nu(G) \neq \emptyset$, we have $R_i^{\nu+1}(G) \neq \emptyset$ by (3). We take any $s_i \in R_i^{\nu+1}(G)$. Then, $\mathrm{Bst}_i(s_i; s_j)$ holds for some $s_j \in R_j^\nu(G)$ by (3). Since $R_j^{\nu-1}(G) \supseteq R_j^\nu(G)$ by the induction supposition, $\mathrm{Bet}_i(s_i; s_j)$ holds and $s_j \in R_j^{\nu-1}(G)$; thus $s_i \in R_i^\nu(G)$.■

Proof of Lemma 3.1 (Only-If): Let (s_1, s_2) be any strategy pair in $E_1 \times E_2$. By N1, $h_1(s_1, s_2)$ is the largest payoff over $h_1(s_1', s_2), s_1' \in S_1$. By the symmetric argument, $h_2(s_1, s_2)$ is the largest payoff over s_2''s. Thus, (s_1, s_2) is an NE in G.

(If): Suppose that any $(s_1, s_2) \in E_1 \times E_2$ is an NE. Let s_1 be arbitrarily taken from E_1. Since for each $s_2 \in E_2$, $h_1(s_1, s_2) \geq h_1(s_1', s_2)$ for all $s_1' \in S_1$, we have N1. We have N2 similarly.■

Proof of Theorem 3.1: Since G is solvable, the set $E(G) \neq \emptyset$ is the Nash solution, which satisfies (2). Let $E = E(G)$. Hence, E is expressed as $E = E_1 \times E_2$ by Lemma 2.1. Since E consists of NE's, (E_1, E_2) satisfies N1-N2 by Lemma 3.1. Since $E = E(G)$ consists of all NE's,

(E_1, E_2) is the greatest pair having N1-N2.■

Proof of Theorem 3.2 (If): Let (E_1, E_2) be a maximal pair satisfying N1-N2, i.e., there is no (E_1', E_2') satisfying N1-N2 with $E_1 \times E_2 \subsetneq E_1' \times E_2'$. By Lemma 3.1, $E_1 \times E_2$ is a set of NE's. Let E' be a set of NE's satisfying (2) with $E_1 \times E_2 \subseteq E'$. Then, E' is also expressed as $E_1' \times E_2'$, which satisfies N1-N2 by Lemma 3.1. Thus, $E_i' \subseteq E_i$ for $i = 1, 2$ by maximality for (E_1, E_2), and $E_1 \times E_2 = E'$. Thus, E is a maximal set satisfying interchangeability (2).

(Only-If): Since E is a subsolution, it satisfies (2) and $E = E_1 \times E_2$ by Lemma 2.1. Thus, (E_1, E_2) satisfies N1-N2 by Lemma 3.1, and its maximality follows from the assumption that since E is a subsolution.■

Proof of Theorem 3.3 (Only-if): Suppose that (E_1, E_2) is the greatest pair satisfying N1(s^*)-N2(s^*). Since (E_1, E_2) satisfies N1-N2, Theorem 3.2 implies that $E_1 \times E_2$ is a maximal subsolution satisfying N1-N2. By N1(s^*)-N2(s^*), it holds that $(s_1^*, s_2^*) \in E_1 \times E_2$. It remains to show that $(s_1^*, s_2^*) \notin E_1' \times E_2'$ for any other subsolution $E_1' \times E_2'$. On the contrary, suppose that $s^* = (s_1^*, s_2^*) \in E_1' \times E_2'$ for some other subsolution $E_1' \times E_2'$. By Theorem 3.2, (E_1', E_2') is a maximal pair satisfying N1-N2, *a fortiori*, N1(s^*)-N2(s^*). This is a contradiction to that (E_1, E_2) is the greatest pair satisfying N1(s^*)-N2(s^*).

(If): Suppose the if part of the assertion. By Theorem 3.2, (E_1, E_2) is a maximal pair satisfying N1(s^*)-N2(s^*). If (E_1, E_2) is not the greatest pair satisfying N1(s^*)-N2(s^*), there is a different subsolution $E_1' \times E_2'$ with $s^* \in E_1' \times E_2'$, which is denied by by the supposition of the if part. Thus, (E_1, E_2) is the greatest pair satisfying N1(s^*)-N2(s^*).■

Proof of Theorem 3.4: We show that if (E_1, E_2) satisfies R1^A-R2^A, then $E_i \subseteq R_i(G)$ for $i = 1, 2$, and the converse that $(E_1, E_2) = (R_1(G), R_2(G))$ satisfies R1^A-R2^A.

Suppose that (E_1, E_2) satisfies R1^A-R2^A. First, we show by induction that $E_1 \times E_2 \subseteq R_1^\nu(G) \times R_2^\nu(G)$ for all $\nu \geq 0$, which implies $E_1 \times E_2 \subseteq R_1(G) \times R_2(G)$. Since $R_i^0(G) = S_i$ for $i = 1, 2$, $E_1 \times E_2 \subseteq R_1^0(G) \times R_2^0(G)$. Now, suppose $E_1 \times E_2 \subseteq R_1^\nu(G) \times R_2^\nu(G)$. Let $s_i \in E_i$. Due to the R1^A-R2^A, there is an $s_j \in E_j$ such that $\mathrm{Bst}_i(s_i; s_j)$ holds. Since $E_j \subseteq R_j^\nu(G)$, we have $s_j \in R_j^\nu(G)$. Thus, $s_i \in R_i^{\nu+1}(G)$.

Conversely, we show that $(E_1, E_2) = (R_1(G), R_2(G))$ satisfies R1^A-R2^A. Let $s_i \in R_i(G) = \bigcap_{\nu=0}^\infty R_i^\nu(G)$. Then, for each $\nu = 0, 1, 2, ...$, there exists $s_j^\nu \in R_j^\nu$ such that $\mathrm{Bst}_i(s_i; s_j^\nu)$ holds. Since S_j is a finite set, we can take a subsequence $\{s_j^{\nu_t}\}_{t=0}^\infty$ in $\{s_j^\nu\}_{\nu=0}^\infty$ such that for some $s_j^* \in S_j$, $s_j^{\nu_t} = s_j^*$ for all ν_t. Then, s_j^* belongs to $R_j(G) = \bigcap_{\nu=0}^\infty R_j^\nu(G)$ by (5). Also, $\mathrm{Bst}_i(s_i; s_j^*)$ holds. Thus, $(R_1(G), R_2(G))$ satisfies R1^A-R2^A.■

References

[1] Bernheim, D., Rationalizable Strategic Behavior, *Econometrica* 52, 1984, 1007-1028. doi.org/10.2307/1911196

[2] Brandenburger, A., Dekel, E, Hierarchies of Beliefs and Common Knowledge, *Journal of Economic Theory* 59, 1993,189-198. doi.org/10.1006/jeth.1993.1012

[3] Chellas, B. F., *Modal Logic, an Introduction*, Cambridge University Press; 1980, Cambridge. doi.org/10.1017/CBO9780511621192

[4] Hu, T.-W., Kaneko, M., Logic of Reciprocal Empathy: Shallow Interpersonal Reasoning and Latent Infinity, DP. No.1392, Department of Policy and Planning Sciences, University of Tsukuba, 2015.

[5] Hu, T.-W., Kaneko, M., N.-Y. Suzuki, N.-Y., Small Infinitary Epistemic Logics, *Review of Symbolic Logic* 12, 2019, 702-735. doi.org/10.1017/S1755020319000029

[6] Grimm, J., Richter, T., Rational Thinking as a General Cognitive Ability: Factorial Structure, underlying Cognitive Processes, and Relevance for University, Learning and Individual Differences 111, 2024, 1-12. doi.org/10.1016/s1041-6080(24)00045-1

[7] Jansen, M. J. M., Maximal Nash Subsets for Bimatrix Games, *Naval Research Logistics Quarterly* 28, 1981, 147-152. doi.org/10.1002/nav.3800280111

[8] Johansen, L., On the Status of the Nash Type of Noncooperative Equilibrium in Economic Theory, *Scandinavian Journal of Economics* 84, 1982, 421-441. doi.org/10.2307/3439426

[9] Kaneko, M., Epistemic Logics and their Game Theoretic Applications: Introduction, *Economic Theory* 19, 2002, 7-62. doi.org/10.1007/s001990100202

[10] Kaneko, M., Kline, J. J., Understanding the Other through Social Roles, *International Game Theory Review* 17, 2015, 1-31. DOI:10.1142/S0219198915400058.

[11] Kaneko, M., Matsui, A. Inductive Game Theory: Discrimination and Prejudices, *Journal of Public Economic Theory* 1, 1999, 101-137. doi.org/10.1111/1097-3923.00005

[12] Maschler, M., Solan, E., Zamir, S., *Game Theory*, Cambridge University Press, 2013, Cambridge, UK. doi.org/10.1017/cbo9780511794216

[13] Mas-Colell, A., Whinston, M. D., Green, J. R., *Mircoeonomic Theory*, Oxford University Press, 1995, New York.

[14] Mendelson, E., (1987), *Introduction to Mathematical Logic (3rd edition),* Wadsworth & BrooksCole Advanced Books & Software. Monterey, California.

[15] Nash, J. F., Non-cooperative Games, *Annals of Mathematics* 54, 1951, 286-295. doi.org/10.1515/9781400829156-008

[16] Pearce, D., Rationalizable Strategic Behavior and the Problem of Perfection, *Econometrica* 52, 1984, 1029-1050. doi.org/10.2307/1911197

[17] Sent, E.-M., Game theorists versus Herbert Simon: Playing Games with Bounded Rationality, *Journal Economic Behavior & Organization* 44, 2001, 129-140. doi.org/10.1016/s0167-2681(00)00160-8

[18] Schelling, T., *The Strategy of Conflict*, Harvard University Press, 1960, Cambridge.

[19] Simon, H., From substantive to procedural rationality, *Method and Appraisal in Economics* 129-148, ed. Latsis, S., 1976. doi.org/10.1017/CBO9780511572203.

[20] Suzuki, N.-Y., Semantics for Intuitionistic Epistemic Logics of Shallow Depths for Game Theory, *Economic Theory* 53, 2012, 85-110. DOI 10.1007/s00199-012-0707-1

[21] Vives, X., Complementarities and Games: New Developments, *Journal of Economic Literature* XLIII, 2004, 437-479. doi.org/10.1017/cbo9780511572203.006

[22] von Mises, R., Human Action: A Treatise on Economics, edited by Greaves, B. B., Liberty Fund, 1949.

[23] von Neumann, J., Morgenstern, O., *Theory of Games and Economic Behavior*, Princeton University Press, 1944, Princeton. doi.org/10.2307/2572550